Categorizing sex and identity from the biological motion of faces

H. Hill and A. Johnston

Head and facial movements can provide valuable cues to identity in addition to their primary roles in communicating speech and expression [1–8]. Here we report experiments in which we have used recent motion capture and animation techniques to animate an average head [9]. These techniques have allowed the isolation of motion from other cues and have enabled us to separate rigid translations and rotations of the head from nonrigid facial motion. In particular, we tested whether human observers can judge sex and identity on the basis of this information. Results show that people can discriminate both between individuals and between males and females from motion-based information alone. Rigid head movements appear particularly useful for categorization on the basis of identity, while nonrigid motion is more useful for categorization on the basis of sex. Accuracy for both sex and identity judgements is reduced when faces are presented upside down, and this finding shows that performance is not based on low-level motion cues alone and suggests that the information is represented in an object-based motion-encoding system specialized for upright faces. Playing animations backward also reduced performance for sex judgements and emphasized the importance of direction specificity in admitting access to stored representations of characteristic male and female movements.

Address: Department of Psychology, University College London, Gower Street, London WC1E 6BT, United Kingdom.

Correspondence: H. Hill
E-mail: harold.hill@ucl.ac.uk

Received: **7 February 2001**
Revised: **9 March 2001**
Accepted: **18 April 2001**

Published: **5 June 2001**

Current Biology 2001, 11:880–885

0960-9822/01/$ – see front matter

Results and discussion

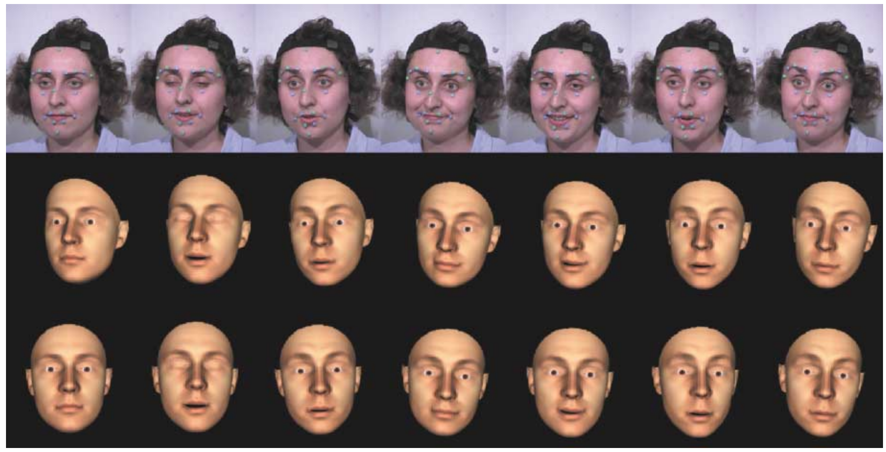
The fact that impersonators can mimic the ways in which famous people move their heads and faces demonstrates that, in addition to the primary role they have in communication, these movements can provide cues to sex and identity independently of the underlying shape and texture of the face. Similarly, in computer-animated films, a character's expressions and voice can be derived from an actor, as in Tom Hanks' performance as Woody in Pixar's Toy Story [10]. The character's face and head movements mimic those of the actor even though their underlying shapes are quite different. We report experiments in which we computer-animated an average head with movements captured from real people in order to investigate whether motion provides useful information for categorizing faces.

The animation process is illustrated and described in Figure 1. Four different movement sequences were captured for each of twelve "actors" (we use the term as a shorthand—the volunteers were not trained actors) and were used to animate the same three-dimensional model of an average face [9]. Each animation was of a person telling a two-line question-and-answer joke to another individual (e.g., "Why do cows have bells? Because their horns don't work!"). This activity was intended to elicit expressive and natural facial gestures, expressions, and speech from the actor.

All the stimuli produced in this manner were physically identical at the start of each animation and differed only in the way that they moved. This allowed us to investigate motion-based information independently of other cues. The technique also allowed the separation of rigid head motion, in which the head translates and rotates but does not change shape, and nonrigid motion, in which the expression changes but the head does not move. Given their different natures, these two components may be processed independently and may make different contributions to the perception of face-based biological motion.

By concentrating on motion, we do not intend to deny the importance of other sources of information for these tasks. Everyday experience and numerous experiments show that we can recognize sex and identity from static photographs that provide no motion information. However, motion is fundamental to vision, and diagnostic differences may be used by a system that makes use of any and all available discriminating information. The available evidence suggests motion may be particularly useful when spatial and other cues are degraded or changed, for example by presentation in photographic negative [5, 6]. In previous studies, researchers have used moving-point light stimuli to look at biological motion in general [11, 12] and facial movement in particular [1, 2], but the current technique provides more natural motion information

Figure 1



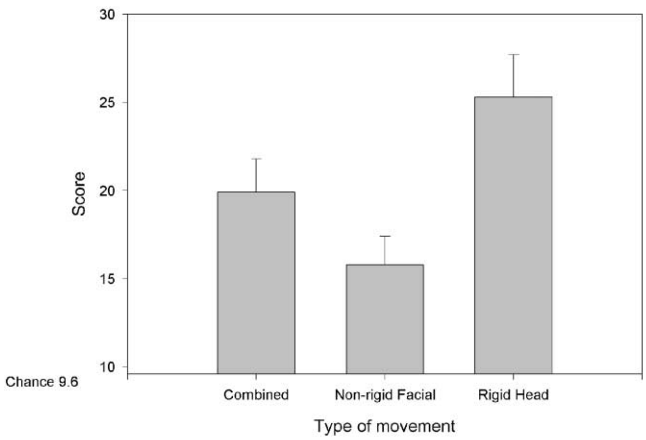
Examples of the animations used as stimuli (please also see the Supplementary material). The movement of the faces was captured by a pair of digital video cameras. The top row shows the original movement sequence from the right hand camera, the middle row shows an animation from the same viewpoint, and the bottom row shows the animation as the observers viewed it. The middle row was not used in any of the experiments, but we include it to facilitate comparison between the original and the animation. In each case, the leftmost image shows the neutral starting position. The following frames were taken at 1 s intervals. This sequence was 6.7 s long, excluding pre- and post-masks. The average length of sequences was 7.2 s (standard deviation, 1.6 s). The motion of the 17 markers and the pupils were automatically tracked with Famous vTracker (Famous Technologies) from video footage taken from the two cameras placed approximately 15° either side of the direction in which the actors were facing and at a distance of 1 m. The cameras were calibrated for each recording with a calibration object of known dimensions so that 2D tracked positions could be converted into 3D positions on the basis of projective geometry. These were used for the animation of a three-dimensional head model produced as an average of 200 heads, 100 male and 100 female [9], with the number of vertices reduced to 65,525 for import into 3D Studio Max. Animation was accomplished with the commercially available Famous animator (Famous technologies). In this system, marker positions are associated with a "hotspot" and an area of influence on the model that inherit the movements of the marker. The markers on the forehead, temples and nose were used for defining rigid translations and rotations of the head. Because rigid motion can be fully characterized by the motion of a few markers in a way that nonrigid motion cannot, the rigid component of the animations was more accurate. In both cases the timing of movement is accurately captured given 25 frames per second temporal resolution, whereas spatial properties cannot be truly veridical given the differences in shape between actor and average head model as well as the limited spatial sampling. However, these limitations were constant for all animations and experimental conditions and so should not have biased the results. Animated head models were texture-mapped with an average texture [9] and rendered with 3D Studio Max (Kinetix) for the production of 640 × 480 pixel 25-frames-per-second avi format movies. Animations were compressed with Radius cinepak, with 90% compression quality and a keyframe every 15 frames. A 10 frame mid-gray mask was added to the beginning and end of each animation. Backward and inverted animations were rendered with the same parameters as for upright forward animations, but with the order or orientation of the images altered. In all experiments, observers could view the animations as many times as they wished. Presentation was controlled by Microsoft Mediaviewer, and responses were recorded manually in the rating and sorting experiments and by programs written with Macromedia Director for the 2-AFC and odd-one-out tasks.

while more fully eliminating residual spatial cues, such as the aspect ratio of the underlying face.

There are a number of ways in which motion may be useful for distinguishing between faces. It can provide indirect cues to three-dimensional shape via structure-from-motion [13]. However, mathematical analysis of this process assumes rigid motion [13], an assumption satisfied by head movements but violated by most facial movements. In previous experiments [1–8] the cues motion provides about shape may have been important, but in the present experiments any such cues would be limited because the underlying shape was always the same. However, differences in shape do result as a consequence of differences in movement, and the resulting differences in shape may provide useful information. This highlights the difficulty of completely separating motion from spatial information. Also, differences in the ways that people move their faces may result in useful and reliable differences in low-level image motion. Some people may move the whole or parts of their faces more than others. Lastly, the spatial changes occurring over time for a moving object produce a "spatio-temporal signature" that in itself may be useful for recognition [14].

In order to investigate whether the biological motion of faces provides cues to identity, we used two tasks. In the first, observers were presented with 16 animations, 4 different animations for each of 4 actors, and were asked to sort these into 4 equally sized groups on the basis of identity. Observers could view the animations in any order as many times as they wished, and they sorted the animations simply by moving their icons into groups on the screen. All the stimuli used in this experiment were generated with movement sequences captured from male actors. This method ensured that the ability to do this task was independent of any ability to categorize sex. Different groups of observers saw rigid head motion (N = 15), nonrigid facial motion (N = 16), or combined rigid head and nonrigid facial motion (N = 16). The observers, like the actors, were recruited from the student population of University College London. The task, like recognition,

Figure 2



The results of the identity sorting experiment, in which four different animations of four "actors" were sorted into equally sized groups on the basis of identity. Different groups of observers saw nonrigid facial motion alone (N = 16), rigid head motion alone (N = 15), or both types of motion combined (N = 16). Performance was scored according to how many other examples of the same actor were put in the same group for each animation. This gives a maximum score of 48 (4 actors × 4 examples × 3 other examples in the group). The minimum score, when each group contains only one example of each actor, is 0. Chance, calculated by the Monte Carlo method of generating and averaging the score for 10,000 of the 16! possible ways to sort 16 animations, was 9.6. We also tested whether average scores for our groups of observers were significantly different from chance by calculating the proportion of times that the observed scores were exceeded by random samples of the same size. Combined rigid and nonrigid and rigid alone stimuli were both sorted significantly better than chance would have allowed ($p < .05$), and nonrigid stimuli alone were sorted marginally better ($p < .1$). Error bars show standard errors.

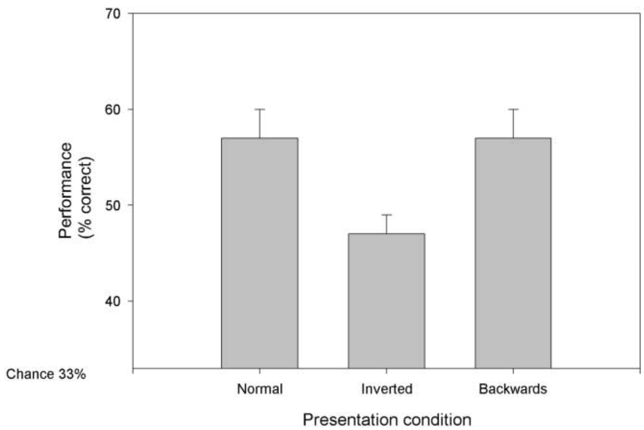
required that the description of the motion that observers recovered was stable enough to generalize over different examples of the same face while sensitive enough to distinguish between examples of different faces [15]. This task allowed us to investigate the motion information essential for recognition independently of memory-based or cue conflict effects that would be involved in recognition per se.

Performance in this task was scored for each animation according to how many other examples of the same person were sorted into the same group (Maximum score = 48, minimum = 0, chance = 9.6 by Monte Carlo simulation). Results are summarized in Figure 2, and details of the scoring system are given in the legend. In order to avoid having to make assumptions about the scoring distribution, we used Monte Carlo and Bootstrap resampling statistical methods to analyze the data [16]. There was a significant effect of type of motion shown ($p < .05$), with performance significantly above chance for rigid motion and for combined rigid and nonrigid motion ($p < .05$). Performance was marginally above chance for nonrigid motion ($p < .1$). Rigid head motion appears particularly useful for categorizing people on the basis of identity. The difference between performance in the rigid head motion alone and nonrigid facial motion alone conditions was significant ($p << .05$), and the other two differences between conditions were marginally significant ($p < .1$). Nonrigid facial motion is less useful than rigid motion for characterizing individuals. It is possible that nonrigid facial movement could interfere with identity judgements in this task given that the facial speech of two people telling the same joke or making the same expression may be more similar in many ways than that of the same person telling different jokes or making different expressions. Although this factor was not fully balanced, two jokes were told by all of the actors used in this experiment, so in order to test whether subjects grouped faces on the basis of the joke told, we scored the data according to whether examples of different actors telling these jokes were grouped together. In this case, sorting scores were no different than what would have been due to chance. This result shows that observers were not sorting on the basis of the joke told and validates the sorting task and scoring method.

We also used an odd-one-out task to test whether motion provides cues to identity. Twelve naive observers were presented with seventy-two trials each consisting of three animations—two different examples of one person and one example of a different person of the same sex. The observers' task was to identify the animation derived from the unique individual—the odd one out. Observers initiated presentation of the stimuli and responded by using an application written in Macromedia Director. In order to provide clues as to the critical properties of the motion information used for this task, we compared inverted and backward play to normal (forward and upright) presentation for stimulus triplets. Inverted presentation leaves low-level motion cues the same but is well known to adversely affect many aspects of face processing [17]. Playing an animation backward uses the same static frames as the same animation played forward but changes the overall pattern of movement. Both inverted and backward play test the extent to which performance can be achieved on the basis of perceptual matching or whether stored knowledge is needed, as they leave the perceptual similarities available for matching the same but might be expected to affect access to stored knowledge about how faces normally appear. All the stimuli used in this experiment contained both rigid head and nonrigid facial motion.

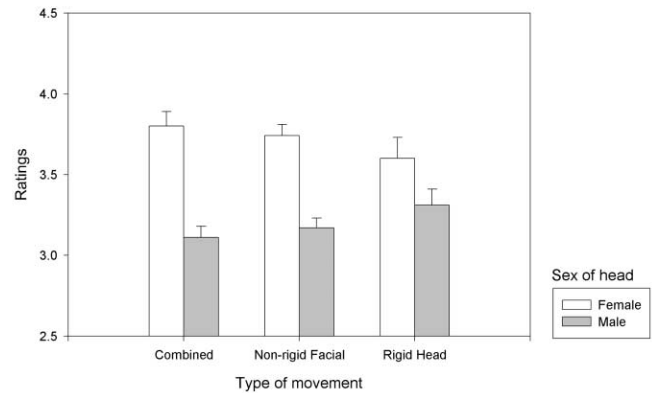
As can be seen from the results summarized in Figure 3, the accuracy with which the odd one out was identified depended upon how the stimuli were presented. A one-way repeated-measures ANOVA on the proportion of correct responses showed a main effect of the presentation

Figure 3



The results of the odd-one-out experiment in which observers (N = 12) had to choose which animation corresponded to the unique individual from a choice of three. The other two animations were two different examples of another individual. Performance was above chance in all conditions but significantly worse when animations were shown inverted. Error bars show standard errors.

Figure 4



The results of an experiment in which observers rated the sex of animations on a 6 point scale with 1 indicating definitely male and 6 indicating definitely female. The same observers took part in the same conditions as described for Figure 2. Ratings for male and female items were significantly different, but the amount of difference depended on the type of movement shown. Rigid head movements alone appear least useful for discriminating sex. Error bars show standard errors.

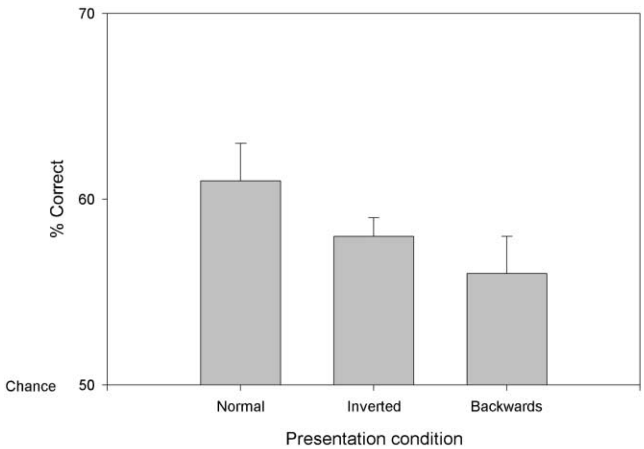
condition [$F(2,22) = 8.7$, $p << .05$]. Post hoc paired t tests showed that playing inverted stimuli produced significantly worse performance than normal [$t(11) = 4.4$, $p << .05$] or backward [$t(11) = 4.4$, $p << .05$] presentation. Normal and backward presentation did not differ from each other ($p > .1$). One-sample t tests showed that performance was significantly above chance (33%) in all conditions [normal: $t(11) = 8.9$, $p << .05$; backward: $t(11) = 7.7$, $p << .05$; and inverted: $t(11) = 5.7$, $p << .05$]. The detrimental effect of inversion shows that performance is not based upon low-level properties of motion alone (for example, gross amount of head motion), as this information would be recoverable as easily from inverted stimuli. Instead, it appears that identity specific-motion information is processed by a system tuned to upright faces. Backward play did not affect performance, and this result shows that even when played backward, motion contains information that allows us to discriminate between individuals. Either discriminating cues are static, direction independent, and/or temporally symmetric, or playing animations backward generates new but equally discriminable patterns of movement. Previous evidence showing that we are less good at recognizing faces that have been learned normally from videos played backward [7] favors the latter explanation.

To extend the evidence obtained from the identity-based tasks used so far, we also tested whether observers could recover information about the sex of the actors from these animations. No training was given, so any ability to do this depended both on there being differences between the ways males and females move their faces and on observers being able to extract these cues from the animations and relate them to their knowledge of sex differences. Sex judgements cannot be achieved on the basis of perceptual matching alone, as they require access to stored knowledge.

In the first sex judgement experiment, 48 observers rated the sex of all 48 animations on a scale of 1 to 6, with 1 indicating definitely male and 6 definitely female (or vice versa for half the observers). Observers controlled presentation of the stimuli by using Microsoft Mediaviewer and responded by hand on a prepared ratings form. Three groups, each with a different set of 16 observers, saw nonrigid facial motion, rigid head motion, or combined rigid head and nonrigid facial motion (the observers were the same as those who subsequently sorted the stimuli according to identity, as reported above).

Results are summarized in Figure 4. Observers rated male and female faces differently, and their ability to do this depended on the type of motion information available. Analysis of variance confirmed this pattern of results by showing a significant interaction between the type of motion and the sex of the face [$F(2,45) = 3.4$, $p < .05$]. There were simple main effects of sex for combined [$F(1,45) = 40.0$, $p << .05$], facial alone [$F(1,45) = 26.9$, $p << .05$] and head alone [$F(1,45) = 7.5$, $p < .05$] conditions, and these results show that all types of motion contained useful cues to sex. One-sample t tests comparing ratings to the theoretically neutral rating value of 3.5 showed that stimuli with combined motion were rated significantly

Figure 5



Results of the 2-AFC sex judgement task. Observers (N = 14) saw pairs of animations, one male and one female, and had to indicate which was which. Performance was significantly better than chance would have allowed only when stimuli were played normally (upright and forward), although inverted faces were also categorized marginally better than chance would have allowed. The detrimental effect of playing animations backward highlights the importance of the pattern of movement for sex judgements as opposed to the low-level motion or static cues that remain the same when animations are played backward. Error bars show standard errors.

differently from neutral for both male [$t(15) = 5.7$, $p << .05$] and female [$t(15) = 3.4$, $p << .05$] items, as were stimuli containing only nonrigid facial motion [male: $t(15) = 5.9$, $p << .05$; female: $t(15) = 3.4$, $p << .05$]. With rigid head motion alone, male stimuli items were rated as marginally differentl from neutral [$t(15) = 1.9$, $p = .07$], but female items were not rated as significantly different from neutral ($p > .1$). For this task, nonrigid motion produced better performance than rigid motion, and this pattern is opposite to that found for the identity-sorting task. This suggests that the differences reported between the two types of information are not simply a function of any limitation in our animation of nonrigid motion. The results contrast with previous evidence from experiments in which point light stimuli were used. In these experiments, nonrigid motions were found to be more useful for sex judgements, with no difference for identity judgements [3]. However, in this previous work rigid motions were posed nods, shakes, and rocks of the head, not naturally occurring movements, and they would have provided clues to underlying differences in shape not available here. Performance in the previous study was also above chance and at a similar level to that reported here, 61.9% [3].

We further investigated sex judgements by using a two-alternative forced-choice task (2-AFC), a task that avoids any bias associated with the shape and texture of the average head. Fourteen observers who had not taken part in any of the previous experiments were presented with pairs of stimuli, one male and one female, and had to decide which was which. Stimuli were presented in two blocks, with upright and inverted stimulus pairs randomly interleaved in one block and forward and backward stimulus pairs randomly interleaved in the other. The order of the blocks was balanced, and both stimuli in a pair were always shown in the same condition. All stimuli contained both rigid head and nonrigid facial movements.

Results are summarized in Figure 5, with the percentage of correct categorizations for normal stimuli collapsed across both blocks. Paired t tests showed no differences between presentation conditions. However, one-sample t tests showed that only for stimulus pairs presented normally was performance significantly above chance (50%); $t(13) = 5.6$, $p << .05$ and $t(13) = 3.2$, $p < .05$ in the blocks with inverted and backward stimuli, respectively. Performance for inverted stimuli was marginally above chance, with $t(13) = 1.9$ and $p < .1$. Levels of performance were not high with these stimuli because most of the normal spatial cues to sex, including color and shape [18], are kept constant. Static or low-level motion cues alone cannot explain the pattern of performance observed, as these remained the same between presentation conditions. Instead, there appear to be direction-specific patterns of movement for upright faces that differentiate between male and female.

Conclusions

The results show that both rigid head and nonrigid facial movements provide useful information for categorizing both sex and identity. There are differences in the ways that people move their heads and their faces, and we can recover and use these identity cues. Rigid head movements appear to be particularly useful for distinguishing between individuals, and nonrigid motion appears to be useful for categorizing on the basis of sex. This may be because rigid head movements can be idiosyncratic, while most nonrigid facial motion is functionally related to specific aspects of speech and expression, which have anyway to be processed independently from identity [19]. Effects of inversion show that low-level motion cues are not sufficient to explain performance and suggest that dynamic information is encoded by a model-based system specialized for upright faces. Backward movement, although discriminable, disrupts sex judgements, and this result shows that the direction of patterns of movement can be critical.

Supplementary material

Examples of animations used as stimuli are available on the internet at http://images.cellpress.com/supmat/supmatin.htm.

Acknowledgements

This work was supported by the the Engineering and Physical Sciences Research Council (MS2715). Thanks go to Peter McOwan, Szonya Durant, Colin Clifford, and Branka Spehar for comments on drafts of the paper and to Glyn Cowe for technical assistance.

References

1. Basilli J: **Facial motion in the perception of faces and emotional expression.** *J Exp Psychol* 1978, **4:**373-379.
2. Basilli J: **Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face.** *J Pers Soc Psychol* 1979, **37:**2049-2058.
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, Morris PE, Sykes RN. Chichester, UK: Wiley; 1988:169-174.
4. Christie F, Bruce V: **The role of dynamic information in the recognition of unfamiliar faces.** *Memory and Cognition* 1988, **26:**780-790.
5. Knight B, Johnston A: **The role of movement in face recognition**. *Visual Cognition* 1997, **4:**265-273.
6. Lander K, Christie F, Bruce V: **The role of movement in the recognition of famous faces**. *Memory and Cognition* 1999, **27:**974-985.
7. Lander K, Bruce V: **Recognizing famous faces: exploring the benefits of facial motion.** *Ecological Psychology* 2000, **12:**259-272.
8. Pike GE, Kemp RI, Towell NA, Phillips KC: **Recognizing moving faces: the relative contribution of motion and perspective view information.** *Visual Cognition* 1997, **4:**409-437.
9. Vetter T, Troje N: **Separation of texture and shape in images of faces for image coding and synthesis.** *J Opt Soc Amer* 1997, **14:**2152-2161.
10. Smith AR: **Digital humans wait in the wings**. *Scientific American* 2000, **283:**72-78.
11. Johannsson G: **Visual perception of biological motion and a model for its analysis.** *Percept Psychophys* 1973, **14:**201-211.
12. Johannsson G: **Studies of perception of locomotion**. *Perception* 1977, **6:**365-376.
13. Ullman S: *The interpretation of visual motion*. Cambridge, Massachusetts: MIT Press; 1979.
14. Stone JV: **Object recognition using spatiotemporal signatures.** *Vision Res* 1998, **38:**947-951.
15. Marr D, Nishihara HK: **Representation and recognition of three-dimensional shapes**. *Proc R Soc Lond B Biol Sci* 1978, **200**:269-294.
16. Efron B: **Computer-intensive methods in statistics**. *Scientific American* 1983, **248:**116-130.
17. Valentine T: **Upside-down faces: a review of the effect of inversion upon face recognition.** *Br J Psychol* 1988, **79:**471-491.
18. Hill H, Bruce V, Akamatsu S: **Perceiving the sex and race of faces: the role of shape and colour**. *Proc R Soc Lond B Biol Sci* 1995, **261:**367-373.
19. Bruce V, Young A: **Understanding face recognition**. *Br J Psychol* 1986, **77:**305-327.