

Recognizing famous people

SEAN BUTLER

University of Victoria, Victoria, British Columbia, Canada

CAROLINE BLAIS AND FRÉDÉRIC GOSSELIN

University of Montreal, Montreal, Quebec, Canada

DANIEL BUB

University of Victoria, Victoria, British Columbia, Canada

AND

DANIEL FISET

*University of Victoria, Victoria, British Columbia, Canada
and University of Montreal, Montreal, Quebec, Canada*

In daily life, face identification requires that the observer select a single representation from hundreds if not thousands in memory. This breadth of choice is nearly impossible to replicate in the laboratory using newly learned faces, especially in the context of a Bubbles experiment (Gosselin & Schyns, 2001). In this study, we obviated this concern by studying the performance of observers in a face-naming task using 210 faces of celebrities. On each trial, we presented a face randomly sampled with Bubbles. We performed least-square multiple linear regressions on the location of the samples and on accuracy to pinpoint the facial features that were used effectively in this task. Correct face identification relied primarily on the eye areas in spatial frequency bands ranging from 4.37 to 70 cycles per face (cpf) and on the mouth and the nose in a spatial frequency band ranging from 8.75 to 17.5 cpf. A comparison with other studies (Caldara et al., 2005; Schyns, Bonnar, & Gosselin, 2002) in which Bubbles was used with a set of 10 newly learned faces revealed that although the eye areas were useful across studies, the mouth area and higher spatial frequencies gained in importance when few newly learned faces were used.

Imagine you are sitting down for dinner in a crowded restaurant when you suddenly recognize the face of Robert DeNiro at a table across the room. As anyone who has had a chance encounter with a celebrity can attest, the human mind is eminently capable of making this sort of spontaneous identification judgment even without any predisposing information or context with which to infer identity. This example of a real-life situation requires the observer to select a single representation from hundreds if not thousands in long-term visual memory. Furthermore, it strongly contrasts with many laboratory experiments—in particular, Bubbles studies of face identification, as we shall see—in which participants were asked to learn a small subset of unfamiliar faces—typically, 10 or fewer—that they had to identify shortly thereafter in a test phase.

Given the considerable distance between the constraints imposed on the visual system in the DeNiro example and a laboratory context, one may ask whether the mechanisms pertaining to face identification remain the same across this span of situations. In the present study, we attempted to answer this question by creating a very challenging situation with respect to face recognition, similar

to the long-term visual memory demands incurred if one were to attend an Oscar award ceremony, in which participants needed to identify faces from a large subset of famous American actors. Since the participants saw each face only once and did not know the extent of the database, each trial, from their perspective, was the recognition of one famous face from all the celebrity faces that they knew. More specifically, our main objective was to reveal the visual information, in terms of facial areas and spatial frequencies, used by our participants to recognize faces in a task that was highly demanding with respect to long-term visual memory, using Bubbles (Gosselin & Schyns, 2001).

On each trial of a typical Bubbles experiment, the photograph of an object is sampled through randomly positioned Gaussian apertures in the image plane, and the participant is instructed to perform a recognition task on this stimulus. If specific visual information is important for this task, depriving the observer of this information will impair his or her performance, and, conversely, showing this information will improve his or her performance. More generally, by performing a multiple linear regression on the random

D. Fiset, daniel.fiset@umontreal.ca

samples and response accuracy, we can pinpoint the visual information correlated with accurate responses.

Bubbles has already been used to examine the visual information correlated with accurate face identification (Caldara et al., 2005; Gosselin & Schyns, 2001; Schyns, Bonnar, & Gosselin, 2002). However, in these studies, participants acquired familiarity with a small set of unknown faces (10 identities) through extensive repeated exposure. In the experiments most comparable to ours (Caldara et al., 2005; Schyns et al., 2002), the learning phase was concluded when a participant perfectly recognized all of the trained faces in two consecutive testing blocks. In the experimental phase, the observers viewed one stimulus from the trained set, provided the identity, and proceeded to the next trial. The experiment comprised between 1,000 (Schyns et al., 2002) and 4,200 trials (Caldara et al., 2005), meaning that over the course of the testing phase, each exemplar was viewed between 50 and 210 times. Although they did not receive feedback during the experiment, the observers had the opportunity to overlearn the stimuli (e.g., Herzog & Fahle, 1997). This learning may however be very specific to the images presented in the task and may not generalize to a real-life situation. For instance, participants may learn to identify one face from the unusual position of its nose relative to the other facial features, another from its remarkably large teeth, and so on. Moreover, trial-by-trial identity judgments required extracting only the visual information sufficient to discriminate between faces from their small and newly learned training set and to successfully find a match. The experimental procedure described above does not tax the face-recognition capacities of the human brain in a way that reflects conditions that generally exist beyond the laboratory.

In light of this argument, it is worth considering whether the design constraints in Bubbles studies of face identification have influenced observer recognition strategies. In the present study, we reexamine the relationship between face information utilization and identity recognition by applying the Bubbles method to famous faces. This alters the procedure in three fundamental ways, each of which contributes to make the task more ecological. First, the initial learning phase can be skipped, because the participants begin testing already knowing most of the face identities; faces judged as unfamiliar are subsequently excluded from the analysis. Second, practical limits on stimulus set size, related to learning face identities in a research setting, can be ignored. Third, the observers do not know which faces are members of the stimulus set, apart from the knowledge that each is famous (and that those already seen will not be seen again). To match a target to memory, a participant has to make judgments within a large set of known celebrities. Note that it is the task (recognition from long-term visual memory) that these modifications to the procedure make more ecological, not the Bubbles technique itself. Here, we examine the diagnostic features of identity recognition using famous faces because they serve as a compromise between an impractically large set containing all known identities and a dubiously small one containing 10 identities or

fewer. We will compare our results with those obtained by Schyns et al. (2002) and Caldara et al. (2005) with a small set of newly learned faces.

METHOD

Participants

Forty students (20–35 years old) from the University of Victoria took part in the experiment. They had normal or corrected-to-normal vision and described themselves as fans of American movies and television. The participants received course credit or were paid for participating.

Stimuli

The set comprised 210 celebrities (135 male, 75 female; 95 neutral, 115 happy) recognized by at least 16 out of 20 participants in a preliminary study. Each grayscale image displayed a frontal view of either a happy or a neutral expression. The images were translated, reoriented, and resized to normalize the positions of the main internal facial features (eyes, mouth, nose, eyebrows); image resolution was 256×256 pixels (image size was 5.7° of visual angle, which is the same as in Schyns et al., 2002), and average face width was 146 pixels. Finally, spatial frequency spectra were equalized using SHINE (Willenbockel et al., in press).

To create a stimulus, a face (Figure 1A) was first bandpass filtered into five different spatial frequency bands (70–35, 35–17.5, 17.5–8.75, 8.75–4.37, and 4.37–2.18 cycles per face [cpf]); the remaining bandwidth served as constant background; see Figure 1B), using the Pyramid toolbox for MATLAB (Simoncelli, 1999). Second, each spatial frequency band was independently and randomly sampled using Gaussian apertures (or *bubbles*) of varying standard deviations; that is, the size of the bubbles was adjusted according to frequency band in order to reveal three cycles in all cases (see Figure 1C; the standard deviations of the bubbles were 0.13° , 0.27° , 0.54° , 1.08° , and 2.15° of visual angle from the finest to the coarsest scale). Because the size of the bubbles increases as the spatial scale becomes coarser, the number of bubbles differed at each scale to maintain a constant probability of revealing a given pixel in each spatial frequency bandwidth. Third and finally, the five randomly sampled images plus the background were then summed to produce the stimulus (see Figure 1D).

Procedure

The stimuli were displayed on a high-resolution 22-in. ViewSonic monitor set with a refresh rate of 85 Hz. The monitor was calibrated to allow a linear manipulation of luminance. The resulting corrected table contained 154 luminance levels, ranging from 0.3 to 98.7 cd/m². The background luminance was 49.3 cd/m². The experiment was run on a dual-core 2.93-GHz PC. The experimental program was written in MATLAB, using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Viewing distance was maintained constant at 57 cm by using a chinrest. Each participant saw each celebrity once in random order. Each trial began with a fixation cross displayed at the center of the screen for 494 msec. It was then immediately replaced by a bubbled face, which remained on the screen until the participant's response. The participant then attempted to verbally retrieve either the name of the actor or any semantic information demonstrating recognition of his or her identity. Valid semantic information included, for example, a character that the actor had portrayed (e.g., Spider-man for Tobey Maguire), a film or television program the actor had starred in (e.g., *he's the lead actor in Schindler's List* for Liam Neeson), or another actor or character whom the actor had starred alongside (e.g., *she's the girlfriend of Spider-man* for Kirsten Dunst). Following the participant's response, the experimenter operating the computer keyboard displayed the same celebrity's face, this time without bubbles. The participant was prompted to confirm or reject the former response. The experimenter coded the response pair as a correct recognition when the participant's first and second

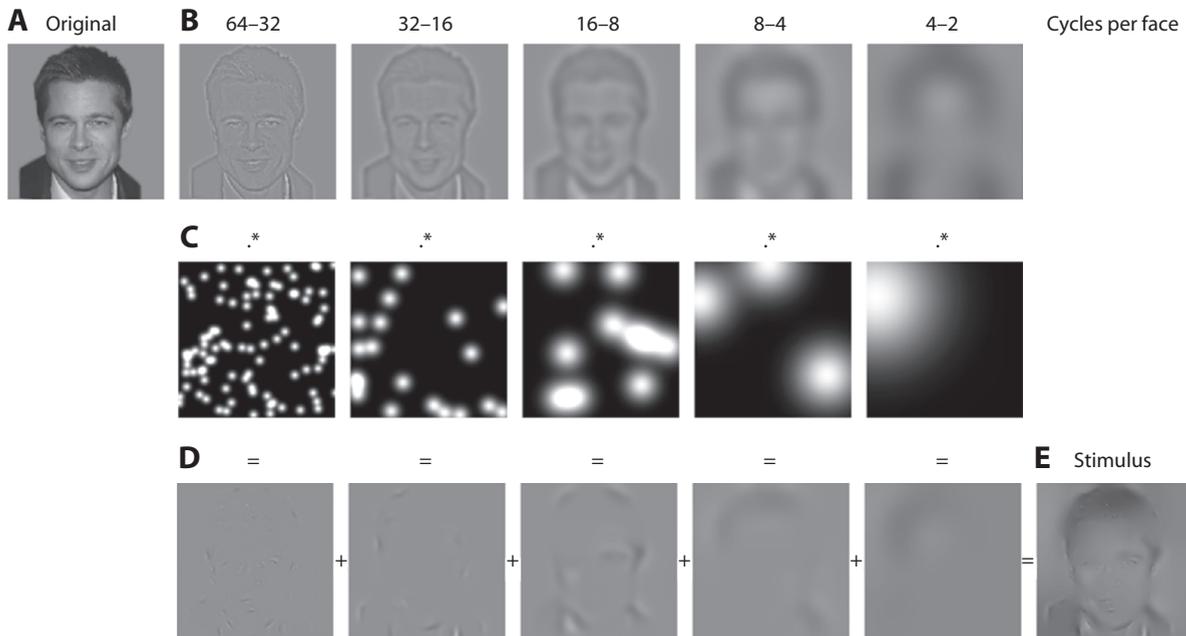


Figure 1. Illustration of the stimuli creation procedure. The face in panel A is filtered into the five spatial frequency bands in panel B plus a constant background (not shown). In each band in panel C, a number of randomly positioned Gaussian apertures puncture a homogeneous black field. Applying the masks reveals the information in each band shown in panel D. Spatially filtered information is then summed, producing a bubbled stimulus (panel E).

responses both matched the identity of the stimulus, an incorrect recognition when only the participant's second response matched the identity of the stimulus, or as an unfamiliar face when the participant failed to recognize the fully visible face. If the participant correctly or incorrectly recognized the bubbled faces on a given trial, the program adjusted the number of bubbles using QUEST (Watson & Pelli, 1983) to maintain a correct identification rate of 60%.

RESULTS AND DISCUSSION

On average, the participants needed 145.18 bubbles (range = 42–195, $SD = 34.16$) to maintain their performance at 60% correct. An average of 29.7 out of 210 famous faces ($SD = 20.4$) were unknown to each participant. Therefore, the analysis was done on a total of 40 participants, with 180.3 trials per participant on average, for a total of 7,212 trials.

To uncover the features that different observers used to recognize famous faces, we performed a least-squares multiple linear regression on the bubble masks and accuracy data for each participant. The plane of regression coefficients yielded by this operation is called a *classification image*. Here, a classification image was obtained on each band of spatial frequencies by summing all of the bubble masks (see Figure 1C) that led to a correct answer and subtracting from this sum all of the bubble masks that led to an incorrect answer. We summed all of the individual classification images, and we transformed the regression coefficients of the resulting group classification images into z scores. To determine whether the visual information significantly correlated with accuracy, we applied the pixel test to these classification images ($p < .05$; $z_{crit} = 4.141, 3.757, 3.362, 2.944, \text{ and } 2.535$, from finer

to coarser scales). The statistical threshold provided by this test corrects for multiple comparisons (for details, see Chauvin, Worsley, Schyns, Arguin, & Gosselin, 2005). Figure 2A shows this effective information—that is, the statistically thresholded classification images superimposed on the five images representing the different bands of spatial frequency information of Brad Pitt's face.

A first glance at the thresholded classification images reveals that the areas significantly correlated with performance were the eyes in the second lowest spatial frequency band (i.e., 4.37–8.75 cpf); the eyes, nose, and mouth in the middle frequency band (i.e., 8.75–17.5 cpf); and the eyes in the two highest frequency bands (i.e., 17.5–35 and 35–70 cpf). Thus, the eyes, as well as the eyebrows, were useful across the four highest spatial frequency bands tested in this experiment. This is in agreement, qualitatively, with the results obtained in other Bubbles experiments (Caldara et al., 2005; Gosselin & Schyns, 2001; Schyns et al., 2002; Vinette, Gosselin, & Schyns, 2004), as well as with other methodologies (Sadr, Jarudi, & Sinha, 2003; Sekuler, Gaspar, Gold, & Bennett, 2004). Interestingly, external features (ears, hair, face contour) were not correlated with accurate famous-face identification. Note that because these external features were not normalized across the stimuli, differentiation based on hairstyle or face contour was a potentially viable strategy. In spite of this, the observers did not utilize these cues in a systematic manner along the x , y , and spatial frequency dimensions. This is consistent with recognition and match-to-sample studies in which familiar-face stimuli were used (e.g., Ellis, Shepherd, & Davies, 1979; Young, Hay, McWeeny, Flude, & Ellis, 1985).

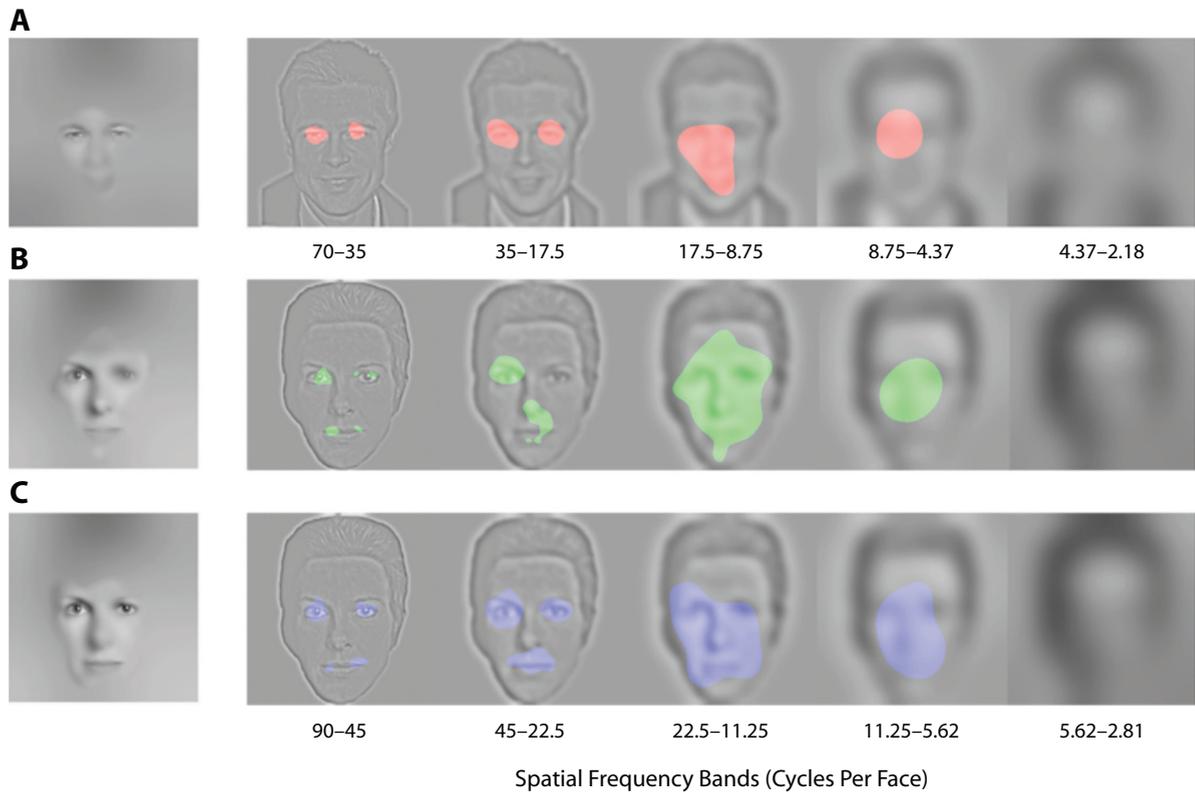


Figure 2. Visual information used effectively to identify faces (A) in our study, (B) in Schyns et al. (2002), and (C) in Caldara et al. (2005).

We precisely compared the results of two studies—Schyns et al. (2002) and Caldara et al. (2005)—in which the facial features important for face identification were examined using the same version of the Bubbles technique that we used but on a set of 10 faces instead of a set of 210 faces. We reanalyzed the data of these two studies ($n = 15$ participants in Schyns et al., 2002; $n = 7$ control participants in Caldara et al., 2005) exactly as described above. The results are shown in Figures 2B and 2C.

First, we contrasted the relative usefulness of each spatial frequency band for each study by calculating the proportion of all of the diagnostic pixels that fell on each frequency band (see Figures 2A–2C). To derive the peak and bandwidth of the spatial frequency tuning for face identification, we best-fitted a Gaussian density function on the proportions of diagnostic pixels (see Figure 3). For our study, spatial frequency tuning peaked at 11.85 cpf

and spanned 2.5 octaves. In Schyns et al. (2002), the spatial frequency tuning peaked at 14.56 cpf and spanned 1.8 octaves, whereas in Caldara et al. (2005), it peaked at 13.94 cpf and spanned 2.5 octaves. Thus, both studies show a spatial frequency tuning slightly higher than in our study. This suggests that when the same stimuli are presented repeatedly, participants shift to a slightly more image-based strategy and rely on more fine-grained information to identify faces.

Second, we examined the full-blown thresholded classification images. Those derived from the data of Caldara et al. (2005) reveal that the eye and mouth areas achieved statistical significance in the two highest spatial frequency bands, whereas all inner facial features are shown in the mid and low-mid spatial frequency bands. The thresholded classification images derived from the data of Schyns et al. (2002) reveal the eyes and the mouth in the

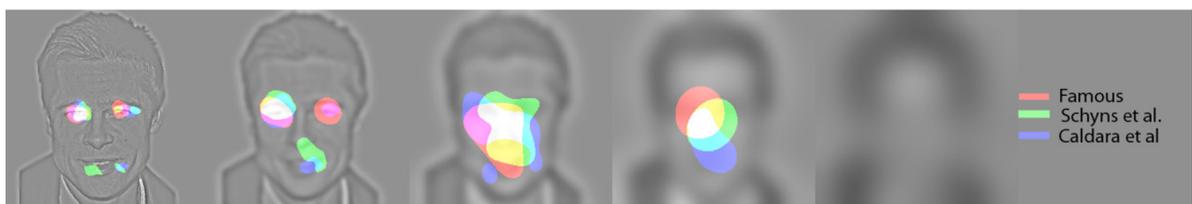


Figure 3. Comparison of the location information used effectively in our study, in Schyns et al. (2002), and in Caldara et al. (2005).

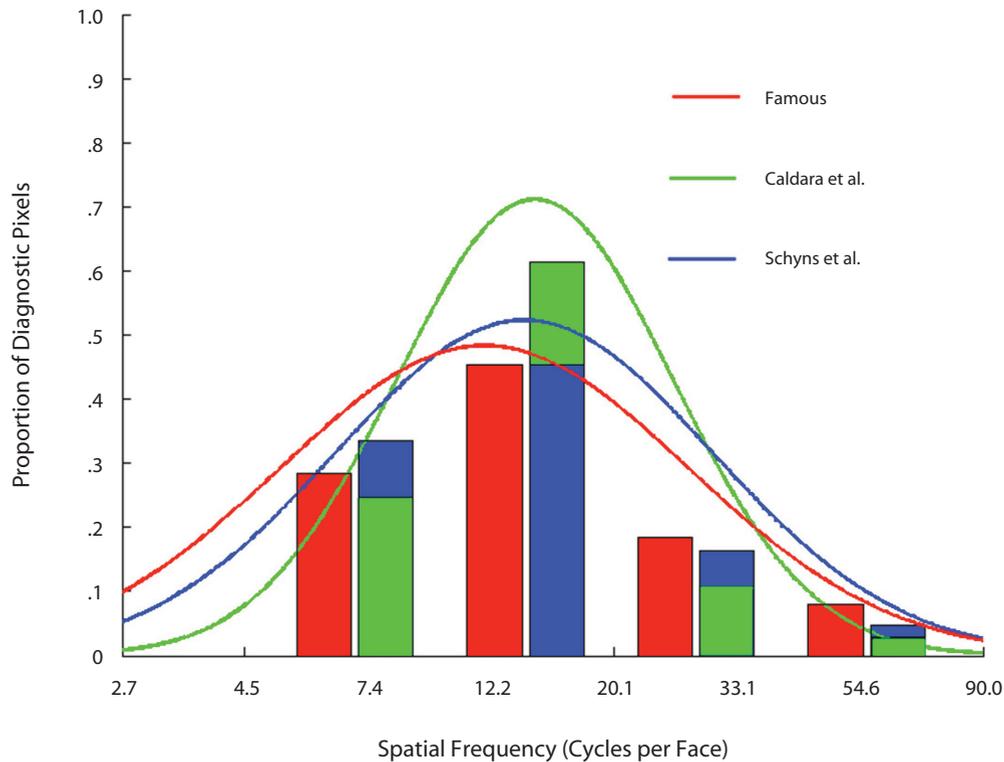


Figure 4. Comparison of the spatial frequency information used effectively in our study, in Schyns et al. (2002), and in Caldara et al. (2005).

highest spatial frequency band, the left eye and the nose/mouth area in the second highest spatial frequency band, all inner features in the mid spatial frequency band, and, finally, the eyes/nose area in the second lowest frequency band. One obvious difference between the results of these two studies and the results of our study is the statistical significance or lack thereof of the mouth in the two highest frequency bands. The total number of trials was higher in Schyns et al. and Caldara et al. than in our study. These studies thus had a greater signal-to-noise ratio. It is thus conceivable that the importance of the mouth has been underestimated in our study.

To do a fairer comparison between the use of information across the three studies, statistical thresholds applied to the classification images of Schyns et al. (2002) and Caldara et al. (2005) were raised independently in each spatial frequency band to match the number of statistically significant pixels found in our study. Furthermore, the thresholded classification images of Schyns et al. and Caldara et al. were aligned with ours on the main facial features by rotation, translation, and homothety. The results are shown in Figure 4.

Interestingly, the statistical importance of the mouth in the two highest spatial frequency bands remains in the studies of Schyns et al. (2002) and Caldara et al. (2005). This discrepancy is not due to our celebrity pictures being relatively less variable in the area of the mouth than the face pictures used in the two previous studies. In fact, the variance in the mouth area relative to the variance in

the eyes area was actually slightly higher in our stimulus set than in the two aforementioned Bubbles studies (see Sekuler et al., 2004, for a similar analysis). Therefore, we believe that the discrepancy between our study and the previous Bubbles studies instead indicates that ecological face processing allocates more weight to the eyes area than to the mouth area.

The mouth is an area of the face that is frequently in motion and that has a variable shape; for example, the size of the lips, the contour of the mouth, and the visibility of the teeth vary as a function of the facial expression of emotions, as well as during the production of language. In contrast, the shape of the eyes remains much more stable. Thus, differences in mouth shape might be blurred in long-term memories because of mouth movements from one visual episode to the next. This could explain why mouths might be less reliable in recognizing faces using long-term memories than in recognizing newly learned faces that have never been seen in motion. A similar logic can be applied to external contours of the face. Interestingly, it has been shown that the external contour information is not used with familiar faces but can be used with unfamiliar ones (Ellis et al., 1979; Young et al., 1985).

Moreover, prosopagnosic patients present a lack of sensitivity to eye information for both familiar and unfamiliar face recognition (Bukach, Bub, Gauthier, & Tarr, 2006; Bukach, Le Grand, Kaiser, Bub, & Tanaka, 2008; Caldara et al., 2005; Rossion, Kaiser, Bub, & Tanaka, 2009), despite a normal sensitivity to the mouth area; these ob-

servations also suggest that mouth information may be difficult to use in a realistic setting, such as the one in our study. Under the hypothesis that prosopagnosia alters normal face recognition mechanisms, these observations suggest that the specific face recognition mechanisms allocate much of their resources to recognition of the eye region and that the mouth area may be processed using more general, unspecific visual mechanisms.

In summary, our results confirm the importance of the eye area for face recognition, as has been shown by other studies using Bubbles, and other classification image studies. However, we show that the mouth area and higher spatial frequencies are less critical for face identification than previously believed on the basis of results obtained with a small set of newly learned stimuli.

AUTHOR NOTE

We thank the participants who took part in this study. This research was supported by a grant from the Canadian Institute of Health Research (CIHR) to Martin Arguin, F.G., and D.B.; by a scholarship from the James S. McDonnell Foundation (Perceptual Expertise Network) and by a postdoctoral scholarship from the Fonds Québécois de Recherche en Nature et Technologies (FQRNT) to D.F.; and by a graduate scholarship from the FQRNT to C.B. Correspondence concerning this article should be addressed to D. Fiset, Département de psychologie, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal, QC, H3C 3J7 Canada (e-mail: daniel.fiset@umontreal.ca).

REFERENCES

- BRAINARD, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, **10**, 433-436. doi:10.1163/156856897X00357
- BUKACH, C. M., BUB, D. N., GAUTHIER, I., & TARR, M. J. (2006). Perceptual expertise effects are not all or none: Spatially limited perceptual expertise for faces in a case of prosopagnosia. *Journal of Cognitive Neuroscience*, **18**, 48-63. doi:10.1162/089892906775250094
- BUKACH, C. M., LE GRAND, R., KAISER, M. D., BUB, D. N., & TANAKA, J. W. (2008). Preservation of mouth region processing in two cases of prosopagnosia. *Journal of Neuropsychology*, **2**, 227-244. doi:10.1348/174866407X231010
- CALDARA, R., SCHYNS, P. G., MAYER, E., SMITH, M. L., GOSSELIN, F., & ROSSION, B. (2005). Does prosopagnosia take the eyes out of face representations? Evidence for a defect in representing diagnostic facial information following brain damage. *Journal of Cognitive Neuroscience*, **17**, 1652-1666. doi:10.1162/089892905774597254
- CHAUVIN, A., WORSLEY, K. J., SCHYNS, P. G., ARGUIN, M., & GOSSELIN, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision*, **5**(9, Art. 1), 659-667. doi:10.1167/5.9.1
- ELLIS, H. D., SHEPHERD, J. W., & DAVIES, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception*, **8**, 431-439. doi:10.1068/p080431
- GOSSELIN, F., & SCHYNS, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research*, **41**, 2261-2271. doi:10.1016/S0042-6989(01)00097-9
- HERZOG, M. H., & FAHLE, M. (1997). The role of feedback in learning a vernier acuity discrimination task. *Vision Research*, **37**, 2133-2141. doi:10.1016/S0042-6989(97)00043-6
- PELLI, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, **10**, 437-442. doi:10.1163/156856897X00366
- ROSSION, B., KAISER, M. D., BUB, D. N., & TANAKA, J. W. (2009). Is the loss of diagnosticity of the eye region a common feature of acquired prosopagnosia? *Journal of Neuropsychology*, **3**, 69-78. doi:10.1348/174866408X289944
- SADR, J., JARUDI, I., & SINHA, P. (2003). The role of eyebrows in face recognition. *Perception*, **32**, 285-293. doi:10.1068/p5027
- 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. doi:10.1111/1467-9280.00472
- SEKULER, A. B., GASPARD, C. M., GOLD, J. M., & BENNETT, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology*, **14**, 391-396. doi:10.1016/j.cub.2004.02.028
- SIMONCELLI, E. P. (1999). Image and multi-scale pyramid tools [Computer software]. New York: Author.
- VINETTE, C., GOSSELIN, F., & SCHYNS, P. G. (2004). Spatio-temporal dynamics of face recognition in a flash: It's in the eyes. *Cognitive Science*, **28**, 289-301. doi:10.1016/j.cogsci.2004.01.002
- WATSON, A. B., & PELLI, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, **33**, 113-120.
- WILLENBOCKEL, V., SADR, J., Fiset, D., HORNE, G. O., GOSSELIN, F., & TANAKA, J. W. (in press). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*.
- YOUNG, A. W., HAY, D. C., MCWEENY, K. H., FLUDE, B. M., & ELLIS, A. W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception*, **14**, 737-746. doi:10.1068/p140737

(Manuscript received November 16, 2009;
revision accepted for publication March 22, 2010.)