Cognitive Science 28 (2004) 289–301
COGNITIVE SCIENCE
http://www.elsevier.com/locate/cogsci

Spatio-temporal dynamics of face recognition in a flash: it's in the eyes

Céline Vinette, Frédéric Gosselin [a,*], Philippe G. Schyns [b]

[a] *Département de psychologie, Université de Montréal, 90 Vincent-d'Indy, C.P. 6128 succ. Centre-ville, Montréal, Que., Canada H3C 3J7*
[b] *Department of Psychology, University of Glasgow, Glasgow G12 8QB, UK*

Received 30 May 2003; received in revised form 8 January 2004; accepted 12 January 2004

Abstract

We adapted the *Bubbles* procedure [Vis. Res. 41 (2001) 2261] to examine the effective use of information during the first 282 ms of face identification. Ten participants each viewed a total of 5100 faces sub-sampled in space–time. We obtained a clear pattern of effective use of information: the eye on the left side of the image became diagnostic between 47 and 94 ms after the onset of the stimulus; after 94 ms, both eyes were used effectively. This preference for the eyes increased with practice, and was not solely due to the informativeness of the eyes for the task at hand. The bias for the eye on the left side of the image is explained in terms of hemispheric specialization. Although there were individual differences, most participants exhibited this pattern of effective use of information. An intriguing finding is that most participants displayed a clear sinusoïdal modulation of effective use of attention through time with a frequency of about 10.6 Hz.

Keywords: Face recognition; Attention; Bubbles

1.

The doorbell rings, but no visitor is expected. As you walk towards the door, you wonder who this might be: a male or a female, someone neutral, happy or even angry, a friendly, an aggressive person, someone beautiful, the postman, your neighbour, maybe even your in-laws?

* Corresponding author.
E-mail address: frederic.gosselin@umontreal.ca (F. Gosselin).

0364-0213/$ – see front matter © 2004 Cognitive Science Society, Inc. All rights reserved.
doi:10.1016/j.cogsci.2004.01.002

As you open the door you recognize in a flash the face of your good old friend Georges, his convivial friendly face and characteristic broad smile. Such effortless face recognition plays a crucial adaptive role in our everyday lives. The main aim of this paper is to better understand the temporal dynamics of the information extracted over the first 282 ms of visual processing of face recognition.

Face recognition "in a flash" is a puzzle for vision researchers: Compared to other objects all faces are very much alike, sharing the same parts organized in a similar configuration. Of course there must be visual information available to distinguish faces, otherwise we would not be able to identify them. The main point is that these differences are subtle and require a system endowed with considerable perceptual expertise to extract them (Gauthier & Tarr, 1997; Tanaka & Farah, 1993; Tanaka & Gauthier, 1997). Speeded recognition is even more surprising when we consider that the visual system samples visual information with saccadic eye movements, and so does not access to the full information contained in the external world (O'Regan, 1992, *Change Blindness* provides powerful demonstrations: e.g., Archambault, O'Donnell & Schyns, 1999; Rensink, O'Regan, & Clark, 1997; Simons & Levin, 1997). Over the first 282 ms of visual processing, the visual system can only sample external visual information with one to two saccades, each of which provide a diameter of about 4–6° of visual angle of full resolution information (Henderson & Hollingworth, 2003).

In sum, the dual constraints of high object similarity and sparse information sampling suggest that face recognition in a flash requires visual routines which, through phylogeny and/or ontogeny, have embedded knowledge of how to extract information most efficiently from face stimuli (Ullman, 1984). In this paper, we apply Gosselin and Schyns (2001) *Bubbles* method to characterize this information extraction routine. We consider simultaneously two aspects of the routine that have not been considered together in the recognition literature (but see Ringach & Shapley, this issue, as well as Tse, this issue, for similar applications in other domains). The first aspect is location of facial features in image space, and the second aspect is the sequential use of features over the time course of recognition.

1.1. Why use Bubbles?

To study the extraction of visual information in space–time, we have adapted *Bubbles* to the time dimension. Imagine a face revealed by an animated sequence of masks very similar to successive thin slices of cheese cut from a brick of Emmenthal. Fig. 1 shows one stimulus. To characterize the space–time use of information of observers, we established a relationship between their face identification responses with the space–time filters used in the experiment (see Section 2 for details).

The advantages of this method over others are twofold. First, we explicitly control in space (i.e., over the face region), and in time (i.e., the latency following onset) the face information that the observer sees. This information is whatever can be captured by a number of 3D Gaussians with fixed a space–time sigma. On any trial, if the observer can identify a face with this particular space–time filter, it is likely that the filter revealed the diagnostic region of the face for identification, at the right time. If the observer could not identify the face, then the

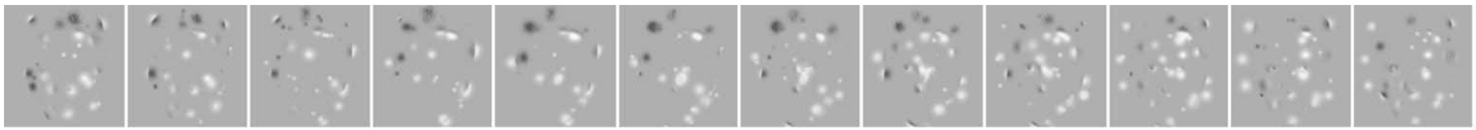

Fig. 1. The 12 frames of a sample movie stimulus. Each frame was presented for about 24 ms, and thus the movie lasted about 282 ms.

diagnostic information was not presented in one of the bubbles, or it was presented in a bubble, but with an offset in time for the visual routine. To the extent that this information display will control face recognition performance, this method enables a characterization of the observer's space–time use that controls performance.

This link between information sampling and performance is particularly important when the respective merits of *Bubbles* and eye movements are compared. Eye movements also sample visual information and the tracking of saccades over time should allow a characterization of a recognition visual routine. Indeed, there is a strong correlation between scan paths and the use of information revealed by the *Bubbles* technique (Pearson, Henderson, Schyns, & Gosselin, 2003). However, *Bubbles* adds the dimension of enabling the dissociation between the automatic versus diagnostic attention to features (Schyns, Jentzsch, Johnson, Schweinberger & Gosselin, 2003; Smith, Gosselin, & Schyns, in press). This simply arises from the observation that there is an explicit link between information sampling and performance in *Bubbles* (Gosselin & Schyns, 2002) whereas there is no explicit knowledge of what observers do with the visual information they have layed their eyes on. Another advantage over eye movement recording is sampling rate. It is now well established that shifts of attention occur when observers fixate a given image location, enabling information extraction within the 140 ms interval between saccades. It is estimated that an attentional shift takes 50–60 ms (Saarinen & Julesz, 1991; Treisman & Gelade, 1980). Analyses of information use with eye movements are bound by a 7–8 Hz sampling rate (7–8 saccades per second) when *Bubbles* in time can apply a sampling rate that matches the fusion sampling frequency (of at least 30 Hz) of the visual system. Yet another advantage of *Bubbles* over eye movements is that the former can reveal a scattered effective use of visual information at any particular time but not the latter because the eyes can only fixate one thing at a time.

The second main advantage of *Bubbles* is that it is the only generic technique that reveals the subset of *available* information that determines the performance of the observer—we call this construct *potent* information, Gosselin & Schyns (2002, 2004). Reverse correlation has also been adapted to reveal the use of information through time in low-level vision (see Neri & Heeger, 2002; Ringach, this issue), but as argued elsewhere (Gosselin & Schyns, 2002), reverse correlation, renders the information *represented* in memory, which is typically more than the stimulus information used in the task. Our interest in potent information stems from the fact that it provides a better measure of visual fitness to the statistics of input information (see Gosselin & Schyns, 2004; Murray & Gold, 2004; Smith, Cottrell, Gosselin, & Schyns, submitted, for discussions).

In sum, we developed and applied *Bubbles* in the space–time domain to characterize the unfolding of visual attention over facial features when observers are instructed to recognize the input faces. From the outset, it is important to emphasize that even though our results will inform the nature of face identification processes, they might sidestep the processes thought by many to precede identification. Our observers know that the experiments only consist of face stimuli in a vacuum, rather than the more naturalistic condition of faces inserted in a natural background. However, as pointed out in the example above, understanding the processes of face identification, even under conditions of reduced uncertainty, presents a significant challenge for vision scientists.

2. Methods

2.1. Participants

Ten paid students from the Université de Montréal with normal or corrected to normal vision participated to this experiment.

2.2. Stimuli

The experiment ran on a Macintosh G4 computer using a program written with the Psychophysics Toolbox for Matlab (Brainard, 1997; Pelli, 1997). The stimuli were presented on a calibrated high-resolution Sony monitor, with a refresh rate of 85 Hz.

A chinrest was used to maintain a constant viewing distance of 100 cm. Stimuli were movies subtending $5.72 \times 5.72^{\circ}$ of visual angle (256×256 pixels) on the screen, and of duration 282 ms (12 frames, each presented for 2 refresh cycles).

Stimuli were constructed from a set of 30 faces (from the grayscale faces data set of Schyns & Oliva, 1999; available from http://www.mapageweb.umontreal.ca/space-time.html): five female faces and five male faces, each displaying three different expressions (neutral, happy and angry). The faces were normalized for hairstyle and global orientation. Also, they were lighted by a standardized light source located on the far right. Because face recognition is sensitive to lighting direction (Braje, Kersten, Tarr, & Troje, 1998), we randomly mirror-reversed the faces on half the trails. This manipulation does not impair face recognition (Brooks, Rosielle, & Cooper, 2002).

During the experimental phase, we sparsely and randomly sampled these static faces in space–time with a certain number of Gaussian "bubbles". Each bubble had a standard deviation of 0.22° of visual angle (45 pixels) in the spatial domain and of 39 ms (1.65 frame) in the temporal domain (see Fig. 1; a movie version of Fig. 1 is available from http://www.mapageweb.umontreal.ca/space-time.html).

2.3. Procedure

The first of 10 sessions began by a learning phase: Participants learnt the 10 identities of the 30 faces for about 15 min. They were then submitted to an identification task during which faces were briefly presented (332 ms) on the screen and followed by a white Gaussian noise mask. Feedback was provided. This learning phase, presented in blocks of 20 trials, ended as soon as subjects attained the criterion of 90% or more correct responses on a block. The learning phase lasted about 20 min or 15 blocks.

The experimental phase consisted in the presentation of 510 dynamically bubbled faces (i.e., 17 presentations of the 30 faces) (see stimuli section for details). The inter-stimuli interval was equal to about 2 s. Participants had to identify the sparsely sampled faces as accurately as possible. To respond, they pressed one of 10 labelled computer-keyboard keys. They were under no time pressure. Again, feedback was provided. The number of space–time bubbles per trial—and thus the proportion of the faces revealed—was adjusted online by a gradient descent procedure to maintain performance at 75% correct. The experimental phase lasted about 45 min.

2.4. Results

On average, participants required 142 bubbles distributed along the temporal stimulus to reach the 75% performance criterion, but by the end of the 10 blocks, only 84 bubbles were required. On average, it took participants 1.32 s per trial ($SD = 0.34$ s per trial) to respond. We rendered the relative diagnosticity of the space–time voxels for face identification for each observer and for each particular experimental block as follows. We first computed the sum of the bubble masks that led to a correct response and another sum of the bubble masks that led to an incorrect response. Then, we subtracted the latter from the former. This procedure amounts to performing a linear multiple regression on the spatio-temporal bubble masks (the explanatory variables) and response accuracy (the predictor variable) (Gosselin & Schyns, 2001). If all regions of the search space were equally effective at determining the accuracy of the response, the regression coefficients would be uniform. The observed divergence from uniformity thus reveals how informative a particular region of space–time is relative to other regions. To establish statistical divergence, we normalized the regression coefficients by transforming them into *Z*-scores.

The 12 frames of Fig. 2a, MEAN (their movie version are available from http://www.mapageweb.umontreal.ca/space-time.html), are the product of this analysis for all observers and all blocks pooled together. The *Z*-scored space–time regression coefficients greater than 1.65 ($p < .05$) were replaced by face parts colored in red: it turned out that the eye regions were used most effectively. Furthermore, the eye on the left side of the image was used effectively earlier than the other one (from about 47 ms vs. 94 ms), and predominated throughout the remainder of the duration of the trial. This is best seen in the plots of Fig. 2a, MEAN, which illustrate the average *Z*-scores in the left-eye (circles) and the right-eye (rectangles) region of the frames over time. We will discuss this bias in Section 3.

At this point, it is quite natural to ask why the eyes were diagnostic. Is it because human observers tend to optimize their use of visual information and the eyes just happen to be the most informative areas for the task at hand? To explore this issue, we first extracted the areas that could best discriminate between faces by using a "super-ideal" observer, i.e., a template-matcher that had perfect knowledge of everything our human observers could have known and, moreover, had perfect knowledge of the location of all bubbles. We sampled space–time with the fixed number of 142 bubbles (i.e., the average number required by human observers for a 75% correct performance), and added enough white Gaussian noise to the stimuli to reduce the performance of the super-ideal to 75% of correct responses. The super-ideal was submitted to a total of 76,500 trials. As this formal observer cannot exhibit any modulation in its use of information over time, a single frame of regression coefficients was computed. The *Z*-scored regression coefficients greater than 1.65 ($p < .05$) are represented in the vignette next to the diamond in Fig. 2a, MEAN. The pattern of information use is almost perfectly symmetrical; this is a direct consequence of randomly mirror-reversing the stimuli in the experiment. What is more interesting is that only two regions reached statistical significance: the eyes and the sides of the face. We can now return to the question that opened the paragraph: Did human observers use the eyes because the eyes are the most informative areas for the task at hand? Qualitatively, the answer is an ambiguous "yes and no" because super-ideal and human effective use of information are both alike and different. Quantitatively, the story is

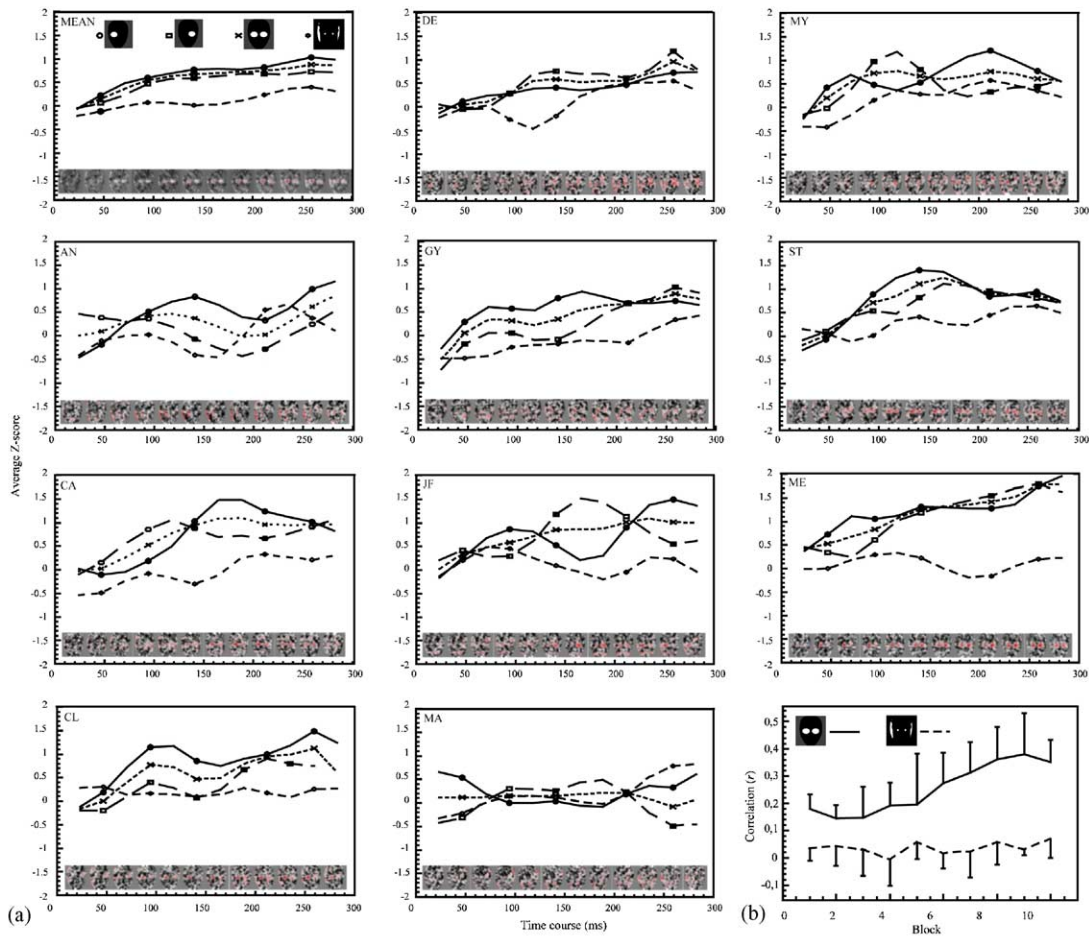

Fig. 2. (a) These pictures depict the *Z*-scored regression coefficients of human observers. They were obtained by regressing space–time information samples with response accuracy per observer (e.g., DE, AN, GY), and across observers (MEAN). These coefficients (varying from black, −1.65, to white, 1.65) indicate the relative usefulness of each region of the face through time, for judgements of face identity—coefficients above 1.65 ($p < .05$) were replaced by a red face marker to facilitate reading. The pictures also plot the average *Z*-scores in the regions of the left eye (circle), the right eye (rectangles), and both eyes (×), as well as the regions most used by a super-ideal observer (diamonds). (b) This panel shows the evolution of the Pearson's correlations between the coefficients of all human observers (average) and that of the super-ideal observer (dotted curve) as well as that of the eye-region model (solid curve), for each one of ten experimental blocks.

told in Fig. 2a, MEAN. We compared the human use of information across the 12 frames, with that of the super-ideal. Specifically, the curve interrupted by diamonds illustrates the evolution of the human (averaged *Z*-scored) regression coefficients in the face regions used by the super-ideal. With successive frames, there is only a slight increase. This is not too surprising given that the super-ideal uses the sides of the faces massively whereas human observers do not. A similar analysis applied to a model which would only uses information from the two eyes (the eye-region model—see ×'s in Fig. 2a, MEAN) reveals a much steadier increase.

Did the observers change their attentional routine over the course of their 5100 trials? We have indirect evidence for this: the number of bubbles required to achieve a 75% correct performance constantly decreased from 244 to 84 bubbles, implying an increase in the efficiency of the use of information. We computed the overall diagnosticity of the space–time voxels for each one of the 10 blocks of 510 trials. Fig. 2b compares the average correlations between the 12 frames of regression coefficients and (solid line) the eye-region model (dashed line) the super-ideal observer, over the 10 blocks of the experiment. The super-ideal line is pretty much flat whereas the eye-region model shows a steep increase with practice. With practice, human observers are therefore increasing their focus on the eyes (a measure of bias), rather than becoming attuned to information relevant in this stimulus set (the sides of the face), possibly not generic (but see the following).

Fig. 2a also depicts the product of the computations described above but applied to individual subject, all blocks pooled together (see the 12 frames for each participants). There are obvious individual differences. For example, two subjects (AN and CA) began by focusing their attention on the right eye region of the image rather than on the left. Another subject (MA) showed an atypical pattern that did not make substantial use of the eyes, and focused instead on few reference points along the contour of the face. Nonetheless, the MEAN results are mostly confirmed. That is, the left eye usually emerges after around 47 ms, followed by the right eye around 94 ms. Again, this is best seen in the individual plots of Fig. 2a: they depict the average *Z*-scores in the left-eye (circles) versus the right-eye (rectangles) region over time for each individual observer, all blocks confounded. It is worth noting that sparse regions around the contour of the face are often used effectively (see the various sequences of 12 frames of Fig. 2a). Importantly, the nose, the cheeks, and the forehead almost never reach statistical significance.

We also computed an average *Z*-scored regression coefficients per frame (see Fig. 3a) for each subject. Surprisingly, we discovered that these patterns of results displayed sinusoïdal pulsations with a mode of 10.6 Hz (see Fig. 3b for the Fourier coefficients of the plots in Fig. 3a); this frequency is a little too high to be linked to the attentional blink phenomenon (about 1–3 Hz; e.g., Raymond, Shapiro, & Arnell, 1992) and too low to be associated with a brain gamma oscillation (about 40–100 Hz; e.g., Engel, König, Kreiter, Schillen, & Singer, 1992). However, recent discussions have related manifestations of discrete "perceptual" (VanRullen & Koch, 2003) and "cognitive" (Ward, 2003) moments (related to encoding, attention and memory) to oscillations broadly identified as covering a range between 6 and 12 Hz (i.e., theta and alpha oscillations). We are currently exploring the relationships between our results and these brain oscillations.

3. General discussion

We adapted the *Bubbles* procedure (Gosselin & Schyns, 2001) to examine the effective use of face information during face identification in a flash (the first 282 ms). Ten participants each viewed a total of 5100 faces sub-sampled with spatio-temporal bubbles. Although we did observe individual differences, most participants displayed the same pattern of information use: the eye on the left side of the image became diagnostic between 47 and 94 ms after the

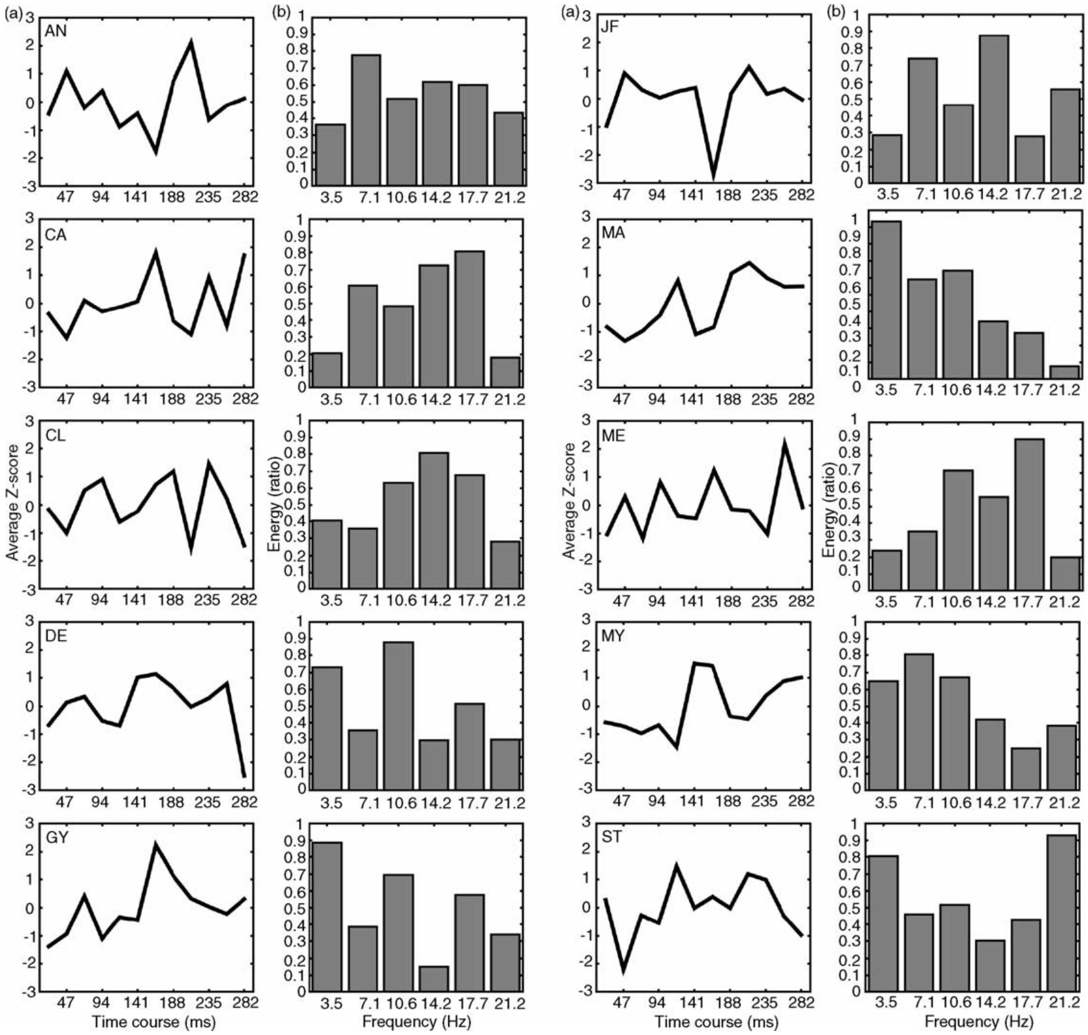

Fig. 3. The plots in (a) depict the average *Z*-scored coefficients of regression per observer (e.g., DE, AN, GY) over time with corresponding Fourier coefficients (b).

onset of the stimulus; after 94 ms, both eyes were used effectively. This preference for the eyes increased with practice.

This main result is consistent with the face recognition literature (e.g., Bentin, Allison, Puce, Perez & McCarthy, 1996; Gold, Bennett, & Sekuler, 2004; Schyns, Bonnar & Gosselin, 2002). For example, event-related potentials (ERP) studies have shown that the upright frontal view of a human face or faces of other species, face photographs, paintings and sketches elicit a larger negative potential at 170 ms (also called N170) than multiple control stimuli such as humans hands, cars, birds, items of furniture (Carmel & Bentin, 2002). Schyns et al. (2003) have recently demonstrated that the amplitude of the N170 is linked to the presence of the eyes within a face, irrespective of task demand, with a later (300 ms, P300 component) tuning on task-dependent information (Smith et al., in press).

Why are we tuned to the eyes? Vessel, Biederman, and Cohen (2003) have suggested that human beings might very well be "infovores", always trying to gain information about our environment. Maybe the eyes are a rich source of real-world information about a person's identity, state of mind, intentions, and so on. For the experiment reported in this article, however, this explanation does not seem to hold fully: although a super-ideal observer (i.e., a template-matcher that has perfect knowledge of everything our human observers could have known and, moreover, has perfect knowledge of the location of all bubbles) did use the eyes of our face stimuli to perform the task, it used the sides of these faces even more effectively. A glance at the individual results in Fig. 2a might help to reconcile the infovore view with the pattern of use of information of our super-ideal. Scattered pieces of the sides of the faces do reach statistical significance in all human observers except one (ME). The location of these scattered pieces varies so much from observer to observer that no area reaches statistical significance in the average of all human observers. Maybe we are infovores after all.

Another aspect of this main pattern of results deserves a discussion: It shows that the eye on the left half of the stimuli was used more effectively and more rapidly than the other one. It is not the first time that a bias for the left half of face stimuli is reported in face recognition literature (e.g., Burt & Perrett, 1997); it is, however, the first time that the precise spacio-temporal dynamics of such a bias is rendered. It is tempting to seek an explanation in properties of the stimuli themselves. To illustrate, because of normalized lighting conditions, the left half of the original faces used in our experiment contained more shadows—and thus more information—than the right. In fact, this is what prompted us to randomly mirror-reverse our stimuli during the experiment.

What is the difference between the left and the right eyes? In an experimental setup very similar to ours, Smith et al. (in press) have demonstrated that the left eye drives the amplitude of the N170 just as well as both eyes in the right hemisphere (P10) and that the same applies to the right eye on the P9 left hemisphere electrode. Therefore, if the eye in the left side of the image is used effectively before the right, it seems that the only possible explanation is that the right hemisphere of the brain processes faces more efficiently than the left. This hypothesis is consistent with functional magnetic reasonance imagery (fMRI) studies that showed that the fusiform face area (FFA) in the left hemisphere is more activated by face stimuli than the one in the right hemisphere (Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997; Sergent, Ohta, & MacDonald, 1992); as well as with event-related potential studies that proved that the right occipitotemporal site (T6) is associated with an earlier and larger N170 than the left one (T5) (e.g., Rossion et al., 1999).

But why is the right hemisphere more efficient at processing faces? It is because it is specialized in face processing per se? This is doubtful. The right hemisphere is better for a whole variety of tasks. For example, Mamassian, Jentzsch, Bacon, and Schweinberger (2003) recently showed that a shape-from-shading task activates the right hemisphere of the brain more than the left. So what is special about the right hemisphere? Ivry and Robertson (1998) proposed that it does not filter task-relevant sensory information in the same manner as the left hemisphere. More specifically, they suggested that the right hemisphere is better and faster than the left at processing the spatial frequencies of an image in the lower range. It just happens that the information contained in faces is concentrated in this lower range (i.e., it follows a $1/f^2$ distribution).

We believe that we have shed new light on the nature of face identification processes. That being said, we are aware of the limitations of our work. Most importantly, our participants were under conditions of reduced uncertainty compared to a more realistic situation like the one described at the onset of the article: they knew that they would only be presented with face stimuli in a vaccum, rather than the more naturalistic condition of faces inserted in a natural background; they knew exactly where and when these faces would appear; they knew which face images they would be presented, they knew that they were in grayscale, that they had normalized hairstyle, and so on; and they also knew what they would have to do with these faces. Because of this reduced uncertainty, we have probably missed a succession of embedded attentional routines preceeding the face identification routine, like the detection of a face (Liu, Harris, & Kanwisher, 2002), the recognition of its expression, gender, age, race, and so on. We are planning to carry out more ecologically valid experiments in future work.

Acknowledgments

This research was supported by an NSERC scholarship to Céline Vinette; by an NSERC (R0010085) and an NATEQ (R0010287) grant awarded to Frédéric Gosselin; and by an ESRC grant (R000237901) awarded to Philippe Schyns.

References

Archambault, A., O'Donnell, C., & Schyns, P. G. (1999). Blind to object changes: When learning one object at different levels of categorization modifies its perception. *Psychological Science*, *10*, 249–255.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perceptions in humans. *Journal of Cognitive Neuroscience*, *8*, 551–565.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

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

Brooks, B. E., Rosielle, L. J., & Cooper, E. E. (2002). The priming of face recognition after metric transformations. *Perception*, *31*, 297–313.

Burt, D. M., & Perrett, D. I. (1997). Perceptual asymmetries in judgements of facial attractiveness, age, gender, speech and expression. *Neuropsychologia*, *35*, 685–693.

Carmel, D., & Bentin, S. (2002). Domain specificity versus expertise: Factors influencing distinct processing of faces. *Cognition*, *83*, 1–29.

Engel, A. K., König, P., Kreiter, A. K., Schillen, T. B., & Singer, W. (1992). Temporal coding in the visual cortex: New vistas on integration in the nervous system. *Trends in Neurosciences*, *15*, 218–226.

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

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

Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research*, *41*, 2261–2271.

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

Gosselin, F., & Schyns, P. G. (2004). No troubles with Bubbles: A reply to Murray and Gold. *Vision Research*, *44*, 471–477.

Henderson, J. M., & Hollingworth, A. (2003). Eye movements, Visual memory, and scene representation. In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes: Analytic and holistic processes*. New York: Oxford University Press.

Ivry, R. B., & Robertson, L. C. (1998). *The two sides of perception* (315 pp.). Massachussett: MIT Press.

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

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience*, *5*, 910–916.

Mamassian, P., Jentzsch, I., Bacon, B. A., & Schweinberger, S. R. (2003). Neural correlates of shape from shading. *NeuroReport*, *14*, 971–975.

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

Murray, R. F., & Gold, J. M. (2004). Troubles with bubbles. *Vision Research*, *44*, 461–470.

Neri, P., & Heeger, D. J. (2002). Spatiotemporal mechanisms for detecting and identifying image features in human vision. *Nature Neuroscience*, *5*, 812–816.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, *46*, 461–488.

Pearson, A. M., Henderson, J. M., Schyns, P. G., & Gosselin, F. (2003). Task-dependant eye movements during face perception. *Abstracts of the Psychonomics Society*, *8*, 84.

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

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 849–860.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*, 368–373.

Rossion, B., Delvenne, J.-F., Debatisse, D., Goffaux, V., Bruyer, R., Crommelinck, M., & Guérit, J.-M. (1999). Spatio-temporal localization of the face inversion effect: An event-related potentials study. *Biological Psychology*, *50*, 173–189.

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science*, *13*, 402–409.

Saarinen, J., & Julesz, B. (1991). The speed of attentional shifts in the visual field. *Proceedings of the National Academy of Sciences of the United States of America*, *88*, 1812–1814.

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

Schyns, P. G., Jentzsch, I., Johnson, M., Schweinberger, S. R., & Gosselin, F. (2003). A principled method for determining the functionality of ERP components. *NeuroReport*, *14*, 1665–1669.

Sergent, J., Ohta, S., & MacDonald, B. (1992). Functional neuroanatomy of face and object processing: A positron emission tomography study. *Brain*, *115*, 15–36.

Simons, D. J., & Levin, D. T. (1997). Change Blindness. *Trends in cognitive sciences*, *1*, 261–267.

Smith, M. L., Cottrell, G., Gosselin, F., & Schyns, P. G. *Transmitting and decoding facial expressions of emotions*, submitted.

Smith, M. L., Gosselin, F., & Schyns, P. G. (in press). Receptive fields for flexible face categorizations. *Psychological Science*.

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

Tanaka, J., & Gauthier, I. (1997). Expertise in object and face recognition. In R. Goldstone, P. G. Schyns, & D. E. Medin (Eds.), *Mechanisms of perceptual learning*. San Diego, CA: Academic Press.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Ullman, S. (1984). Visual routines. *Cognition*, *18*, 97–159.

VanRullen, R., & Koch, C. (2003). Is perception discrete or continuous. *Trends in Cognitive Sciences*, *7*, 207–213.

Vessel, E.A., Biederman, I., & Cohen, M. S. (2003). The neural basis of spontaneous perceptual selection. *Vision Sciences*.

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences*, *7*, 553–559.