Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research 44 (2004) 1493–1498
Vision Research
www.elsevier.com/locate/visres

What makes Mona Lisa smile?

Leonid L. Kontsevich *, Christopher W. Tyler

Smith-Kettlewell Eye Research Institute, 2318 Fillmore Street, San Francisco, CA 94115, USA

Received 19 March 2002; received in revised form 1 October 2003

Abstract

To study the ability of humans to read subtle changes in facial expression, we applied reverse correlation technique to reveal visual features that mediate understanding of emotion expressed by the face. Surprising findings were that (1) the noise added to a test face image had *profound* effect on the facial expression and (2) in almost every instance the new expression was meaningful. To quantify the effect, we asked naïve observers to rank the face of Mona Lisa superimposed with noise, based on their perception of her emotional state along the sad/happy dimension. Typically, a hundred trials (with 10 or more samples for each rank category) were sufficient to reveal areas altering the facial expression, which is about two orders of magnitude less than in the other reverse correlation studies. Moreover, the perception of smiling in the eyes was solely attributable to a configurational effect projecting from the mouth region.

1. Introduction

Reading facial expressions is of high importance for humans as social beings. Expressions are often the result of subtle changes in facial features, so humans indeed develop impressive sensitivity to these changes (Webster & MacLin, 1999). Over the centuries, artists have excelled in depicting facial expressions, although it is often hard to formulate explicitly the exact changes that make a particular expression (Ekman & Friesen, 1984; Ekman, Friesen, & O'Sullivan, 1988; Hess, 1975). We discovered that added noise has profound effect on the facial expression, which seemed to have a meaningful interpretation in almost every instance. Fig. 1 shows a randomly picked sequence of eight noise instances to illustrate how the expression can be (although the present illustration lacks the strength and vividness of the images shown in the experiments on monitor screen due to grain distortion and smaller contrast range). In the experiments the effect also was enhanced by animated changes of the expression when a new image replaces the previous one in the same aperture.

We combined this effect of noise on facial expressions with a spatial reverse correlation technique (Ahumada & Lovell, 1971; DeAngelis, Ohzawa, & Freeman, 1995; Neri, Parker, & Blakemore, 1999; Ringach, Hawken, & Shapley, 1997; Sutter, 1975), which here provides a method for studying the spatial features distinguishing emotions in facial expressions. The results demonstrate that the method has great potential to identify features of which the observer may have only an intuitive awareness.

2. Methods

In the reverse correlation experiments, the observers were asked to rank the emotional expression of the Mona Lisa face modified by noise into four categories: SAD, SLIGHTLY SAD, SLIGHTLY HAPPY, HAPPY. The noise samples from each trial were accumulated separately for each category. Each of 12 normally-sighted observers conducted 100–120 trials, with the results averaged across the observers.

The portrait of Mona Lisa was chosen because it is the best-known example of an expression at the ambiguity point between a happy and a sad dimension. The reverse correlation experimental results to be described had been replicated with a photograph depicting a face with subtle expression. However, the person depicted declined to give permission to publish her photograph, so that our illustrations will be limited to the Mona Lisa portrait only.

* Corresponding author.
E-mail addresses: lenny@ski.org (L.L. Kontsevich), cwt@ski.org (C.W. Tyler).
URL: http://www.ski.org/cwt.

0042-6989/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2003.11.027

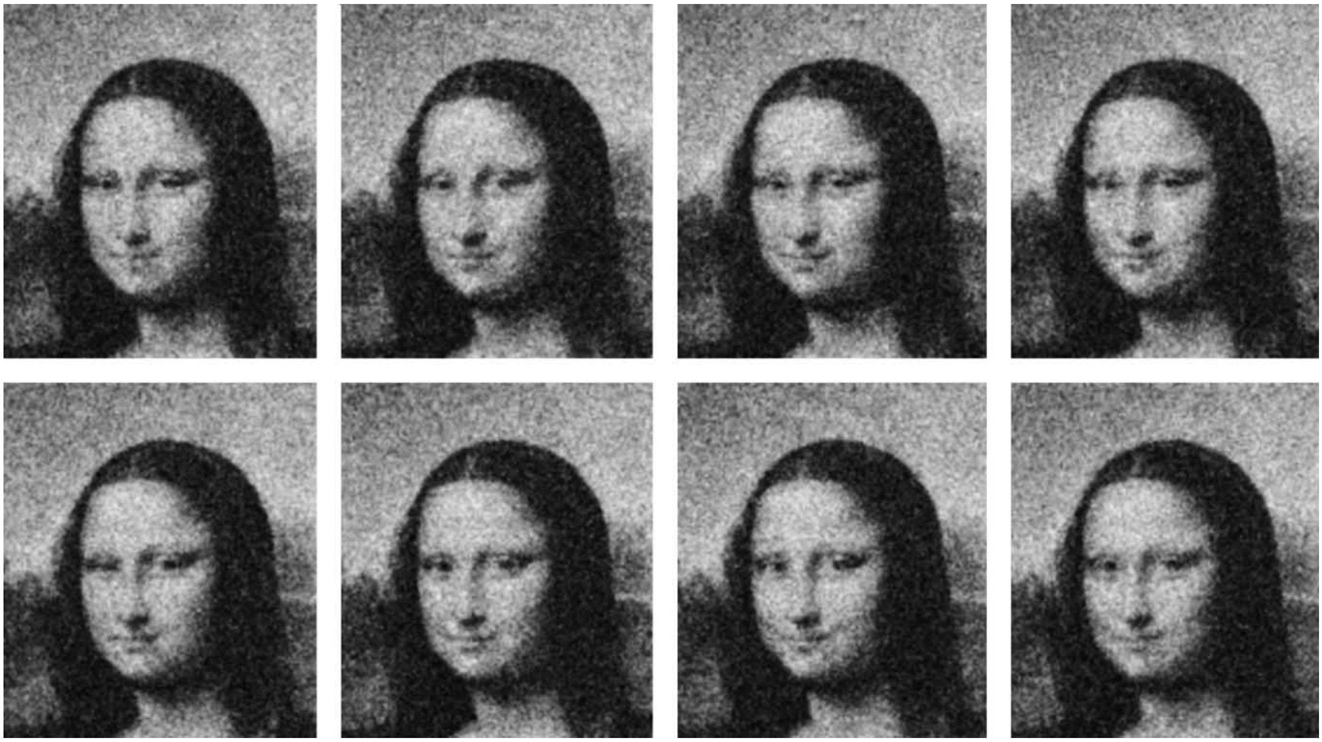

Fig. 1. The base stimulus was a gray-scale detail of the Mona Lisa painting by Leonardo da Vinci, superimposed with binary, one-pixel noise to alter the facial configuration of the original face. A series of eight examples as they were generated by the experimental program are shown to exemplify the range of emotional expressions induced by the addition of the noise.

The images were presented on a monitor screen at 114 cm distance with dimensions 10.3 cm width by 11.8 cm height (260 by 300 pixels). The monitor look-up table was linearized for 256 gray levels. In this linear scale, the added noise had a uniform distribution in the range between −50 and +50, and it was added to the gray-scale image whose luminance was scaled between 51 and 206.

Summation of the noise instances coherently accumulates luminance information in the locations relevant to a particular facial expression and tends to average it out in the irrelevant locations. It should be noted that not all features can be revealed by this noise-averaging effect, which is sensitive only to the features determined by luminance at stable locations. Features that do not "lock" a particular luminance at a particular location, such as wrinkles, will tend to be averaged out in the same way as the noise in irrelevant locations, so the method will selectively enhance average location-specific features.

To evaluate the specific effects in the eye region, we then presented the cumulated noise for the HAPPY and SAD categories (1) over the whole image, and (2) only in the upper part of the face. Twelve observers were asked to rank the difference in degree of happiness specifically in the region of the eyes on a 21-point scale from 'VERY SAD' (−10) to 'VERY HAPPY' (10) with 'NEUTRAL' at zero.

3. Results

When the result of the reverse correlation for a particular category is added to the original face image, it changes the facial expression toward the average expression from that category. On the original Mona Lisa face reproduced in Fig. 2A, the effect of the reverse correlation noise for the two extreme categories, i.e., the SAD and HAPPY facial expressions averaged for all observers, are shown in Fig. 2B and C, respectively. The percent of instances for SAD and HAPPY category for all 12 observers is given in Table 1.

The most prominent difference between the altered images and the original is in the mouth shape: in the sad face the mouth is flat (Fig. 2B) and in the happy face it is curved upward at the corners (Fig. 2C). This feature hardly can be considered surprising, although it is impressive that the reverse correlation method extracts this feature in about only 100 trials (and only about 10 examples in the HAPPY category). The cumulated noise for SAD and HAPPY judgments is shown in Figs. 2D and E correspondingly. Red ovals correspond to the outline of Mona Lisa's face. Somewhat complimentary features for the SAD and HAPPY noises may be seen on the two sides of the mouth.

Many observers who looked at the altered images noted an apparent change in the eye expression. There is also a possibility that there are other locations revealed

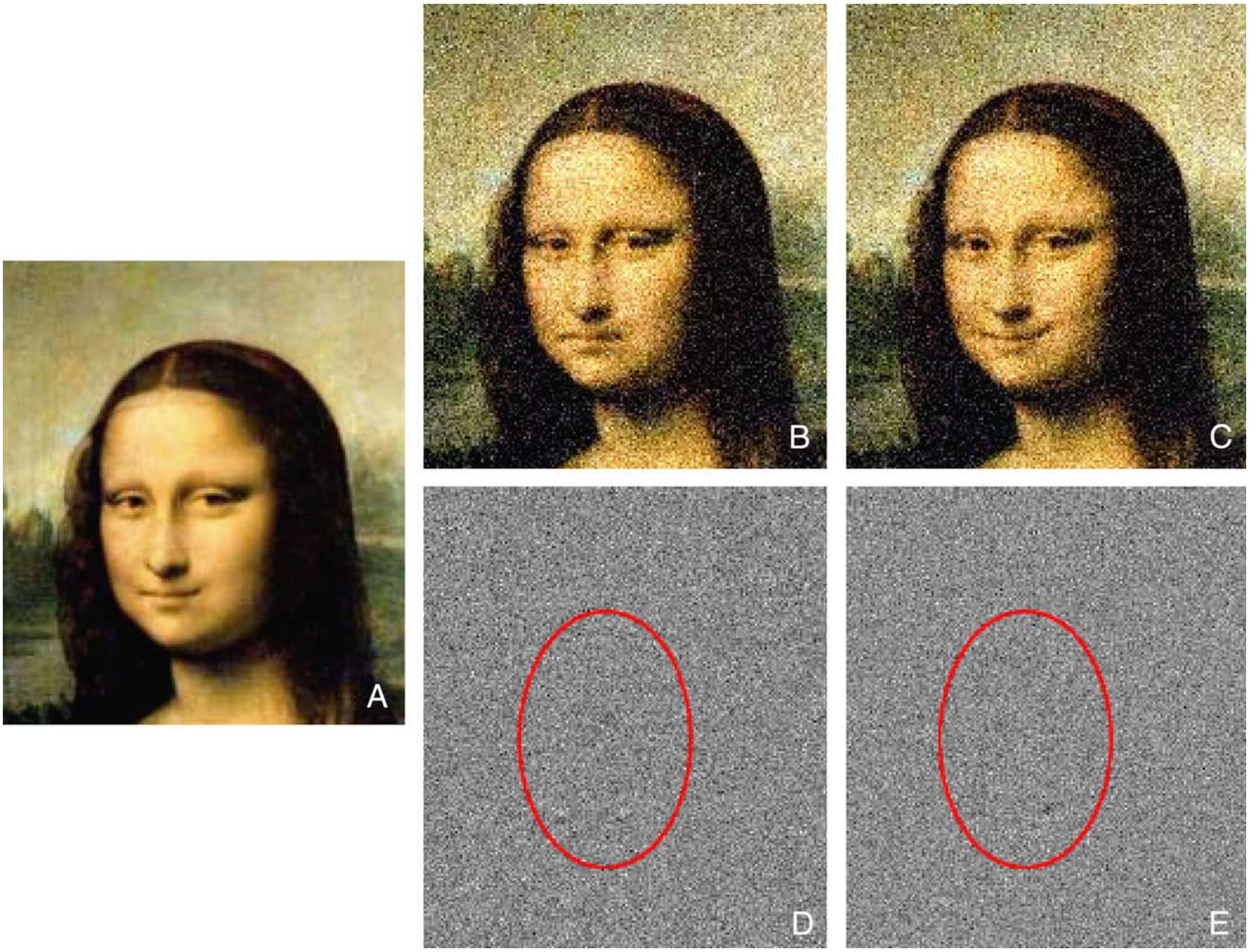

Fig. 2. (A) The original detail of Mona Lisa painting used in the present study is shown in color for reference. The averaged noise samples for 12 observers are shown for the two extreme categories (sad and happy) superimposed on the full-color original in B and C. The original noise samples are shown on the bottom (D and E).

Table 1
Percent of SAD and HAPPY category entries for each observer

	Observer											
	1	2	3	4	5	6	7	8	9	10	11	12
SAD (%)	17	26	21	24	28	21	30	44	19	46	30	24
HAPPY (%)	11	11	13	8	9	12	9	11	12	5	14	6

by reverse correlation that remained unnoticed to visual inspection. There is, therefore, a need for a rigorous statistical procedure that would help to extract relevant to the task locations with a confidence.

The approach we developed to address this problem is based on Kolmogorov–Smirnov goodness-of-fit test for a single sample, which compares an empirical sample distribution with a known distribution (Shleskin, 2000). This test computes the discrepancy measure between two distributions and then compares this measure with a criterion confidence level. If the measure exceeds the criterion value, one can conclude that the distributions are different.

If a sufficient number of noise samples is accumulated in the relevant locations, they are likely to produce a different noise distribution than for irrelevant locations: the deviation from the mean of the cumulated noise in relevant locations should increase linearly whereas the deviation from the mean in irrelevant locations should be proportional to the square root of the number of noise instances summed. This discrepancy can be detected by the Kolmogorov–Smirnov test. The cumulated noise distribution for irrelevant regions can be obtained with high precision by taking the noise samples from the large area outside the face: this distribution will serve as a reference. (The high accuracy of this distribution

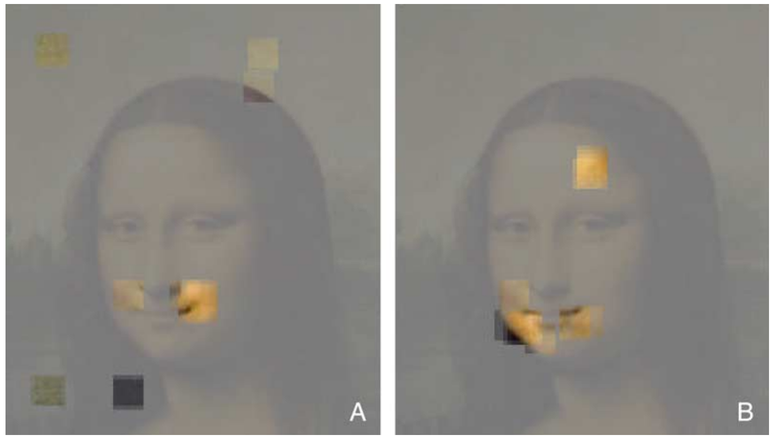

Fig. 3. The areas potentially relevant to (A) SAD and (B) HAPPY expressions as revealed by Kolmogorov–Smirnov test. The areas are shown by contrast enhancement from the low-contrast background. Both images reveal significant regions near the mouth corners. The other regions are a result of false detections, because the confidence criterion ($p < 0.001$) allows for a few such events per image.

allowed us to use the simple Kolmogorov–Smirnov test for a single sample instead of more complicated test for two independent samples.) The tested distribution was obtained for square areas of 21 by 21 pixels; there were

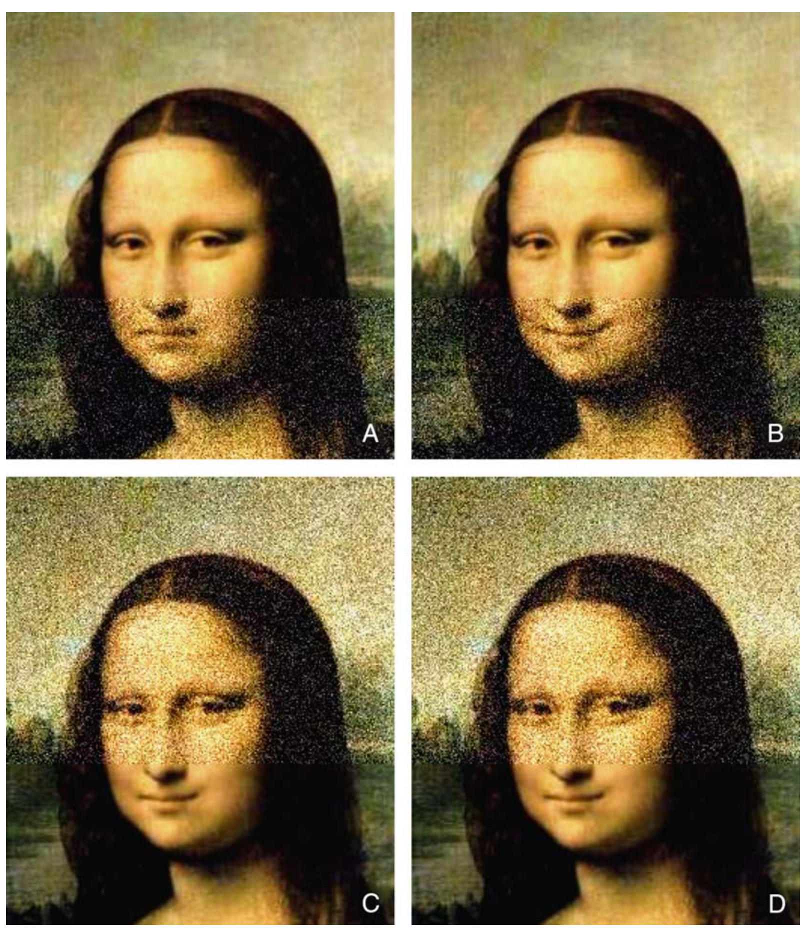

Fig. 4. The top pair of images (A and B) has SAD and HAPPY cumulated noise added to the lower part of the face. Most observers see the eye expression as much happier than on the left although the eyes are identical in both images. In the lower pair (C and D) the cumulated noise was added to the upper part of the image. In this case the mouth expression was the same, and most observers report no difference in the eye expression, despite the differences in the overlaid noise.

about 7000 such areas within each image. The criterion level was set at $p < 0.001$, which allows for about seven falsely detected locations within each image.

The results of the Kolmogorov–Smirnov analysis are shown in Fig. 3 by contrast-enhanced squares that depict the areas where the cumulated noise distribution deviated significantly from the cumulated noise distribution from irrelevant locations. Both categories consistently reveal relevant areas at the corners of the mouth that conform to the results of visual inspection described earlier. Most of the potentially relevant regions are located outside the face area, which indicates that they are false detections for facial expression. One of the potentially relevant regions is located in the right side of the forehead in Fig. 3B, which depicts the results for HAPPY responses. This location is also likely to be a false detection. The only detected regions that coincide for the SAD and HAPPY analysis are near the two corners of the mouth. It is important to note that the eyes contain many features, such as pupils, irises, eyelids and eyebrows that would phase-lock the cumulated deviations, but none of these showed significant deviations on the test.

Thus, the statistical analysis indicates that there are no potentially meaningful areas consistently related to the SAD–HAPPY change of facial expression other than the mouth corners. This result was somewhat surprising since it does not indicate any involvement of the eyes, which are often characterized as the "window of the soul". At first glance, the eyes show an obvious change. They seem to be smiling in Fig. 2C and serious in Fig. 2B. To evaluate this observation we set up the pair of images shown in Fig. 4: the upper parts of the images of the top one (Fig. 4A and B) are identical since SAD and HAPPY noise was added to the lower part of the portrait while in the lower pair (Fig. 4C and D) images have identical lower parts. Most observers report that the eye expressions in the top pair (Fig. 4A and B) are different: the left look more sad and the right looks more happy. We asked the observers to rate specifically the eye expressions in Fig. 2A and B. The rating across the observers for the happy expression in the eyes was higher by 4.9 units ($t(22) = 2.58$, $p < 0.01$) for the happy than for the sad one, despite the fact that the eyes are identical in the two images. Similar ratings showed no significant difference (0.0 ± 2.1) for eye expressions

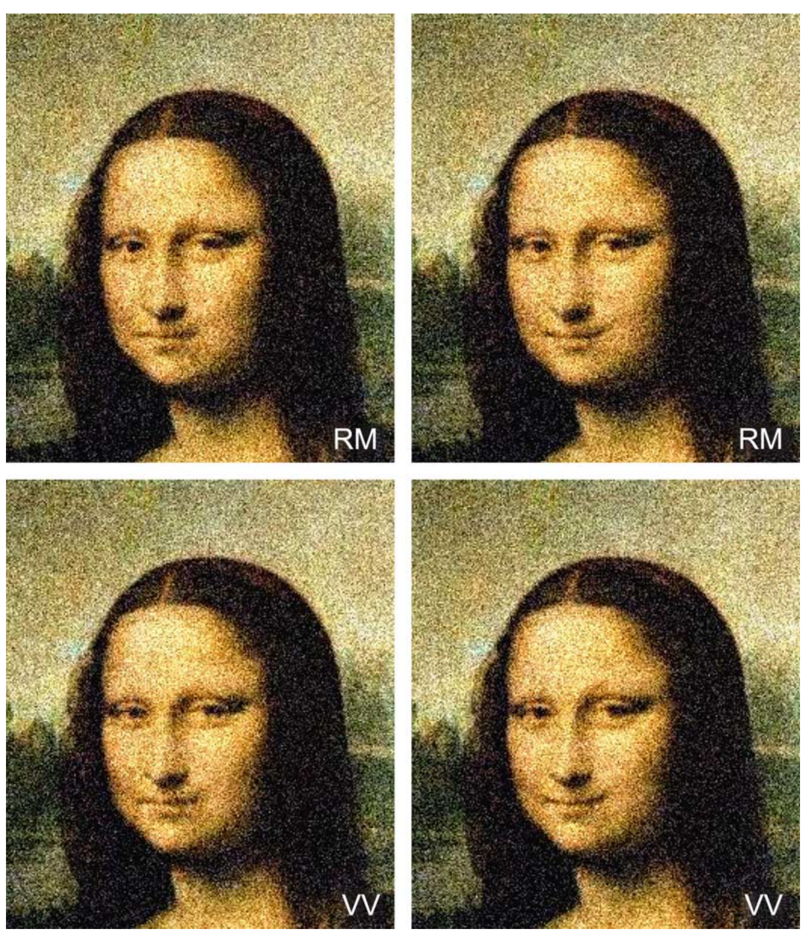

Fig. 5. The 'sad' (left) and 'happy' (right) categories obtained from 100 trials. Top—observer RM, bottom—observer VV.

between the images with the SAD and HAPPY cumulated noise over the upper half of the face (Fig. 4C and D). The difference in emotional expression perceived in the eyes from the change in the lower face is significantly stronger than that carried by the upper face containing the eyes themselves. We conclude, therefore, that the perceived emotional expression in the eyes is evoked entirely by long-range influences from the change in the mouth expression, at least to the level of statistical evaluation used in this example. (We cannot, of course, rule out a minor role of eye configuration if more extensive testing was conducted.)

4. Discussion

From the methodological standpoint, the reverse correlation technique reveals both the locations and the changes required to represent a certain emotional state in an analyzed image. In this respect it appears to be more powerful than the ''Bubbles'' technique of Gosselin and Schyns (2001), which is limited to identifying only the locations. Our application resembles an approach taken by Gosselin and Schyns (2003) who revealed the unobservable memory representations in memory by asking observers to judge presence of a certain object in visual noise. However, our observers were identifying the object features expressing the changes underlying the perception of emotions rather than the basic structure of the object itself.

Unlike many reverse correlation studies reported in the literature, our experiment required a relatively small number of trials (hundreds vs. tens of thousands) to reveal locus of the relevant feature. (To illustrate this point, the Mona Lisa portrait overlaid with noise accumulated for 2 individual observers in 100 trials is shown in Fig. 5. For observer RM the SAD rating was given by 44 noise instances and HAPPY by 11, for observer VV the corresponding numbers were 30 and 14.) Such small numbers of samples indicate that the emotional state in the Mona Lisa portrait is encoded by only a few pixels, which have high probability of coming out in the noise with the right luminance bias. These pixels are, as for the average, in the corners of Mona Lisa mouth.

The lack of effect of the cumulated noise on the mood of the eyes (Fig. 4C and D) demonstrates that an evident difference in the eye expressions in Fig. 4A and B is due to long-range induction from information in the mouth region. This result does not imply, however, that eyes have no role in perception of human emotion. They may act as emotional intensifiers for an expression whose mood is set by the mouth. Also, we cannot exclude the possibility that the eyes convey the expression itself with features undetectable by the reverse correlation technique.

In conclusion, we demonstrate that the reverse correlation technique provides a convenient means to reveal the cues employed in subtle changes of emotional content of facial expression. These cues can be identified with Kolmogorov–Smirnov test as described. The same method could be applied to evaluation of the spatial cues for a full range of emotional expressions. Beyond its scientific value, the spatial reverse correlation method provides a general method of converting internal imagery into real images.

References

Ahumada, A. J., Jr., & Lovell, J. (1971). Stimulus features in signal detection. *Journal of the Acoustical Society of America, 49*, 1751–1756.

DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1995). Neuronal mechanisms underlying stereopsis: how do simple cells in the visual cortex encode binocular disparity? *Perception, 24*, 3–31.

Ekman, P., & Friesen, W. (1984). *Unmasking the face. A guide to recognizing emotions from facial cues*. Englewood Cliffs, NJ: Prentice Hall.

Ekman, P., Friesen, W. V., & O'Sullivan, M. (1988). Smiles when lying. *Journal of Personality and Social Psychology, 54*, 414–420.

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

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

Hess, E. (1975). *The tell-tale eye*. New York: Van Nostrand Reinhold.

Neri, P., Parker, A. J., & Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. *Nature, 401*, 695–698.

Ringach, D. L., Hawken, M. J., & Shapley, R. (1997). Dynamics of orientation tuning in macaque primary visual cortex. *Nature, 387*, 281–284.

Shleskin, D. J. (2000). *Handbook of parametric and nonparametric statistical procedures*. Chapman & Hall/CRC.

Sutter, E. E. (1975). A revised conception of visual receptive fields based on pseudorandom spatiotemporal pattern stimuli. In *Proceedings of the first symposium on testing an identification of nonlinear systems, March 17–20, 1975, Pasadena, CA*.

Webster, M. A., & MacLin, O. H. (1999). Figural aftereffects in the perception of faces. *Psychonomics Bulletin Review, 6*, 647–653.