Research Report

Gender Recognition of Human Faces Using Color

Adrian Nestor and Michael J. Tarr

Brown University

ABSTRACT—***A continuing question in the object recognition literature is whether surface properties play a role in visual representation and recognition. Here, we examined the use of color as a cue in facial gender recognition by applying a version of reverse correlation to face categorization in CIE L*a*b* color space. We found that observers exploited color information to classify ambiguous signals embedded in chromatic noise. The method also allowed us to identify the specific spatial locations and the components of color used by observers. Although the color patterns found with human observers did not accurately mirror objective natural color differences, they suggest sensitivity to the contrast between the main features and the rest of the face. Overall, the results provide evidence that observers encode and can use the local color properties of faces, in particular, in tasks in which color provides diagnostic information and the availability of other cues is reduced.***

The contribution of color to face recognition has been largely dismissed (Bruce & Young, 1998; Kemp, Pike, White, & Musselman, 1996) or restricted to low-level processing such as image segmentation (Yip & Sinha, 2002). This mirrors the traditional downplay of color cues in visual object recognition (Biederman & Ju, 1988; Davidoff & Ostergaard, 1988) and their alleged restriction to early segmentation processing (Wurm & Legge, 1993). More recently, this view has been challenged by studies relating the use of color in scene and object recognition to task diagnosticity and cue reliability (Naor-Raz, Tarr, & Kersten, 2003; Oliva & Schyns, 2000; for a review of this debate, see Tanaka, Weiskopf, & Williams, 2001). In particular, the results of these studies support a role for color in the visual representation of objects and provide evidence for the use of color in high-level visual processing.

Address correspondence to Michael J. Tarr, Brown University, Cognitive and Linguistic Sciences, 190 Thayer St. (CLS), Providence, RI 02912, e-mail: Michael_Tarr@brown.edu.

Similarly, the contribution of surface properties has been noted in studies of visual face processing (Bruce & Langton, 1994; O'Toole, Vetter, & Blanz, 1999). However, it is less clear whether this contribution is mainly an aspect of low-level face processing, notably segmentation, or reflects a genuine advantage of encoding surface properties into high-level object representations. Our present study provides evidence in favor of the latter, and takes a further step in attempting to unravel the specific role of color information as used in a particular face categorization task: gender recognition.

As a special case of face processing, gender categorization has been extensively studied, with much of this research aimed at sorting out the contribution of various visual cues. Unsurprisingly, shape, both in-plane and three-dimensional, provides significant information about gender that human observers are able to exploit (Bruce et al., 1993; Burton, Bruce, & Dench, 1993). However, there is also evidence that observers are able to exploit other types of cues, including surface properties (Bruce et al., 1993; Bruce & Langton, 1994) and motion (Hill & Johnston, 2001). Beyond these general findings, the specific surface properties that support facial gender recognition have also been investigated. For instance, the luminance contrast between the main features—the eyes and the mouth—and the rest of the face appears to generate a pattern more typical of female rather than male faces (Russell, 2003). As far as color is concerned, although there is some evidence that it does facilitate gender recognition, it is less clear whether the chrominance components of color provide additional help beyond information from luminance alone (Hill, Bruce, & Akamatsu, 1995). However, some evidence has been presented that the global red:green ratio within faces can serve as a reasonably reliable cue to gender (Tarr, Kersten, Cheng, & Rossion, 2001)—the caveat being that its contribution was restricted to conditions in which the contribution of shape information was minimized by blurring.

Taken together, the results reviewed above raise an array of questions regarding the concrete pattern of color differences used by human observers to categorize gender, the relative contribution of different color channels to gender categorization,

 Copyright © 2008 Association for Psychological Science

and the extension of color-based recognition to different viewing conditions. To address these issues, we adopted a version of the reverse correlation method (Beard & Ahumada, 1998) known as superstitious perception (Gosselin & Schyns, 2003). This method is unique in that it requires observers to categorize stimuli composed (unknowingly to them) entirely of noise or, in our case, of ambiguous signals embedded in noise (Kontsevich & Tyler, 2004; Mangini & Biederman, 2004; Martin-Malivel, Mangini, Fagot, & Biederman, 2006). One of the merits of this method, critical for the goals of our study, is that it allows the experimenter to examine and characterize the internal representations of the categories discriminated by observers. The properties of these internal representations are revealed by the classification images constructed from the experimental stimuli sorted according to the observers' responses.

This method was applied to color face images represented in CIE L*a*b*, the color space that best approximates the way we perceive and process chromatic information (Brainard, 2003). The three dimensions of this space correspond to the three color channels of the human visual system: luminance (L*), red:green (a*), and yellow:blue (b*). It is important to note that the euclidean distance between two points in CIE L*a*b* space mirrors the color differences perceived by humans (Wyszecki & Stiles, 1982). In the present study, the involvement of color in facial gender recognition was assessed by constructing and analyzing classification images separately for the three color channels as represented by the three dimensions of this color space.

METHOD

Three young adults (J.G., A.G., and S.L.) were recruited from the Brown University community to participate in the experiment in exchange for pay. All reported normal or corrected-to-normal vision and no color blindness.

A single androgynous base face image was computed from the Max Planck Institute, Tübingen (MPIK) color face data set (the current version of the database is available at http://faces.kyb.tuebingen.mpg.de; see Fig. 1). This database contains 200 faces (100 males, 100 females), with one frontal, color image per individual. The stimuli were collected under controlled, consistent lighting conditions. All subjects have a neutral expression, and none of them is wearing makeup, glasses, or other accessories. The faces have no hair or facial hair other than stubble. We further normalized the MPIK faces with the position of the eyes and the nose and reduced the resolution of the faces to 46×38 pixels. The androgynous base face image was obtained by averaging the resulting faces along with their mirror symmetric versions.

To construct a given stimulus, white Gaussian noise was added independently to the L*, a*, and b* components of the base image. Noise was sampled from a distribution with standard deviation matching the average pixel-wise standard deviation found across MPIK faces for the L* component (euclidean distance in CIE L*a*b* space: $\Delta E^*_{ab} = 11.7$). Although this leads to larger variations for the a* and b* components than those found in the original MPIK images, it allows equating the contribution of the three components in the design of the stimuli. The pattern of noise added to the images was symmetric with respect to the facial symmetry axis, thus reducing the dimensionality of the search space. Each stimulus spanned $5.2^\circ \times 6.1^\circ$ from a distance of 75 cm after doubling the size of the image by pixel replication.

We note that, although stimulus construction was performed in CIE L*a*b* color space, in order to display images on a conventional monitor, L*a*b* values had to be converted to RGB color space. To deal with this, values in L*a*b* space were manipulated so as to ensure that different L*a*b* triplets corresponded to different displayable RGB triplets. Stimuli were presented on a gamma-corrected Sony Trinitron 20-in. monitor. Stimulus design and presentation relied on Matlab 7.2 and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) running on a Dell PC.

Observers were told that the images would contain either a male or a female face and were asked to categorize each face as male or female. Every observer completed 20,000 trials over the course of 10 1-hr experimental sessions.

RESULTS

The observers categorized the stimuli as male on 48% (J.G.), 52% (A.G.), and 63% (S.L.) of the trials. Classification images per observer and per component were constructed by subtracting the

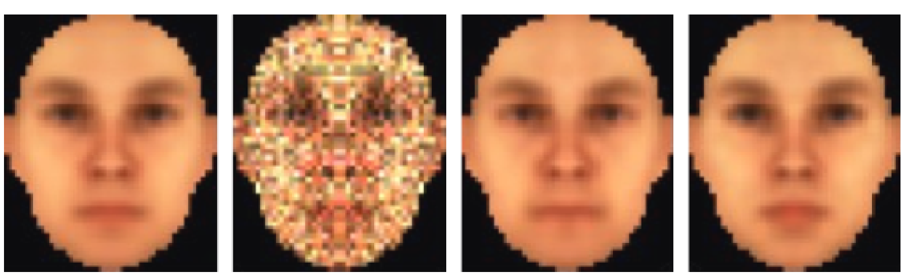

Fig. 1. From left to right: androgynous base image, one example of an experimental stimulus (base image plus noise), and male and female prototypes reconstructed from classification results across observers.

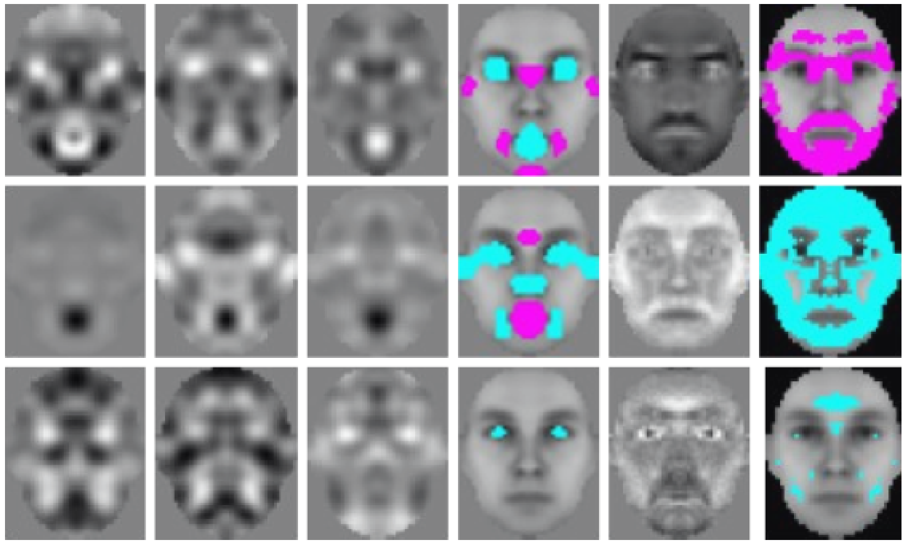

Fig. 2. Classification images for gender categorization and their analysis. Results are separately presented for L*, a*, and b*, from top to bottom. Each row displays classification images for the 3 observers (J.G., A.G., and S.L.), significant clusters across the 3 observers superimposed on a gray-scale version of the base image, a map of objective color differences, and the analysis of this map. Lighter areas in classification images mark regions that are brighter, redder, or more yellow in males than in females, and darker areas indicate regions that are brighter, redder, or more yellow in females than in males. Cyan and magenta mark significant clusters in the same directions for males and females, respectively.

average noise patterns of the stimuli classified as females from those classified as males. The three L*, a*, and b* images for each observer represent an approximation across space and color channels of the layout of information driving stimulus categorization (Fig. 2). These images were smoothed with a Gaussian filter with full width at half maximum of 5 pixels to allow their analysis with random field theory-based tests (Chauvin, Worsley, Schyns, Arguin, & Gosselin, 2005). Note that our data analyses were performed only on the left half of the images (46×19 pixels) in that the right half of each image was a mirror replica of the left and thereby contained no additional information.

Classification images were analyzed using the pixel method (Chauvin et al., 2005), a technique that takes into account the spatial contiguity of the information in classification images. Significant clusters ($p_{\text{rep}} < .05$) distinguishing males from females across the three channels are shown for classification images obtained by pooling results from the 3 observers. To reconstruct male prototypes from the experimental data, we added the three L*a*b* classification images to the base image (Fig. 1); to reconstruct female prototypes, we subtracted the three L*a*b* classification images from the base image.

Maps of objective color differences for each channel were derived by subtracting the average female face from the average male in the MPIK data set. The analysis of these differences was performed using pixel-wise pair-wise comparisons between the male and female face averages ($p < .05$, Bonferroni-corrected for multiple comparisons).

Cross-correlations of smoothed classification images and objective color-difference maps are displayed in Figure 3 (left panel). Although some correlations across the three channels reach significance ($p < .01$), most values are small, which is typical of those found for classification images using faces (Sekuler, Gaspar, Gold, & Bennett, 2004). Furthermore, one of the classification images was negatively correlated with the objective color-difference map (the b* map of observer S.L.).

Significant areas revealed by the pixel test were relatively stable across subjects (for this reason, we only present the analyzed group maps). However, individual differences do exist, especially at the level of the overall patterns revealed by classification images. The (dis)similarity of these patterns was quantified by computing pair-wise correlations between the classification images of different subjects (Fig. 3, right panel). Interestingly, these correlations yielded larger positive values than the one computed between the objective color-difference maps and the classification images, suggesting that the observers are more in agreement with each other than with the objective color differences derived from the original face stimuli.

Examination of objective color-difference maps reveals diagnostic information across all color channels. Objectively, males tend to be darker and redder than females, but not across the mouth and the eyes. In contrast, our observers seem to focus on these features, where objectively we find less of a difference. For instance, females are judged to have darker, bluer eyes and redder lips, whereas the red:green ratio is larger in males around the lower nose. This focus on main facial features makes human and objective color maps visibly dissimilar from each other and suggests poor use by observers of objective absolute color differences between male and female faces. However, the results

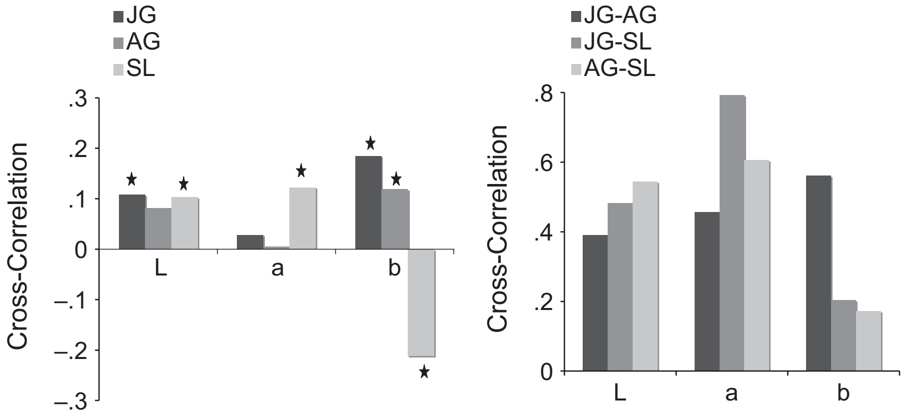

Fig. 3. Cross-correlation of smoothed classification images with objective color-difference maps (left panel; asterisks mark significant correlations, $p < .01$) and pair-wise correlations between the classification images of different subjects (right panel; all correlations significant, $p < .01$).

are consistent with a recognition process based on the contrast between the main features and the rest of the face, such as the luminance contrast between the eyes and the rest of the face, which tends to be larger in females (Russell, 2003).

Along with color differences, we also note differences in the shape of the features. For instance, females are perceived to have narrower, rounder mouths, whereas males have more protuberant noses.

In sum, our results show that observers are sensitive to local color differences when the availability of other cues, notably shape, is reduced by degrading the images. Critically, the independence of the noise masks for the three channels precludes the possibility that categorization was driven exclusively by luminance. This finding suggests that the use of color in facial gender recognition is a real phenomenon, at least when shape is unavailable or degraded. However, the classification images of the observers do not accurately represent the natural color differences between genders.

DISCUSSION

The role of surface properties in visual representation and recognition has been a topic of significant debate in the object recognition literature and especially in the face recognition literature. Here, we posited that color, because it is relatively stable across facial gender, can be an effective cue for gender classification. Critically, because we used a paradigm in which no learning was possible, we argue that any significant role for color in either task must have arisen from observers' pre-existing mental representations of the categories in question.

Overall, our results indicate a role for color information in gender recognition. Interestingly, the local use of color does not always mirror the diagnosticity pattern revealed by objective color differences. In particular, human observers tend to focus on the photometric properties of the main facial features, the eyes and the mouth, even though actual differences are found almost everywhere except across the main features! This difference is resolved by noting that the contrast between these features and the rest of the face is relatively preserved with respect to the diagnostic pattern. It is important to note that sensitivity to this type of contrast has previously been found for luminance in gender categorization with undegraded stimuli (Russell, 2003) and is also consistent with a general face-recognition strategy based on the contrast between pairs of features (Balas & Sinha, 2006). Furthermore, our results suggest such a strategy extends to other color channels beyond luminance. We also note that the color contrast between the main features and the rest of the face can be augmented by the application of makeup in a direction that makes female faces more attractive (Russell, 2003) and, thus, closer to the female face prototype our classification images aim to reconstruct.

Another point regarding our objective color-difference maps is that they do not represent the only true difference between the categories discriminated. This would be the case if observers discriminated two different faces embedded in noise after becoming familiar with the noise-free versions of those faces, as in Sekuler et al. (2004). One problem faced by such an approach is that it runs the risk of testing image differences rather than face differences. In contrast, superstitious perception, relying on a match with previously acquired internal templates, eliminates this risk. However, superstitious perception introduces relativity because of observers' pre-existing biases. For example, different expectations regarding pose, expression, the use of cosmetics, or the presence of facial hair can all affect the internal representations tested. Despite the variability introduced by such factors, we found relative agreement between the classification images of different observers.

Finally, we note that the contribution of color to face recognition was assessed in conditions of noise. Previous results (Tarr et al., 2001) suggest that color tends to be ignored in gender recognition under optimal viewing conditions. Thus, we are not claiming that color is used exclusively or even as the primary cue for any particular recognition task on a regular basis, only that it is an available, mentally represented diagnostic cue that can be exploited when needed.

In sum, our results establish that observers do encode specific patterns of color information about faces, including but not limited to luminance, and that, when needed, they can use this information for facial gender categorization.

Acknowledgments—This research was supported by National Science Foundation (NSF) Award 0339122 and by the Temporal Dynamics of Learning Center (NSF Science of Learning Center SBE-0542013).

REFERENCES

Balas, B.J., & Sinha, P. (2006). Region-based representations for face recognition. *ACM Transactions on Applied Perception*, *3*, 354–375.

Beard, B.L., & Ahumada, A.J. (1998). A technique to extract the relevant features for visual tasks. In B.E. Rogowitz & T.N. Pappas (Eds.), *Human vision and electronic imaging III (SPIE Proceedings Vol. 3299)*. Bellingham, WA: International Society for Optical Engineering.

Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology*, *20*, 38–64.

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

Brainard, D.H. (2003). Color appearance and color difference specification. In S.K. Shevell (Ed.), *The science of color* (pp. 191–216). Washington, DC: Optical Society of America.

Bruce, V., Burton, A.M., Hanna, E., Healey, P., Mason, O., Coombes, A., et al. (1993). Sex discrimination: How do we tell the difference between male and female faces? *Perception*, *22*, 131–152.

Bruce, V., & Langton, S. (1994). The use of pigmentation and shading information in recognizing the sex and identities of faces. *Perception*, *23*, 803–822.

Bruce, V., & Young, A. (1998). *In the eye of the beholder: The science of face perception*. New York: Oxford University Press.

Burton, A.M., Bruce, V., & Dench, N. (1993). What's the difference between men and women? Evidence from facial measurement. *Perception*, *22*, 153–176.

Chauvin, A., Worsley, K.J., Schyns, P.G., Arguin, M., & Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *Journal of Vision*, *5*, 659–667.

Davidoff, J.B., & Ostergaard, A.L. (1988). The role of colour in categorical judgements. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *40*, 533–544.

Gosselin, F., & Schyns, P.G. (2003). Superstitious perceptions reveal properties of internal representations. *Psychological Science*, *14*, 505–509.

Hill, H., Bruce, V., & Akamatsu, S. (1995). Perceiving the sex and race of faces: The role of shape and color. *Proceedings of the Royal Society of London Series B: Biological Sciences*, *261*, 367–373.

Hill, H., & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Current Biology*, *11*, 880–885.

Kemp, R., Pike, G., White, P., & Musselman, A. (1996). Perception and recognition of normal and negative faces: The role of shape from shading and pigmentation cues. *Perception*, *25*, 37–52.

Kontsevich, L.L., & Tyler, C.W. (2004). What makes Mona Lisa smile? *Vision Research*, *44*, 1493–1498.

Mangini, M.C., & Biederman, I. (2004). Making the ineffable explicit: Estimating the information employed for face classifications. *Cognitive Science*, *28*, 209–226.

Martin-Malivel, J., Mangini, M.C., Fagot, J., & Biederman, I. (2006). Do humans and baboons use the same information when categorizing human and baboon faces? *Psychological Science*, *17*, 599–607.

Naor-Raz, G., Tarr, M.J., & Kersten, D. (2003). Is color an intrinsic property of object representation? *Perception*, *32*, 667–680.

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

O'Toole, A.J., Vetter, T., & Blanz, V. (1999). Three-dimensional shape and two-dimensional surface reflectance contributions to face recognition: An application of three-dimensional morphing. *Vision Research*, *39*, 3145–3155.

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

Russell, R. (2003). Sex, beauty, and the relative luminance of facial features. *Perception*, *32*, 1093–1107.

Sekuler, A.B., Gaspar, C.M., Gold, J.M., & Bennett, P.J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology*, *14*, 391–396.

Tanaka, J., Weiskopf, D., & Williams, P. (2001). The role of color in high-level vision. *Trends in Cognitive Sciences*, *5*, 211–215.

Tarr, M.J., Kersten, D., Cheng, Y., & Rossion, B. (2001). It's Pat! Sexing faces using only red and green [Abstract]. *Journal of Vision*, *1*, 337.

Wurm, L.H., & Legge, G.E. (1993). Color improves object recognition in normal and low vision. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 899.

Wyszecki, G., & Stiles, W.S. (1982). *Color science: Concepts and methods, quantitative data and formulae*. New York: Wiley.

Yip, A.W., & Sinha, P. (2002). Contribution of color to face recognition. *Perception*, *31*, 995–1003.

(RECEIVED 2/14/08; REVISION ACCEPTED 6/21/08)