PERGAMON

Vision Research 41 (2001) 2261–2271

Vision Research

www.elsevier.com/locate/visres

Bubbles: a technique to reveal the use of information in recognition tasks

Frédéric Gosselin *, Philippe G. Schyns

Department of Psychology, University of Glasgow, 58, Hillhead Street, Glasgow G12 8QB, Scotland, UK

Received 31 August 2000; received in revised form 3 December 2000

Abstract

Everyday, people flexibly perform different categorizations of common faces, objects and scenes. Intuition and scattered evidence suggest that these categorizations require the use of different visual information from the input. However, there is no unifying method, based on the categorization performance of subjects, that can isolate the information used. To this end, we developed Bubbles, a general technique that can assign the credit of human categorization performance to specific visual information. To illustrate the technique, we applied Bubbles on three categorization tasks (gender, expressive or not and identity) on the same set of faces, with human and ideal observers to compare the features they used. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Bubbles; Recognition tasks; Categorizations

1. Introduction

Even casual observers would have no problem to classify the two faces of Fig. 1. They would say that the face in Fig. 1a is a woman, with a happy expression, who is called 'Anne,' if this was her identity. In contrast, Fig. 1b is a man, called 'Simon,' with a neutral expression, and who is comparatively older. These different judgements of similar images reveal the impressive versatility of face categorization mechanisms (e.g. Etcoff & Maggee, 1992; Calder, Young, Perrett, Etcoff, & Rowland, 1996; Schyns & Oliva, 1999). That is, observers can make subtle judgments of gender, identity, age and expression, based on the same visual input.

Versatile categorizations are not restricted to faces. People can typically classify a given object as a car, at the basic-level, a vehicle at the superordinate level, and as a Porsche at the subordinate level, when they know this expert categorization (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). In a related vein, one scene can be an outdoor scene, a city, or New York, depending on the level of category precision (Oliva & Schyns, 2000). Flexible categorizations of objects and scenes at different levels of abstraction have become central to modern theories of categorization and recognition (Tarr & Bülthoff, 1995; Murphy & Lassaline, 1997; Cutzu & Edelman, 1998; Schyns, 1998; Gauthier, Tarr, Moylan, Anderson, Skudlarski, & Gore, 2000; Gosselin & Schyns, 2001).

Such flexible categorizations tend to require different visual information from the same input. For example, the information presented in Fig. 2 (EXNEX and human observer) is sufficient to determine whether the underlying face is expressive or not. However, could you as confidently determine its gender? Fig. 2 (GENDER and human observer) reveals supplementary face information that should improve a gender judgement (i.e. male).

Even though we might have many good intuitions (but fewer data) about the information required for different visual categorizations, there is a need in recognition studies for a principled method that reveals the stimulus information that is diagnostic of a given categorization task. To this end, we introduce Bubbles, a general technique that can assign the credit of a categorization performance to specific visual information.

* Corresponding author. Tel.: +44-141-3304937.

E-mail addresses: gosselif@psy.gla.ac.uk (F. Gosselin), philippe@psy.gla.ac.uk (P.G. Schyns).

PII: S0042-6989(01)00097-9

We will illustrate several properties of Bubbles with three experiments on face stimuli. Starting from one set of male and female faces displaying two expressions (neutral and happy), in experiment 1 Bubbles will isolate the spatial location of the visual cues that are responsible for the gender and expressive categorizations (Fig. 2, human observer, is the outcome of this experiment). The second experiment will investigate the more challenging task of face identity. It will also illustrate the generality of Bubbles by localizing diganostic cues in a larger dimensional (3D) space (2D spatial location x spatial scales). In these two experiments, we will contrast the information humans used with the optimal information available to resolve the tasks. Experiment 3 will seek to show, in a typical recognition experiment, that the identity cues extracted in experiment 2 have a general validity.

From the outset, it is important to emphasize that the aim of this paper is to illustrate the fundamental principles of Bubbles in the context of simple, but nevertheless challenging experiments, not to resolve face gender, expression, and identity in optimal conditions of ecological validity. Moreover, we used faces because they are good stimuli for our illustrations, but the principles of Bubbles should generalize to other objects and scenes.

a.

b.

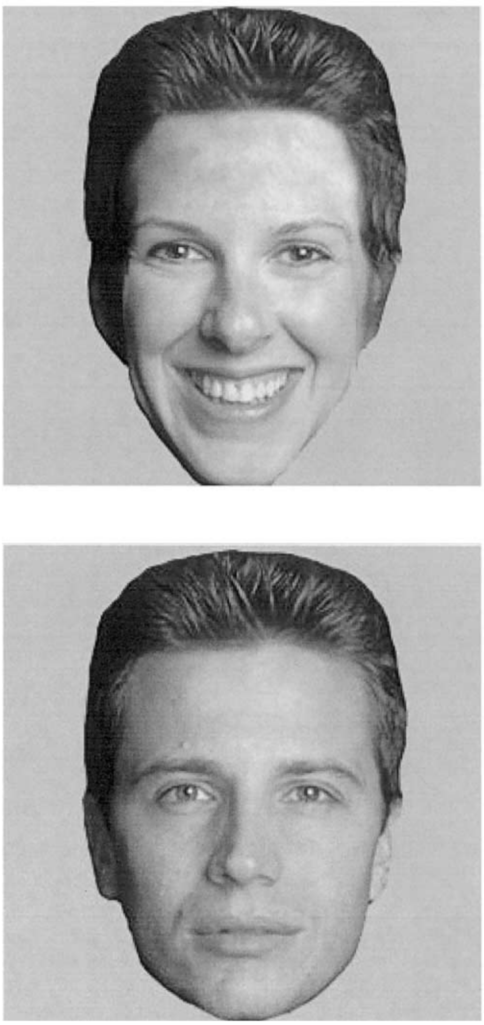

Fig. 1. This figure shows two of the faces used in experiment 1. Note that hairstyle, pose and lighting were normalized.

2. Experiment 1

2.1. Method

All experiments reported in this paper ran on a Macintosh G4 using a program written with the Psychophysics Toolbox for Matlab (Brainard, 1997; Pelli, 1997). Participants were five paid University of Glasgow students, with normal, or corrected to normal vision. In a within-subjects design, each participant was sequentially submitted to two independent tasks (male vs. female, GENDER; and expressive or not, EXNEX) on the same stimulus set. Order of task changed randomly across participants.

Stimuli were computed from the 32 greyscale faces of Schyns and Oliva (1999) (eight males, eight females, each displaying either a neutral or happy expression, with normalized hairstyle, global orientation and lighting, see Fig. 1). Each face was partly revealed by a mid-grey mask punctured by a number of randomly located Gaussian windows (henceforth called ‘bubbles’) with standard deviation of 0.22° of visual angle, see Fig. 3c for examples. We chose bubbles with a Gaussian shape because it is smooth and symmetrical (see Marr, 1982).

During the experiment, the number of bubbles per image was automatically adjusted, using an adaptive procedure, to reveal just enough face information to maintain a 75% correct categorization criterion (Bubbles is a self-calibrating technique). The size of the bubbles and the self-calibration are important aspects of the technique that we will discuss in the results section.

In a given trial, one sparse face computed as described above appeared on the screen. To respond, subjects pressed labelled computer-keyboard keys. It is important to stress that subjects were not under any time pressure to respond and so could freely explore each stimulus. The experiment comprised a total of 512 trials (16 presentations of the 32 faces). A chinrest was used to maintain a constant viewing distance of 100 cm. Stimuli subtended $5.72 \times 5.72°$ of visual angle on the screen.

2.2. Results

An average of 15 and 23 bubbles (S.D. = 4 and 9), respectively, in the EXNEX and the GENDER condi-

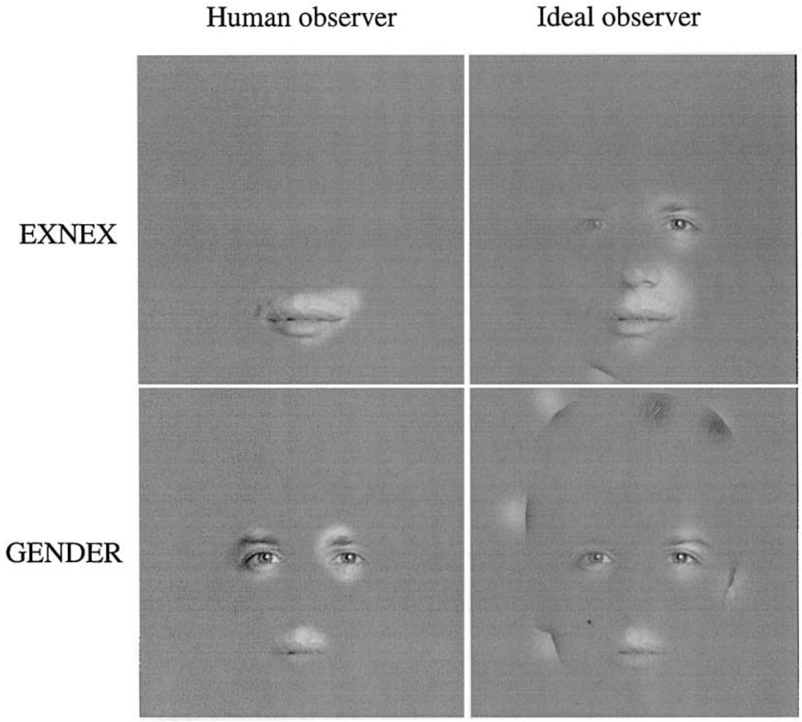

Fig. 2. This figure illustrates diagnostic face information for judging whether a face is expressive or not (EXNEX), or its gender (GENDER). The pictures are the outcome of Bubbles in the EXNEX and GENDER categorizations of experiment 1 on human (left column) and ideal observers (right column).

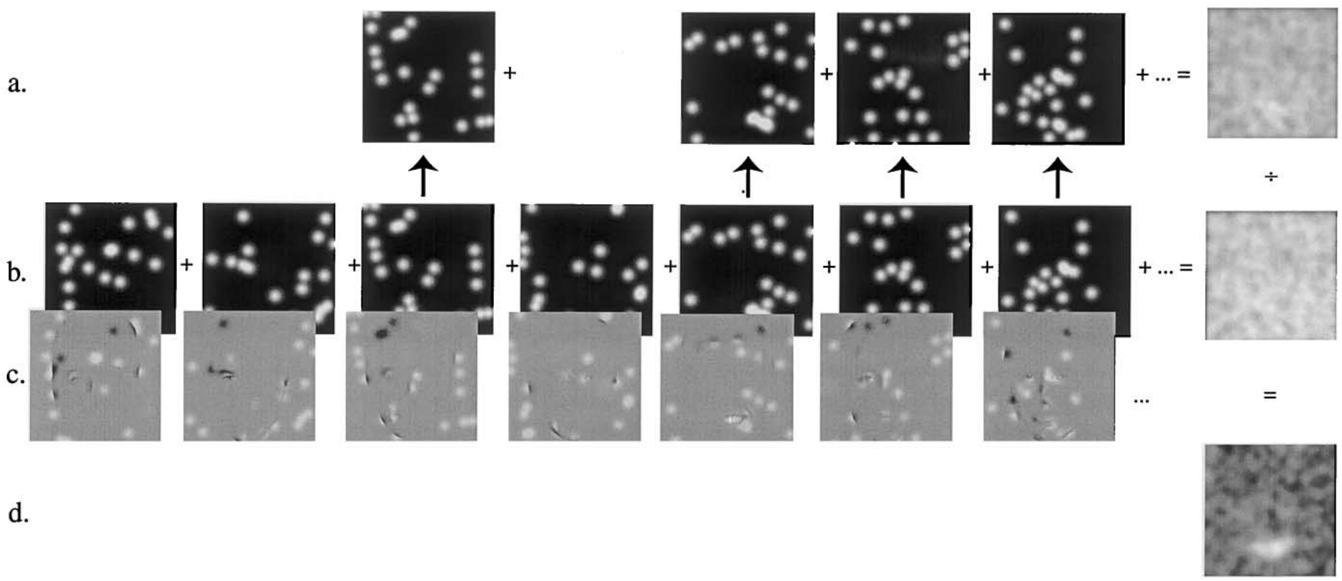

Fig. 3. This figure illustrates Bubbles in experiment 1 for the EXNEX task. In (a), the bubbles leading to a correct categorization are added together to form the CorrectPlane (the rightmost greyscale picture). In (b), all bubbles (those leading to a correct and incorrect categorizations) are added to form TotalPlane (the rightmost greyscale picture). In (c), examples of experimental stimuli as revealed by the bubbles of (b). It is illustrative to judge whether each sparse stimulus is expressive or not. ProportionPlane (d) is the division of CorrectPlane with TotalPlane. Note the whiter mouth area (the greyscale has been renormalized to facilitate interpretation). See Fig. 2 for the outcome of experiment 1.

tions, were required for subjects to reach the performance criteria. A correct response meant that the bubbles (or a subset of them) revealed enough face information to correctly categorize the sparse face. When this happened, we added the mask made of bubbles to CorrectPlane. Across trials, CorrectPlane sums all the masks leading to successful categorizations (see Fig. 3a and c). We also added the successful masks to TotalPlane. Across trials, TotalPlane sums all of the masks, thus summing both the masks, leading to a successful categorization and all the masks leading to a miscategorization (see Fig. 3b).

Remember that the location of all bubbles in a mask changes randomly across trials. They randomly reveal a portion of the tested space (here, the image plane) to an observer who must then use this information for a categorization. Hence, the interaction between the random bubbles and the observer can be depicted as a random search for diagnostic task information. With enough trials, a random search is exhaustive and all the search space is explored.

For each subject, we derived a ProportionPlane by dividing CorrectPlane by TotalPlane. We then computed the mean ProportionPlane of GENDER and EXNEX by averaging across subjects. The averaged ProportionPlane is a measure of the relative importance of the regions of the 2D image for the task at hand. If no region had any special status, ProportionPlane would be homogeneously grey. That is, the probability that the information revealed by any bubble led to a correct categorization would be 0.75, the performance criterion. In contrast, the more diagnostic regions should be significantly above the criterion (i.e. whiter). Fig. 3d illustrates the ProportionPlane of EXNEX. Note the salient region corresponding to the mouth.

To derive the statistical significance of diagnostic regions, we construct around the mean of the ProportionPlane a confidence interval for each proportion ($P < 0.05$). The DiagnosticPlane is a task-specific mask that removes all information below the confidence interval. The DiagnosticPlanes in Fig. 2 were smoothed out with a Gaussian bubble identical to the experimental bubble.

This simple experiment has demonstrated that two distinct categorizations of the same faces do indeed require different visual information. Fig. 2, Human observer, reveals that the mouth is the only diagnostic region of EXNEX, however, the eyes and the center of the mouth are used in GENDER.

At this stage, it is worth expanding on the dynamics of the technique. We stated earlier that Bubbles was a self-calibrating technique. In fact, Bubbles is a gradient-descent algorithm (Hertz, Krogh, & Palmer, 1991) that constantly adjusts the number of bubbles (i.e. the total face area revealed) to minimize an error term — the difference between subject and target performance. This self-calibration has one important side-effect with regards the size of bubbles. Simply put, if subjects require information represented at a scale larger than that of a bubble, the technique will recalibrate and automatically increase the number of bubbles. Consequently, the density of bubbles will increase, they will start to form clusters at larger scales, subjects' performance will improve, and this will in turn stabilize the number of bubbles. This self-calibration implies that Bubbles is relatively insensitive to the size of the bubbles. For example, all the images of Fig. 2 illustrate that the diagnostic masks are much larger than the size of the small bubbles.

2.3. Human versus ideal observer

In Bubbles, the observer determines the informative subset of a randomly, and sparsely sampled search space. To highlight this unique property, we here contrast human and ideal observers (Tjan, Braje, Legge, & Kersten, 1987). The ideal observer will provide a benchmark of the information available in the stimulus set to resolve each task. In the tasks of experiment 1, the ideal will capture all the regions of the image that have highest local variance between the considered categories (male vs. female, and neutral vs. expressive). This ideal considers the stimuli as images (not faces composed of eyes, a nose and a mouth, as humans do) and it might not necessarily be sensitive to the regions that humans find most useful (the diagnostic regions), but rather to the information that is mostly available in the data set for the task at hand.

We constructed a different ideal observer for EXNEX and GENDER and submitted them to Bubbles, using the same parameters as those humans used in experiment 1. Specifically, the number of bubbles was held constant (equal to the average numbers humans required in EXNEX and GENDER, respectively), and we added to the faces a varying percentage of Gaussian white noise to maintain performance at 75% correct. In a winner-take-all algorithm (Hertz et al., 1991), the ideal matched the information revealed by the bubbles of the input with the same bubbles applied to the 32 memorized face pictures. The gender or expression of the best match was the categorization response. CorrectPlanes, TotalPlanes, ProportionPlanes and DiagnosticPlanes were computed as explained before.

Fig. 2 shows that the DiagnosticPlanes of the ideal and human observers are only partially correlated ($r =$ .75 and .55 for the EXNEX and GENDER DiagnosticPlanes, respectively). For GENDER, human and ideal observers use similar information (e.g., the eyes and the central upper part of the mouth). However, the ideal also uses supplementary information from the silhouette of the head. Similarly, for EXNEX, the human and ideal observers both use information around the mouth. However, the ideal also uses lateralized information from the eyes.

In sum, the ideal and human observers revealed that the EXNEX and GENDER tasks require different information from the same face set. The partial correlation between human and ideal use of information demonstrates the unique property of Bubbles: it is a human, partially efficient, not a formal, optimally efficient, feature extraction algorithm.

3. Experiment 2

Experiment 2 applies Bubbles to the more challenging task of face identity. We want to demonstrate that the technique is versatile and can be applied to a more complex, abstract 'image generation' space. Bubbles will here search a 3D space comprising the two dimensions of the image plane and the third abstract dimension of spatial scales.

It is now well established that the identity of faces is represented at multiple spatial scales (see Morrison & Schyns, 2001, for a review). However, research on face recognition has so far lacked a technique that identifies the specific aspects of identity that humans locally represent at different scales. Experiment 2 applies Bubbles to a simple face identification task.

3.1. Method

This application of Bubbles is very similar to that of experiment 1. Participants were twenty paid University of Glasgow students, with normal, or corrected to normal vision. Stimuli were computed from ten of the

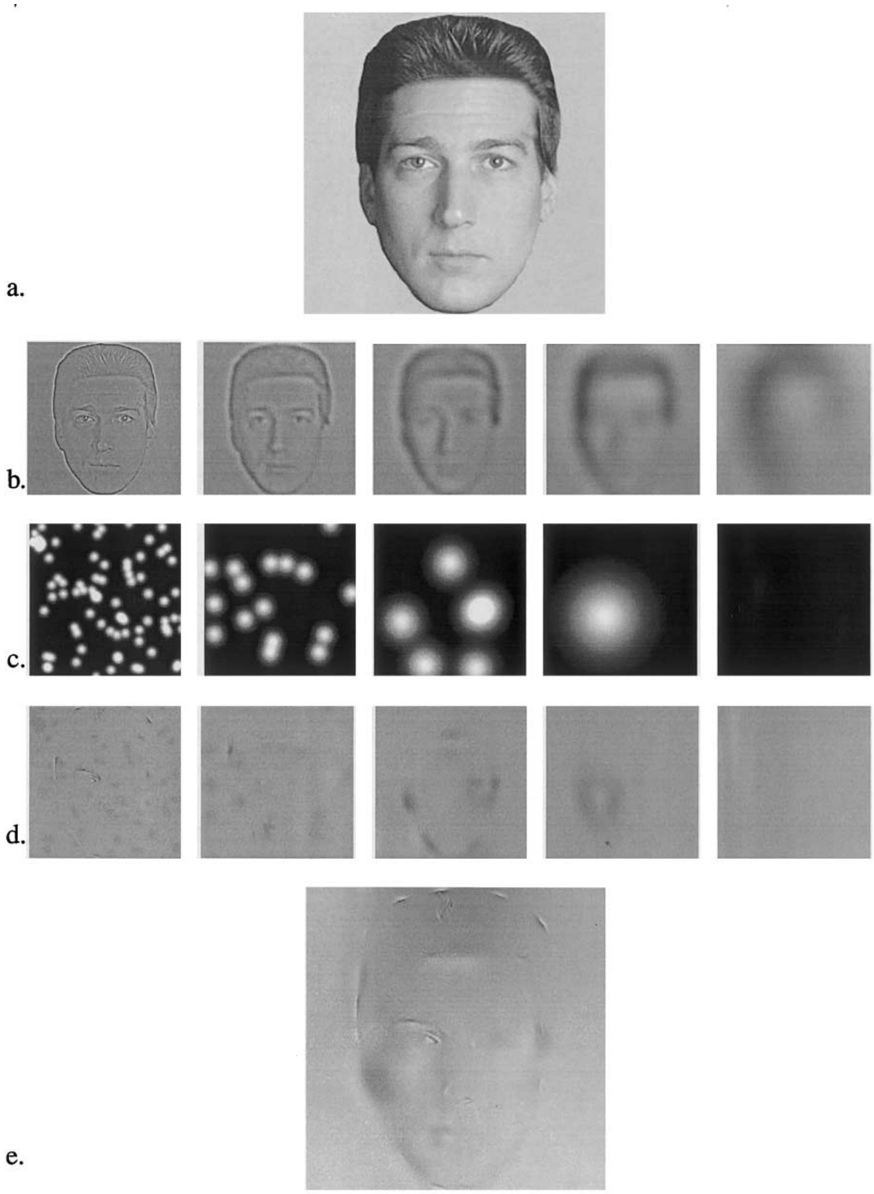

Fig. 4. This figure illustrates the application of Bubbles in experiment 2. Pictures in (b) represent five different scales of (a); (c) illustrate the bubbles applied to each scale; (d) are the revealed information of (b) by the bubbles of (c). Note that on this trial there is no revealed information at the fifth scale. By integrating the pictures in (d) we obtain (e), a stimulus subjects actually saw.

greyscale faces (five males and five females all displaying a neutral expression) used in experiment 1 (see Fig. 4a). Prior to experimentation, all subjects learned to criterion (perfect identification of all faces twice in a row) the name attached to each face from printed pictures with corresponding name at the bottom.

To compute the experimental stimuli, we decomposed the original faces into six bands of spatial frequencies of one octave each — at 2.81, 5.62, 11.25, 22.5, 45 and 90 cycles per face, from coarse to fine (computations were made with the Matlab Pyramid Toolbox, Simoncelli, 1999). The coarsest band served as a constant background, as a prior study revealed that it does not contain face identification information (see Fig. 4b). We applied different Gaussian windows to each one of the five spatial frequency bands, to normalize to 3 the number of cycles per face that any bubble revealed (standard deviations of bubbles were 2.15, 1.08, 0.54, 0.27 and 0.13° of visual angle, from coarse to fine scales, see Fig. 4c). Pilot testing revealed that three cycles per bubble was the smallest integer choice leading to naturalistic sparse faces.

The multiplication of scale-specific face information (Fig. 4b) with its respective bubbles (Fig. 4c) produced the information revealed at each scale (Fig. 4d). To generate an experimental stimulus, we simply added the information revealed at each scale (Fig. 4e). As in experiment 1, the total subspace revealed by the bubbles was self-calibrated to maintain identification of the sparse faces at a 75% correct criterion.

In a given trial, one sparse face appeared on the screen. Subjects identified it by pressing the keyboard-key tagged with the appropriate name. To allow for complete inspection of the revealed information, subjects were under no time pressure to respond. The experiment comprised two sessions of 500 trials each (50 presentations of the ten faces), but we only used the data from the last 500 trials, when subjects were familiar with the faces and experimental procedure. A chin-rest was used to maintain subjects at a constant viewing distance (of 100 cm). Stimuli subtended $5.72 \times 5.72°$ of visual angle on the screen.

3.2. Results

An average of 47 bubbles (S.D. = 16) were needed for subjects to reach the performance criterion. The correct identification of a sparse stimulus indicates that the bubbles used in its construction (or a subset of them) revealed enough information about the face for its identification. In experiment 2, this information can reside at different scales of the same stimulus. To compute CorrectPlane, we must memorize the locations of the bubbles at each scale (i.e. all those of Fig. 4c). To this end, we recorded an independent CorrectPlane for each scale — henceforth called CorrectPlane(scale), with scale = 1 to 5. A similar argument applies to TotalPlane — henceforth, TotalPlane(scale), with scale = 1 to 5. Whenever a stimulus was correctly identified, its bubbles were added to their respective CorrectPlane(scale) and TotalPlane(scale). When the input was misidentified, bubbles were only added to TotalPlane(scale).

To derive diagnostic information, we computed a different ProportionPlane for each scale by dividing CorrectPlane(scale) by TotalPlane(scale), for each subject. We then averaged ProportionPlane(scale) across subjects. The result enables a much finer analysis of information than that of experiment 1: ProportionPlane(scale) weighs the importance of the regions of each scale for face identification. To derive the DiagnosticPlane(scale), we constructed a confidence interval ($P < 0.05$) around the mean of each ProportionPlane(scale), for each proportion. Fig. 5c reveals the diagnostic regions of face identification at different scales.

It is interesting to step back from the computations to observe the interaction between spatial scales and information use. To do this, we multiply the scale information of Fig. 5b with the diagnostic masks of Fig. 5c to derive Fig. 5d. At the finest scale, the eyes and a corner of the mouth appear to stand out (see the leftmost picture in Fig. 5d). At the next to finest scale, the diagnostic information is a mask comprising the eyes, the nose and the mouth. The next scale is consistent with the information that face recognition researchers would call a configural representation of the face. Together, the eyes, the nose, the mouth and the chin appear to form a meaningful recognition unit, but in isolation, these features do not diagnose the identity of the face (Sergent, 1986; Gauthier & Tarr, 1997; Tanaka & Sengco, 1997; Schyns & Oliva, 1999). At the next meaningful scale, the left side of the face silhouette is used. It is worth pointing out that the lighting was always coming from the right side of the faces. Therefore, the left sides of the faces were more shaded and thus more informative. This is also apparent in the third diagnostic plane.

To visualize the diagnostic information of a face identification task, we can now reconstruct the 'effective face.' The effective face (Fig. 5e) is the sum of the face information revealed by the diagnostic filters in Fig. 5d.

3.3. Human versus ideal performance

To compare the human versus ideal features of face identity, we ran an ideal observer similar to that in experiment 1. The ideal was exposed to faces punctured with bubbled masks at different scales (the number of

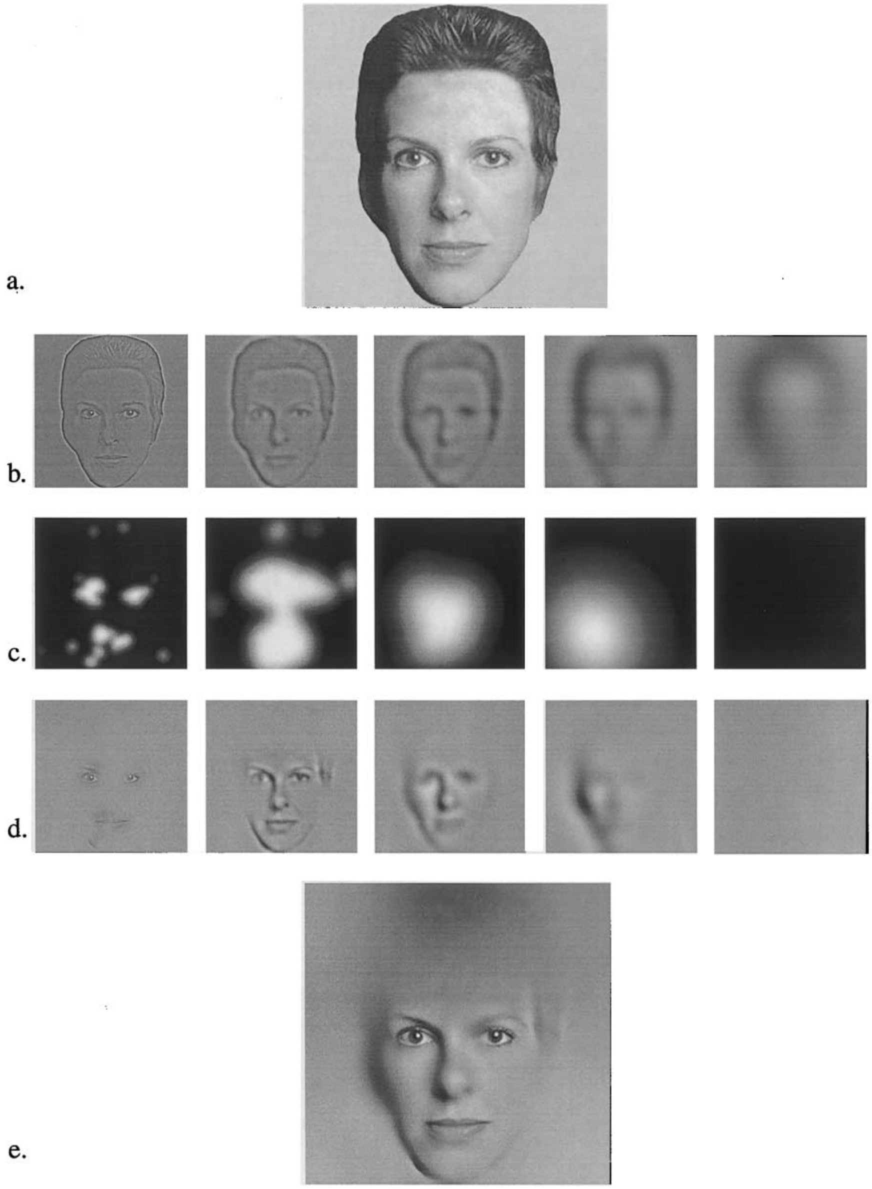

Fig. 5. This figure illustrates the outcome of Bubbles in experiment 2 with human observers. Pictures in (b) represent five scales of (a); (c) represent the statistically significant diagnostic regions at each spatial scale of the face (see discussion in text); (d) multiply (b) with (c). The bottom picture is the effective (or diagnostic) stimulus: a depiction of the information used to identify faces in experiment 2.

bubbles per scale was normalized to the average number humans needed) and correlated the sparse face with the pictures in memory. The best match constituted the categorization response. Performance was maintained at 75% correct by adding a varying percentage of Gaussian white noise to the input face. CorrectPlanes, TotalPlanes, ProportionPlanes and DiagnosticPlanes were computed as explained before. Fig. 6 illustrates the ideal diagnostic masks.

As in experiment 1, the diagnostic masks of the human observers were only partially correlated with those of the ideal ($r = 1$, 0.48, 0.12, 0.01 and 0.05, from coarse to fine scales), revealing again the specific human contribution to the feature extraction process.

4. Experiment 3

Bubbles is a technique that presents sparse stimuli to determine the diagnostic visual information of categorization tasks. This information takes the form of diagnostic masks whose general validity we now turn to. Two separate issues must be addressed. The first one stems from the way that sparse stimuli reveal visual information (i.e. via bubbles). Subjects could adopt an atypical recognition strategy elicited by the presence of local information. In a related vein, stimuli were displayed on the screen for an unlimited time, and this might also have elicited atypical strategies (when compared to typical recognition experiments that restrict presentation time). Consequently, the information revealed by the diagnostic masks might not be used in more typical situations of face recognition (i.e. when face information is complete, not sparse, and briefly, not indefinitely presented).

To address this issue, experiment 3 was set up as a typical recognition task. Subjects had to identify faces presented on the screen for brief, varying durations. Each face could come in one of three possible versions: original, filtered with diagnostic masks, and filtered with nondiagnostic masks (Fig. 7; nondiagnostic masks are simply the complement of the diagnostic masks). The diagnostic masks would be validated if recognition performance was similar for the original faces and for those filtered with diagnostic masks, and was hindered for faces filtered with nondiagnostic masks.

The second issue of validation concerns the restricted number of faces used in experiment 2 to derive the masks. With few faces, the masks might be idiosyncratic to this stimulus set, instead of capturing a more generic information about face identity. If the masks were idiosyncratic then they would not transfer to a new set of faces. That is, they would not reveal the identity information of the new faces. To address this issue, we also ran the recognition task described

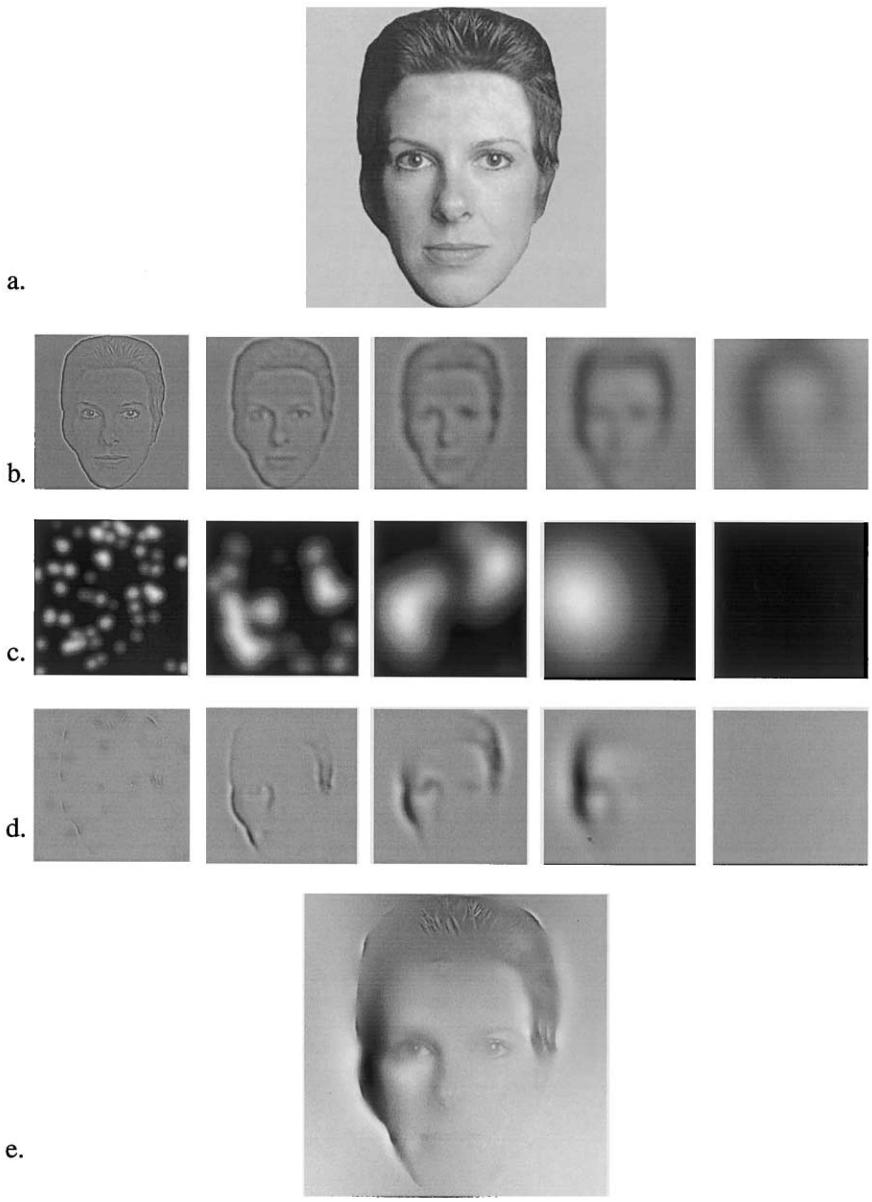

Fig. 6. This figure illustrates the outcome of Bubbles in experiment 2 with an ideal observer. Pictures in (b) represent five scales of (a); (c) represent the statistically significant available regions at each spatial scale of the face; (d) multiply (b) with (c). The bottom picture is the available stimulus: a depiction of the information that is most informative to identify faces in experiment 2.

Fig. 7. This figure illustrates the three conditions of experiment 3: DIAGNOSTIC is a filtered version of the ORIGINAL with the diagnostic masks derived in experiment 2. NONDIAGNOSTIC is the same face filtered with the complement of the diagnostic masks (1 − Diagnostic-Plane(scale), for scale = 1 to 5). Average energy per scale is identical in all conditions of a face.

above on a new set of ten faces (from Gold, Bennett, & Sekuler, 1999a,b). If the masks revealed generic identity information, we would expect a transfer of recognition performance. That is, we expect recognition performance to be similar with the 'new,' and the 'old' face sets, although the diagnostic masks were derived from only the 'old' face set.

4.1. Method

Participants were 20 paid University of Glasgow students, with normal, or corrected to normal vision. They were randomly split between the conditions of OLD and NEW face sets.

For each face set, we computed three versions of each greyscale face: original (ORIGINAL), filtered with diagnostic masks (DIAGNOSTIC), and filtered with non-diagnostic masks (NONDIAGNOSTIC, see Fig. 7). The computation of the DIAGNOSTIC faces was already presented in Fig. 5. By definition of a diagnostic mask, its complement (1 — DiagnosticPlane(scale), for scale = 1 to 5) reveals the less diagnostic information. NONDIAGNOSTIC faces were filtered with these nondiagnostic masks. For each face and scale, we then normalized contrast energy across conditions of ORIGINAL, DIAGNOSTIC and NONDIAGNOSTIC.

In a given trial, one face (either ORIGINAL, DIAGNOSTIC or NONDIAGNOSTIC) appeared on the screen for a varying duration (either 13, 27, 53, 107 or 213 ms). This was immediately followed by a bit noise mask that remained on the screen until subjects responded. Subjects identified the face by pressing the keyboard-key tagged with the appropriate name. In total, there were 450 such trials (ten faces × three types of stimuli × five durations × three repetitions = 450 trials). A chinrest was used to maintain subjects at a constant viewing distance of 100 cm. Stimuli subtended $5.72 \times 5.72°$ of visual angle on the screen.

4.2. Results

To measure recognition performance, we computed the average percent correct identification per subject for each of the five presentation times (13, 27, 53, 107 and 213 ms), in the three conditions of face stimulus (ORIGINAL, DIAGNOSTIC and NONDIAGNOSTIC). This was done for the two conditions of face sets (OLD and NEW). The recognition curves are plotted in Fig. 8, bestfitted with Weibull distributions (smallest $R^2 = 0.88$).

As expected, for both the OLD and NEW face sets, performance with the ORIGINAL and DIAGNOSTIC faces evolved similarly, whereas performance was hindered with NONDIAGNOSTIC faces. Remember that the first goal of experiment 3 was to use a typical time constrained face identification experiment to validate the diagnostic masks which were extracted in conditions of sparse stimulation and unlimited stimulus presentation. The similarity of performance between ORIGINAL and DIAGNOSTIC in contrast to NONDIAGNOSTIC faces validates that the information revealed by the diagnostic masks does drive the process of recognizing full faces under time pressure. When this information was removed in the NONDIAGNOSTIC condition, subjects' performance was significantly hindered across all durations. Note that DIAGNOSTIC faces were consistently better recognized than the ORIGINAL faces. This was expected because the energy normalization reduced the strength of the DIAGNOSTIC subspace of the ORIGINAL image. If people must use the diagnostic information, the ORIGINAL stimulus must then be less effective than the DIAGNOSTIC stimulus.

A comparison between the three curves for OLD and NEW face sets reveals that the evolution of performance was very similar, but scaled down for the NEW set. This probably occurred because the NEW faces were more similar (subjects took more time to learn them). The comparison between OLD and NEW suggests that the masks captured generic information about face identity, not only the idiosyncrasies of the OLD face set.

In sum, experiment 3 was designed to validate the generality of the diagnostic masks derived by Bubbles in conditions of a restricted stimulus set, sparsely presented, for an unlimited duration. In a recognition experiment where full faces were presented under time pressure, we found similar performance for the original faces and those filtered by the diagnostic masks. Moreover, we found a degradation of performance with nondiagnostically filtered faces. This pattern was replicated on a new set of faces. Together, the evidence suggests that the masks derived from Bubbles captured generic information for face identification.

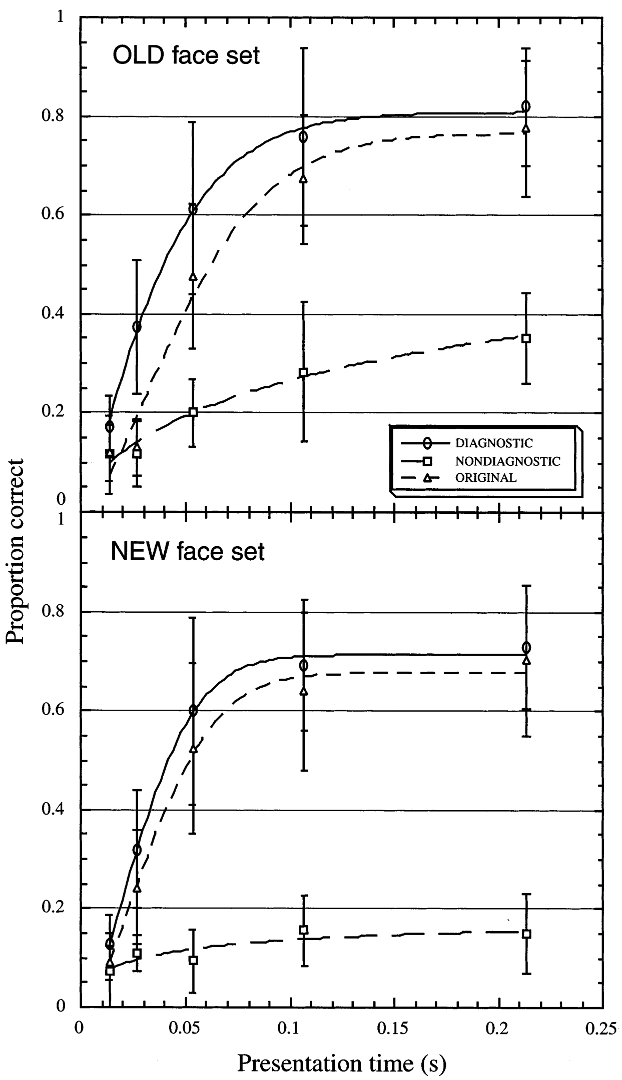

Fig. 8. This figure illustrates the results of experiment 3. The curve plots the average proportions of correct responses (with error bars) against presentation time in DIAGNOSTIC, ORIGINAL, and NONDIAGNOSTIC using the OLD face set, and the NEW face set (from Gold et al., 1999a, b).

5. Concluding remarks

Experiments 1 and 2 have demonstrated that Bubbles can be used to isolate the diagnostic information of face recognition tasks. Experiment 3 validated that the masks of experiment 2 captured generic identity information. Bubbles applied to human and ideal observers produced different diagnostic masks, and so it is advisable to use a method based on human performance to derive the features humans use. Note that the principles of Bubbles are not limited to faces but are also applicable to other object and scene categorizations. The technique is a human search for diagnostic features in any specified n-dimensional image generation space, even if the space is abstract.

Acknowledgements

The authors wish to thank Dr Paula Niedenthal from the Psychology Department at Indiana University for lending us the original face stimuli that were used in our experiments. Thanks also to Lizann Bonnar for having helped us with running the experiments.

References

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

Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L., & Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition*, *3*, 81–117.

Cutzu, F., & Edelman, S. (1998). Representation of object similarity in human vision: Psychophysics and a computational model. *Vision Research*, *38*, 2229–2257.

Etcoff, N. L., & Maggee, J. J. (1992). Categorical perception of facial expressions. *Cognition*, *44*, 227–240.

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., Moylan, J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (2000). Does visual subordinate-level categorisation engage the functionally defined fusiform face area? *Cognitive Neuropsychology*, *17*, 143–163.

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999a). Signal but not noise changes with perceptual learning. *Nature*, *402*, 176–178.

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999b). Identification of band-pass filtered faces and letters by human and ideal observers. *Vision Research*, *39*(21), 3537–3560.

Gosselin, F., & Schyns, P. G. (2001). Why do we slip to the basic level? Computational constraints and their implementation. *Psychological Review* (in press).

Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the theory of neural computation*. Redwood City, CA: Addison-Wesley.

Marr, D. (1982). *Vision*. New York: Freeman.

Morrison, D., & Schyns, P. G. (2001). Usage of spatial scales for the categorization of faces, objects and scenes: a review. *Psychological Bulletin and Review*, in press.

Murphy, G. L., & Lassaline, M. E. (1997). Hierarchical structure in concepts and the basic level of categorization. In K. Lamberts, D. R. Shanks, et al., *Knowledge, concepts and categories: studies in cognition* (pp. 93–131). Cambridge, MA: MIT Press.

Oliva, A., & Schyns, P. G. (2000). Colored diagnostic blobs mediate scene recognition. *Cognitive Psychology*, *41*, 176–210.

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

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 352–382.

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

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. (1986). Microgenesis of face perception. In D. H. Ellis, M. A. Jeeves, F. Newcombe, & A. Young, *Aspects of face processing* (pp. 17–73). Dordrecht: Martinus Nijhoff.

Simoncelli, E. P. (1999). *Image and multi-scale pyramid tools* [*computer software*]. New York: Author.

Tanaka, J., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory and Cognition*, *25*, 583–592.

Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1494–1505.

Tjan, B. S., Braje, W. L., Legge, G. E., & Kersten, D. (1987). Human efficiency in for recognizing 3-D objects in luminance noise. *Vision Research*, *35*, 3053–3069.