Trigger Features for Conveying Facial Expressions: The Role of Segmentation

Vilayanur S Ramachandran
Center for Brain and Cognition, University of California, San Diego, La Jolla, CA, USA

Chaipat Chunharas
Center for Brain and Cognition, University of California, San Diego, La Jolla, CA, USA; King Chulalongkorn Memorial Hospital, Chulalongkorn University, Thailand

Michael Smythies
Center for Brain and Cognition, University of California, San Diego, La Jolla, CA, USA

Abstract
Primates are especially good at recognizing facial expression using two contrasting strategies—an individual diagnostic feature (e.g., raise eyebrows or lower mouth corner) versus a relationship between features. We report several novel experiments that demonstrate a profound role of grouping and segmentation—including stereo—on recognition of facial expressions.

Keywords
face perception, segmentation, trigger feature, emotion

The idea of trigger features is well known in ethology. Instead of responding to the tout ensemble or gestalt—as is usually the case—the organism responds to certain diagnostic features that circumvent the need for high-level processing. Subsequent activations of limbic structures result in rapid evocation of certain emotions and corresponding behavior. Baboons responding to purple red rumps of females in estrus (the mainstay of many nature films), seagull chicks pecking the red spot at the tip of a disembodied beak, and nightingales regurgitating food into the huge gaping mouths of grotesquely large cuckoo chicks are good examples. In humans, best examples of trigger features are perhaps facial expressions as demonstrated by face inversion effect (Thompson, 1980; other body parts like

Corresponding author:
Vilayanur S Ramachandran, Department of Psychology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA, USA.
Email: vramacha@ucsd.edu
breasts may also qualify—male *homo sapiens* have a pronounced mammo-ocular reflex; Stuart Anstis, personal communication).

The trigger-like action of facial expressions can be seen in Figure 1(a), which shows that a wide range of expressions can be evoked using simple tilted lines, curves, and dots (e.g., compare a smiley face with a frown). Despite the fact that all details have been removed and the stimulus has been stripped to a bare minimum, it nonetheless evokes recognition of emotions almost as powerfully as real photos. Yet the mere presence of features is inadequate (e.g., Figure 1(b)), and the relationship between features is critical (Kosslyn et al., 1989). A more striking demo of this is shown in Figure 2, which is bistable. If the tilted eye brow line is assigned to the left circle, it becomes unmistakably imbued with anger, but if assimilated into the right circle, it triggers sadness. Even though the feature itself has not changed physically, it becomes imbued with the emotion that is ascribed to the whole face. An equally striking example of the powerful role of segmentation is shown in Figure 3. In Figure 3(a), the features within the circle are seen as mere fragments. But the addition of an occluding mustache (Figure 3(b)) results in amodal completion of the mouth behind a mustache, thereby conveying a smile.

While exploring a wide range of stimuli, we discovered something quite odd, illustrated by comparing the two faces in Figure 4(a) and (b). Figure 4(a) clearly conveys a smile which is consistent with the *coarse coding hypothesis*: The mere presence of a curve with two tips pointing upward—in this case depicted by a banana—signal smile despite being recognized as a banana at a conscious cognitive level. The convexity bulging down contour is an ethological releasing signal. Yet this simple explanation is contradicted by Figure 4(b); the same banana placed at the correct location on a full color photo of a face does not evoke

![Figure 1](image.png)

**Figure 1.** (a) A wide array of complex facial expressions can be conveyed by adjusting the locations and relative distances of simple lines and dots. (b) Illustration indicating that the mere presence of features is not enough, and the relationship between them is critical.
recognition of a smile—even though there is more information that it is a face than the mere circle and dots and eyebrow lines of Figure 4(a).

How does one interpret this result? Upon seeing the real face rich in detail, the neural network responding to the face is powerfully activated and is perhaps more demanding of equally strong confirmation from features conveying expressions—a banana will not do, and a mouth is needed. In the cartoon, on the other hand, nearly all the features are trimmed down to a bare minimum and therefore the need for distinctiveness of the one feature in question (mouth) is substantially relaxed—so even the fruit is effective despite complete lack of resemblance to lips.

We believe that these findings illustrate a new principle in perception. Imagine you wander into a completely new visual scene and it is very foggy—all objects are seen blurred. If there is a single road sign saying “MANCHESTER ROAD” seen in fragments partially hidden by fog, for example, “..AN..HEST..ER R..AD,” you are more willing to accept degraded information from the sign, lowering your perceptual threshold. But if there is no fog and

\textbf{Figure 2.} The features of overlapping zone of the two faces, when attributed to one of the two faces, give a distinct impression of belonging to that face. The features themselves do not change but become tainted with the emotion that is ascribed to the whole face.

\textbf{Figure 3.} Two images demonstrate that perceptual continuity of the mouth behind the opaque mustache evokes a smile (b) whereas the same fragments fail to do so without mustache (a).
Figure 4. (a) A representation of a smiling face, depicted by two dots and a banana. (b) A photograph of a real person with a banana superimposed on the mouth. The result is a more convincing smile in (a) than in (b).

Figure 5. A house with a banana mouth evokes the appearance of a smile.
all the other objects are clear, you would be reluctant to read it as “MANCHESTER ROAD.”

Figure 5 also illustrates this; a house with a banana mouth is paradoxically more effective as a face than a face photograph with a banana mouth! Again, what counts is the consistency in how indistinct or impoverished the different cues are, rather than the presence of a few diagnostic features. The house with the banana mouth is so unlikely an accidental juxtaposition that the brain is willing to accept that the only explanation is that it is a face. As Sherlock Holmes said “When you have ruled out the impossible, my dear Watson, whatever remains—however improbable—must be the correct answer” (Doyle, 2010).

As noted earlier, knowing which feature belongs to which object (which is constrained by image segmentation) is a necessary prelude to the recognition of emotional expression. To explore the role of segmentation further, we optically grafted a smiling real mouth on the photo of a face and simply put it in a different stereo plane floating out in front of the paper (Figure 6). Four subjects confirmed this loss of emotion. The smile goes away since the brain no longer groups the lips with the face.

Taken collectively, our results together with those from Nakayama, Shimojo, and Silverman (1989) introduce a novel probe for understanding the manner in which the brain parses the visual image into objects—such as faces—and assigns different values to different features defining the object. Face detecting cells are abundant in fusiform gyrus
(Kanwisher, McDermott, & Chun, 1997; Pinsk, DeSimone, Moore, Gross, & Kastner, 2005; Tsao, Freiwald, Tootell, & Livingstone, 2006) and our stimuli could be used to explore the underlying circuitry.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Doyle, A. C. (2010). The sign of four. Broadview Press.
Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 17, 4302–4311.
Kosslyn, S. M., Koenig, O., Barrett, A., Cave, C. B., Tang, J., & Gabrieli, J. D. (1989). Evidence for two types of spatial representations: Hemispheric specialization for categorical and coordinate relations. Journal of Experimental Psychology: Human Perception and Performance, 15, 723–735. doi: 10.1037/0096-1523.15.4.723.
Nakayama, K., Shimojo, S., & Silverman, G. H. (1989). Stereoscopic depth: Its relation to image segmentation, grouping, and the recognition of occluded objects. Perception, 18, 55–68. doi: 10.1068/p180055.
Pinsk, M. A., DeSimone, K., Moore, T., Gross, C. G., & Kastner, S. (2005). Representations of faces and body parts in macaque temporal cortex: A functional MRI study. Proceedings of the National Academy of Sciences of the United States of America, 102, 6996–7001. doi: 10.1073/pnas.0502605102.
Thompson, P. (1980). Margaret Thatcher: A new illusion. Perception, 9, 483–484. doi: 10.1068/p090483.
Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. Science, 311, 670–674. doi: 10.1126/science.1119983.

Author Biography

Vilayanur S Ramachandran is a professor and director of Center for Brain and Cognition, UCSD. Ramachandran’s accolades include receiving the Henry Dale Medal of the Royal Institution, London; giving the BBC Reith lectures in the UK and Gifford lectures in Glasgow.