***Psychological Research* (2006). *70*, 273–292.**
DOI 10.1007/s00426-005-0215-z

Running head: FACE PERCEPTION AND SPATIAL FREQUENCIES

Face Perception: An Integrative Review of the Role of Spatial Frequencies

Marcos Ruiz-Soler	Francesc S. Beltran

University of Malaga	University of Barcelona

Authors' Note

Marcos Ruiz-Soler, Department of Psychobiology and Methodology for the Behavioral Sciences, University of Malaga, Spain. Francesc S. Beltran, Adaptive Behavior and Interaction Research Group, University of Barcelona.

This project was supported by a grant awarded to the first author by the Directorate General for Higher Education and Culture, Spain (PB96-0689). The project was also supported in part by a grant awarded to the second author by the Directorate General for Research of the Government of Catalonia (2001SGR-00140). The authors would like to thank Manuel J. Molina Valencia for his useful explanations about spatial frequencies and the Fast Fourier Transform. Our thanks also go to two anonymous reviewers for their comments on an earlier version of the manuscript.

Correspondence concerning this article should be addressed to Marcos Ruiz-Soler, Department of Psychobiology and Methodology for the Behavioral Sciences, University of Malaga, Campus de Teatinos, 29071 Málaga, Spain. Email: ruizsoler@uma.es.

Abstract

The aim of this article is to reinterpret the results obtained from the research analyzing the role played by spatial frequencies in face perception. Two main working lines have been explored in this body of research: (a) the critical bandwidth of spatial frequencies that allows face recognition to take place (the masking approach) and (b) the role played by different spatial frequencies while the visual percept is being developed (the microgenetic approach). However, results obtained to date are not satisfactory in that no single explanation accounts for all the data obtained from each of the approaches. We propose that the main factor for understanding the role of spatial frequencies in face perception depends on the interaction between the demands of the task and the information in the image (the diagnostic-recognition approach). Using this new framework, we review the most significant research carried out since the early 1970s to provide a reinterpretation of the data obtained.

Face Perception: An Integrative Review of the Role of Spatial Frequencies

Human beings recognize each other mainly by means of the face. Knowing which factors make it possible or impossible to recognize a face is of great interest to both basic and applied research. It is of interest to basic research because it would explain a fundamental human perceptual process and the underlying physiological mechanisms of a complex visual task. Moreover, the results obtained could probably be applied to the perception of other visual stimuli with important configurational properties. It is of interest to applied research because it would make it possible to improve procedures for working with eyewitnesses (Rakover & Cahlon, 2001), design video-surveillance and automatic identification systems and develop better person/machine interfaces in the near future. The importance of this process has motivated the study of these underlying mechanisms for more than three decades. However, after so much effort, there is still no theory that offers a full explanation of all the results obtained.

The research aimed at providing an explanation of the face-recognition process is basically focused on three approaches: cognitive, psychophysical and neurophysiological (see Table 1). The cognitive approach has tried to identify the variables that affect successful perceptual tasks (similarity of stimuli, observation time, level of processing, etc.) and to describe as far as possible the different stages in the process in order to generate a high-level symbolic explanatory model (Bruce & Humphreys, 1994; Bruce & Young, 1986; Burton, Bruce & Hancock, 1999; Ellis, 1992). The psychophysical approach manipulates the physical characteristics of the image, usually by filtering specific spatial frequencies (SFs), to learn how these characteristics affect recognition processes (Costen, Parker & Craw, 1996). The neurophysiological approach focuses on determining which cerebral structures are activated during face perception and, very particularly, whether the inferotemporal zone is the specialized area where this kind of visual stimuli is processed (Desimone, Albright, Gross & Bruce, 1984; Gross, Bender & Rocha-Miranda, 1969; Gross, Rocha-Miranda & Bender, 1972; O'Scalaidhe, Wilson & Goldman-Rakic, 1997; Rolls, 1992). In this study we will focus on the second approach and will describe an explanatory model capable of accounting for the results obtained on face perception within the framework of the role played by SFs.

Spatial Frequencies in Face Perception

Any image, whether of a human face or any other visual object, can be described in terms of SFs, i.e. it can be described as the sum of a set of sinusoidal grids with different frequencies and orientations. Psychophysical research into contrast detection and adaptation to specific SFs has proven that our perceptual system analyzes visual input on multiple scales or frequencies (see De Valois & De Valois, 1988; Graham, 1989; and Westheimer, 2001 for an overview). It is now generally agreed that spatial filtering is the basic mechanism for extracting visual information from luminance contrasts in early visual processes, including edge detection (Marr & Hildreth, 1980), stereopsis (Legge & Gu, 1989), movement (Morgan, 1992) and depth perception (Marshall, Burbeck, Ariely, Rolland & Martin, 1996). In light of all this, one of the main approaches involves manipulating the SF bands in the luminance spectrum of images (see Figure 1) and observing how these changes affect performance on visual tasks.

One of the first questions asked when investigating face perception was: What range of SFs is necessary to recognize a face? However, the results of different experiments did not provide a clear answer, since an extensive range of SFs seems to play a role in recognition. The second question was: in what order are low spatial frequencies (LSFs) and high spatial frequencies (HSFs) integrated in face perception and how does this order affect recognition? Unfortunately, the results obtained here have not been definitive either because they do not always point toward the same length of time or order of integration (see Hoeger, 1997 or McSorley & Findlay, 2002). Nevertheless, a group of experiments on scene perception (Oliva & Schyns, 1997) seems to provide an explanation for the disparity of these results. These researchers found that when the already integrated early perceptual representation is formed it may be used flexibly in a topcontrolled manner permitting selective use of LSFs or HSFs depending on how "diagnostic" they are for the task. Taking this into account, we suggest that a key question for determining what role SFs play in face perception is not really which SFs are necessary or in which sequential order they are integrated, but rather how LSFs and HSFs are made use of in face perception depending on the demands of the task involved. Therefore, the role of different SFs is critically modulated by the subject's visual task and it is only when there is not a specific visual task that the mandatory aspects of SF-processing work by default.

A New Framework for Research into Visual Perception: The Diagnostic-Recognition Approach

The fact that the importance of SFs varies depending on the demands of the task was reported by Schyns (1998) in a new framework that attempts to explain how sensorial information is adjusted to the information stored in memory. The most important idea in this framework is that the information required to place the same object in one category or another will change depending on the categorization criterion in use (i.e. the interaction between task constraints and object information). Task constraints are related to the information needed to place the perceptual object in the category required by the task, e.g. given the question: "Is this object a car?", it will be necessary to find certain visual information (such as wheels, rearview mirrors, a steering wheel, etc.) before giving an answer. Object information is related to the informative-perceptual structure available for placing the perceptual object in the category demanded by the task. If the image is an object and it is possible to observe wheels, rearview mirrors, a steering wheel, etc., then we have the information we need for categorization and the question can be answered. Therefore, given a specific perceptual task, a group of visual characteristics of the object becomes especially useful (diagnostic) since it provides the necessary information to place the object in the category that resolves the task.

It is our belief that the diagnostic-recognition approach can be used to account completely for the empirical evidence supporting the notion that face processing is mainly holistic (based on configurational aspects or based on all the information about the entire image) and the empirical evidence supporting the idea that face processing is occasionally analytical (based on local features). Since the types of features in a face are always the same or very similar, as a general rule they cannot be used as a criteria for fast differentiation of faces (or differentiation with limited conditions), i.e. such aspects will not be diagnostic, but only the particular configuration will be diagnostic (the interrelationship between facial features) since such a configuration is sufficient, unique and significant and therefore useful for recognition (Schyns, 1998). However, holistic processing is probably not efficient when we are asked to recognize people we have just met, such as when we are introduced to several people at a crowded party. Categorization of these people will be faster if it is based on specific features such as a beard, moustache, mole, glasses, hair, earrings and so forth than on the use of configurational information. In other words, in this specific case, such features become diagnostic because of the kind of categorization required by the task and some people will prefer to make use of them. This does not mean that the best strategy is always to select a specific feature, but to select the element that is most relevant to oneself, including holistic elements (such as an enlarged or reddish face).

Although previous research has pointed out that default SF-processing works by integrating LSFs to HSFs, the possible importance of task demands in face perception has been explicitly affirmed by several researchers. For instance, when referring to the long-running controversy about the importance of LSFs and HSFs in face perception, Sergent (1986, 1994) said: “the controversy may be resolved if one considers the particular requirements in terms of the spatial-frequency information that needs to be processed for optimal performance. While a matching task may be achieved equally well by processing the high or the low frequencies, an identification benefits from processing the high frequencies.” (Sergent, 1986, p. 23).

Furthermore, Costen, Parker & Crow (1996) said, “A third explanatory [about the discrepancy between studies that have varied the effective SF range of the images] might be in terms of the task that the subjects were asked to perform […]. This suggests that different tasks are supported by different spatial frequencies and, thus, that the results from the constant-frequency-range studies may not reflect face identification, but rather perceptual matching of some sort” (p. 603). And, more recently, McSorley & Findlay (1999) stated that “The pattern of results found […] could be interpreted as reflecting the differential use of spatial-frequency information according to the task or training […]; further work is needed to clarify whether spatial-frequency integration is indeed taking place in the tasks reported by Schyns and Oliva (1997) and Parker et al. (1996). If integration does occur, then it is possible that these results simply reflect a flexible integration mechanism which depends upon task demands” (pp. 1047-1048).

The results obtained by Schyns & Oliva (1997) indicate that subjects perform as well as they can to solve the task, selecting the most diagnostic SFs for the task from among the SFs available in the stimuli. In other words, the subjects will be expected to select other SFs when they are asked to perform the same task and the SFs available in the image are different (e.g. recognizing a friend within a few meters in broad daylight is not the same as recognizing her at a distance on a dark street). Likewise, the subjects will be expected to select different SFs when the task is different and the same SFs are available (e.g. determining whether a person coming toward us is a man or a woman is not the same as determining whether the person is someone we know or don’t know). Therefore, if we are interested in understanding how face recognition occurs, the question to be asked is: what SFs are diagnostic for the task? And the answer will be the particular combination of task requirements of a specific categorization and the SFs available in the image, i.e. through task constraints and object information. Recently, while reviewing the role of SFs in visual processing, including faces, objects and scenes, Morrison & Schyns (2001) pointed out that “mechanisms of categorization can modulate the usage of different scales, according to the presence of task-dependent, diagnostic information. Further research is required to unravel the nature of this diagnostic information for different categorization tasks and the same object and how this information depends on the scale" (p. 467). In this article, we propose an extensive framework for the investigation of face perception based on the diagnostic-recognition approach. We will first review the key findings on face recognition and show that if these very diverse results are interpreted in terms of this new framework for visual perception, a clear picture is provided of how the process happens. Then, based on the compatibility between the diagnostic-recognition approach and the empirical results, we will indicate the main directions the research on face recognition should go and specify some of the hypotheses to be tested in each area.

To show how the research carried out to date is compatible with the diagnostic-recognition approach, we will provide an exhaustive review of the empirical studies done since the 1970s, grouped according to their main objective: (a) finding the optimum SFs for face recognition, (b) determining how SFs are involved in the recognition process, and (c) analyzing how image information and task requirements make some SFs diagnostic. Each section will include a brief introduction to the experimental paradigm used, a review of the experiments grouped according to their conclusions and a discussion of the theory with the aim of providing a comprehensive analysis of the experimental results' compatibility with the postulates of the diagnostic-recognition approach.

Which SFs are Critical? The Masking Approach

Initially, the main objective of research focusing on image filtering was the search for the range of SFs that are critical for face recognition. Although many results seem to indicate that middle-range SFs are the critical ones (Costen, Parker & Craw, 1994, 1996), it has been demonstrated that a large group of SFs, some of which are very far from the middle range, are needed to resolve recognition tasks with a good level of efficiency. Likewise, other results suggest that HSFs can also play an important role in face recognition (Fiorentini, Maffei & Sandini, 1983). Despite the disparity of these results, they can all be accounted for by the predictions of the diagnostic-recognition approach.

Experimental Paradigm

Research aimed at finding the critical range of SFs for face recognition uses a masking paradigm, i.e. a procedure that produces an impairment of the perceptual quality of the image of the face used for recognition. There are two kinds of masking: one we will call classic-procedure masking and the other is known as critical-band masking (the name used in the pioneering work by Harmon & Julesz, 1973). Classic-procedure masking consists of displaying a stimulus (mask) milliseconds (ms) before or after the target stimulus, thus interfering with recognition. When masking follows the target stimulus it is called backward masking and when it precedes the target it is called forward masking. Critical-band masking consists of generating an image with less informative value than the corresponding "original" image due to some kind of transformation that selectively affects the critical band. There are basically four ways to transform an image: (a) pixelization / quantization (mosaic effect), (b) noising (snow effect), (c) gridding (grid effect) and (d) filtering (fuzzy or sharp effect). The most commonly used transformation technique is filtering, which has three varieties: low-pass, high-pass and band-pass filtering. Low-pass filtering consists of obtaining a new image composed only of the SFs in the original image that are below a specified cut-off value, thus resulting in a fuzzier image than the original one. Similarly, high-pass filtering consists of obtaining a new image composed only of the SFs in the original image that are over a specified cut-off value, thus resulting in a sharper image than the original one. Band-pass filtering consists of keeping the SFs in the original image that are between two cut-off values, thus resulting in an image defined by fine-scale contours and edges only. (See Figures 1 and 2)

Since filtering produces impairment of the perceptual quality of the image, if a critical component for face recognition is contained in the SFs, removing that component before testing will be more disruptive than removing a non-critical range. In the procedure used for data collection, subjects are asked to recognize faces they have previously learned in which some SFs have been removed. Subjects' different levels of efficiency in recognizing some faces but not others allow us to deteremine what eliminations will be most disruptive so we can then deduce which SFs are critical.

A Framework for Studying the SFs that are Critical for Face Identification

Harmon and Julesz (1973) tested an explanation for a well-known phenomenon: a pixelized image with large blocks is easier to identify when a blurring operation is applied to it, such as squinting. Using a digitized picture of President Lincoln, two low-pass-filtered images (12.6 cycles/deg and 40.6 cycles/deg) and one high-pass-filtered image (from 121.2 to 39.4 cycles/deg) were obtained[1]. The results showed that the low-pass-filtered image at 12.6 cycles/deg was recognized better, but the low-pass-filtered image at 40.6 cycles/deg was not. However, filtering of the critical band of frequencies between 1.22 cycles/deg and 3.94 cycles/deg while sustaining HSF above 3.94 cycles/deg did lead to easy recognition, despite the presence of HSFs in the filtered picture.

This study demonstrated that the SFs within images play a key role in the recognition of visual objects. A main line of research focused on determining whether some SFs are critical for face recognition and, if this was found to be true, identifying the specific values of these SFs. The image-filtering technique became the basic experimental paradigm of this line of research. Subjects were asked to recognize faces they were familiar with, either because the images were of celebrities or because the subjects were shown their pictures in a learning phase. Images of faces made up of a wide range of SFs were manipulated by filtering spatial frequencies and then displayed in the test phase. Accuracy and reaction time (RT) were recorded to find out how the results were affected by the kind of filtering used to manipulate the images[2].

Empirical Evidence Supporting the Critical Role of Medium-Range SFs

Although some results have shown that LSFs are efficiently used for face recognition (Sinha, 2002), many results also indicate that one range of SFs is critical for face recognition, including middle-range SFs. An initial proposal came from the results of a classic recognition task (Tieger & Ganz, 1979). A gridded mask with four sinusoidal values was used in the learning phase: 0.54, 0.82, 2.2 and 3.9 cycles/deg. Since discrimination (d') was most interfered with by the 2.2 cycles/deg mask, i.e. about 17.6 cycles/face width (cycles/fw), the conclusion was reached that this SF is more important for face identification than other SFs with higher or lower values.

From that seminal paper, an extensive line of research has since provided empirical evidence to determine the boundaries of the critical SF range more accurately. Costen, Parker and Craw (1994) instructed subjects to learn six faces (which were equivalent in size, position and physical features) that were displayed on a screen for one second. During the test phase the subjects were instructed to recognize faces that were shown on the screen for only 100 ms. Each face in the test phase had 11, 21, and 42 pixels (5.5, 10.5 and 21 cycles/fw, respectively), had been low-pass filtered and had been blurred using Gaussian filters. The results indicated a critical value (8 cycles/fw), below which recognition decreased dramatically (similar results were obtained in parametric studies carried out by Bachmann [1991] and Bhatia, Lakshminarayanan, Samal & Welland [1995]).

A second group of experiments carried out by Costen, Parker and Craw (1996) employed the same learning and test phases as in their 1994 experiments except that the images in the test phase were manipulated by pixelization (45, 23, 12 and 9 pixels per face), low-pass filtering and high-pass filtering (22.5, 11.5, 6 and 4.5 cycles/fw). The results of the first experiment indicated that a specific band of SFs (between 8 and 16 cycles/fw) was the most useful for perceptual recognition. It was therefore confirmed that the face-recognition process is based on the information provided by medium-range SFs (between 8 and 16 cycles/fw).

All these studies showed a decrease in recognition rates when medium-range SFs were eliminated, so it was reasonable to predict better recognition of images with medium SFs (more right answers and shorter RTs) than of images containing only SFs from each extreme (Parker & Costen, 1999). Photographs of six faces were shown at five rotation angles (0, 22, 45, 65 and 90 deg) and band-pass filtered at five values (centered at 2.46, 5.22, 11.1, 23.6 and 50.15 cycles/fw, and a width of 1 octave). The results showed no effects due to the rotation angle, but showed effects due to SF filtering. In accordance with the hypothesis, the highest number of right answers was obtained for images with medium-range SFs. Differences were also observed between medium-range SFs and SFs at each extreme, as well as between the SFs at both extremes (2.46 and 50.15 cycles/fw). RTs were shorter for medium-range SFs (5.22 and 11.1 cycles/fw), which differed from the RTs of all other SFs.

Although these results confirmed the fact that there is a privileged range of SFs for face recognition between 7.85 cycles/fw and 15.69 cycles/fw (with a harmonic mean of 11.10 cycles/fw), the images with the lowest recognition rate (2.46 cycles/fw) showed a mean success rate of over 70%. Moreover, for the same range of SFs, the RTs were only 120 ms slower than the fastest RTs (corresponding to 11.1 cycles/fw filtered images), i.e. although medium-range SFs proved to be ideal for face recognition, lower or higher SFs were also useful for acceptable recognition.

These results, which highlight the extraordinary flexibility of the visual system when it comes to recognizing faces, were also confirmed by four experiments measuring the visual system's relative sensitivity to different SFs (Näsänen, 1999). In the test phase faces were displayed, but information in narrow bands centered at different SFs was either selectively eliminated or preserved. The subjects were given feedback on their accuracy, in that the contrast was decreased 1.26 times after four consecutive successes and increased 1.26 times after four consecutive failures. The results showed that the main information used by the visual system for recognition was in a range of between 8 and 13 cycles/fw with a bandwidth below 2 octaves. Moreover, using a dynamic face-recognition task requiring eye movements, Ojanpää and Näsänen (2003) found that the most important information for locating a face is also found in a limited band of middle-range SFs.

Empirical Evidence Supporting the Significant Role of HSFs

Experiments have proven that image recognition is also possible with pictures containing SFs that are far from the medium range, so it remains to be ascertained whether or not these SFs play a specific role in recognition. One group of studies has pointed out the possible role of HSFs. In Fiorentini, Maffei & Sandini's (1983) experiments, subjects performed the test phase with faces filtered below 5 cycles/fw and faces filtered above 5 cycles/fw, or with faces filtered above and below 8 cycles/fw. The results were interpreted as showing that the information contained in HSFs does not overlap the information contained in LSFs and that HSFs contain sufficient information to produce visual recognition.

The specific role of LSFs and HSFs was also explored using positive and negative images (Hayes, Morrone & Burr, 1986). It is a known fact that it is more difficult to recognize a negative image than a positive one. Negative images show a 180-degree change in phase, but keep the amplitude constant. We can predict that displaying positive or negative images will produce different effects on recognition, because LSFs are more phase sensitive than HSFs. The faces were shown at three angles (front, 30 deg right and 30 deg left) and band-pass filtered (3.2, 6.4, 12.5, 25 and 50 cycles/fw). The results suggest that HSFs play a different role than LSFs depending on the perceptual process in progress, e.g. HSFs could be informative enough to help the subject locate edges. Moreover, the smallest number of errors was recorded with the images filtered at 25 cycles/fw, regardless of the visual angle, i.e. the results of this study indicate that

SFs over 20 cycles/fw (which is higher than the value usually found) are the most useful for recognition in that particular condition.

Empirical Evidence Supporting Configurational Properties: Is Anything Used Besides SFs?

If the information provided by a wide range of SFs allows for efficient face recognition and the information provided by HSFs produces different effects depending on the specific perceptual process, we can conclude that all SFs contained in an image's luminance spectrum contribute in some way to making recognition possible. But are some elements in the spatial domain represented in the information provided by the luminance spectrum of the image associated with some combinations of SFs? Costen, Shepherd, Ellis & Craw (1994) showed faces with different masks, none of which the subjects were familiar with. The masks were retroactively displayed with an inter-stimuli interval (ISI) equal to 0 ms and for the same exposure time as the known faces. It was found that there was strong masking with the masks containing faces, there was medium-level masking with inverted faces, disorganized faces and faces without internal elements, and there was no masking with masks containing objects or noise. These results can be explained by the fact that masks hide configurational information about the face, which seems to be essential to recognize complex stimuli (Collishaw & Hole, 2000; Haig, 1984; Hole, 1994; Sergent, 1984; Tanaka & Farah, 1993; Young, Hellawell & Hay, 1987; Sinha, 2002).

Uttal, Baruch & Allen (1997) tried to generalize the effect obtained by Harmon and Julesz (1973), i.e. filtering following pixelization improves face recognition. The faces were shown in two sizes (3.5 x 6.05 deg and 0.75 x 1 deg) and with different low-pass-filtered values (0.43, 0.35, 0.26 and 0.17 cycles/deg for large stimuli; and 3.04, 2.6, 2.17 and 1.74 cycles/deg for small stimuli). Every face was pixelized in several different numbers of blocks, both for large stimuli (10 x 10, 15 x 15 and 20 x 20) and small stimuli (2 x 2, 3 x 3, 4 x 4, 5 x 5 and 6 x 6). If low-pass filtering of a previously pixelized image improves recognition because it removes the masking produced by HSFs over LSFs, then no improvement will be expected when pixelization follows low-pass filtering because pixelization introduces HSF components that interfere with the lowest SFs. The results agree with this prediction for large stimuli, but do not agree for small stimuli. Hence, the results show effects of perceptual organization on the recognition process, which are as important as the energetic distribution of SFs.

More recently, using a forward masking paradigm, Bachmann, Luiga & Põder (2004) and Bachmann & Põder (2002) showed that pixelized noise with the face-typical spectrum of SFs causes less masking of broad-band gray-scale images of faces in comparison with pixelized versions of the different faces. This result confirms the high importance of configural information, because pixelisation and spatial frequency values of noise and different-face masks were the same, but their configuration was different. Similar results on the role played by the configural information of the face were also found by McKone, Martini & Nakayama (2001).

Theoretical Discussion

The results of the studies designed to determine whether or not there is a critical set of SFs for face recognition indicate that (a) recognition decreases when images contain only SFs below about 8 cycles/fw (between 6-9 cycles/fw), and (b) elimination of the SF range between 8 and 16 cycles/fw produces a greater disruption than elimination of SFs outside this range. Hence, the information contained in a small medium range of SFs contributes more to the face-recognition process than the information contained in all the other SFs (Costen, Parker & Craw, 1994, 1996; Näsänen, 1999; Parker & Costen, 1999). However, though all these results indicate that privileged information can be found in medium-range SFs, the role of the SFs outside that range cannot be overlooked. There are at least three arguments that support this: (a) the high level of efficient recognition using SFs outside the medium range, (b) the bimodal representation of HSFs (Ivry & Robertson, 1998; Sergent, 1986), and (c) the effect of perceptual-organization properties as a result of some combinations of SFs.

Evidence for (a) can be found in the same studies that identified the optimal medium range of SFs and which also showed acceptable performance by subjects when SFs above and below the medium range were used. Images of faces made with SFs centered at 50.15 cycles/fw or 2.46 cycles/fw (which is extraordinarily far from the medium range) showed a recognition efficiency only 15% lower than the efficiency when recognizing images of faces made with medium-range SFs (Parker & Costen, 1999). Moreover, the tails obtained in the sensitivity function for images of faces indicate that an extensive range of SFs contribute to recognition (Näsänen, 1999).

Evidence for (b) can be found in the studies that concluded that SFs higher than the proposed medium range were critical for recognition: 20 cycles/fw (Hayes, Morrone & Burr, 1986), i.e. HSFs can play an important role that differs from the role of LSFs. Therefore, given the fact that low-pass-filtered images are seen as blurred versions of the original and high-pass-filtered images are seen as line drawings, it would appear that subjects are able to interpret the two altered versions of the originals (Fiorentini, Maffei & Sandini, 1983).

Evidence for (c) can be found in the results that indicate there is information in the properties of perceptual organization that could contribute to face recognition (Uttal, Baruch & Allen, 1997). Given these three factors, the idea of a “critical range” of SFs for face recognition should be replaced with the notion of an “optimal range” of SFs for face recognition: a preferred, but not exclusive, tendency to use the information contained in a given range of SFs depending on the size of the facial image.

Based on a model using a diagnostic-recognition approach, diagnostic SFs are the ones preferentially used to solve a given task, depending on the task constraints and object information. Given the fact that the most useful information for a face-recognition task can be found in the optimal range of SFs, if that range is available in the image, it will be the diagnostic range used for the task and will be used first from among all the SFs available in the image after the mandatory integrative stage. However, the range of diagnostic SFs also depends on the information provided by the face. Hence, if the SFs from the optimal range are not available, then the diagnostic SFs will either be those containing the information that makes it possible to solve the task efficiently from among the SFs available or the SFs that make it possible to retrieve similar information to that provided by the SFs in the optimal range. In this case, given the fact that what is diagnostic guides the process of interpreting the information in the stimuli, even images containing a minimum amount of information about the face can be recognized if one knows the face comes from a limited domain of faces. For example, if subjects know they will be shown well-known active politicians, such as the President of the United States or the British Prime Minister, the task imposes important restrictions on any information shown, and even images containing LSFs can help make recognition possible (Parker & Costen , 1999). Hence, according to predictions based on the diagnostic-recognition approach, knowledge about the task (task constraints) indicates which image-specific informative characteristics will be useful to solve the task. For this reason, even when information about the face is poor, the observer can exploit any cue it provides for recognition, e.g. the configurational traits of the stimuli. This also explains why the configurational traits of the stimuli can be exploited in the recognition process (Uttal, Baruch & Allen, 1997). Furthermore, if the information provided on the faces is varied, the diagnostic SFs are no longer in the optimal range, e.g. if the images are line drawings, the critical SFs are higher than the ones in the optimal range (Hayes, Morrone & Burr, 1986).

When Are SFs Selected? The Microgenetic Approach

The main body of research into face recognition has focused on the temporary integration process necessary for recognition, i.e. the framework called the microgenesis of perception. However, like research into critical SFs, the results obtained do not completely adapt to the theoretical models proposed. A great deal of empirical evidence points toward a fixed integration process that starts with the coarse information (contained in LSFs) and then moves on to the fine information (contained in HSFs), similar to object or form perception in general (e.g. Hughes, Fendrich & Reuter-Lorentz, 1990; Hughes, Nozawa & Kitterle, 1996; Kimchi, 1992; Navon, 1977; Sanocki, 1993) . However, some empirical evidence has also indicated processing can be flexible, as processing is sometimes coarse to fine, while others it is fine to coarse.

The ability to identify a movie actor's face as the lead in The Mask of Zorro or as the face of Antonio Banderas would appear to be immediate, but it actually involves a complex process of sequentially integrating information, which can be analyzed by studying the activation process of SF channels. A great deal of research has suggested there is interaction between channels, so one channel's activity will affect the selective responsiveness of SFs in another channel (Henning, Hertz & Broadbent, 1975; Nachmias, Sansbury, Vassiley & Weber, 1973; Nachmias & Weber, 1975; Tolhurst & Barfield, 1978; and, more recently, Field, Hayes & Hess, 1993; Hess & Dakin, 1997; Polat & Sagi, 1993, 1994). One of the interactions involves the time domain, i.e. interactions between channels depend on the point in time at which each channel is activated, which makes different SFs available at different times. The result is a visual percept that gradually increases in brightness, detail and metrical resolution: it is first perceived as a diffuse, nebulous whole that allows figure-ground discrimination to be made, and finally the edges and details are perceived clearly (Sergent, 1986). Thus, seeing a face implies initially perceiving the shape and configuration of the parts, then the parts themselves (eyes, nose, mouth, ears, forehead, chin, etc.) and finally the details that make it possible to identify it as a unique visual object.

Experimental Paradigm

Tests based on the microgenetic approach attempt to manipulate the SF-integration process by controlling exposure time and resolution. The exposure time is manipulated to extraordinarily low values, ranging from a few ms to tens of ms, and the resolution level is manipulated in terms of the number of cycles/degree (or cycles/face width) using a pixelization or filtering process. The development of the visual percept is studied through the interaction between image information (in SFs) and exposure time for processing. The microgenetic approach predicts an initially fast increase in the recognition rate, followed by a variation depending on the kind of stimulus: (a) stabilization of the LSFs in the stimuli, and (b) a gradual increase in the HSFs in the stimuli (including full-bandwidth stimuli). The literature does not give a specific name to this procedure, so we will call it the incremental-exposure paradigm. Besides this paradigm, two more experimental paradigms are also used: the priming paradigm and the backward-masking paradigm. (see Figure 3)

A Framework for Studying Microgenesis in Visual Perception: The Integration of SFs

Calis, Sterenborg and Maarse (1984), and also Bachmann (1989), introduced the exposure time and resolution level of stimuli as critical variables in face recognition. In terms of SFs, the resolution level can be "translated" into LSFs for low resolution and HSFs for high resolution, i.e. the problem of microgenesis can be seen as the problem of the order in which SFs are integrated. There are two kinds of integration: (a) integration that depends completely on a fixed sequence from LSFs to HSFs, where any possible interruption would damage the integration process (the anisotropic-integration hypothesis), and (b) integration of spatial frequencies that does not depend on any pre-set sequence, but where all available frequencies at a given time are integrated (the isotropic-integration hypothesis). The empirical research supports both hypotheses, as well as the interaction of other variables.

Empirical Evidence Supporting the Hypothesis of Anisotropic Integration

For the anisotropic-integration hypothesis, the integration order is fixed and the majority of empirical data indicates that LSFs are integrated faster than HSFs (Hughes, Fendrich & Reuter-Lorenz, 1990; Hughes, Nozawa & Kitterle, 1996; LaGasse, 1993; Parker, Lishman & Hughes, 1992, 1996; Watt, 1987, 1988), though one study found that HSFs were integrated faster (McSorley & Findlay, 2002).

Bachmann (1987) showed two target images (an eye with an eyebrow and a face) mixed up with eight images of different objects. All of them were pixelized at 128 x 128, 56 x 56 and 36 x

36 pixels, and shown for 1, 20 and 100 ms. According to the results, when the exposure time increased, the ratio of correct responses also increased for the 128 x 128 pixelization, but for the interval between 20 and 100 ms the rate stayed constant for the 56 x 56 pixelization and decreased for the 36 x 36 pixelization. The results were interpreted as proof of the hypothesis that, as the processing time increases, local information also begins to be processed. Given the fact that no useful local information is contained in a stimulus with 36 x 36 pixelization, self-backward masking by the blocks in the pixelized image obstructs efficient recognition of the original face information contained in the LSFs of the pixelized image. These results agree with the anisotropic-integration hypothesis for visual information starting with large-scale properties (LSFs) and moving on to fine-scale properties (HSFs).

In accordance with the preceding results, Bachmann (1991) found similar data when he pixelized faces at different values (15, 18, 21, 24, 27, 32, 44 and 74 pixels/fw) and showed them at 6 different exposure times (1, 4, 8, 20, 40 and 100 ms). The percentage of correct identifications increased with the increase in the SFs, but only when the images with 15 pixels/fw were compared with the remaining pixelization conditions, which did not show differences among them. The identification of images with more than 15 pixels/fw monotonically increased with exposure duration. The images with 15 pixels/fw produced a marked decrease in the percentage of identifications and recognition efficiency did not increase with exposure time, but surprisingly decreased when the exposure time was increased, probably because of HSF masking produced by the block edges. The results agree with a microgenetic model, which describes initial processing of global information (coarse) and the progressive inclusion of local information (fine), whereas global information is no longer used. The decrease in recognition for the images with 15 pixels/face, despite the increased exposure time, indicates that these images did not contain the fine information necessary to allow for processing at later stages, but rather provided misleading, detailed information, which decreased the recognition rate.

Although Parker, Lishman and Hughes (1992) found empirical support for the anisotropic hypothesis using a procedure that involved displaying a three-image sequence containing different frequencies of the same scene, shown for an exposure time of 40 ms, they also argued that the task (judging the quality of an image) was not precise enough and it probably would have been better to ask to the subjects to perform direct discriminations. Hence, Parker, Lishman and Hughes (1997) tested subjects' ability to classify full-bandwidth images into coarse-to-fine and fine-to-coarse sequences containing distorted versions of the originals, i.e. only LSFs or HSFs. It was critical to force the subjects to look at the whole information spectrum and not only at a single band of SFs, i.e. it was necessary to ensure the subjects did not use ad hoc strategies because that would make it impossible to know the order in which the SFs had been integrated. The procedure showed a three-image sequence: two LSF versions of the images and one full-bandwidth version, or two HSF versions and one full-bandwidth version for 40 ms each. The results made it possible to conclude that the responses were not determined by the final quality of the image in the sequence, but by the direction of the flow of information. For this reason, there is evidence of an anisotropic mechanism of temporal and spatial integration that operates more efficiently with a coarse-to-fine sequence (LSFs to HSFs) than with a fine-to-coarse sequence (HSFs to LSFs).

<u>Empirical Evidence Permitting the Hypothesis of Isotropic Integration</u>

Although "evidence for the temporal precedence of coarse-scale information in the recovery of information from a visual stimulus in the integration of information across spatial scales and in the allocation of attention shows it to be a robust phenomenon" (Parker, Lishman and Hughes, 1996, p. 1464), the role of LSFs is not clear. There are two main approaches, which are referred to as the <u>pre-processing approach</u> and the <u>coarse-processing approach</u>. According to the pre-processing approach, described mainly by Marr (1982), LSFs are necessary for pre-processing the image, but are not necessary for higher-level tasks, such as pattern classification and object recognition. According to the coarse-processing approach, described mainly by Ginsburg (1978), LSFs play a key role in the recognition and classification processes, providing the basis for coarse categorization that is progressively refined when the scales of higher frequency are integrated (e.g. an image is successively categorized as “an animal”, “a dog”, “a cocker spaniel” and finally “Sultan” [his name]). This approach is consistent with Eriksen and Schultz’s continuous-flow model (1979), which has been supported by several empirical studies.

Given the fact that each theoretical approach implies different predictions, the objective of the four experiments carried out by Parker, Lishman and Hughes (1996) was to determine which approach provided the best explanation for the early stages of face-recognition processing. All experiments used a face-matching task and a priming experimental paradigm, according to which spatially filtered images were used as a prime of a second image. If the images were of the target stimulus, recognition was facilitated. Otherwise, recognition was disrupted. The researchers hypothesized that two things could happen. If the prediction of the coarse-processing approach was correct, the related LSF primes should provide for facilitation and the unrelated LSF primes should cause disruption. If the pre-processing approach was correct, the related HSF primes should provide for facilitation and the unrelated HSF primes should cause disruption. In order to test this hypothesis, six stimuli containing frontal male faces were high-pass filtered (32 cycles/fw) and low-pass filtered (5 cycles/fw), where the last filtering eliminated the SFs above 7.5 cycles/fw, the bandwidth that appears to contain configurational information that is very useful for recognizing faces. The results showed that if relevant information was shown in primes containing only LSFs or HSFs, they were both effective for facilitation, but if irrelevant information was shown, the HSF primes were more disruptive. For this reason, the results did not clearly support a single approach and did not support “the view that the natural path to object recognition is initially via coarse-scale information” (Parker, Lishman & Hughes, 1996, p. 1462). In accordance with these results, the temporal priority of the integration of coarse visual information was also questioned, but that study probably involved an artefact: judgements were made based on integrated images kept in the visual memory, which were therefore equally well stored and available for visual processing.

Finally, although McSorley & Findlay (1999) did not use facial stimuli, their results supported an anisotropic model for integration of SFs. However, if we bear in mind the inconclusive results of the experiments carried out by Parker, Lishman and Hughes (1996), Oliva and Schyns (1997) and Schyns and Oliva (1997), we can see they are compatible with an approach based on a flexible integration process: “the pattern of results found could be reinterpreted as reflecting the differential use of spatial-frequency information according to the task or training” (McSorley & Findlay, 1999, p. 1048).

Empirical Evidence Supporting the Hypothesis of Interactive Integration

How can these results be made compatible with the robust effect of faster processing of LSFs compared to HSFs? To answer this question some researchers have argued that the order of SF integration involves other factors, a proposal that we will call the interactive-integration hypothesis. This hypothesis rejects a fixed integration order (from LSFs to HSFs) and a completely variable integration order, and states that integration is always in a given order (LSFs to HSFs or HSFs to LSFs), but will depend on the values of other factors.

Hoeger (1997) studied stimulus complexity as a possible interaction factor and tried to show that the temporal precedence of processing LSFs did not necessary imply a dominance of LSF processing. One interesting finding made by Hoeger was that the information-processing speed using HSFs depended on the image’s degree of complexity, but when LSFs were used, the speed was similar for all images, regardless of their complexity. These results agree with an interpretation favoring both approaches, the coarse-to-fine processing approach and the pre-processing approach. Grouping is faster when the image contains easily organized elements such as emergent traits (e.g. symmetry or closure), whereas grouping is slower when it is done in different stages, as for example when information that describes the LSF content is synthesized from the HSFs available (Hoeger, 1997). For this reason, both the information in LSFs and the information in HSFs seem to be temporarily available in equal measure, even though the human visual system processes LSFs faster than HSFs.

Bachmann and Kahusk (1997) examined the role of selective attention as a possible way to explain the changes in recognition rates using pixelized images. Based on the paradox of exposure time (an increase in the image's exposure time from the lowest value to 25 ms involves a decrease in the identification rate), they argued that if this phenomenon was related to attention, then significant interaction should be obtained between attention manipulation and the SF level in pixelized images. But if the phenomenon was not related to attention, no interaction should be found or only additive effects should be found. An experiment was carried out in which the localization cue (pre-cue) (present vs. absent), kind of image (original vs. pixelized) and exposure time (28, 44 and 76 ms) were manipulated. Pictures of male faces (from 9 to 16 pixels of inter-auricular distance) were shown with a visual angle of 5.3 x 7.44 degrees, which could appear in four places (the upper-left, upper-right, bottom-left or bottom-right corners). For local pre-cues, as well as for global pre-cues, the results seem to support gradual coarse-to-fine tuning levels. The localization cue (pre-cue) could initiate and run the action of the LSF filters ahead in time and prepare the HSF filters. If the following target provides fine facial information (12-16 pixels), the identification process will be facilitated, but if the stimulus only provides coarse facial information (9-11 pixels), the potentially useful information for identification will not be present, and the mechanism for tuning HSFs activated by the localization cue will only receive HSF information from the edges of the pixelization blocks. This will disrupt the identification process or mislead the subject. In fact, the results showed significant interaction between the level of pixelization and the cuing condition, which produced a drop in identification efficiency when stimuli were coarse pixelized.

Theoretical Discussion

It is a well-known fact that the first integrative stage (including impenetrable modular units) and early sensory SF integration work virtually in an automatic manner to build up the initial representation; but what happens inmediatly after this? How are SFs integrated in a second stage as carriers of the spatial information of a face?

The results of the experiments carried out within the context of the microgenetic approach would appear to contradict each other: some favor the hypothesis of anisotropic integration, whereas others point to a third interaction factor that might explain why one order of integration is used instead of another. This third factor could be the complexity of the stimulus or the focus of attention. However, all of these results can be explained using the diagnostic-recognition approach, mainly when it is understood that: (a) The experiments described in this section were biased to rule out the possible influence of knowledge structures; (b) the results of Hoeger's (1997) experiments gave empirical evidence of the effects of object information (which was manipulated by making the stimuli more complex), which is an essential element in the diagnostic-recognition approach, and (c) the results of Bachmann and Kahusk's (1997) experiments gave empirical evidence of task constraints (by manipulating selective attention), which are also essential elements of the diagnostic-recognition approach. Now we will develop each of these ideas to show how the results of the studies described in this section are compatible with an explanation based on the diagnostic-recognition approach.

From the diagnostic-recognition approach the use of SFs conveying configural or featural cues of facial image will depend on the task. If you are looking for the overweight members of the family, you may begin by integrating the LSFs, but if you are looking for the people who wear glasses, you may begin by integrating the HSFs. In short, depending on the information available in the image and the constraints of the task, SFs can be integrated from LSFs to HSFs or from HSFs to LSFs. However, when there are no special task demands, the natural sequence of integration tends to be from LSFs to HSFs.

Hoeger's (1997) research empirically shows the importance of the information furnished by the objects, i.e. the use of SFs at a specific moment. He does not use faces as stimuli, but his results can probably be applied to different kinds of images. One consequence is that specific perceptual tasks can be solved by starting either with LSFs or HSFs, which means there are two ways to reach the same goal. The subjects' task was to categorize the target stimuli as a motorcycle, tree, ant or hedgehog. The main result of Hoeger's study is that LSF primes more or less facilitate the categorization task, whereas HSF primes more or less facilitate the categorization task, depending on the complexity of the prime. These results can be explained in terms of the diagnostic-recognition approach. It was basic level categorization of very different objects and the task could be solved based on the overall shape of the stimuli, i.e. from the information provided by LSFs. The diagnostic-recognition approach predicts that the subjects will select the most diagnostic information for the kind of categorization, i.e. in this case, LSFs. And this is exactly what happens. If the LSFs are available, they are used to solve the task and greatly facilitate the process. But if LSFs are not available (because the prime contains only HSFs), the SFs available are used to obtain "derivate LSFs," which explains the fact that HSF primes for more complex stimuli provide for slower facilitation, because more time is needed to obtain the "derivate LSFs". This result is also compatible with the findings of Parker, Lishman and Hughes (1996).

Bachmann and Kahusk's (1997) study empirically showed the importance of task constraints, i.e. the manipulation of subjects' tendency to use one kind of SF over the other, depending on whether or not they were informed about the location of the target stimuli. If subjects had received no information about where the target stimuli would appear, they generally began processing LSFs, which provided for quick detection of the stimuli. But if the subjects had received information about where the target stimuli would appear, then HSF filters are ready to work as soon as the target appeared and provided more detailed information about the face. It can therefore be observed that the demands of the task determine the use of one kind of SF or the other.

In summary, all these results, together with the results obtained from different experiments about the categorization of natural scenes (Oliva & Schyns, 1997; Schyns & Oliva, 1994), indicate that the critical question for predicting subjects' performance, after the first integrative stage from LSFs to HSFs, is: which SFs provide the information required to solve the on-going task? According to the main thesis of this article, the diagnostic role of SFs provided as stimuli for the on-going task is the key to understanding the role of the different SFs, as will be shown in the next section (see Table 2).

How Are SFs Selected in a Visual Task? The Diagnostic-Recognition Approach

We saw above that the anisotropic point of view used to explain the second stage in early visual processes, i.e. that recognition should be coarse to fine, has not always been supported by empirical data. It is therefore possible to state that "…this scenario neglects one important aspect of any recognition task: the information demands of the considered categorization. … different categorizations of an identical picture could require different perceptual cues from the input. If these cues were associated with different spatial resolutions, then an identical stimulus might have been flexibly encoded at the scale that optimizes the information demands of the considered categorization." (Schyns & Oliva, 1997, p. 1029). However, to prove the possibility of a different use of LSFs and HSFs depending on the task demands, it was necessary to develop a procedure capable of showing the human visual system's flexibility when it comes to selecting different kinds of information. This implies that different categorizations of an identical stimulus can change the information scale, or SF, used, i.e. it implies the analysis of the interaction between task constraints (kind of categorization) and the availability of object information (SFs).

Experimental Paradigm

The procedure developed by Schyns and Oliva (1994) shows the visual system's flexibility when selecting SFs in an integrated image. This procedure can be summarized as follows: (a) using an image composed of: $I_h = I_{lf} + I_{hf}$, where I_h is a hybrid image obtained by superimposing the I_{lf} image (low-pass-filtered) on the I_{hf} image (high-pass-filtered); (b) using a very brief exposure time of the I_h image (between 30 and 150 ms); and (c) using a task to categorize the I_h image, which allows us to know whether the information from the I_{lf} image (and only this information) was used or whether the information from the I_{hf} image (and only this information) was used. This procedure is called the hybrid-pictures paradigm. The use of hybrid stimuli solves the problem of conventionally filtered stimuli where SFs could not compete. There are also two variants of this paradigm: (a) showing a hybrid image as a single stimulus (simultaneous exposure), and (b) showing the hybrid image as a sequence of images containing different filtering levels (sequential exposure). Thus, for example, to obtain a hybrid image composed of a female face and a male face, in case (a), 2-cycle/deg low-pass filtering can be done on the female face and 6-cycle/deg high-pass filtering on the male face and then both filtered faces can be digitally superimposed. In case (b), a three-stimuli hybrid may be used (< 2 + > 6, < 3 + > 5 and < 4 + > 4 cycles/deg) and shown at a fast exposure, usually at 45 ms per stimulus and without any ISI between them. (see Figure 4 for details)

Empirical Research

Schyns and Oliva (1997) hypothesized that different categorizations of the same stimulus changed the SFs selected. They carried out a group of three experiments, in which the task was to indicate the gender or expression of some faces. The results of all three experiments indicated that the information necessary to carry out each different task is in different SFs, in accordance with the idea that, for categorization, the SFs with the most informative value for that task are the ones chosen.

Schyns and Oliva (1999) also obtained evidence of a flexible use of the information from the SFs using hybrid stimuli (a face containing LSFs and a face containing HSFs) and the task was to categorize them in terms of gender, expression, look and identity. Taking into consideration that the information contained in a given SF is sufficient for a given categorization and that this does not necessarily imply that these SFs are selectively used when all SFs are present, the experiments were designed to study whether the information contained in any SFs is selectively accessed and used when information about other SFs is also present. The results showed significant bias in the categorization task by look, because the subjects used 38% of the faces in the hybrid containing LSFs. Significant bias was also obtained in the categorization task by expression, because the subjects used 66% of the faces in the hybrid containing LSFs. However, significant bias was not obtained in categorization by gender, where 52% of the faces in the hybrid containing LSFs were used.

Theoretical Discussion

The preceding experiments involving the diagnostic-recognition approach highlight three main features of face-perception research. First, the studies done within this framework have demonstrated that top-down processes (those manipulated through task demands) influence how the integrated image generated after the early perceptual representation is formed (Oliva & Schyns, 1997; Schyns, 1998; Schyns & Oliva, 1994, 1997, 1999).

Second, the experimental paradigms reviewed in the preceding pages (critical-band masking, increased exposure, backward masking and priming) were only useful for determining whether SFs were sufficient for recognition, which is very different from when SFs are used in everyday conditions. However, the hybrid-picture paradigm makes it possible to study selective access to specific SFs depending on task demands and this fact is more representative of everyday conditions, given that all SFs are available when we see images in the real world.

Finally, generalization of the results obtained should not be limited only to images of faces. Quite the contrary: the diagnostic-recognition approach was introduced as a general theory to explain categorization of every kind of visual stimuli (Gosselin & Schyns, 2002). Though it does not claim to do so, the theory offers an answer to the frequently asked question of whether or not human faces are processed in a qualitatively different way compared with any other kind of visual stimuli. Processing differences found in relation to object perception can be easily explained as a result of the interaction between task demands and image properties. In the case of face identification, subordinate categorization occurs for images with important configurational properties and a high degree of intersimilarity, while in the case of the identification of objects, though not for all of them, basic categorization occurs, and this consequently makes it necessary for different SFs to be selected to satisfy categorization requirements.

In summary, we can say that the diagnostic-recognition approach has led to a change in the perspective of research into mid-level visual processes from a bottom-up perspective to an interactive perspective between stimulus properties and task demands (see Gosselin & Schyns, 2002; Smith, Gosselin & Schyns, 2004). This means that studying the physical properties of an image (spatial and temporal) to understand how it is processed by the visual system is not enough. It is also necessary to include task demands, as we have tried to demonstrate in this article.

General Conclusions and Open Questions for Future Research

This review of the research on face perception, focused on studies involving spatial filtering, has shown that one of the most analyzed aspects has been the physical properties of images. In the masking approach, the aspects most frequently studied were the spatial effects of the representation of the face, while from the microgenetic approach, the temporal effects of face representation were the aspects most commonly studied. However, in both cases, the tests done to discover which SFs are necessary for face recognition have not produced conclusive results. The diagnostic-recognition approach, on the other hand, demonstrates that people tend to use the most diagnostic SFs for the visual task at hand. Though these studies were not focused on how face perception takes place, certain physical characteristics of faces led to these images being used as experimental stimuli and valuable data were obtained to help understand face perception. The main value of this approach is that it makes it necessary to take task demands into account even in the middle stages of visual processing.

In keeping with the thesis of this article, we have pointed out that when the results of the research into face perception carried out over the last thirty years are examined from the diagnostic-recognition approach, some of the contradictions disappear. And this is because questions like "Which SFs are critical?" and "In what order are SFs integrated?" should be considered within the framework of the demands of the task at hand. These questions should therefore be transformed into "Which SFs are diagnostic for recognition/identification of an image?" In order to build a complete model of face processing, there are three explanatory sources that cannot be ignored: object information, task constraints and the characteristics of the subject. With regard to image properties, we know there is a range of SFs (between 8 and 16 cycles/fw) that is especially important for face identification and the development of a face percept, though it was initially thought that this process took place by concentrating on coarse information and then moving on to fine details. We now know that task demands are important enough to determine the selection of specific SFs (high or low) to facilitate visual recognition as much as possible. In terms of subject characteristics, we know that the mental representation of faces (conceptualized in an SF format) and processing strategies (conceptualized as a preferential selection of LSFs over HSFs) are two subjects that research into face recognition should explore in the future. We will now present some new ideas about how these three explanations can be studied.

How Do Task Demands and Image Information Specifically Interact in Terms of the Use of SFs?

The role played by specific SFs in different perceptual tasks is not completely understood. In some detection tasks, such as when someone tries to see if there are portraits in a painting exhibit, coarse information like that supplied by LSFs is sufficient to determine whether or not there are faces in the paintings. In an identification task, such as when someone tries to name the members in a picture of a music group, it will be necessary to have fine information like that supplied by HSFs to determine the name of each member. However, in a categorization or discrimination task, deciding which SFs are necessary and/or sufficient for diagnosis is not quite so clear: is it done by comparing local elements in the image or based on the configuration of the image as a whole? Therefore, it seems necessary to begin researching in a context where task demands are manipulated for the same set of images so that the different use of SFs in the image can be analyzed. It would be of interest for these experiments to be designed with the preceding experimental paradigms in mind so that new results could be compared with those produced in recent decades.

How Can it be Explained that the Same Visual Task Can be Solved Using Different SFs?

The observed fact that certain perceptual tasks can be solved using different SFs (Sergent, 1985) makes it necessary to include another factor besides image properties and task demands to explain this phenomenon. This factor could be the subject’s characteristics as an observer, characteristics that result in differences between individuals and that can be classified in two categories: (a) the types of mental representation of faces, conditioned by the subject’s level of familiarity or expertise in relation to faces, and (b) a preferential strategy for visual processing, which is conditioned by the subject’s hemispheric dominance, specific learned skills and cognitive style. The subject’s characteristics should be taken into account because the use of some SFs rather than others seems to depend very critically on aspects (a) and (b). For different reasons, many studies do not allow these aspects to play a relevant role. Mental representation is usually a controlled variable: all subjects have a very similar representation, either due to their previous knowledge of the faces or the learning phase used to familiarize the subjects with the faces, whereas the preferential strategy for processing is a randomized extraneous variable whose influence does not affect the main objective of the experiment. Therefore, studying mental representation, as has been done by some researchers (Wenger & Townsend, 2000), and cognitive style (Ruiz-Soler, López, Pelegrina, Videra & Wallace, 2000) will probably provide two good ways of finding out how the observer's representational formats for faces and/or his/her preferential ways of processing them can affect how SFs in images are used differently.

What Empirical Evidence Supports the Idea of Considering Mental Representation and Processing Strategy as Being New Explanatory Factors?

Concerning mental representation, research into memory using faces as stimuli has reported different codification of faces depending on the amount of previous knowledge (Coin, Versace and Tiberghien, 1992; Liu, Collin, Rainville & Chaudhuri, 2000; O'Toole, Millward & Anderson, 1988; Tong & Nakayama, 1999; Wenger & Townsend, 2000). Furthermore, studies of cognitive development and face recognition have shown important differences between children and adults (Barrera & Maurer, 1981; Kuchuk, Vibbert & Bornstein, 1986; Mondloch et al., 1999). Finally, studies involving experts and novices using stimuli with perceptual characteristics very similar to faces (complex, symmetrical, 3D, intersimilars, etc.) have proven that there are different mental representations (Coin, Versace & Tiberghien, 1992; Harvey & Sinclair, 1985; Millward & O'Toole, 1986).

Concerning processing strategies, research taking into account hemispheric cerebral dominance (Keenan, Whitman & Pepe, 1989, 1990 and especially Ivry & Robertson, 1998) should be considered, as well as other works aimed at studying the development of expert skills in perceptual discrimination (Gauthier, Behrmann & Tarr, 1999; Gauthier & Logothetis, 2000; Gauthier & Tarr, 1997; Gauthier, Tarr, Anderson, Skudlarski & Gore, 1999; Gauthier, Williams, Tarr & Tanaka, 1998) and the reinterpretation of data from specific research studies on visual perception [3]. The results indicate that processing is linked to cognitive styles, in that some subjects are basically analytical (field-independent subjects) while others are basically holistic (field-dependent subjects). This is probably related to the fact that some subjects prefer to process HSFs, while others prefer LSFs. Although some previous studies have not shown the relationship between these two styles (Bruce, 1998), this is a field we have begun to explore after drawing up some procedural controls, in which we are looking for ways to classify field-dependent subjects but not merely by including the people who do not belong in the group of field-independent subjects, as has generally been done (Ruiz-Soler, López, Pelegrina, Videra & Wallace, 2000).

Finally, it is important to point out that although the conclusions of our work are focused on face perception, they can be generalized to very different visual objects, particularly those with important configurational properties (e.g. cars and animals) and those where different kinds of categorization are possible. In fact, the diagnostic-recognition approach has been introduced as a general theory for the visual identification/categorization of objects (Schyns, 1998).

References

Bachmann, T. (1987). Different trends in perceptual pattern microgenesis as a function of the spatial range of local brightness averaging. Psychological Research, 49, 107-111.

Bachmann, T. (1989). Microgenesis as traced by the transient paired-forms paradigm. Acta Psychologica, 70, 3-17.

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 (1), 87-103.

Bachmann, T. & Kahusk, N. (1997). The effects of coarseness of quantisation, exposure duration, and selective spatial attention on the perception of spatially quantised ("blocked") visual images. Perception, 26, 1181-1196.

Bachmann, T., Luiga, I. & Põder, E. (2004). Forward masking of faces by spatially quantized random and structured masks: On the roles of wholistic configuration, local features and spatial-frequency spectra in perceptual identification. Psychological Research, 69, 11-21.

Bachmann, T. & Põder, E. (2002). Forward masking of faces with three types of spatially quantised (pixelised) masks: evidence for configuration microgenesis. Perception, 31 (supplement), 22.

Bhatia, S.K., Lakshminarayanan, V., Samal, A. & Welland, G.V. (1995). Human face perception in degraded images. Journal of Visual Communication and Image Representation, 6, 280-295.

Barrera, M.E. & Maurer, M. (1981). The perception of facial expressions by the three-month-old. Child Development, 52 (1), 203-206.

Bruce, V. (1998, December). Personal communication.

Bruce, V. & Humphreys, G.W. (1994). Recognizing objects and faces. Visual Cognition, 1 (2/3), 141-180.

Bruce, V. & Young, A. (1986). Understanding face recognition. British Journal of Psychology, 77(3), 305-327 .

Burton, A.M., Bruce, V. & Hancock, P.J.B. (1999). From pixels to people: A model of familiar face recognition. Cognitive Science, 23(1), 1-31.

Calis, G., Sterenborg, J. & Maarse, F. (1984). Initial microgenetic steps in single-glance face recognition. Acta Psychologica, 55, 215-230.

Coin, C., Versace, R. & Tiberghien, G. (1992). Role of spatial frequencies and exposure duration in face processing: potential consequences on the memory format of facial representations. Cahiers de Psychologie Cognitive, 1, 79-98.

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

Costen, N. P., Parker, D. M. & Craw, I. (1994). Spatial content and spatial quantisation effects in face recognition. Perception, 23, 129-146.

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

Costen, N. P., Shepherd, J. W., Ellis, H. D. & Craw, I. (1994). Masking of faces by facial and non-facial stimuli. Visual Cognition, 1 (2/3), 227-251.

Desimone, R., Albright, T.D., Gross, C.G. & Bruce, C. (1984). Stimulus selective properties of inferior temporal neurons in the macaque. Journal of Neuroscience, 4, 2051-2062.

De Valois, R.L. & De Valois, K.K. (1988). Spatial vision. New York: Oxford University Press.

Ellis, A.W. (1992). Cognitive mechanisms of face processing. Philosophical Transactions of the Royal Society of London B, 335, 113-119.

Eriksen, C.W. & Schultz, D.W. (1979). Information processing in visual search: A continuous flow conception of experimental results. Perception & Psychophysics, 25, 249-263.

Field, D. J., Hayes, A. & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local "association field.". Vision Research, 33, 173-193

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

Gauthier, I., Behrmann, M. & Tarr, M.J. (1999). Can face recognition really be dissociated from object recognition? Journal of Cognitive Neuroscience, 11 (4), 349-370.

Gauthier, I. & Logothetis, N. (2000). Is face recognition not so unique after all?. Journal of Cognitive Neuropsychology, 17 (1-3), 125-142.

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 (6), 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). Visual information processing based on spatial filters constrained by biological data. Unpublished doctoral dissertation, University of Cambridge, Cambridge, England.

Gosselin, F. & Schyns, P.G. (2002). RAP: A new framework for visual categorization. Trends in Cognitive Sciences, 6 (2), 70-77.

Graham, N.V.S. (1989). Visual pattern analyzers. New York: Oxford University Press.

Gross, C.G., Bender, D.B. & Rocha-Miranda, C.E. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. Science, 166 (3910), 1303-1306.

Gross, C.G., Rocha-Miranda, C.E. & Bender, D.B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. Journal of Neurophysiology, 35, 96-111.

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

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

Harvey, L.O. & Sinclair, G.P. (1985). On the quality of visual imagery. Investigative Ophthalmology & Visual Science, Suppl. 26, 281.

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

Henning, G.B., Hertz, B.G., Broadbent, D.E. (1975). Some experiments bearing on the hypothesis that the visual system analyzes patterns in independent bands of spatial frequency. Vision Research, 15, 887-899.

Hess, R.F. & Dakin, S.C. (1997). Absence of contour linking in peripheral vision. Nature, 390, 602-604.

Hoeger, R. (1997). Speed of processing and stimulus complexity in low-frequency and high-frequency channels. Perception, 26, 1039-1045.

Hole, G. (1994). Configurational factors in the perception of unfamiliar faces. Perception, 23, 65-74.

Hughes, H.C., Fendrich, R. & Reuter-Lorenz, P.A. (1990). Global versus local processing in the absence of low spatial frequencies. Journal of Cognitive Neuroscience, 2, 272-282.

Hughes, H.C., Nozawa, G. & Kitterle, F. (1996). Global precedence, spatial frequency channels, and the statics of natural images. Journal of Cognitive Neuroscience, 8, 197-230.

Ivry, R.B. & Robertson, L. (1998). The two sides of perception. Massachusetts, NJ: The MIT Press.

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

Kimchi, R. (1992). Primacy of holistic processing and global/local paradigm. Psychological Bulletin, 112, 24-38.

Kuchuk, A., Vibbert, M. & Bornstein, M. H. (1986). The perception of smiling and its experiential correlates in three-month-old infants. Child Development, 57 (4), 1054-1061.

LaGasse, L. L. (1993). Effects of good form and spatial frequency on global precedence. Perception & Psychophysics, 53, 89-105.

Lander, K., Bruce, V., & Hill, H. (2001). Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. Applied Cognitive Psychology, 15, 101-116.

Legge, G.E. & Gu, Y. (1989). Stereopsis and contrast. Vision Research, 29, 989-1004.

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

Marr, D. & Hildreth, E.C. (1980). Theory of edge detection. Proceedings of the Royal Society of London B, 207, 187-217.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. San Francisco: Freeman.

Marshall, J.A., Burbeck, C.A., Ariely, J.P., Rolland, J.P. & Martin, K.E. (1996). Occlusion edge blur: A cue to relative visual depth. Journal of the Optical Society of America A, 13, 681-688.

McKone, E., Martini, P. & Nakayama, K. (2001). Categorical perception of face identity in noise: A method for isolating configural processing. Journal of Experimental Psychology: Human Perception & Performance, 27, 573-599.

McSorley, E. & Findlay, J. M. (1999). An examination of a temporal anisotropy in the visual integration of spatial frequencies. Perception, 28, 1031-1050.

McSorley, E. & Findlay, J.M. (2002). Are spatial frequencies integrated from coarse to fine? Perception, 31 (5), 955-967.

Millward, R.B. & 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 (pp. 34-44). Dordrech, The Netherlands: Martinus Nijhoff.

Mondloch, C.J., Lewis, T.L., Budreau, D.R., Maurer, D., Dannemiller, J.L., Stephens, B.R. & Kleiner-Gathercoal, K.A. (1999). Face perception during early infancy. Psychological Science, 10 (5), 419-422.

Morgan, M.J. (1992). Spatial filtering precedes motion detection. Nature, 355, 344-346.

Morrison, D.J. & Schyns, P.G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. Psychonomic Bulletin & Review, 8 (3), 454-469.

Nachmias, J., Sansbury, R., Vassiley, A. & Weber, A. (1973). Adaptation to square-wave gratings: in search of the elusive third harmonic. Vision Research, 13, 1335-1342.

Nachmias, J. & Weber, A. (1975). Discrimination of simple and complex gratings. Vision Research, 15, 217-223.

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

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. Cognitive Psychology, 9, 353-383.

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

Oliva, A. & Schyns, P. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. Cognitive Psychology, 34, 72-107.

O'Scalaidhe, S.P., Wilson, F.A.W. & Goldman-Rakic, P.S. (1997). Areal segregation of face-processing neurons in prefrontal cortex. Science, 278 (5340), 1135-1138.

O'Toole, A., Millward, R.B. & Anderson, J.A. (1988). A physical system approach to recognition memory for spatially transformed faces. Neural Networks, 1, 179-199.

Parker, D.M. & Costen, N.P. (1999). One extreme or the other or perhaps the golden mean? Issues of spatial resolution in face processing. Current Psychology: Development, Learning, Personality, Social, 18 (1), 118-127.

Parker, D.M., Lishman, J.R. & Hughes, J. (1992). Temporal integration of spatially filtered visual images. Perception, 21, 147-160.

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 and Performance, 22, 1448-1466.

Parker, D.M., Lishman, J.R. & Hughes, J. (1997). Evidence for the view that temporospatial integration in vision is temporally anisotropic. Perception, 26, 1169-1180.

Polat, U. & Sagi, D. (1993). Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. Vision Research, 33, 993-999.

Polat, U. & Sagi, D. (1994). The architecture of perceptual spatial interactions. Vision Research, 34, 73-78.

Rakover, S.S. & Cahlon, B. (2001). Face recognition: Cognitive and computational processes. Amsterdam/Philadelphia: John Benjamins.

Rolls, E.T. (1992). Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. In V. Bruce, A. Cowey, W. Ellis & D.I. Perrett (Eds.), Processing the facial image (pp. 11-21). Oxford, England: Oxford University Press.

Ruiz-Soler, M., López, E., Pelegrina, M., Videra, A. & Wallace, A. (2000, July). Face recognition and response criterion: Moderating effects of cognitive style. Paper presented at XXVII International Congress of Psychology, Stockholm, Sweden.

Sanocki, T. (1993). Time course of object identification: Evidence for a global-to-local contingency. Journal of Experimental Psychology: Human Perception and Performance, 19, 878-898.

Schyns, P. G. (1998). Diagnostic recognition: task constraints, object information and their interactions. Cognition, 67, 147-179.

Schyns, P. G. & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale dependent scene recognition. Psychological Science, 5 (4), 195-200

Schyns, P. & Oliva, A. (1997). Flexible diagnosticity, rather than fixed, perceptually determined scale selection in scene and face recognition. Perception, 26, 1027-1038.

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

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. British Journal of Psychology, 75 (2), 221-242.

Sergent, J. (1985). Influence of task and input factors on hemispheric involvement in face processing. Journal of Experimental Psychology: Human Perception and Performance, 11, 846-861.

Sergent, J. (1986). Microgenesis in face perception. In H.D. Ellis, M.A. Jeeves, F. Newcombe & A. Young (Eds.), Aspects of face processing (pp. 17-73). Dordrecht, The Netherlands: Martinus Nijhoff.

Sergent, J. (1994). Brain-imaging studies of cognitive functions. Trends in Neurosciences, 17(6), 221-227.

Sinha, P. (2002). Recognizing complex patterns. Nature Neuroscience. Supplement, 5, 1093-1097.

Smith, M.L., Gosselin, F., & Schyns, P.G. (2004). Receptive fields for flexible face categorizations. Psychological Science, 15 (11), 753-761.

Tanaka, J. & Farah, M. (1993). Parts and wholes in face recognition. Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 46A (2), 225-245.

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

Tolhurst, D.J. & Barfield, L.P. (1978). Interactions between spatial frequency channels. Vision Research, 18, 951-958.

Tong, F. & Nakayama, K. (1999). Robust representation for faces: evidence from visual research. Journal of Experimental Psychology: Human Perception and Performance, 25, 1016-1035.

Uttal, W. R., Baruch, T. & Allen, L. (1997). A parametric study of face recognition when image degradations are combined. Spatial Vision, 11 (2), 179-204.

Watt, R. J. (1987). Scanning from coarse to fine spatial in the human visual system after the onset of a stimulus. Journal of Optical Society of America A, 4, 2006-2021.

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

Wenger, M.J. & Townsend, J.T. (2000). Spatial frequencies in short-term memory for faces: A test of three frequency-dependent hypotheses. Memory & Cognition, 28, 125-142.

Westheimer, G. (2001). The Fourier theory of vision. Perception, 30, 531-541.

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

Footnotes

[1] In face-recognition research, the unit of measure is the number of cycles per face width (cycles/fw), i.e. “the number of sinusoidal repetitions of a given width that can be placed within the eye-level width of the face” (Costen, Parker & Craw, 1996; p. 602). However, some authors give measurements in cycles per degree of visual angle (cycles/deg), and sometimes the information available makes it impossible to provide an exact conversion of cycles/deg into cycles/fw. We therefore provide measurements in cycles/fw whenever possible; otherwise, we give them in cycles/deg.

[2] This framework is also used for some applied proposals. Bhatia, Laksminarayanan, Samal & Welland (1995) degraded images of faces by pixelization in order to find the minimum image quality required for retrieval of facial records in databases. More recently, Lander, Bruce & Hill (2001) used both pixelization and blurring to mask the identity of familiar faces in short naturalistic television clips.

[3] Oliva & Schyns’ second experiment (1997) describes the results of the control group that was exposed to 12 stimuli made up of LSFs and HSFs. Of these 12 subjects, none perceived that there were two scenes in the same stimulus. Surprisingly, four of them were “HSF categorizers”, four were “LSF categorizers” and four successfully categorized using both types. Since there was no sensitivity phase in this group in order to bias subjects toward LSF or HSF processing, these results are compatible with the interpretation that, in the absence of specific task demands, there are subjects whose preferential strategy is to use LSFs and others whose preferential strategy is to use HSFs.

Table 1
Distinguishing features of the different approaches in face-recognition research

Approach	Objective	Research fields	Main research subjects
Cognitive	Which structures and processes make face recognition possible?	Stimulus variables	How do size, exposure time, orientation, color, etc. affect face recognition?
		Subject variables	How do age, long-term memory, cognitive style, etc. affect face recognition?
		Stages in the process	What are the stages in the face-recognition process?
Psychophysical	What effect do the SFs in an image of a face have on recognition?	Image masking	What SFs are necessary and/or sufficient for face recognition?
		Perceptual microgenesis	In what order are SFs integrated during the face-recognition process?
		Diagnostic recognition	What SFs are diagnostic for a specific categorization task?
Neurophysiological	What are the underlying biological mechanisms of face recognition?	Microanalysis (neural specialization)	Does the inferotemporal region specialize exclusively in face recognition?
		Macroanalysis (hemispheric specialization)	How are the cerebral hemispheres involved when face perception is going on?
		Neuropsychology (prosopagnosia)	What does prosopagnosia tell us about the underlying neural mechanisms of face recognition?

Table 2

Summary of the main results on face perception using spatial frequencies. Papers are listed by research approach and publication year. Since different topics are often studied in the same paper, the articles are classified according to the approach associated with the main ideas of our present article.

	Main objective	Exp. [a]	Image properties	Task demands [b]	Conclusions	Compatibility
Harmon & Julesz (1973)[c]	What is the role of spatial frequencies in recognition?	3	Band-pass and low-pass filtering. A block portrait of President Lincoln.	Level of identification of President Lincoln.	Spatial frequencies play a key role in the recognition of faces.	Masking approach
Tieger & Ganz (1979)	Is there a critical range of spatial frequencies?	1	Full bandwidth and grid masking. Unknown faces.	Recognizing faces in a classic recognition paradigm.	The critical spatial frequency for face recognition is about 17.6 cycles/fw [d].	Masking approach
Fiorentini, Maffei & Sandini (1983)	What is the role of high spatial frequencies in recognition?	1	Full bandwidth and high- and low-pass filtering. Unknown faces.	Identification (naming).	High spatial frequencies contain sufficient information to produce visual recognition.	Masking approach
Hayes, Morrone & Burr (1986)	What is the specific role of low and high spatial frequencies using positive and negative images?	1	Full bandwidth and high- and low-pass filtering. Unknown faces. Negative and positive images. Three rotation angles.	Identification (naming) while full bandwidth faces were shown on a panel.	The most useful spatial frequencies for face recognition are those over 20 cycles/fw.	Masking approach

Costen, Shepherd, Ellis & Craw (1994)	Does the masking process that makes face recognition impossible intervene in the face-processing system?	3	Full bandwidth. Masks containing normal faces, inverted faces, disorganized faces, faces without internal elements, objects and noise. Known famous faces.	Naming.	Normal face masks produce the highest masking because they hide configurational information about the face.	Masking approach
Costen, Parker & Craw (1994)	What are the accurate bounds for the critical range of spatial frequencies?	2	Full bandwidth, quantizing, low-pass filtering, blurring and jumbling. Unknown faces.	Naming.	The range of medium spatial frequencies contains the most efficient information for recognizing faces.	Masking approach
Costen, Parker & Craw (1996)	What are the accurate bounds for the critical range of spatial frequencies?	2	Full bandwidth, quantizing, low-pass and high-pass filtering. Unknown faces.	Naming.	The range of spatial frequencies between 8 and 16 cycles/fw contains the most efficient information for recognizing faces.	Masking approach
Uttal, Baruch & Allen (1997)	Does filtering following pixelization improve face recognition?	8	Full bandwidth, quantizing and low-pass filtering. Unknown faces.	Naming while full-bandwidth faces were shown on a panel.	The perceptual organization is as important for recognition as the energetic distribution of the spatial frequencies.	Masking approach

Parker & Costen (1999)	Do faces containing only medium-range spatial frequencies produce better recognition than faces containing only spatial frequencies from the extremes?	1	Full bandwidth. Band-pass filtering. Five rotation angles. Unknown faces.	Naming.	Although medium-range spatial frequencies allow for better recognition, lower or higher spatial frequencies are also useful for acceptable recognition.	Masking approach
Näsänen (1999)	Is the visual system sensitive to different spatial frequencies when recognizing faces?	4	Full bandwidth. Narrow-band noise masks. Band-pass filtering. Unknown faces. Synthetic faces.	Pointing to and clicking an array of buttons with the mouse to indicate previously learned faces.	The main information is between 8 and 13 cycles/fw (a band-width below 2 octaves), but the sensitivity function showed long tails.	Masking approach
Calis, Sterenborg & Maarse (1984)	Does more specific classification increase in the first phases of seeing an object?	1	Full bandwidth. ¾ position.	Transient paired forms.	There is a set of temporary brief sub-stages in the face-recognition process, which shows a hierarchic sequential process.	Microgenetic approach
Bachmann (1987) e	Are any spatial resolutions and temporal exposures critical for the visual integration process?	1	Quantizing.	Writing the name of the object shown on a screen on a response sheet.	There is anisotropic integration from global to local properties.	Microgenetic approach

Bachmann (1989) [f]	Is classification more specific in the first phases of seeing and classifying an object?	2	Full bandwidth. Frontal position.	Transient paired-forms.	A microgenetic half-cycle shows a time limit between 50 and 70 ms.	Microgenetic approach
Bachmann (1991)	Are any spatial resolutions and temporal exposures critical for the visual-integration process?	2	Quantizing.	Naming.	There is anisotropic-integration from global to local properties.	Microgenetic approach
Parker, Lishman & Hughes (1992) [g]	Does the integration of visual images follow an anisotropic order?	5	Full bandwidth. Low-pass and high-pass filtering.	Rating the quality of an image from 1 to 4 by pressing a key.	There is anisotropic integration from global to local properties.	Microgenetic approach
Parker, Lishman & Hughes (1996)	Which model explains the early stages of face recognition better: pre-processing or coarse processing?	4	Full bandwidth. Low-pass and high-pass filtering.	Face-matching.	The pattern of results did not clearly support only one approach (pre-processing or coarse processing).	Microgenetic approach
Parker, Lishman & Hughes (1997) [h]	Is the integration of visual images anisotropic?	2	Full bandwidth. Low-pass and high-pass filtering.		There is anisotropic-integration from global to local properties.	Microgenetic approach

Bachmann & Kahusk (1997)	Does selective attention by pre-cues modulate the use of SFs in face processing?	2	Full bandwidth. Quantizing. Unknown male faces shown in four places (upper-left, upper-right, bottom-left, bottom right).	Naming of faces that were placed on the wall during the experiment.	Selective attention interacts with the face-recognition process.	Microgenetic approach
Hoeger (1997) [i]	Do low spatial frequencies compete with high spatial frequencies?	1	Drawings of objects. Quantizing. Low-pass and high-pass filtering.	Naming objects that were previously associated with an object.	The processing speed of high spatial frequencies depends on the level of image quality.	Microgenetic approach
McSorley & Findlay (1999) [j]	Which temporary model does the visual integration of spatial frequencies follow: the anisotropic or isotropic model?	3	Full bandwidth.	Naming.	The visual integration of spatial frequencies follows a time-based anisotropic model.	Microgenetic approach
Schyns & Oliva (1997)	Do different categorizations of the same stimuli change the spatial frequencies selected?	3	Low-pass and high-pass filtering. Hybrid faces.	Categorizing faces by gender and expression that were or were not previously associated with a name.	Each categorization chooses the spatial frequencies with the most informative value for the task.	Diagnostic-recognition approach

Schyns & Oliva (1999)	Does categorization flexibly modify the preference for spatial frequencies in face perception?	3	Low-pass and high-pass filtering. Hybrid faces.	Categorizing faces by gender, expression, look and identity.	Task demands change the preference for spatial frequencies in face perception.	Diagnostic-recognition approach

Notes. [a] The number of experiments. [b] Though there have been many differences in procedure, task demands before the diagnostic-recognition approach always involved one of the following: identification (by naming, i.e. assigning a name, letter or number, or pressing a key previously associated with a face shown to the subject) or recognition (by matching). [c] No experimental subjects were used in this study because it was more a visual demonstration than an empirical experiment. [d] The metrics reported in this study were cycles/deg, so 2.2 cycles/deg was approximately 17.6 cycles/fw. [e] Only one of the stimuli used in this study was a face. [f] This study only used faces as stimuli in the first experiment. [g] This study used scenes as stimuli in the first experiment and faces in the second. [h, i, j] These studies did not use faces as stimuli, but their results are closely related to the face-processing theory.

Figure Captions

Figure 1. Examples of the visual stimuli in face-recognition research based on the role of different SFs: (a) full bandwidth image, (b) low-pass image, (c) high-pass image; and (d) band-pass image. The faces used for experiments are usually square images (256 x 256 or 512 x 512 pixels) with 256 gray levels in order to facilitate calculation of the Fast Fourier Transform.

Figure 2. Examples of the visual stimuli used in face-recognition research based on masking procedures: (1) pixelization (a) large (b) small; (2) noising (a) high (b) low; (3) gridding (a) few cycles/deg (b) many cycles/deg; (4) filtering (a) low-pass (b) high-pass. Visual masking increases the recognition threshold for perceptual objects by adding another object or noise. Such masking reaches the maximum level when the target and the mask are structurally similar, i.e. when they share edges, orientation or outlines.

Figure 3. Two of the main experimental paradigms in the microgenetic approach. Two features characterize early visual processing: (a) processing of the entry information begins before the precept is fully developed, hence some cognitive processes can only use information based on the first SFs processed, and (b) the visual integration of stimuli does not end when the stimuli are removed, which is why backward masking can effectively keep subjects from capturing information from SFs that should be integrated later. Each experimental paradigm has an advantage for visual-processing research: the perceptual-priming paradigm indicates the order in which SFs are integrated when images filtered using different SFs are used; and the backward-masking paradigm indicates the amount of information integrated before a time limit is reached if delayed images filtered using different SFs are shown while the target stimuli are being processed.

Figure 4. The main experimental paradigm based on hybrid faces in the research following the diagnostic-recognition approach. The fine spatial scale (HSF) represents a female face, while the coarse spatial scale (LSF) represents a male face. To see the LSF face in the hybrid image, it is necessary to squint, blink or step back from the picture.

(a) (b) (c) (d)

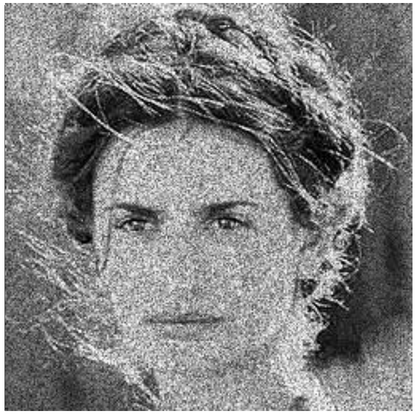
(1 a)

(2 a)

(3 a)

(4 a)

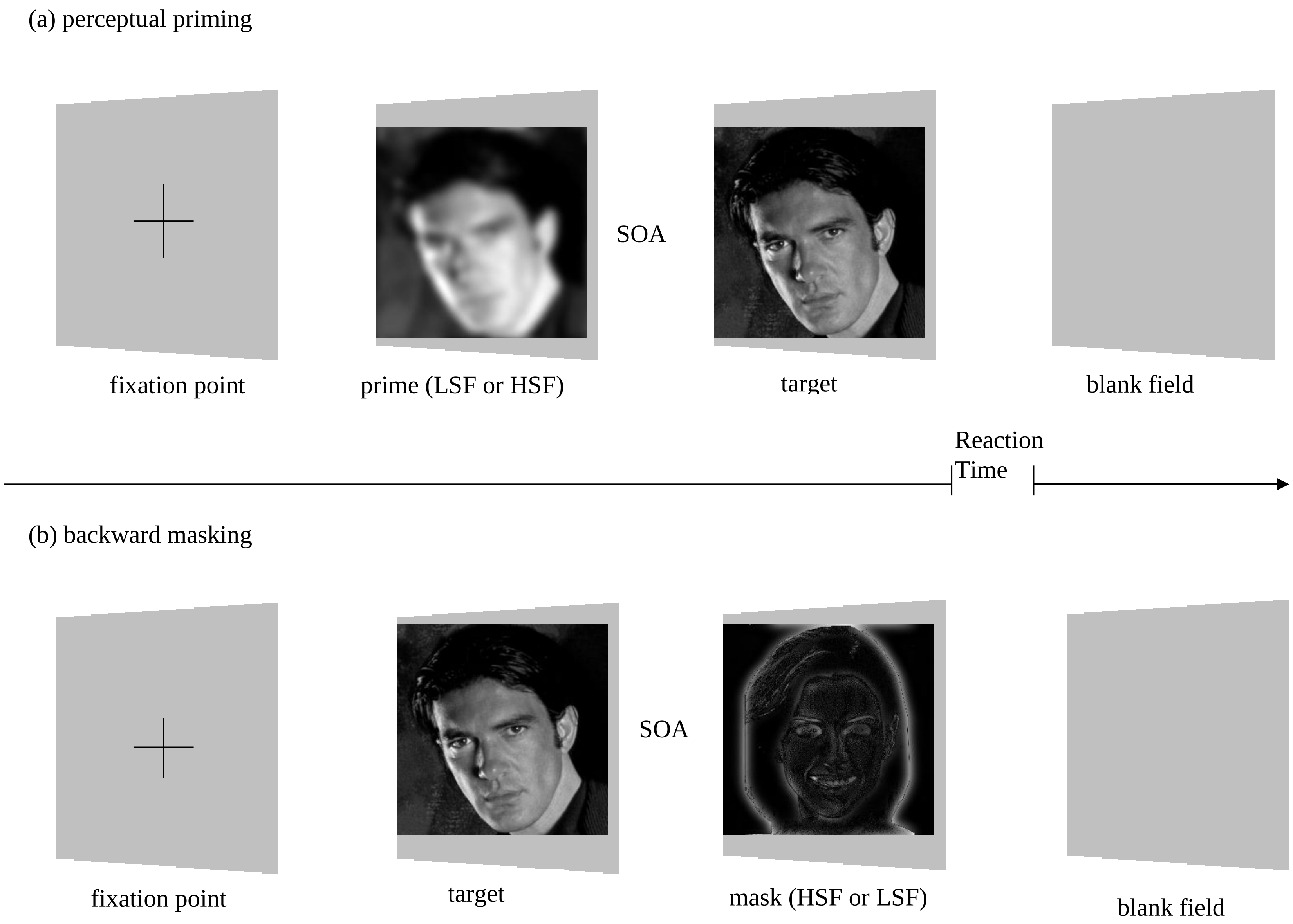
(1 b)

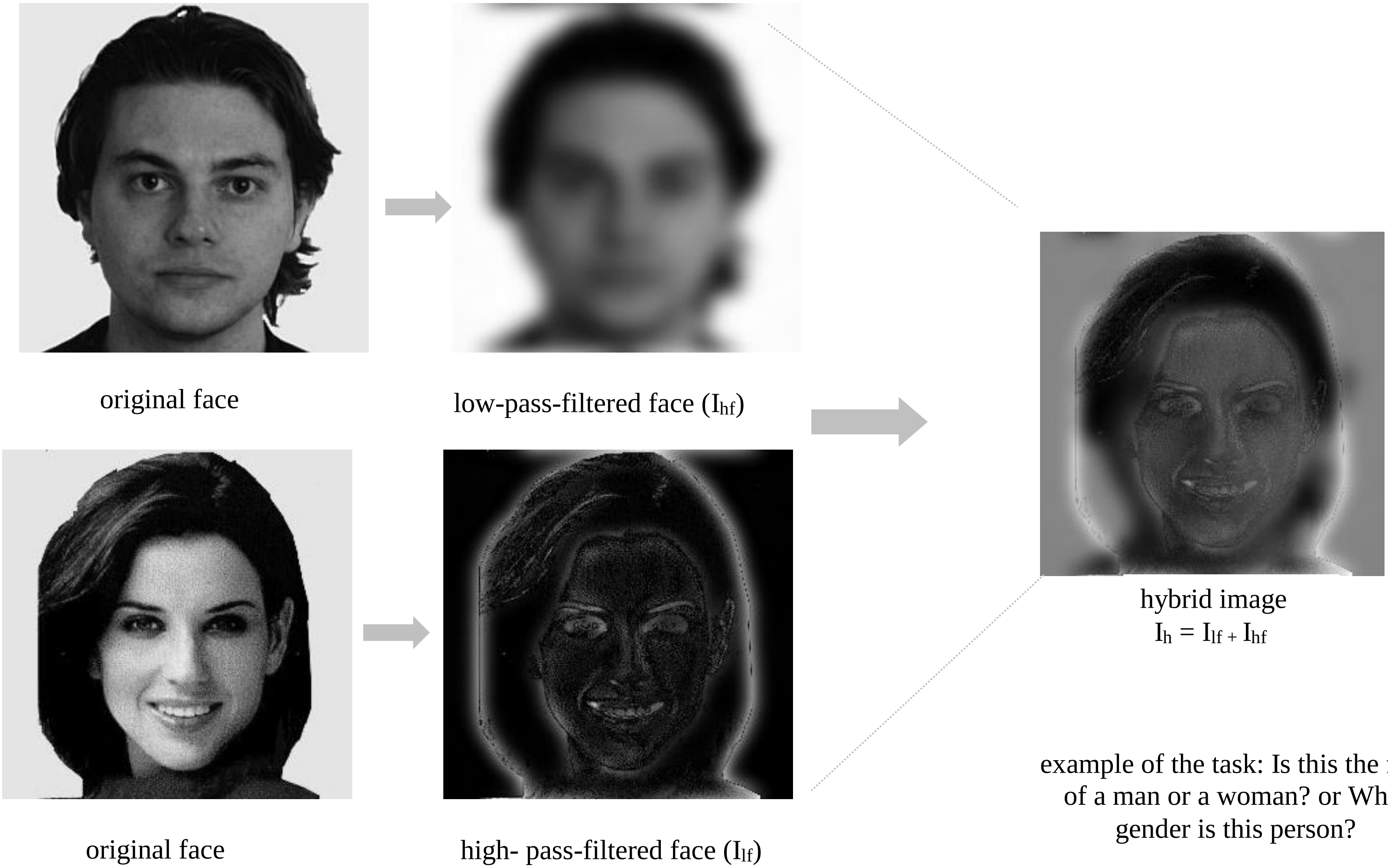


(2 b)

(3 b)

(4 b)


(a) perceptual priming
SOA
fixation point
prime (LSF or HSF)
target
blank field
Reaction
Time
(b) backward masking
SOA
fixation point
target
mask (HSF or LSF)
blank field



example of the task: Is this the face of a man or a woman? or What gender is this person?