Perception, 2003, volume 32, pages 561 – 566

DOI:10.1068/p3435

Comparing solid-body with point-light animations

Harold Hill

Department 2, Human Information Science Laboratories, 2-2-2 Hikaridai, Keihanna Science City, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan; e-mail: hill@atr.co.jp

Yuri Jinno, Alan Johnston

Department of Psychology, University College London, London, WC1E 6BT, UK

Received 5 August 2002, in revised form 29 January 2003; published online 23 April 2003

Abstract. The movement of faces provides useful information for a variety of tasks and is now an active area of research. We compare here two ways of presenting face motion in experiments: as solid-body animations and as point-light displays. In the first experiment solid-body and point-light animations, based on the same motion-captured marker data, produced similar levels of performance on a sex-judgment task. The trend was for an advantage for the point-light displays, probably in part because of residual spatial cues available in such stimuli. In the second experiment we compared spatially normalised point-light displays of marker data with solid-body animations and pseudo-random point-light animations. Performance with solid-body animations and normalised point-light displays was similar and above chance, while performance with the pseudorandom point-light stimuli was not above chance. We conclude that both relatively few well-placed points and solid-body animations provide useful information about facial motion, but that a greater number of randomly placed points does not support above-chance performance. Solid-body animations have the methodological advantages of reducing the importance of marker placement and are more effective in isolating motion information, even if they are subsequently rendered as point-light displays.

1 Introduction

We recover a lot of useful information from the movement of people's faces, including cues to speech, emotional state, and attention. This knowledge has made studying how this information is perceived and encoded an active area of research in perception. While many studies of the motion of faces have made use of video presentation, where spatial cues are present even if deliberately degraded (eg Lander et al 1999), other studies have attempted to isolate motion information (eg Bassili 1979). The aim of this paper is to compare the relative effectiveness of different ways of presenting motion information in isolation.

The first attempts to isolate motion information adapted Johansson-type (Johansson 1975) point-light displays of whole-body movements to faces (Bassili 1979; Bruce and Valentine 1988; Rosenblum et al 1996; Rosenblum and Saldaña 1996). These stimuli are produced by filming bright or reflective dots attached to the face and presenting the resulting stimuli at high contrast so that only the dots can be seen. The frames from such stimuli are often not recognisable when presented statically but are informative when presented in motion. This is interpreted as meaning that they contain only motion information and little or no static spatial information. While point-light displays have been found to provide useful information for a number of tasks, there are at least two disadvantages associated with such stimuli. First, they provide only a limited sampling and representation of the motion information available in natural facial movement. Second, unless explicitly normalised, point-light displays contain residual cues to spatial configuration as well as motion information. For example, the aspect ratio of a face, a spatial cue, is recoverable from a point-light display as are structure-from-motion-based cues to differences in 3-D shape. Structure-from-motion, although not recoverable from a single frame, is also an essentially spatial cue in the sense that it provides information about shape rather than about motion per se.

Solid-body computer animation provides a means to remove these spatial cues, at the same time presenting face-based motion in a more natural-looking way. Movements derived from different people or different events can be mapped to the same 3-D model, in our case an average head model (Vetter and Troje 1997), and used to generate animations where all differences are a function of differences in motion and not differences in shape. While the sampling of the original motion is still limited, these samples can be used to generate a continuous-motion field on the model through the use of weighting functions (Hill and Johnston 2001; Knappmeyer et al 2001). Whether the continuous-motion field generated by the mapping of motion information to the face model adds useful information or not will depend on the extent to which it captures correlations between the movement of the markers and the movement of neighbouring areas of the faces. One of the aims of the experiments reported here is as an explicit test whether the additional information provided by solid-body animations facilitates performance.

Previous work has shown that these solid-body animations do provide useful information for face-processing tasks, including judging whether a face is male or female. That task requires access to prior knowledge about sex differences in facial movement. Task performance depends on the stimuli being presented upright and played forwards, suggesting that the movement information used is face and direction specific and is not just based on low-level motion cues (Hill and Johnston 2001). We can also judge sex from point-light animations, though in that case much of the useful information appears to be associated with the cues provided by rigid rotations of the whole head to the underlying 3-D shape (Berry 1991; Bruce and Valentine 1988).

In this paper we use performance on the sex-judgment task to compare the usefulness of the same motion information presented either as point-light displays or as solid-body animations. Previous work has shown advantages for solid-body animations over point-light displays on a speech-reading task, but in that work all movements were synthetic and had been developed specifically for the solid-body animations (Cohen et al 1996). Our animations are driven by motion data captured from real faces and, as such, are not tailored to any particular means of presentation. In experiment 1 the unnormalised movement of the original markers was compared with solid-body animations, while in experiment 2 solid-body animations, normalised markers, and stimuli similar to those described by Bassili (1979) were all compared.

2 General methods

2.1 *Materials*

The stimuli used in these experiments were based on dual video recordings of faces with 17 coloured markers attached. The videos were of twelve people, six male and six female, telling four short ($\sim$5 s) jokes to a friend sitting in front of them. This was designed to elicit natural and expressive speech. There were 6 markers around the vermilion lip border, 3 on each eyebrow, and 1 each on the chin, left and right temples, the top centre of the forehead, and the bridge of the nose. The last 4 markers listed were on parts of the face that do not move much relative to the head, and were used to estimate the rigid rotations and translations of the whole head. The pupils of the eyes were also tracked. The number of markers on the lips was limited by the resolution of the tracking system; while some additional markers on the cheeks could also be tracked, we were not able to use the information from these to enhance the animations.

The 2-D positions of the points were tracked in each video with FamousFaces vTracker software. Both cameras were calibrated with a calibration object of known dimensions enabling recovery of 3-D marker positions. For the 'unnormalised-markers' condition of experiment 1, x and y coordinates from the calculated 3-D marker positions were plotted for each frame. For the solid-body and other point-light conditions, we

used the 3-D positions to generate motion files as input to FamousFaces Animator software. In this software weighting functions are used to link the movement of the markers to the movement of a 3-D model, in this case an average face (Vetter and Troje 1997). The animated model was then rendered in 3-D StudioMax. For solid-body animations we used the average texture while, for the Bassili-type stimuli, faces were texture-mapped with a bitmap of fifty 8 mm diameter white circles (Bassili 1979). For the normalised-markers condition, individual pairs of triangular facets, in approximately the same positions on the average head as for the markers on the original faces, were made white while the rest of the head model was left black. All these conditions are effective 'spatially normalised' by being projected onto the same underlying shape. Examples of all the stimuli are shown in figures 1 and 2 and corresponding animations can be found at http://www.his.atr.co.jp/~hill/ptLgt.html and on the *Perception* website at http://www.perceptionweb.com/misc/p3435/.

2.2 *Experiment 1*

In this experiment we compared performance on a 2AFC sex-judgment task when viewing either the original, unnormalised markers, or solid-body animations generated from the same motion-capture data. The two types of animation were presented in different blocks to the same sixteen observers, with order counterbalanced. The 4 examples of each of the 6 same-sex faces gave a total of 24 trials when randomly paired with an animation of a face of the opposite sex.

2.2.1 *Results and discussion.* The mean (and standard error) proportions correct for unnormalised markers and solid-body animations were 66.9% (4%) and 58.4% (3%), respectively. A paired samples t-test showed no significant difference between conditions, $t_{15} = 1.7$, $p > 0.1$. One sample t-test showed that both conditions were significantly better than chance (50%) ($t_{15} = 2.6$, $p < 0.05$ for animations and $t_{15} = 4.8$, $p < 0.005$ for unnormalised markers). The trend was for dots to have an advantage over animations, perhaps because of residual spatial cues including the aspect ratio and relative proportions of the face available only in the unnormalised-markers condition. There was no effect of the order in which solid-body and point-light animations were presented (ps > 0.1).

2.3 *Experiment 2*

In experiment 2, the motion information available in all conditions was spatially normalised by mapping it onto the same average head. This provided a test of the extent to which such cues are critical.

The point-light animations used in experiment 1 also differed from traditional point-light stimuli in that the face and even features were clearly identifiable even from static frames (see figure 1). To test whether this was critical, in this experiment we introduced a condition based on the stimulus description given in Bassili (1979), where marker placement was pseudorandom and marker positions did not correspond with the positions of any clearly defined facial features. The task was again the 2AFC sex-judgment task with three different stimulus conditions: solid-body animations, normalised markers, and Bassili-type stimuli (see figure 2). The three conditions were presented in separate blocks, with the six possible block orders counterbalanced across subjects. There were 24 trials in each block (as for experiment 1) and the pairing of male with female animations was fully randomised.

2.3.1 *Results and discussion.* Mean proportions correct (and standard errors) were 58.0% (2.5%) for animations, 60.5% (2.7%) for normalised dots, and 54.1% (3.1%) for Bassili-type stimuli. A one-way repeated-measures analysis of variance showed no effect of presentation type, $p > 0.1$. However, one sample t-tests showed that performance was above chance with solid-body animations and normalised markers ($t_{11} = 3.3$, $p < 0.05$; and $t_{11} = 4.0$, $p < 0.005$ respectively), but not with the Bassili-type stimuli ($p > 0.1$).

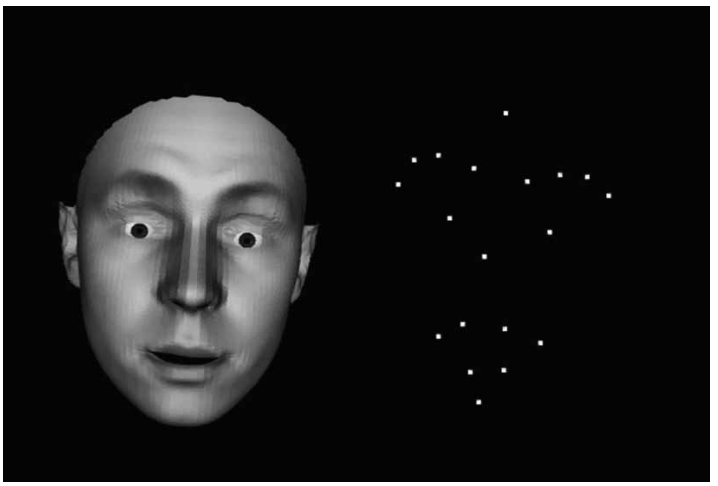

Figure 1. The stimuli used in experiment 1 with the solid-body animation on the left and the orthographic projection of the unnormalised markers on the right. Both images show the same frame from the same sequence.

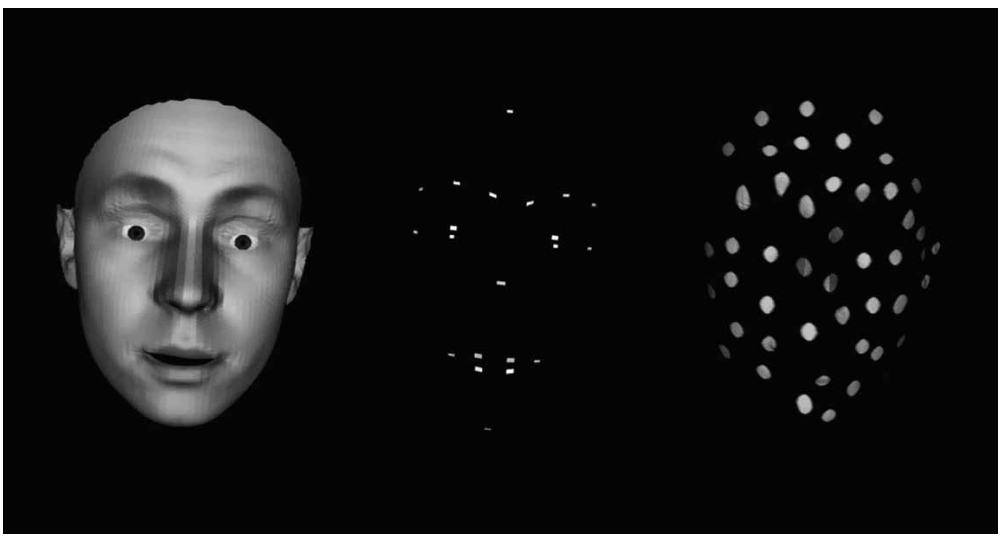

Figure 2. The stimuli used in experiment 2. The left frame shows the solid-body animation, the central frame the normalised markers, and the right frame the Bassili-type stimuli. All images are of the same frame from the same sequence as used for figure 1.

As with experiment 1, there was no effect of presentation order on performance, suggesting that practice without feedback does not facilitate this task.

3 General discussion

In the experiments reported here, performance was as good for dots corresponding to the original markers as for solid-body animations, even when residual spatial differences had been normalised. This suggests that our animation techniques did not add useful information to that available from the original markers, at least for this task. It seems the mind is as at least as good at 'filling in' the missing motion information and generating an overall impression of facial movement as our animation system is. The placement of dots seems more important than their density in that a limited number of dots, on clearly identifiable features (14 on features), supported above-chance

performance, while a greater number (50) of randomly placed dots did not. The animation system may even add noise in the form of the nonveridical movement that is added by the weights to the signal derived directly from the movement of the markers. Cues to sex from the shape and texture may also act as noise in that, being held constant, they are not a signal for the task.

This finding contrasts with a previously reported advantage for solid-body animations over point-light displays (Cohen et al 1996). This study differed from ours in a number of ways. In particular, the previous study used synthetic movements that had been optimised for display as solid-body animations. In our study, with real movements, this advantage did not appear to hold. Other possible reasons for the differences between the studies include the task—speech reading as opposed to sex judgments—the placement of dots, and the limited amount of head movement in the previous study. A greater amount of rigid head movement, for example, may facilitate the recovery of structure-from-motion from point-light displays and thus improve performance when motion information is presented in this way. This information would, anyway, be expected to be more important for judging sex, where there are known differences in underlying structure, than for facial speech where the relevant information has to be recovered independently of individual or sex-based differences in structure. Other studies, with nonfacial movements, have shown task-dependent advantages of solid-body animations (Paterson et al 2002). It also appears that point-light stimuli may be sufficient when there is existing knowledge to tap on related movements, as in our experiments, but not when explicit comparisons, that have to be made solely on the basis of the stimuli, are involved (Hodgins et al 1998).

One reason why we might have expected better performance with solid-body animations is that they provide at least the approximation of a continuous-motion field. However, it is clear from the results that a continuous field is not essential for the recovery of useful motion information—even when normalised, relatively few dots corresponding to the original markers conveyed useful motion information as effectively as solid-body animations. Also, increasing the number and size of dots in the Bassili condition did not support above-chance performance despite providing a fuller sampling of the velocity field. In the original experiments, performance was above chance (Bassili 1979). This difference may have been in part a function of the task: emotional expressions perhaps involve more of the face than the speech-related movements used here. Also, here the trend was for the highest level of performance with unnormalised markers, suggesting that structure-from-motion cues may have played an important part in the original studies (see also Bruce and Valentine 1987). The perspective distortions of the relatively large dots used (see figure 2) may have provided additional information in the original experiments but would not have been informative here.

The results reported here suggest that the placement of point lights rather than their density may be a critical determinant of performance. Results of experiments with point-light displays will always be a function of the particular placement of dots to some extent, and this detracts from their aim of studying the effects of movement. Although the same argument can be applied to the placement of markers, for marker-based motion capture, the estimation of all the intermediate points should reduce the absolute differences between the resulting stimuli. For example, we use a spline to estimate the movement of the lips. Thus, the resulting motion should not change dramatically with the placement or even the number of markers, although these factors would result in obvious changes to the appearance of point-light displays of those markers.

The results clearly demonstrate the importance of marker placement, in that performance was above chance with a set of markers located on important facial features but not with a greater number of markers place pseudorandomly. The particular parts of the face that were tracked—the eyebrows, eyes, and the vermilion border of the

lips—may be particularly important for the perceptions of facial motion. All these facial features have high-contrast borders that may make motion easier to recover, as compared with areas of relatively uniform contrast like the cheeks. From these results it does not appear that the extra information about these areas, provided by the Bassili or solid-body-type stimuli, facilitated performances, although this may have been in part a limitation of the way in which these areas were animated.

For the rigid head movements known to be useful for some tasks (Hill and Johnston 2001) a very limited number of dots should be sufficient for recovery, so long as those dots are placed on relatively rigid parts of the face. The rigid movements of our animations are based on the movement of 4 markers and, in theory, 3 should be sufficient if their 3-D positions can be recovered.

In conclusion, useful movement information can be derived even from a limited number of point lights. Relatively few points may be sufficient to recover overall head movement, the movement of the eyes and, if they are captured, the movement of the mouth and eyebrows—critical aspects of nonrigid facial motion. It does not appear that a continuous-motion field is necessary for the recovery of this information, so long as key areas are represented, and approximating a dense field through solid-body animation does not significantly improve performance, at least for this task. Indeed, increasing the number of motion samples available may actually reduce performance when these additional samples are randomly distributed, perhaps because they mask the perception of the movement of key facial features. However, the animation of a solid-body model has the methodological advantages of normalising the spatial cues available and reducing the importance of marker placement.

Acknowledgments. This work was funded by the UK Engineering and Physical Sciences Research Council and by the Japanese Telecommunications Advancement Organization and Communications Research Laboratories.

References

Bassili J N, 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** 2049 – 2058

Berry D S, 1991 "Child and adult sensitivity to gender information in patterns of facial motion" *Ecological Psychology* **3** 349 – 366

Bruce V, Valentine T, 1988 "When a nod's as good as a wink: The role of dynamic information in face recognition", in *Practical Aspects of Memory: Current Research and Issues* Eds M M Gruneberg, P E Morris, R N Sykes (Chichester, Sussex: John Wiley) pp 169 – 174

Cohen M M, Walker R L, Massaro D W, 1996 "Perception of synthetic visual speech", in *Speechreading by Humans and Machines* Eds D G Stork, M E Hennecke (Berlin: Springer) pp 153 – 168

Hill H, Johnston A, 2001 "Categorizing sex and identity from the biological motion of faces" *Current Biology* **11** 880 – 885

Hodgins J K, O'Brien F O, Tumblin J, 1998 "Perception of human motion with different geometrical models" *IEEE Transactions on Visualization and Computer Graphics* **4** 307 – 316

Johansson G, 1975 "Visual motion perception" *Scientific American* **232**(6) 76 – 88

Knappmeyer B, Thornton I M, Bülthoff H H, 2001 "Facial motion can determine facial identity" *Journal of Vision* **1** 338

Lander K, Christie F, Bruce V, 1999 "The role of movement in the recognition of famous faces" *Memory and Cognition* **27** 974 – 985

Paterson H M, Pollick F E, Ude A, 2002 "Shaping biological motion: Adding realistic form cues to biological motion displays" *Journal of Vision* **2** 336a

Rosenblum L D, Johnson J A, Saldaña H M, 1996 "Visual kinematic information for embellishing speech in noise" *Journal of Speech and Hearing Research* **39** 1159 – 1170

Rosenblum L D, Saldaña H M, 1996 "An audiovisual test of kinematic primitives for visual speech perception" *Journal of Experimental Psychology: Human Perception and Performance* **22** 318 – 331

Vetter T, Troje N F, 1997 "Separation of texture and shape in images of faces for image coding and synthesis" *Journal of the Optical Society of America A* **14** 2152 – 2161

ISSN 0301-0066 (print) ISSN 1468-4233 (electronic)

PERCEPTION
VOLUME 32 2003

www.perceptionweb.com