## Recognizing famous people

<table>
<thead>
<tr>
<th>Journal:</th>
<th><em>Attention, Perception, &amp; Psychophysics</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>PP-ORIG-09-266.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Short Report</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>24-Feb-2010</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Butler, Sean; University of Victoria, Psychology Blais, Caroline; University of Montreal, Psychology Gosselin, Fred; Université de Montréal, Psychologie Bub, Dan; University of Victoria, Psychology Fiset, Daniel; University of Montreal, Psychology</td>
</tr>
<tr>
<td>Keywords:</td>
<td>face perception, face recognition, Methods: Classification Images</td>
</tr>
</tbody>
</table>
Recognizing famous people

Sean Butler¹, Caroline Blais², Frédéric Gosselin²,
Daniel Bub¹, & Daniel Fiset¹-²,

¹ Department of Psychology, University of Victoria, PO Box 3050 STN CSC, Victoria, BC, V8W 3P5, CANADA

² Centre de Recherche en Neuropsychologie et Cognition, Département de Psychologie,
Université de Montréal, CP 6128, succ. Centre-ville, Montréal H3C 3J7, CANADA

Address correspondence to
Daniel Fiset, Ph.D.
Département de psychologie,
Université de Montréal,
C.P. 6128 Succ. Centre-ville,
Montréal, QC,
H3C 3J7, Canada
Email : daniel.fiset@umontreal.ca
Abstract

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 laboratory using newly learned faces, especially in the context of a Bubbles experiment (Gosselin & Schyns, 2001). This study 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 accuracy to pinpoint the facial features that are 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 (Schyns, Bonnar & Gosselin, 2002; and Caldara et al., 2005) that used Bubbles with a set of ten newly learned faces revealed that while the eye areas are useful across studies, the mouth area and higher spatial frequencies gain in importance when few newly learned faces are used.
Recognizing famous people

Imagine you’re sitting down for dinner in a crowded restaurant when you suddenly recognize the face of Robert DeNiro from a table across the room. As anyone that 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. Further, it strongly contrasts with many laboratory experiments—in particular, Bubbles studies of face identification, as we shall see—in which subjects are asked to learn a small subset of unfamiliar faces—typically ten or less—that they will have to identify shortly after 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 participants saw each face only once and didn’t know the extent of the database, each trial from their perspective is the recognition of one famous face from all the celebrity faces they know. More specifically, our main objective is 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/her performance and conversely showing this information will improve his/her performance. More generally, by performing a multiple linear regression on the random 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 (Gosselin & Schyns, 2001; Schyns, Bonnar & Gosselin, 2002; Caldara et al., 2005). 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 participant recognized perfectly all trained faces in two consecutive testing blocks. In the experimental phase, observers viewed one stimulus from the trained set, provided the identity, and proceeded to the next trial. The experiment comprised between 1000 (Schyns et al., 2002) to 4200 trials (Caldara et al., 2005), meaning that over the course of the testing phase each exemplar is viewed between 50 to 210 times each. Though they did not receive feedback during the experiment, observers had the opportunity to “over-learn” 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 nostrils; 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 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, it is worth considering whether or not the design constraints in Bubbles studies of face identification have influenced observer recognition strategies. The present study reexamines 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 that each contributes to make the task more ecological. First, the initial learning phase can be skipped as 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, 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 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 ten identities or less. We will compare our results with those obtained by Schyns et al. (2002) and Caldara et al. (2005) with a small newly learned set of faces.

**Methods**

Forty (20-35 years old) students 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. Subjects received course credit or were paid for participating.

The set comprised 210 celebrities (135 male, 75 female; 95 neutral and 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 neutral expression. Images were translated, reoriented and resized to normalize position of main internal facial features (eyes, mouth, nose, eyebrows); image resolution was 256 x 256 pixels (image size = 5.7 deg of visual angle, which is the same as 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 was first bandpass filtered into five different spatial frequency bands (70-35, 35-17.5, 17.5-8.75, 8.75-4.37, 4.37-2.18 cycles/face; the remaining bandwidth served as constant background; see Figure 1A), 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 3 cycles in all cases (see Figure 1B; standard deviations of the bubbles were 0.13, 0.27, 0.54, 1.08, and 2.15 degree 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.

Stimuli were displayed on a high-resolution 22-inch 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 cd/m$^2$ to 98.7 cd/m$^2$. The background luminance was 49.3 cd/m$^2$. The experiment was run on a dual core 2.93 GHz PC computer. The experimental program was written in Matlab, using the Psychophysics Toolbox.
(Brainard, 1997; Pelli, 1997). Viewing distance was maintained constant at 57 cm 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 ms. It was then immediately replaced by a "bubblized" face, which remained on the screen until the subject’s response. The participant then attempted to verbally retrieve either the name of the actor or any semantic information demonstrating recognition of their identity. Valid semantic information included, for example, a character the actor portrayed (e.g. Spiderman for Tobey Maguire), a film or television program the actor starred in (e.g. he’s the lead actor in Schindler’s List for Liam Neeson), or another actor or character the actor starred alongside (e.g. she’s the girlfriend of Spiderman for Kirsten Dunst). Following the participant’s response, the experimenter operating the computer keyboard displayed the same celebrity’s face, but 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 both the participant’s first and second response 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 bubblized 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, participants needed 145.18 bubbles (between 42 and 195, $SD = 34.16$) to maintain their performance at 60% correct. An average of 29.7 out of 210 famous faces (with a standard deviation of 20.4) were unknown to each participant. Therefore, the analysis was done on a total
of 40 participants * 180.3 trials per participants = 7,212 trials (i.e. 180 trials per participant on average).

To uncover the features that different observers used to recognize famous faces, we performed a least-square multiple linear regression on the bubble masks and accuracy data for each subject. The plane of regression coefficients yielded by this operation is called a classification image. Here, a classification image is obtained on each band of spatial frequencies by summing all the bubble masks (see Fig. 1, second row) that led to a correct answer, and subtracting from this sum all the bubble masks that led to an incorrect answer. We summed all individual classification images, and we transformed the regression coefficients of the resulting group classification images into Z-scores. To determine the visual information significantly correlated with accuracy, we applied the Pixel test to these classification images ($p < .05$; $Z_{crit} = 4.141, 3.757, 3.362, 2.944,$ and $2.535$, from finer to coarser scales). The statistical threshold provided by this test corrects for multiple comparisons (for details, see Chauvin et al., 2005). Fig. 2a shows this effective information, that is the statistically thresholded classification images superimposed to the five images representing 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 cycles per face), the eyes, nose and mouth in the middle frequency band (i.e. 8.75-17.5 cycles per face), and the eyes in the two highest frequency bands (i.e. 17.5-35, and 35-70 cycles per face). 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 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 faces identification. Of note, because these external features were not normalized across stimuli, differentiation based on hairstyle or face contour was a potentially viable strategy. In spite of this, 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 using familiar face stimuli (e.g. Ellis, Shepherd, & Davies, 1979; Young, Hay, & McWeeny, 1985).

We compared precisely the results of two studies—Schyns et al, 2002, and Caldara et al., 2005—that examined the facial features important for face identification 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 re-analyzed the data of these two studies (N = 15 participants in Schyns et al., 2002; and N = 7 control participants in Caldara et al., 2005) exactly as described above in the Results and Discussion section. Results are shown on Figure 2b and 2c.

First, we contrasted the relative usefulness of each spatial frequency band for each study by calculating the proportion of all the diagnostic pixels that fell on each frequency band (see Fig. 2). 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 the red, green and blue lines on Fig. 3). For our study, spatial frequency tuning peaked at 11.85 cycles per face, and spanned 2.5 octaves. In the study of Schyns et al. (2002), the spatial frequency tuning peaked at 14.56 cycles per face and spanned 1.8 octaves, whereas in Caldara et al (2005), it peaked at 13.94 cycles per face and spanned 2.5 octaves. Thus both studies show a slightly higher spatial...
frequency tuning than 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 the eyes and the mouth 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 highest spatial frequency band; the left eye and the nose/mouth area in the second highest spatial frequency band; all inner features were significant 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 presence/absence of the mouth in the two highest frequency bands. The total number of trials was higher in the studies of Schyns et al. (2002) and Caldara et al. (2005) than in our study. These studies thus had greater signal-to-noise. 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. and Caldara et al. 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 homothecy. Results are shown on Figure 4.
Interestingly, the usefulness 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). Thus, 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 which 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 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 observations also suggest that mouth information may be difficult to use in a realistic setting such as the one used 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 on 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 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 based on results obtained with a small set of newly learned stimuli.
REFERENCES


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, Frédéric Gosselin, and Daniel Bub; by a scholarship from the James S. McDonnell Foundation (Perceptual Expertise Network) and by a post-doctoral scholarship from the Fonds Québécois de Recherche en Nature et Technologies (FQRNT) to Daniel Fiset; and by graduate scholarship from the FQRNT to Caroline Blais.

Correspondence concerning this article should be addressed to Daniel Fiset, Département de psychologie, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal, QC, H3C 3J7, Canada. Electronic mail may be sent to daniel.fiset@umontreal.ca.
Figure Captions

Figure 1. Illustration of the stimuli creation procedure. The face in (a) is filtered into the five spatial frequency bands in (b) plus constant background (not shown). (c) In each band, a number of randomly positioned Gaussian apertures puncture a homogenous black field. Applying the masks reveals the information in each band shown in (d). Spatially filtered information is then summed, producing a "bubblized" stimulus (e).

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).

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.
Illustration of the stimuli creation procedure. The face in (a) is filtered into the five spatial frequency bands in (b) plus constant background (not shown). (c) In each band, a number of randomly positioned Gaussian apertures puncture a homogenous black field. Applying the masks reveals the information in each band shown in (d). Spatially filtered information is then summed, producing a "bubblized" stimulus (e).
Visual information used effectively to identify faces (a) in our study, (b) in Schyns et al. (2002), and (c) in Caldara et al. (2005).

59x36mm (600 x 600 DPI)
Comparison of the location information used effectively in our study, in Schyns et al. (2002), and in Caldara et al. (2005).

63x10mm (600 x 600 DPI)
Comparison of the spatial frequency information used effectively in our study, in Schyns et al. (2002), and in Caldara et al.

101x72mm (300 x 300 DPI)