The Role of Movement in Face Recognition

Article *in* Visual Cognition · September 1997

DOI: 10.1080/713756764

Alan Johnston
University College London

VISUAL COGNITION, 1997, *4* (3), 265–273

The Role of Movement in Face Recognition

Barbara Knight and Alan Johnston
Psychology Dept., University College London, UK

The movement of the face may provide information that facilitates recognition. However, in most situations people who are very familiar to us can be recognized easily from a single typical view of the face and the presence of further information derived from movement would not be expected to improve performance. Here the effects of movement on face recognition are investigated for faces presented under non-optimal conditions. Subjects were required to identify moving or still videotaped faces of famous and unknown people. Faces were presented in negative, a manipulation which preserved the two-dimensional shape and configuration of the face and facial features, while degrading face recognition performance. Results indicated that moving faces were significantly better recognized than still faces. It was proposed that movement may provide evidence about the three-dimensional structure of the face and allow the recognition of characteristic facial gestures. When the faces were inverted, no significant effect of movement on recognition was found. This was interpreted as reflecting difficulties in the recovery of changing configurations of the face and facial gestures in upside-down faces.

INTRODUCTION

Studies in which a direct comparison has been made between recognition of faces presented as moving video sequences and still photographs have not, in the main, been able to show any benefits of movement (Bruce & Valentine, 1988; Shepherd, Ellis, & Davies, 1982). However, we should note that Schiff, Banker, and Galdie (1986) have reported an advantage of dynamic presentation, a rigid rotation of the head, over static presentation, at the test phase of a recognition memory experiment. In an attempt to reduce the non-motion cues in video sequences, several studies examining the role of facial movement have used an adaptation of Johansson's (1973, 1977) point-light technique to isolate

Requests for reprints should be addressed to Dr A. Johnston, Psychology Dept., University College London, Gower Street, London WC1E 6BT, UK.

information provided by facial motion from feature-based information (Bassili, 1978, 1979; Bruce & Valentine, 1988). Bassili (1979) proposed that, under good viewing conditions, the observer had access to a superfluous or redundant system of facial information which could be revealed and assist recognition when viewing conditions were poor. Results of these point-light studies demonstrated that movement of the face increased the likelihood of the recognition of basic expressions. Bruce and Valentine (1988) also showed a significant improvement in identity judgements for moving dot displays, although this was less marked than in the case of recognition of expression. They concluded that motion may provide information in situations where the static cues are impoverished by contributing to perception of the three-dimensional (3D) structure of the face.

One of the problems with the point-light technique is that it provides a sparsely sampled representation of the velocity fields generated by the moving face. Therefore, in the point-light display, both the spatial information about the facial structure and the dynamic information about the changes in the facial structure are degraded. Ideally, if we wish subjects to rely on dynamic cues, we should attempt to degrade spatial information while leaving cues from image motion intact. Shading provides important information about the 3D structure of the face and this source of information can be disrupted by presenting faces as photographic negatives. Presenting faces in negative reverses the sign of brightness contrast between image points but retains the 2D shape of the face and features, along with the position of features within the face (Hayes, Morrone, & Burr, 1986; Johnston, Hill, & Carman, 1992; Luria & Strauss, 1978). If motion cues are utilized when the spatial information about the face is degraded, in this case by presentation in negative, we should expect to see an improvement in performance for dynamic against static presentation. As faces are readily recognized from static cues under normal viewing conditions we would not predict any great benefit of moving images over still frames for normal image sequences. This is particularly true if there is little change in pose which is largely the case in the experiment described here.

It has been shown that the encoding of the configuration of the face is important for identification of faces (Rhodes, 1988; Sergent, 1984; Young, Hellawell, & Hay, 1987). The difficulty in processing inverted faces is well documented (Carey & Diamond, 1977; Diamond & Carey, 1986; Goldstein & Chance, 1981; Valentine, 1988) and it has been proposed that this problem is related to the difficulty of extracting the configuration from an inverted face (Bartlett & Searcy, 1993; Sergent, 1984; Tanaka & Farah, 1993; Young et al., 1987). The term "configuration" is used in a number of ways in face recognition research. It is most often used to refer to the spatial distribution of features. This is sometimes called a first order representation of the face. A second order representation relates the position of features to a facial prototype (Bruce, Doyle, Dench, & Burton, 1991; Diamond & Carey, 1986). Alternatively, Bartlett and Searcy (1993) and Tanaka and Farah (1993) refer to holistic versus component or featural information. The face may be encoded by an explicit hierarchy of parts. Recognition based on this representation would be featural or component based. Recognition from holistic information would be on the basis of the undifferentiated facial image taken as a whole.

The use of the term "configuration" adopted in this paper has been developed in more detail elsewhere (Johnston, 1992, 1996). In this account recognition is thought of as an inverse process in which we recover the parameters controlling the mechanisms which generate the shape the face. This places the burden of the development of a representational theory of face perception on defining the forward process, i.e. specifying the way in which faces are constructed, describing natural sources of variation in the face and constraints on variation (Bruce, 1988). Configuration is used here to refer to the set of parameter values that define the shape of a particular face. Some of the parameters would be expected to have a global action and therefore influence the shape of the whole face. Others may have a limited scope and therefore influence only part of the face. It is proposed that we encode faces in terms of dimensions which capture the major sources of variation in facial shape. These are likely to reflect growth processes, muscular action, and other structural influences (Bruce, 1988). This perspective implies a very close relationship between the encoding of facial configurations and the analysis of dynamic information.

The experiment investigates the effects of movement on recognition for photographic negative stills and movies presented upright and inverted. If the benefits of motion are related to configurational encoding we would only expect to see differences in performance for the upright stimuli. If the benefits are more general, for example if they involve an improvement in the representation of the three-dimensional structure of the stimulus, then motion should aid recognition for both upright and inverted stimulus presentation.

METHOD

Subjects

Forty subjects (21 male, 19 female) with an age range of 20–44 years (mean age 24.1 years) participated in the experiment. All subjects were graduates and undergraduates at University College London and were fluent English speakers who had lived in Britain for at least five years.

Design

This experiment had a mixed design with one between-subjects factor of *presentation* (images presented in negative or positive) and two within-subjects factors of *movement* (still or moving) and *orientation* (upright or inverted presentation). Twenty subjects saw the faces presented in negative and twenty

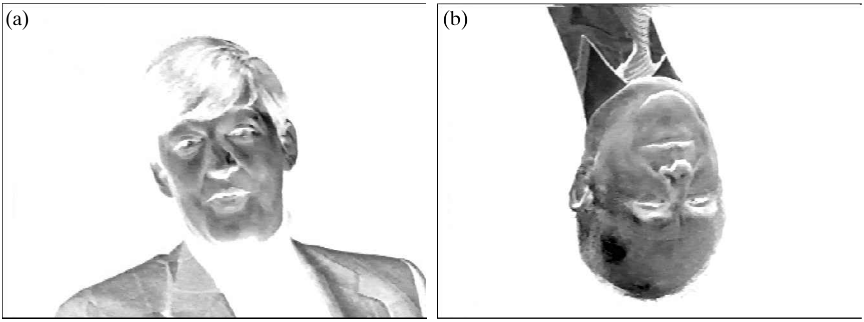

FIG. 1. Examples of upright and inverted negative faces: (a) Stephen Fry, (b) Clive James.

saw the faces in positive. Still and moving faces were grouped, with order effects controlled for by counterbalancing the sequence of presentation of the still and moving conditions. Within each sequence inverted and upright faces were randomly interleaved (Figure 1). Subjects were scored on the number of famous faces correctly identified, calculated as a proportion of those faces known to the subject. Each subject was presented with 80 faces in the same order and orientation.

Materials

Four VHS videotapes were constructed. Each contained 80 images of 40 famous and 40 unknown people presented in a random order with two example faces included at the beginning of each videotape. The duration of each image was 5s with 10s of blank tape separating the image sequences. Out of the corpus of faces, 20 of the famous faces and 20 of the unknown faces were chosen at random and inverted. The order in which the faces appeared was the same for all four tapes. In the first tape the first 40 faces were still and the second set were moving faces. In the second tape this was reversed and the first 40 faces were moving and the second set were still. Each subject in the positive group saw one of these two tapes. This counterbalances for both order of presentation of still and moving conditions and also for the allocation of faces to each condition. We then constructed two new tapes which were photographic negative copies of the two original tapes. No subject saw the same face twice.

The 40 famous faces were selected from a larger pool of 60 individuals by eight independent judges, on the criterion that they would be likely to be recognized by most people. These comprised 28 male and 12 female faces. The 40 unknown faces were selected to match approximately the famous faces on age, sex, and general appearance. The images were videotaped from current news and talk programmes and revealed head and shoulders only against a neutral background. All moving faces were shown speaking towards the camera in an interview situation. Exaggerated facial expressions were avoided. The still views were selected from the moving sequences with the mouth in a closed or almost closed position, again avoiding exaggerated facial expressions. We chose a frame which was close to a full face view and which we considered to be a typical member of the image sequence.

Apparatus

Faces videotaped on VHS off air were copied to umatic hiband (Sony 9800), edited through a Mickey edit controller and a Roger timebase corrector and copied to umatic loband (Sony 5800). Images were transformed into negative and inverted through a FOR 9 VEC-400 video effect controller and a FOR 9 FA PS digital timebase corrector to the final VHS videotapes. The videotaped stimuli were presented to subjects on a 20" Mitsubishi television through a JVC video recorder. Photographs were taken from the video pictures with a Nikon F2 camera using Kodacolor 35mm 100 ASA film (exposure $^1/_8$s at f5.6).

Procedure

Subjects were tested in groups of up to three. The following instructions were issued:

> You are about to see some faces. Some of these faces belong to famous or well-known people. Some faces will be inverted, some will be still and some will be moving. Your task is to identify the person on the screen. If the face is unfamiliar to you, tick the NO column. If the face is familiar to you, (1) tick the YES column and (2) identify the person either (a) by name or (b) by other distinguishing characteristics, such as occupation, nationality, etc., if you cannot recall the name. Two faces will be shown at the beginning of the experiment to demonstrate the presentation and timing of the following stimuli.

Subjects recorded their responses in the 10s interval between trials.

When all 80 images had been presented, subjects were asked to look at photographs of the 40 famous people and indicate any who were unknown to the subject. Subjects who were unfamiliar with more than three famous faces in any one condition were not included. Scores were calculated as a proportion of those faces which were known to the subject. For example, if in any one condition only eight of the faces were known to the subject and they identified those eight correctly they were given a score of 100%.

Results

A 2 (presentation) × 2 (movement) × 2 (orientation) ANOVA with repeated measures on the last two factors reached significance (Table 1). Overall, faces

TABLE 1
Percentage of Correct Responses for All Conditions

	Moving–Up	*Still–Up*	*Moving–Inverted*	*Still–Inverted*
Negative				
Percentage correct	61.4	42.0	34.4	26.3
SD	27.8	21.6	18.0	19.4
No. of subjects	20	20	20	20
Positive				
Percentage correct	88.0	88.4	74.9	69.4
SD	16.3	11.5	22.2	18.5
No. of subjects	20	20	20	20

were better recognized when presented in positive than in negative ($F(1, 38) = 65.58$, $p < 0.0001$), moving faces were better recognized than still faces ($F(1,38) = 7.15$, $p = 0.011$), and upright faces yielded higher recognition rates than inverted faces ($F(1, 38)=78.77$, $p = < 0.0001$). The significant effects of negation and inversion were expected and confirm the findings of previous experiments. There were no significant two-way interactions (presentation × movement: $F(1, 38) = 3.37$, $p = 0.074$; presentation × orientation: $F(1, 38) = 1.57$, $p = 0.218$; movement × orientation: $F(1, 38) = 0.41$, $p = 0.525$). However, the three-way interaction of presentation × movement × orientation was close to significance ($F(1, 38) = 4.05$, $p = 0.051$) and so we investigated the effect of movement in more detail.

The main question to be addressed in the current study was whether movement would facilitate recognition when the faces were presented in negative. Statistical analysis of this planned comparison revealed that when presented in negative, significantly more moving–upright faces were recognized than still–upright faces ($F(1, 68) = 13.62$, $p = 0.0004$). The next question was whether movement facilitated recognition when the negative faces were inverted. Comparison of moving–inverted faces with still–inverted faces revealed no significant effect of movement on recognition ($F(1, 68) = 2.37$, $p = 0.128$). When presented as positive, there was no significant effect of movement on recognition for either the upright ($F(1, 68) < 1$, $p = 0.944$) or the inverted images ($F(1, 68) = 1.08$, $p = 0.301$).

Discussion

These results suggest that movement significantly improves recognition only in *upright–negative* faces. As expected, movement played no role in recognition when the faces were presented as positive, whether upright or inverted. However, when the faces were presented in negative, moving faces were significantly better recognized than still faces. The absence of an advantage of moving sequences over still images for positives supports the view that for many tasks the information provided by movement is largely redundant. This does not of course mean that subjects do not utilize dynamic cues in information-rich displays.

A motion sequence provides information about dynamic change but also results in the presentation of more views of a face or more examples of possible configurations of the face. One might expect better recognition for moving stimuli than static stimuli because of this increase in the richness of the stimulus material without regard to the encoding of dynamic information. However, there is no benefit from the additional views in the case of inverted negatives or positives even though both manipulations lead to a degradation of performance with respect to the *upright–positive* condition. This suggests that the benefits in performance with motion found for the upright–negative faces reflect the processing of dynamic information and do not arise from a trivial effect of numbers of views.

Bruce and Valentine (1988) suggested that movement contributed to the three-dimensional structure of the face, and indeed several subjects reported that the negative moving faces appeared to be more three-dimensional than the still faces. Depth cues are considerably reduced when faces are presented in negative but movements in the negative image would allow the recovery of three-dimensional structure from motion (Hildreth, Norberto, Grzywacz, Adelson, & Inada, 1990; Johansson, 1973, 1977, 1985; Ullman, 1984). The introduction of this structure from motion cue should compensate for the degradation of shape-from-shading cues and facilitate the recognition of the face. Thus it is possible that the benefits of dynamic information act through the better representation of the three-dimensional structure of the face.

However, in experiments using the point-light technique of Johansson (1973, 1977), Bruce and Valentine (1988) found that although the recognition of facial expressions was performed quite accurately the identification of familiar faces was relatively poor. This then raises the question of why familiar moving faces were recognized with considerably more accuracy when presented in negative in comparison to point-light displays. Moving photographic negatives generate a pattern of image velocities which are identical to the normal dynamic signals generated by the face. Motion information is not degraded in this manipulation in the way that it is in sparse displays. It is likely that the degraded information in point-light displays is sufficient to recognize general categories of events such as those generated by a change in expression but insufficient to distinguish between two similar but different individuals.

If the motion advantage we see is derived from the better representation of three-dimensional structure it is difficult to understand why movement does not increase recognition in the case of inverted faces, since the recovery of 3D spatial structure should be unaffected by inversion. The explanations may lie in the difficulty of accessing configural information in inverted faces. It may be that, whatever form the information about the face takes, it is more difficult to recover configurational information from inverted faces. If, as suggested by Sergent (1984), subjects resort to analysing the features individually because the configuration is not easily accessed in inverted faces, this would occasionally lead to recognition of those faces with very unusual or distinguishing features. Most faces, however, would remain difficult to recognize. Alternatively moving faces may be recognized on the basis of characteristic dynamic events or facial gestures. These are in effect constrained changes in the configuration of the face. The extraction and reconstruction of these facial gestures or mannerisms forms the basis of mimicry. If subjects are recognizing famous faces on the basis of these highly characteristic changes in expression or movements of the head then one would expect improvement in performance from the static to the dynamic case for upright faces. No improvement would be expected for the inverted case, however, because the recognition of configural change would require the same conditions for optimal performance as required for recognition of facial configuration.

REFERENCES

Bartlett, J., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology*, *25*, 281–316.

Bassili, J. (1978). Facial motion in the perception of faces and emotional expression. *Journal of Experimental Psychology*, *4*(3), 373–379.

Bassili, J. (1979). Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology*, *37*(11), 2049–2058.

Bruce, V. (1988). *Recognising faces.* London: Lawrence Erlbaum Associates Ltd.

Bruce, V., Doyle, A., Dench, N., & Burton, M. (1991). Remembering facial configurations. *Cognition*, *38*, 109–144.

Bruce, V., & Valentine, T. (1988). When a nod's as good as a wink: The role of dynamic information in facial recognition. In M.M. Gruneberg, P.E. Morris, & R.N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (Vol. 1, pp. 169–174). Chichester, UK: Wiley.

Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of faces. *Science*, *195*, 312–314.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, *115*, 107–117.

Goldstein, A., & Chance, J. (1981). Laboratory studies of face recognition. In G. Davies, H. Ellis, & J. Shepherd (Eds.), *Perceiving and remembering faces.* London: Academic Press.

Hayes, T., Morrone, M., & Burr, D. (1986). Recognition of positive and negative bandpass-filtered images. *Perception*, *15*, 595–602.

Hildreth, E., Norberto, M., Grzywacz, N., Adelson, E., & Inada, V. (1990). The perceptual buildup of three-dimensional structure from motion. *Perception and Psychophysics*, *48*, 19–36.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, *14*(2), 201–211.

Johansson, G. (1977). Studies of perception of locomotion. *Perception*, *6*, 365–376.

Johansson, G. (1985). About visual perception. In C. Warren & R. Shaw (Eds.), *Persistence and change.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Johnston, A. (1992). Object constancy in face processing: Intermediate representations and object forms. *Irish Journal of Psychology*, *13*(4), 426–439.

Johnston, A. (1996). Surfaces, objects and faces. In D.W. Green et al., *Cognitive science: An introduction.* Oxford, UK: Basil Blackwell.

Johnston, A., Hill, H., & Carman, N. (1992). Recognising faces: Effects of lighting direction, inversion and brightness reversal. *Perception*, *21*, 365–375.

Luria, S., & Strauss, M. (1978). Comparison of eye movements over photographic positives and negatives. *Perception*, *7*, 349–358.

Rhodes, G. (1988). Looking at faces: First-order and second-order features as determinants of facial appearance. *Perception*, *17*, 43–63.

Sergent, J. (1984). An investigation into component and configural processes underlying face recognition. *British Journal of Psychology*, *75*, 221–242.

Shepherd, J.W., Ellis, H.D., and Davies, G.H. (1982). *Identification evidence: A psychological evaluation.* University of Aberdeen Press, Aberdeen.

Schiff, W., Banka, L., and Gauldie, G.D. (1986). Recognising people seen in events via dynamic "mug shots". *American Journal of Psychology*, *99*, 219–231.

Tanaka, J., & Farah, M. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, *46A*, 225–245.

Ullman, S. (1984). Maximising rigidity: The incremental recovery of 3-D structure from rigid and non-rigid motion. *Perception*, *13*, 255–274.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion on face recognition. *British Journal of Psychology*, *79*, 471–491.

Young, A., Hellawell, D., & Hay, D. (1987). Configural information in face perception. *Perception*, *16*, 747–759.

Manuscript received 1 March 1995
Revised manuscript received 5 March 1996