Differential Processing of Facial Motion

Tamara L. Watson[1], Alan Johnston[1], Harold C.H Hill[2] and Nikolaus Troje[3]

[1]Department of Psychology, University College London, Gower Street, London WC1E 6BT
[2]ATR Human Information Science Laboratories, Kyoto, 619-0288
[3]Fakultät für Psychologie, Ruhr-Universität-Bochum, 44780 Bochum

Abstract. To investigate viewpoint dependence in dynamic faces an avatar was animated using actors' movements. In Experiment 1 subjects were shown a full-face animation. They were then asked to judge which of two rotated moving avatars matched the first. Test view, orientation and the type of motion were manipulated. In a second experiment subjects were shown two views of the same facial animation and were asked which of the two avatars was the same as the initial animation. Initial views could be rotated to 15° and 45° or 45° and 75° while test views were presented at 30° or 60°. Learnt view, test view, orientation and type of movement (rigid + non-rigid vs non-rigid) were manipulated. Both experiments and movement conditions produced an advantage for upright over inverted matching demonstrating subjects were encoding facial information. Non-rigid movement alone showed no effect of view for both experiments demonstrating viewpoint invariance. Rigid and non-rigid movement presented together produced a decline in performance for larger test rotations in Experiment 1, while Experiment 2 produced a differential advantage for 30° test rotation when initially viewed upright faces were rotated to 15° and 45° however no difference was found in the 45° and 75° condition or with inverted faces. These experiments suggest that non-rigid facial movement is represented in a viewpoint invariant manner whereas the addition of rigid head movements encourages a more viewpoint dependent encoding when the initial orientation of the head is not rotated further than the half profile (45°).

What role does motion play in the recognition of faces? Two types of motion, rigid transformations of the head and non-rigid deformations that occur during speech and changes in expression, are available to the viewer during social interaction. Research to date suggests that rigid motion of a head does provide beneficial information for the viewer. Pike et al. [1] have shown that this additional motion information presented at learning can enhance recognition. It is suggested that this advantage is affected by the ability to build up or access a 3-dimensional representation. The extra structural information provided by the rigid transformational motion of the head offers more opportunity to encode or access this information. However it is rare that when we are introduced to a person we see their face moving in the highly controlled way that was adopted by Pike et al. [1]. During most social interaction we will also be exposed to the face moving in a non-rigid manner.

The advantages of non-rigid motion for recognition have been the subject of debate. It has been shown that a degraded representation of a face will benefit from the addition of non-rigid motion particularly for faces the viewer is familiar with. Knight and Johnston [2] have demonstrated that recognition of degraded famous faces will be significantly enhanced by the addition of non-rigid motion. Lander, Christie and Bruce [3] have demonstrated the same advantage with degraded famous faces. Christie and Bruce [4] have studied the effects of presenting non-rigid motion at training and at test with unfamiliar faces. They found no advantages for presenting motion at training or at test and suggest that non-rigid motion may only be beneficial when accessing existing representations.

Recently Thornton and Kourtzi [5] have used a sequential matching task rather than a recognition task in the study of non-rigid facial motion. They demonstrate that presentation of a short video sequence aided matching when the face differed in expression or viewpoint between prime and test images. The demonstration of an advantage in sequential matching of unfamiliar faces after presentation of a face moving non-rigidly in this study is interpreted with the view that mechanisms responsible for representing change over time are established and maintained in working memory and show little transference to long term memory over the course of the study.

All of the above studies have presented spatial layout cues alongside motion cues and have therefore not studied the role of facial motion alone. The question of whether facial motion can be represented independently of spatial cues remains open. However, Hill and Johnston [6] have shown that both rigid head movements and non-rigid head movements in the absence of spatial cues provide sufficient information to allow observers to categorize faces on the basis of both identity and gender. On the basis of differences between accuracy of categorization depending on the type of motion, Hill and Johnston [6] suggest that rigid movements are idiosyncratic and provide the basis for performance in identity categorization while non-rigid movements provide independent cues to speech and expression. These results would appear to complement the findings discussed above in that a more permanent representation is possibly mediated by encoding rigid motion while speech and expression are both encoded in a more transient manner.

The recognition of static faces has typically been found to be viewpoint dependent. Results of studies such as that by Hill, Schyns and Akamatsu [7] suggest that when a single view is presented during a learning stage, recognition of the same face from other views is impaired. They also found that the addition of cues that do not vary over view, such as facial colouring, greatly enhanced the accuracy of the results to the extent that learning presentation times need to be reduced. Recognition for the reduced presentation time was also found to be view dependent for conditions except in the case of the ¾ learnt views. These results suggest that generalized prior knowledge of the 3-dimensional structure of faces does not allow a view invariant representation of a face to be accessed when generalizing from a single static view.

As non-rigid facial motion is specifically a property of the object in motion it cannot be mimicked by movement of the viewer in the same way that rigid transformations can. Since this non-rigid motion is a change in the intrinsic shape of an object, it would make sense for the visual system to encode the motion in a viewpoint independent way if possible.

The first experiment was designed to assess view dependence when matching non-rigid facial movement as opposed to both rigid and non-rigid movement together from a full-face view.

Stimuli for both experiments reported consisted of a total of 64 animations based on motion capture recordings of 8 males and 8 females, each telling 4 question and answer type jokes. Recordings were made with an eight camera Oxford Metrics' Vicon motion capture system with the cameras placed in a semicircle at different heights in front of the head. Forty markers were used to capture facial movement and a headband with 4 markers was used to capture rigid movements. The resulting motion information was used to animate an average 3-dimensional facial model created from 100 male and 100 female faces [8]. Animation of the 3d model was achieved in Famous Animator where 'areas of influence' around each marker placed on the face inherit the movement of the marker (see also [6]). As no eye movements were captured the eyes were made to "look at" a point straight ahead of the face. The three-dimensional head model was texture mapped with a corresponding average texture and the resulting animated sequences rendered using 3DS Max. Two versions of each sequence were rendered; one with just non-rigid facial movements and the other with both types of movements combined.

Two groups of 20 subjects were presented with animations containing only non-rigid motion or rigid and non-rigid together. During one trial participants were first shown a learning animation sequence oriented at 0° (where 0° is a full face and 90° a profile). This was followed by a target and distracter animation presented sequentially at an orientation in depth of 0°, 15°, 30°, 45°, 60°, 75° or 90°. Participants were asked to indicate which animation was shown in the learning stage. The target animation was the same sequence as the learning animation while the distracter was randomly chosen with the constraint that it contains an actor telling the same joke as the target stimulus. Both target and distracter animations were shortened such that the video sequence would start at a random point within the first half of the animation and run for half the length of the full animation. Shortening the animation was required in order to lower performance from ceiling. Each animation could only be viewed once and all animations were required to have been viewed before response. Subjects controlled the speed of presentation.

Faces were also presented upside down as a control in order to assess the likelihood that subjects were utilizing extraneous cues in order to carry out the task. It has been shown previously that presenting inverted facial motion reduces the accuracy of gender and identity judgments suggesting that upright facial motion is represented in a object-motion encoding system specialized for faces [6]. Inversion constituted a within-subjects condition. During the upright condition all faces were presented upright. During the inverted condition all faces were presented upside down. Initial orientation was randomized. Each condition contained 64 trials.

A within subjects analysis was carried out for each type of motion. Data are shown in Table 1. The effect of test viewpoint for non-rigid motion was not found to be statistically significant, $f(6,114)=2.163$, $p>0.05$ although a significant overall effect of inversion was found, $f(1,19)=5.834$, $p=0.03$. Rigid and non-rigid motion together displayed a significant effect of test rotation $f(6,114)=2.311$, $p=0.04$ which was found to produce a linear trend, $f_{(1,19)}= 14.468$, $p<0.01$. An effect of inversion was also found, $f(1,19)=5.819$, $p=0.03$.

Table 1. Mean percent correct for Experiment 1

	Non-Rigid Motion						
	0°	15°	30°	45°	60°	75°	90°
Upright	82.5	77.2	77.8	75.0	78.3	74.5	73.9
Inverted	76.5	65.0	78.3	67.8	68.1	71.1	77.2
	Rigid + Non-Rigid Motion						
	0°	15°	30°	45°	60°	75°	90°
Upright	82.8	82.1	81.5	77.2	74.1	75.3	69.8
Inverted	76.1	73.5	74.1	73.5	72.2	69.8	69.1

These results suggest that non-rigid facial motion is less viewpoint dependent than non-rigid and rigid facial motion presented together when generalizing from a full face view. Both the statistics and inspection of the data suggests that for upright non-rigid animation there is some decline in performance as test viewpoint rotates away from the target view. However this is not as pronounced as when all motion information is presented within the animation. The advantage displayed for upright animations is suggestive of a specialized face processing module that is not available when inverted animations are presented.

The next experiment was designed to incorporate a larger range of initial views. Non-rigid movement alone was presented to one group of 40 participants while another group viewed animations containing both rigid and non-rigid movements. Two different groups of subjects were shown ‘learning faces’ consisting of two different views of the same animation rotated by either 15° and 45° or 45° and 75° and were then tested on faces rotated to 30° or 60°. The target face was the same as the learnt face while the distracter was chosen randomly with the constraint that it would be at the same rotation as the test and of the same gender. Subjects were asked to indicate which animation had been shown in the learning stage. Both target and distracter animations were shortened and viewing conditions were as in Experiment 1. Inversion acted as an added within-subjects condition.

Table 2. Mean percent correct for Experiment 2

		Non-Rigid Motion		Rigid + Non-Rigid Motion	
	Target Rotation	15° + 45°	45° + 75°	15° + 45°	45° + 75°
Upright Test	30°	81.3	81.6	85.0	88.1
	60°	79.4	81.2	76.3	90.6
Inverted Test	30°	74.3	75.0	74.4	78.1
	60°	70.0	77.8	71.9	78.4

Data are shown in Table 2. Non-rigid motion did not display an interaction between the initial and test rotation of faces $f(1,38)=0.152$, $p=0.70$, however an effect of inversion was found $f(1,38)=14.324$, $p<0.01$. The condition showing all available motion information displayed a significant interaction between the initial and test rotation of faces, $f(1,38)=9.215$, $p<0.01$. This interaction was not present when faces were inverted, $f(1,38)=0.599$, $p=0.45$. Again an overall inversion effect was found, $f(1,38)=33.485$, $p<0.01$.

The results of Experiment 2 suggest that non-rigid motion when presented alone displays less viewpoint dependence than when all motion information is presented in an animation. Further investigation of the data also suggests that viewpoint dependence is stronger for the learnt animations closest to the full face view when compared to those closest to the profile in the full motion condition. This effect could be due to a larger angular difference between the more frontal views presented in this condition compared to the views closer to profile. The advantage for upright animations over inverted, as in Experiment 1, is suggestive of specialized processing of the upright facial motion.

The results reported here suggest that non-rigid transformations of a face may initially be encoded in a less viewpoint dependent manner than faces that are transforming and translating rigidly. It is also noted that while viewpoint dependence has been found in the latter case it does not seem as pronounced as that found when investigating view dependence in static faces. This may suggest that motion does have a large role to play in the perception of faces across different views. That non-rigid motion alone displays no view dependence unless teamed with rigid motion may suggest that the addition of rigid motion into the stimulus adds a layer of information that may make the representation of the motion that is formed less than optimal for the task at hand. Whether this is specific to the rigid motion or occurs as a combination of non-rigid and rigid motion is not clear from these experiments. The inversion effect displayed suggests that the results found in this study are due to specialized processing of the upright facial motion.

References

1. Pike, G., Kemp, R., Towell, N., Phillips, K.: Recognizing Moving Faces: The Relative Contribution of Motion and Perspective View Information. Visual Cognition 4. (1997) 409-437.
2. Knight, B., Johnston, A.: The Role of Movement in Face Recognition. Visual Cognition 4. (1997) 265-273.
3. Lander, K., Christie, F., Bruce, V.: The Role of Movement in the Recognition of Famous Faces. Memory and Cognition. 27. (1999) 974-985.
4. Christie, F., Bruce, V.: The Role of Dynamic Information in the Recognition of Unfamiliar Faces. Memory and Cognition 26. (1998) 780-790.
5. Thornton, I.M., Kourtzi, Z.: A Matching Advantage for Dynamic Human Faces. Perception 31. (2002) 113-132.
6. Hill, H.C.H, Johnston, A.: Categorizing Sex and Identity From the Biological Motion of Faces. Current Biology 11. (2001) 880-885.
7. Hill, H.C.H., Schyns, P.G., Akamatsu, S.: Information and Viewpoint Dependence in Face Recognition. Cognition 62. (1997) 201-222.
8. Vetter, T., Troje, N.: Separation of Texture and Shape in Images of Faces for Image Coding and Synthesis. Journal of the Optical Society of America 14. (1997) 2152-2161