Face recognition: **When a nod is better than a wink**

Jim Stone

A recent study has shown that, when people talk, their changing facial expressions and head movements provide dynamic cues for recognition.

Address: Psychology Department, Sheffield University, Sheffield S10 2UR, UK.
E-mail: j.v.stone@sheffield.ac.uk

Current Biology 2001, 11:R663–R664

0960-9822/01/$ – see front matter

How do we recognise faces? Is it by fixed facial features, such as the eyes and nose, or by dynamic personal characteristics, such as head movements and changes in facial expression? In a series of experiments reported recently in *Current Biology*, Hill and Johnston [1] show that characteristic changes in facial expression and head movements can be used for recognition of both the identity and sex of a person. This work suggests that different dynamic cues provide independent sources of information for face recognition. But whereas recognition of identity seems to be mediated mainly by characteristic head movements, recognition of sex seems to be mediated mainly by changes in facial expression.

Until recently, static photographic pictures were the main tool used to probe the mechanisms of face recognition. Unfortunately, such stimuli can be used to investigate only the relative salience of different fixed facial features in recognition. Early work using minimal 'Johansson' figures [2] and faces [3] — where the coherent movements of seemingly arbitrary sets of dots allow unambiguous object recognition — suggested that motion plays a key role in the recognition of dynamic biological stimuli. Using image processing techniques, it is now possible to create realistic computer-generated dynamic faces without resorting to sparse Johansson stimuli for investigating the role of dynamic cues in recognition.

Hill and Johnston [1] used these image processing techniques to investigate the role of motion in recognition of realistically animated faces. They induced natural motion in actors by getting them to tell simple jokes whilst being filmed. They then effectively separated the motion and identity of each actor by measuring the motion signal in each video sequence. The motion signal had two separate components, corresponding to the rigid motion of the head, and the non-rigid, relative motion of individual facial features caused by changes in facial expression. These rigid and non-rigid motion signals were then imposed on a computer generated canonical head (a morph-average of 200 actors), so that the motion (but not the facial features) of the original actor was preserved in the reconstituted (silent) video sequence (Figure 1). An intriguing side-effect of this motion-extraction process is that the extent to which rigid and non-rigid motion affects face recognition can be tested by imposing either the rigid motion, or the non-rigid motion, or both, on a canonical head. Example video clips can be viewed at the website http://spectrum.psychol.ucl.ac.uk/haroold/stim.html.

A key finding of this study [1] is that the recognition of identity depends on rigid motion more than on non-rigid motion. Subjects were asked to sort a set of 16 video sequences into four sets of four, such that each set contained videos of one individual. Note that all of the video sequences contained images of the same morph-average actor, but each video sequence displayed this morph-average actor with the dynamics obtained from only one individual actor. Subjects were asked to sort the silent video sequences containing rigid motion only, non-rigid motion only, or combined rigid and non-rigid motion. The first surprise is that subjects can successfully sort video sequences according to identity using combined rigid and non-rigid motion as the only cue. It appears that recognising a friend may not depend purely on the shape of their face.

The second surprise is that subjects performed best when given only rigid motion cues, and performed at chance when given only non-rigid motion cues. Performance with combined head and facial dynamic cues was intermediate between performance with rigid motion cues only and non-rigid motion cues only. This suggests that attending to non-rigid motion may disrupt ability to judge identity on the basis of rigid motion. This finding requires confirmation, however, because the motion-extraction process appears to be more accurate for rigid motion than for non-rigid motion; the latter are quite subtle, and tend to be attenuated by the extraction process.

Even more intriguing is the finding that, whereas spatial inversion of video sequences — turning them upside down — disrupts the recognition of both sex and identity, temporal reversal — playing them backwards — disrupts the recognition of sex, but not identity. Inversion and temporal reversal of video sequences are useful standard tricks for isolating different component processes in recognition, because neither transformation affects the content of images being presented. So any change in performance induced by these transformations cannot be attributed to changes in the visual cues such as the image grey level or speed, or the perception of structure of motion, shading or

Figure 1

Hill and Johnston [1] captured the motion associated with changes in facial expression and head position by tracking seventeen markers during natural speech (left). This motion could then be superimposed on a morph-average head (right).

texture, because these cues remain invariant with respect to changes in orientation and temporal direction.

These findings are based on a pair of experiments carried out by Hill and Johnston [1]. First, in an experiment on identity recognition, subjects specified the 'odd one out' of three video sequences; only two sequences were of the same individual, and all sequences had both rigid and non-rigid motion. Temporal reversal had no impact on the performance of this test. In contrast, inversion caused a substantial decrease in performance. These results suggest that the dynamic cues used for identity recognition are independent of temporal direction, but are not independent of spatial orientation.

Second, in an experiment on sex recognition, subjects classified pairs of video sequences according to whether they contained individuals of the same sex or not. Subjects performed well above chance for sequences with rigid and non-rigid motion, but performed at chance levels when these videos were played backwards. Inverted videos yielded intermediate performance. Thus, recognition of sex is reduced to chance levels by temporal reversal, but not by inversion.

The results of this pair of experiments suggest the intriguing possibility of a double dissociation, such that dynamic cues to identity are affected by inversion but not temporal reversal, whereas dynamic cues to sex are affected by temporal reversal but not inversion. Unfortunately, whilst the statistical results permit this as a possibility, they are tantalisingly not quite sufficient to permit it as a conclusion. It is also possible that such apparent double dissociation may be due to task difficulty; that is, the different dynamic cues apparently used for recognising identity and sex may be an artifact of the relative difficulty associated with recognition of sex and identity. Attempting to refute such generic 'task-difficulty' hypotheses is notoriously difficult, and usually generates more heat than light.

Aside from a few animal studies on low-level vision (for example, [4,5]), it used to be thought that visual recognition depends principally on two-dimensional and three-dimensional shape. Dynamic cues were considered to be useful only insofar as they could be used to estimate three-dimensional shape [6], which could then be used for recognition. The new work of Hill and Johnston [1] contributes to a growing body of evidence that dynamic cues provide an independent source of information for the recognition of faces [7–9], people [10] and even solid objects [11–13].

Whilst professional impersonators may claim that the above conclusion is common knowledge, common knowledge is not evidence. Having said this, even in Shakespeare's time it was accepted that dynamic cues contribute to recognition, and that if a man is to be mimicked then it is by "the manner of his gait, the expressure of his eye, forehead, and complexion, he shall find himself most feelingly personated" (*Twelfth Night*).

Acknowledgements

I would like to thank Nikki Hunkin and Tim Valentine for discussions during the preparation of this paper.

References

1. Hill H, Johnston A: Categorising sex and identity from the biological motion of faces. *Curr Biol* 2001, 11:880-885.
2. Johansson G: Visual perception of biological motion and a model for its analysis. *Percept Psychophys* 1973, 14:201-211.
3. Bruce V, Valentine T: When a nod's as good as a wink: The role of dynamic information in facial recognition. In *Practical Aspects of Memory: Current Research and Issues. Vol 1.* Edited by Gruneberg MM, Morriss PE, Sykes RN. Chichester: Wiley; 1988:169-217.
4. Tinbergen N: *The Study of Instinct.* Oxford University Press, 1951.
5. Lettvin JV, Maturana HR, McCulloch WS, Pitts WH: What the frog's eye tells the frog's brain. *Proc IRE* 1959, 1940-1951.
6. Ullman S: *The Interpretation of Visual Motion.* MIT Press, 1979.
7. Christie F, Bruce V: The role of dynamic information in the recognition of unfamiliar faces. *Memory Cognit* 1998, 26:780-790.
8. Lander K, Christie F, Bruce V: The role of movement in the recognition of famous faces. *Memory Cognit* 1999, 27:974-985.
9. Wallis G, Bulthoff HH: Effects of temporal association on recognition memory. *Proc Natl Acad Sci USA* 2001, 98:4800-4804.
10. Mather G, Murdoch K: Gender discrimination in biological motion displays based on dynamic cues. *Proc R Soc London Series [Biol]* 1994, 258:273-279.
11. Stone JV: Object recognition using spatiotemporal signatures. *Vision Res* 1998, 38:947-951.
12. Stone JV: Object Recognition: view-specificity and motion-specificity. *Vision Res* 1999, 39:4032-4044.
13. Wallis G, Bulthoff H: Learning to recognize objects. *Trends Cognit Sci* 1999, 3:22-31.