University of Wollongong

Research Online

Faculty of Health and Behavioural Sciences -
Papers (Archive)

Faculty of Science, Medicine and Health

2003

Recognising facial expression from spatially and temporally modified movements

Frank E. Pollick
University of Glasgow

Harold C. Hill
University of Wollongong, harry@uow.edu.au

Andrew Calder
MRC Cognition and Brain Sciences Unit

Helena Paterson
University of Glasgow

Follow this and additional works at: <https://ro.uow.edu.au/hbspapers>

Part of the [Arts and Humanities Commons](#), [Life Sciences Commons](#), [Medicine and Health Sciences Commons](#), and the [Social and Behavioral Sciences Commons](#)

Recommended Citation

Pollick, Frank E.; Hill, Harold C.; Calder, Andrew; and Paterson, Helena: Recognising facial expression from spatially and temporally modified movements 2003, 813-826.
<https://ro.uow.edu.au/hbspapers/965>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Recognising facial expression from spatially and temporally modified movements

Abstract

We examined how the recognition of facial emotion was influenced by manipulation of both spatial and temporal properties of 3-D point-light displays of facial motion. We started with the measurement of 3-D position of multiple locations on the face during posed expressions of anger, happiness, sadness, and surprise, and then manipulated the spatial and temporal properties of the measurements to obtain new versions of the movements. In two experiments, we examined recognition of these original and modified facial expressions: in experiment 1, we manipulated the spatial properties of the facial movement, and in experiment 2 we manipulated the temporal properties. The results of experiment 1 showed that exaggeration of facial expressions relative to a fixed neutral expression resulted in enhanced ratings of the intensity of that emotion. The results of experiment 2 showed that changing the duration of an expression had a small effect on ratings of emotional intensity, with a trend for expressions with shorter durations to have lower ratings of intensity. The results are discussed within the context of theories of encoding as related to caricature and emotion.

Keywords

temporally, modified, movements, recognising, spatially, facial, expression

Disciplines

Arts and Humanities | Life Sciences | Medicine and Health Sciences | Social and Behavioral Sciences

Publication Details

Pollick, FE, Hill, HC, Calder, A & Paterson, H, Recognising facial expression from spatially and temporally modified movements, *Perception*, 32(7), 2003, p 813-826.

Recognising facial expression from spatially and temporally modified movements

Frank E Pollick, Harold Hill¶, Andrew Calder§, Helena Paterson

Department of Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, Scotland, UK; e-mail: frank@psy.gla.ac.uk; ¶ ATR Human Information Science Laboratories, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan; § MRC Cognition and Brain Sciences Unit, Cambridge CB2 2EF, UK
Received 7 December 2002, in revised form 19 March 2003; published online 6 June 2003

Abstract. We examined how the recognition of facial emotion was influenced by manipulation of both spatial and temporal properties of 3-D point-light displays of facial motion. We started with the measurement of 3-D position of multiple locations on the face during posed expressions of anger, happiness, sadness, and surprise, and then manipulated the spatial and temporal properties of the measurements to obtain new versions of the movements. In two experiments, we examined recognition of these original and modified facial expressions: in experiment 1, we manipulated the spatial properties of the facial movement, and in experiment 2 we manipulated the temporal properties. The results of experiment 1 showed that exaggeration of facial expressions relative to a fixed neutral expression resulted in enhanced ratings of the intensity of that emotion. The results of experiment 2 showed that changing the duration of an expression had a small effect on ratings of emotional intensity, with a trend for expressions with shorter durations to have lower ratings of intensity. The results are discussed within the context of theories of encoding as related to caricature and emotion.

1 Introduction

The ability to perceive human action through point-light motion displays (Johansson 1973) has long been used as a way to explore motion processing and the recognition of human movement. Owing to its unique function in human communication, one question which has received attention is whether information about emotional state can be derived from movement. Investigations of dance movements indicate that the emotions of surprise, fear, anger, disgust, grief/sadness, and joy/happiness can be recognised at above chance levels (Walk and Homan 1984; Dittrich et al 1996). Even arm movements alone performing simple actions have been shown to convey information about affect (Pollick et al 2001b). Of particular interest for communication of emotion is the nonrigid motion of the face, and research by Bassili (1978, 1979) has indicated that point-light displays of facial motion can be effective stimuli for the recognition of emotion. Here, we investigate whether the recognition of emotion from point-light displays can be enhanced through the manipulation of spatial and temporal properties of facial movement.

It is well known that exaggerating the spatial features of static images of faces can enhance recognition of a number of facial characteristics (eg facial identity, facial expression, age, and attractiveness). The initial research in this area looked at facial-identity recognition, and showed that line-drawing or photographic-quality caricatures of faces produced better recognition than the original (undistorted) faces (Rhodes et al 1987; Benson and Perrett 1991; Calder et al 1996). In addition, they showed that images in which the facial features were made less distinctive, so-called anticaricatures, were less readily recognised than the original faces. The images in these different studies were prepared by the same basic computer-based procedure in which the positions of a set number of anatomical feature points on each face were exaggerated relative to the locations of corresponding points on a reference norm face. Following on from this research, Calder et al (1997) applied a similar methodology to facial expressions and

found that caricatured expressions were recognised faster than their original counterparts, which, in turn, were more readily identified than anticaricatured expressions. In line with these findings, Calder et al (2000) found a strong linear relationship between the amount by which a facial expression was caricatured and subjects ratings of its emotional intensity.

For movement, there are several possible ways in which a motion can be exaggerated, and two of these have been explored for the exaggeration of human movement. Spatial exaggerations have been studied by Pollick et al (2001a) who showed that recognition of the style of a tennis serve (eg flat, slice, or topspin) could be enhanced by exaggeration of joint locations, relative to a grand-average movement, for every frame of the tennis-serve movement. This technique is thus roughly equivalent to frame-by-frame application of the facial caricaturing technique explained above. Temporal exaggerations have been studied by Hill and Pollick (2000) who demonstrated that recognition of identity from the arm movements involved in drinking was enhanced by exaggeration of temporal properties of movement. To obtain movement exaggerations, they broke the drinking movement down into parts and exaggerated the duration of these parts relative to the average durations. These new exaggerated movements had the identical spatial pattern as the original movement, but different temporal durations for the sub-parts of the movement. After training on a set of original movements, recognition was tested on the original movements trained upon and the exaggerated movements. It was found that the identities of the point-light drinkers were better recognised from the exaggerated movements than from the original movements that had been learned.

Here, we investigate how recognition of facial emotion is influenced by the manipulation of both spatial and temporal properties of 3-D point-light displays of facial motion. Related work by Kamachi and colleagues (Kamachi et al 2001) has used artificially created morphed image sequences between pairs of photographs to study the effect of movement speed on the perception of facial emotion. In our work, we start with real measurements of 3-D position of multiple locations on the face during posed expressions, and then manipulate the spatial and temporal properties of the measurements throughout the movement to obtain new versions of the motion sequences. We then examine categorisation of the original and spatially and temporally modified versions of these facial expressions.

Specifically, we examine a spatial manipulation that consists of exaggerating differences from a neutral expression. This is equivalent to statically caricaturing an expression relative to the neutral expression for each frame in a sequence. It differs from Pollick et al's (2001a) technique which exaggerated each frame relative to the average *for that frame*. The current technique will always increase the amount of movement, whereas exaggerating relative to an average will actually decrease movement if the original movement is less than the average. The temporal manipulation we used altered the duration of the period between the onset and the maximum spatial extent of our facial expressions by resampling the original movement trajectories, by means of cubic interpolation. Again, this differed from previous temporal exaggerations (Hill and Pollick 2000) where movements were segmented and the durations of the segments exaggerated independently. The technique differed from that of Kamachi et al (2001) in that all intermediate frames were based on interpolations of nearby frames from an actual movement sequence, and not just upon the neutral and eventual expression. As in the work of Kamachi et al (2001), the current techniques have the advantage of simplicity in that they do not involve segmentation or the calculation of averages. Our goal is to use these manipulations to understand the relative contributions of spatial and temporal information to the recognition of facial expression.

2 Methods of movement collection, construction and presentation

2.1 Collection of movement data

Facial motion was captured with an Optotrak (Northern Digital) motion capture system by attaching infrared light-emitting markers to the surface of the face and recording their 3-D position at 30 Hz. For the purpose of the visual display, thirty markers were used; these included seven markers placed pseudo-randomly on each of the forehead and the left and right cheeks (a total of twenty-one markers), two markers on each eyebrow, four on the mouth, and one marker on the tip of the nose. The marker locations on the mouth and eyebrows were chosen at random from the complete set to minimise the featural cues of these structures. An additional two reference markers were placed at the top of the forehead and on the bridge of the nose. Actors were asked to produce the four expressions of anger, happiness, sadness, and surprise, after having been shown examples of the target expressions from the Ekman and Friesen (1976) pictures of facial affect series. Each facial movement began from a neutral expression, and ten repetitions were recorded for each expression. Data from four actors were used in the experiments. Examples of the displays can be seen in figure 1.

Figure 1. Examples of displays.

Rigid head movements were removed from the data by fixing the position and orientation of the triangle defined by the two reference markers and the marker at the tip of the nose. Only the onset of each expression was used, that is from the start of movement to the time when the expression reached its most extreme point, defined as the point where displacement of the central marker on the bottom lip reached its maximum.

Examination of the time duration required to pose the expressions indicated strong influences of both actors and expressions. Averaging across expressions revealed mean durations among the different actors of 800, 883, 1008, and 1266 ms (SD 221, 190, 310, 205 ms, respectively). Averaging across actors revealed mean durations among the expressions: surprise, 858 ms; anger, 933 ms; sadness, 1067 ms; happiness, 1100 ms (SD 302, 387, 194, 242 ms, respectively).

Finally, the placement of Optotrak markers and wires on the face does provide some constraint to performing facial movement. Thus, we performed pilot studies to examine whether categorisation of emotional displays of our four actors was comparable to previous results published by Bassili (1978, 1979) and Dittrich (1991), who used less invasive methods to obtain point-light video displays. Comparable results were obtained and, as can be seen in the present results, were sustained across the two experiments.

2.2 Construction of spatially exaggerated displays

The spatial structures of the displays were altered by using an exaggeration technique that constructed exaggerated expressions relative to initial, neutral positions. For each frame, the spatial difference of corresponding points from the neutral position was scaled by an exaggeration factor. The levels of the scaling factor used were $-\frac{2}{3}$, $-\frac{1}{3}$, 0, $\frac{1}{3}$, $\frac{2}{3}$, and 1. That is:

$$\text{exaggerated frame} = \text{original frame} + \text{scale factor} \times (\text{original frame} - \text{neutral}).$$

These six different levels of exaggeration used are termed $s - 2$, $s - 1$, $s0$, $s1$, $s2$, and $s3$, respectively.

2.3 Construction of temporally manipulated displays

Temporal alteration of the facial motion was obtained by having the expressions played within either a shorter or longer duration. This was achieved by taking the original measurements of the face and resampling each coordinate of each marker independently by one-dimensional cubic spline approximation to the originally sampled data so that the total number of frames contained in the display either decreased or increased. Thus, when the resampled movements were played back at the constant frame rate of 30 Hz, the speed of the movement would have a shorter duration (be faster) if the number of frames decreased and have a longer duration (be slower) if the number increased. The presentation durations used in the experiment included the original duration plus two shorter durations of $\frac{1}{2}$ and $\frac{3}{4}$ of the original duration and three longer durations of $\frac{3}{4}$, $\frac{3}{2}$, and $\frac{7}{4}$ of the original. These six levels of duration were termed $t - 2$, $t - 1$, $t0$, $t1$, $t2$, and $t3$ from shortest to longest duration, respectively.

2.4 Presentation of movement displays

Expressions were presented as moving point-light displays by means of a perspective projection of the full-face view that simulated a viewing position of 1.3 m from the face. These point-light animations were displayed on a Silicon Graphics, Octane computer with MXI graphics and viewed in a dimly lit experimental booth. The frame rate achieved by the computer system was approximately 30 Hz. Subjects were seated approximately 1.3 m away from the computer monitor and viewed the display under binocular conditions. The approximate size of a facial display in the neutral position was a height of 7.0 deg and a width of 4.4 deg of visual angle. Before and after the presentation of the facial expression a random mask of points was presented for 667 ms.

3 Experiment 1a

In experiment 1a we examined whether exaggeration of spatial properties of a recorded facial expression facilitated the perception of emotion from point-light displays. Exaggerations were obtained by amplifying the differences between the facial configuration at all points during an expression and the neutral configuration at the beginning of the expression.

3.1 Participants

A total of thirteen participants were run; all were naïve to the purpose of the experiment and were paid for their participation.

3.2 Design

The experiment was a 4 (expression) $\times$ 6 (level of spatial exaggeration) within-subjects factorial design. The expressions used were anger, happiness, sadness, and surprise. The six levels of exaggeration were those described above in the section on construction of spatially exaggerated displays. The dependent variable was the rated intensity of emotional expression on a scale of 1 to 100, with 1 indicating low emotional intensity and 100 indicating high emotional intensity.

3.3 Procedure

There were 4 (expressions) $\times$ 6 (levels of exaggeration) $\times$ 4 (actor) = 96 trials in each experiment. Participants saw each stimulus once with order randomised and were instructed to choose which of the four expressions (happiness, sadness, anger, surprise) they thought was depicted in the display and to rate it on a scale from 1 to 100, with 1 indicating very unemotional and 100 indicating highly emotional. Participants were instructed to guess the expression if they were unsure and to always choose only one expression.

The judgments of participants were obtained via a dialog box that appeared on the computer screen at the end of each trial and had a mouse-operated slider bar for each of the possible expressions. When participants were satisfied with their setting they clicked the mouse on the 'OK' button and the experiment proceeded to the next trial. The entire experiment took approximately 15 min.

3.4 Results

The overall results for experiment 1a are summarised in table 1 as a matrix of intensity ratings and response rates, collapsed across actors, for all emotional expressions, at each of the six levels of spatial manipulation—one matrix for each level. From table 1 it can be seen that both intensity ratings and response rate were raised for veridical responses (those upon the diagonal of the matrix). However, there were some instances of confusions that, although uncommon, were rated as intense. For example, at the highest level of exaggeration, confusions between happiness and surprise were infrequent, but produced high-intensity ratings when obtained. Caution is advised in the interpretation of rating data based on few responses.

For analyses of ratings, data were collapsed across actors. In addition, only when an expression was judged correctly was the rating response considered. Any missing cells resulting from a participant never correctly identifying an expression at a particular level of spatial manipulation (19 of a total of 312) were replaced by a hot-deck imputation technique that used a Euclidean distance similarity measure (Roth 1994). A 4 (expression) $\times$ 6 (level of exaggeration) within-subjects analysis of variance on ratings of emotional intensity showed a significant main effect of expression ($F_{3,36} = 19.1$, $p < 0.0001$) and a significant main effect of level of exaggeration ($F_{5,60} = 37.8$, $p < 0.0001$). No significant interaction was obtained. Figure 2a shows the rating data plotted for each expression and level of exaggeration. For the variable of expression, pairwise comparison of means (Tukey) revealed significant differences ($p < 0.05$) between anger versus

Table 1. Response matrices of response rates and intensity ratings obtained for spatial manipulation of facial movements in experiment 1a.

Level	Presented emotion	Response rate/Intensity rating			
		anger	happiness	sadness	surprise
−2	anger	0.35/27	0.25/22	0.21/21	0.19/22
	happiness	0.06/28	0.77/20	0.06/41	0.12/39
	sadness	0.10/27	0.25/24	0.50/21	0.15/33
	surprise	0.10/37	0.21/39	0.25/28	0.44/27
−1	anger	0.44/40	0.23/30	0.19/19	0.13/40
	happiness	0.12/41	0.83/32	0.06/31	0.00/−
	sadness	0.21/26	0.25/20	0.48/22	0.06/24
	surprise	0.13/23	0.12/48	0.21/45	0.54/36
0	anger	0.60/41	0.17/25	0.06/53	0.17/39
	happiness	0.10/34	0.83/36	0.04/39	0.04/59
	sadness	0.33/31	0.15/45	0.48/30	0.04/63
	surprise	0.10/27	0.12/41	0.12/44	0.67/48
1	anger	0.40/54	0.19/47	0.15/29	0.25/35
	happiness	0.15/36	0.77/46	0.04/13	0.04/28
	sadness	0.21/30	0.19/32	0.54/37	0.06/35
	surprise	0.13/49	0.12/71	0.12/53	0.63/58
2	anger	0.60/58	0.15/42	0.13/41	0.12/32
	happiness	0.04/25	0.85/49	0.06/34	0.06/72
	sadness	0.25/45	0.15/32	0.48/43	0.12/26
	surprise	0.13/58	0.12/59	0.13/61	0.62/59
3	anger	0.67/65	0.17/34	0.08/63	0.08/34
	happiness	0.10/40	0.81/50	0.04/10	0.06/76
	sadness	0.23/41	0.08/24	0.54/40	0.15/30
	surprise	0.08/64	0.12/86	0.15/45	0.65/69

Figure 2. Ratings (a) and accuracy (b) obtained in experiment 1a.

sadness and happiness, happiness versus surprise, and sadness versus surprise. For the variable of level of exaggeration, pairwise comparison of means (Tukey) revealed significant differences ($p < 0.05$) between $s - 2$ and all other levels; $s - 1$ versus $s1$, $s2$, and $s3$; as well as $s0$ versus $s1$, $s2$, and $s3$.

In addition to the analysis of the rating data we also examined the data on accuracy, defined as correctly identifying an expression and visible in the main diagonals of table 1. Overall, the proportion correct on identifying an expression was 60% (64% for condition $s0$ only, chance 25%), and is consistent with previously published results

of 33%, 60%, and 45% reported by Bassili (1978), Bassili (1979), and Dittrich (1991), respectively. A 4 (expression) $\times$ 6 (level of exaggeration) within-subject ANOVA gave a main effect of expression ($F_{3,36} = 19.5$, $p < 0.0001$) and a main effect of level of exaggeration ($F_{5,60} = 3.8$, $p < 0.01$). No significant interaction was obtained. Figure 2b shows the accuracy data plotted for each expression and level of exaggeration. For the variable of expression, pairwise comparison of means (Tukey) revealed significant differences ($p < 0.05$) for anger versus happiness as well as happiness versus surprise and sadness. For the variable of level of exaggeration, pairwise comparison of means (Tukey) revealed significant differences ($p < 0.05$) between $s - 2$ and $s0$ as well as between $s - 2$ and $s3$.

3.5 Discussion

The results of experiment 1a showed that level of exaggeration modulated the recognition of facial expression. Ratings of emotional intensity were diminished for negative exaggerations and were enhanced for positive exaggerations. The increase of intensity ratings from the 0 level of exaggeration indicated by the a posteriori analysis supports the claim that the spatial exaggeration technique was effective for enhancing the recognition of expression.

4 Experiment 1b

Although experiment 1a showed that for each of the facial expressions subjects' ratings of emotional intensity increased with level of exaggeration, it is possible that this was due to subjects equating intensity with exaggeration in general. In other words, subjects did not actually perceive the extreme exaggerations of angry movements as more angry, but instead adopted a strategy where they rated the spatially exaggerated images as more intense, regardless of the perceived emotion. Another issue is that there appeared to be an increase in rated intensity with exaggeration for expressions which were confused with each other, such as happiness and surprise, but owing to the infrequent occurrence of such confusions the effect on ratings for confused expressions was unclear. To examine the role of confusions more closely and to test this alternative interpretation of participants broadly confounding intensity of exaggeration with intensity of expression, we conducted an additional control experiment. In this experiment, examples of all four facial expressions were rated individually on each of the four emotional rating scales. If the effect of exaggeration is non-specific, we would expect to see increase in ratings on all scales and not just the scale appropriate for the particular expression shown on a trial. However, if the effect of exaggeration is specific to the emotion, there should be an interaction between exaggeration and expression, with an effect of exaggeration only on the scale appropriate to that expression.

4.1 Materials

In order to fully test for any effects of exaggeration while keeping the number of trials at a minimum to offset the increased complexity of the response, a subset of actors and levels of exaggeration from experiment 1a were used. This subset contained the two actors whose emotions had been most accurately recognised overall in experiment 1a and were displayed at unexaggerated and maximum positive and negative levels of exaggeration (ie -2 , 0 , $+3$).

4.2 Participants

A total of sixteen participants were run; all were naïve to the purpose of the experiment and were paid for their participation.

4.3 Design

The experiment was a 4 (expression) × 3 (level of exaggeration) × 4 (rating scale) within-subjects factorial design. The expressions used were anger, happiness, sadness, and surprise. The three levels of exaggeration were the maximal negative (s – 2), original (s0), and maximal positive (s3) levels. The four ratings scales were anger, happiness, sadness, and surprise. The dependent variable used was rated intensity from 1 to 100.

4.4 Procedure

There were 4 (expression) × 3 (level of exaggeration) × 4 (rating scale) × 2 (actor) = 96 trials in each experiment. Participants saw each stimulus once in a random order for each of the four scales, with the order of the scales randomised for each participant. For a given scale, they were instructed to rate the example shown for emotional intensity on a scale from 1 to 100; for example, for the happy scale 1 indicated not happy at all and 100 indicated very happy.

The judgments of participants were obtained via a dialog box that appeared on the computer screen at the end of each trial and had a mouse-operated slider bar for one expression. When participants were satisfied with their setting, they clicked the mouse on the ‘OK’ button and the experiment proceeded to the next trial. Each block of 96 stimuli took approximately 6 min to complete and the entire experiment was finished in around 20 min.

4.5 Results

The overall results for experiment 1b are summarised in table 2 as a matrix of intensity ratings containing results for each of the four expressions and three levels of spatial manipulation. From table 2, it can be seen that, while intensity ratings increased with level of exaggeration for veridical on-diagonal entries, there were also increases for some nonveridical responses. For example, at the highest level of exaggeration there were high ratings of anger for surprise expressions as well as high ratings of surprise for angry expressions. Overall, it appeared that the off-diagonal intensity ratings with large magnitudes were roughly symmetric about the diagonal, suggesting that certain pairs of expressions were confusable.

To further analyse the results, we recorded the data into ten categories that included the four categories of veridical responses (anger responses to anger displays, happiness responses to happiness displays, sadness responses to sadness displays, and surprise responses to surprise displays), and six categories of nonveridical responses

Table 2. Response matrices of intensity ratings obtained for spatial manipulation of facial movements in experiment 1b.

Level	Presented emotion	Intensity rating			
		anger	happiness	sadness	surprise
–2	anger	18	11	7	14
	happiness	13	27	10	18
	sadness	20	14	17	17
	surprise	17	15	11	24
0	anger	41	11	18	30
	happiness	22	47	10	32
	sadness	23	12	28	15
	surprise	30	24	13	51
3	anger	52	19	23	51
	happiness	29	63	17	32
	sadness	28	12	37	28
	surprise	52	27	20	77

corresponding to the possible confusions between pairs of stimuli. These six categories of nonveridical responses contained the pairs of confusions and were obtained by averaging across the diagonal of the response matrix. For example, one such category includes the average intensity ratings of the happiness responses to sadness displays and sadness responses to happiness displays. A final note is that we did consider recoding into twelve nonveridical categories, rather than six, by not averaging across the diagonal. However, a paired t -test revealed no significant difference ($t_{17} = 0.6$, $p = 0.55$) across the diagonal and it was thus considered more parsimonious to use the six pairs of confusions rather than the twelve individual cases.

A 10 (stimulus category) $\times 3$ (level of exaggeration) within-subjects analysis of variance was performed that revealed a main effect of category ($F_{9,135} = 22.2$, $p < 0.0001$), a main effect of level ($F_{2,30} = 91.6$, $p < 0.0001$), and a significant interaction ($F_{18,270} = 9.1$, $p < 0.0001$). The results plotted for each stimulus category as a function of level of exaggeration are shown in figure 3. Further examination of the interaction through analysis of simple main effects revealed significant simple main effects for category at the levels of s0 ($F_{9,135} = 8.5$, $p < 0.0001$) and s3 ($F_{9,135} = 18.7$, $p < 0.0001$), as well as effects of level at the veridical responses for anger ($F_{2,30} = 21.1$, $p < 0.0001$), happiness ($F_{2,30} = 23.2$, $p < 0.0001$), sadness ($F_{2,30} = 7.2$, $p < 0.01$), and surprise ($F_{2,30} = 48.9$, $p < 0.0001$), expressions, plus anger–surprise ($F_{2,30} = 23.0$, $p < 0.0001$), and happiness–surprise ($F_{2,30} = 3.5$, $p < 0.05$) confusions. A closer examination of these simple main effects was obtained by calculation of pairwise comparisons (Tukey). These revealed that for happiness, surprise, and anger–surprise categories all pairwise comparisons among levels were significant. For anger displays there were significant differences for s – 2 versus s0, and s – 2 versus s3. For sadness displays there was a significant difference for s – 2 versus s3.

Figure 3. Ratings obtained in experiment 1b.

4.6 Discussion

The results of the control experiment 1b showed that ratings of emotional intensity showed a significant increase with level of exaggeration for all four cases when the facial expression and ratings scale were veridical *and* for the two of the six cases of confusion. A posteriori analysis revealed the clearest effects of exaggeration for expressions of happiness, surprise, and the confusions between anger and surprise. It thus appears unlikely that the results of experiment 1a can be explained broadly by a general tendency to rate exaggerated displays as more intense. Rather it seems that the effect of exaggeration is specific to the particular expression being exaggerated as well as other expressions that are confused with that expression.

5 Experiment 2

The results of experiment 1 showed that exaggeration of individual points relative to a neutral expression could enhance recognition of facial emotion. While this manipulation is purely spatial in nature, it clearly has an effect on spatiotemporal properties such as the velocity of individual points. In experiment 2 we wished to see whether manipulation of temporal properties alone, through uniformly speeding up or slowing down the same spatial configuration, would also affect the perception of emotion.

Kamachi et al (2001) have previously looked at the influence of speed of facial expressions on the identification of the emotion displayed. The stimuli used were morphed (blended) colour images of faces, interpolated between neutral and another facial expression (eg anger), in rapid succession. The results showed that increasing the speed of presentation optimised the recognition of happiness expressions, whereas decreasing the speed of presentation optimised the recognition of sadness expressions. The images used by these authors were full images; hence, we were interested to determine whether similar effects would be found for the point-light displays used in experiment 1. We obtained differences in speed using a similar method to that used by Kamachi et al (2001) by keeping the frame rate constant and increasing or decreasing the number of frames by resampling to adjust the total duration of the expression. The experiments differed in that our resampling was based on all the frames originally recorded, not just neutral and extreme positions.

5.1 Participants

A total of twenty participants were run; all were naïve to the purpose of the experiment and were paid for their participation.

5.2 Design

The experiment was a 4 (expression) $\times$ 6 (level of duration) within-subjects factorial design. The expressions used were anger, happiness, sadness, and surprise. The six levels of duration ($t - 2$, $t - 1$, t_0 , t_1 , t_2 , t_3) were those described above in the construction of temporally manipulated displays. Recall that the t_0 display is the movement played at the recorded duration, $t - 2$ corresponds to the same spatial pattern played with the shortest duration, and t_3 corresponds to the same spatial pattern played with longest duration. The dependent variable used was rated intensity of emotional expression on a scale 1 to 100.

5.3 Procedure

There were 4 (expression) $\times$ 6 (level of temporal manipulation) $\times$ 4 (actor) = 96 trials in each experiment. Participants saw each stimulus once with order randomised. They were instructed to choose which of the four expressions they thought each trial depicted, and to rate the example shown for emotional intensity on a scale from 1 to 100, with 1 indicating very unemotional and 100 indicating highly emotional. Participants were instructed to guess the expression if they were unsure and to always choose only one expression.

The judgments of participants were obtained via a dialog box that appeared on the computer screen at the end of each trial. They had a mouse-operated slider bar for each of the possible expressions. When participants were satisfied with their setting they clicked the mouse on the 'OK' button and the experiment proceeded to the next trial. The entire experiment took approximately 15 min.

5.4 Results

The overall results for experiment 2 are summarised in table 3 as a matrix of response rates and intensity ratings, collapsed across actors, for each combination of the four expressions and the six levels of temporal manipulation—one matrix for each level. From table 3 it can be seen that although the expressions did seem discriminable from

Table 3. Response matrices of response rates and intensity ratings obtained for spatial manipulation of facial movements in experiment 2.

Level	Presented emotion	Response rate/Intensity rating			
		anger	happiness	sadness	surprise
-2	anger	0.48/46	0.20/43	0.14/48	0.19/39
	happiness	0.14/39	0.74/40	0.06/50	0.06/49
	sadness	0.33/39	0.15/54	0.35/32	0.15/33
	surprise	0.14/38	0.08/48	0.11/46	0.69/53
-1	anger	0.48/50	0.30/40	0.14/39	0.10/31
	happiness	0.06/43	0.89/44	0.01/20	0.04/38
	sadness	0.25/36	0.19/40	0.43/36	0.13/27
	surprise	0.14/36	0.08/60	0.11/49	0.69/55
0	anger	0.58/55	0.23/50	0.16/37	0.04/33
	happiness	0.14/52	0.78/48	0.03/44	0.06/39
	sadness	0.25/31	0.20/42	0.49/40	0.06/26
	surprise	0.10/52	0.13/47	0.14/47	0.64/58
1	anger	0.60/50	0.18/31	0.09/40	0.14/35
	happiness	0.10/36	0.76/46	0.06/48	0.08/48
	sadness	0.26/33	0.16/37	0.44/44	0.11/18
	surprise	0.09/52	0.11/45	0.15/44	0.65/58
2	anger	0.53/54	0.14/35	0.16/50	0.18/38
	happiness	0.19/38	0.75/46	0.04/31	0.03/28
	sadness	0.24/23	0.18/45	0.49/46	0.10/32
	surprise	0.13/53	0.14/46	0.19/59	0.54/60
3	anger	0.55/57	0.15/42	0.19/52	0.11/58
	happiness	0.11/36	0.78/45	0.04/37	0.08/58
	sadness	0.24/30	0.15/36	0.53/41	0.09/44
	surprise	0.05/34	0.11/57	0.18/52	0.66/60

one another as evidenced by higher on-diagonal than off-diagonal entries of response rate, there was a tendency for rating data to not appear to be substantially modulated by whether or not the response was correct.

For analyses of ratings, data were collapsed across actors. In addition, only when an expression was judged correctly was the rating response considered. Any missing cells resulting from a participant never correctly identifying an expression at a particular level of temporal manipulation (19 of a total of 480) were replaced by a hot-deck imputation technique that used a Euclidean distance similarity measure (Roth 1994). A 4 (expression) × 6 (level of duration) within-subjects analysis of variance of rated emotional intensity gave main effects of level of duration ($F_{5,95} = 3.1, p < 0.05$) and expression ($F_{3,57} = 22.3, p < 0.0001$) but no interaction ($p > 0.1$). The results are displayed in figure 4a. A posteriori analysis (Tukey) revealed that only two levels of duration differed significantly from one another (t – 2 versus t2). Surprise expressions were rated as most intense (mean rating 57, SE 2), followed by anger (52, 2), happiness (45, 2), and lastly sadness (39, 2). A posteriori analysis (Tukey) gave a significant difference for anger versus sadness and happiness, sadness versus surprise, as well as happiness versus surprise.

In addition to analysing the rating data, we also examined the data on accuracy, defined as correctly identifying an expression and visible in the diagonals of table 3. Overall, the proportion correct on identifying an expression was 60% (62% for condition t0 only, chance 25%). A 4 (expression) × 6 (level of duration) within-subjects ANOVA provided a significant main effect of expression ($F_{3,57} = 17.3, p < 0.0001$) and an interaction between expressions and level of duration ($F_{15,285} = 1.76, p < 0.05$).

Figure 4. Ratings (a) and accuracy (b) obtained in experiment 2.

These results are displayed in figure 4b. Further analysis of the interaction revealed no simple main effects of level of duration for any of the expressions but did show simple main effects of expression at the level of duration of $t-2$ ($F_{3,57} = 4.7$, $p < 0.01$) and $t-1$ ($F_{3,57} = 6.4$, $p < 0.001$).

5.5 Discussion

The results of experiment 2 showed only limited effects of movement speed on recognised expression. Ratings of emotional intensity did show effects of level of duration and expression, but from the results of Kamachi et al (2001) we might have expected significant interactions of level of duration and expression with some movements increasing in emotional intensity with longer duration and some decreasing. Such effects would be consistent with exaggeration from the average in terms of duration. However, no such differential effects were found in the ratings of emotional intensity. For the results on accuracy there was a significant interaction of level of duration with expression, but there was no clear pattern except for the trend that sadness expressions were more accurately recognised at longer durations, consistent with Kamachi et al (2001). One possible reason why no strong effect of level of duration was found in experiment 2 is that, for the stimuli used, speed was not diagnostic for emotion. Evidence in support of this comes from examination of the variability of duration of the veridical expressions. This revealed that the range of average duration of the four expressions was 242 ms (surprise, 858 ms; anger, 933 ms; sadness, 1067 ms; happiness, 1100 ms) and the range of average duration of the four actors was 466 ms (800, 883, 1008, and 1266 ms). This suggests that individual differences among the actors would have compromised the diagnosticity of duration for specifying expression. This is in contrast to the stimuli of Kamachi et al (2001) which had fixed durations of 200, 867, and 3357 ms for all six identities viewed.

6 General discussion

The present results indicate that, for facial movements represented as point-light displays, spatial exaggeration has a substantial effect on the perception of emotion while the manipulation of movement duration has little effect. These results can help to inform the question of what role motion plays in the perception of emotion from moving faces. Although spatial exaggeration was successful, interpretation of these results is not perfectly clear since the effectiveness of spatial exaggeration could have been due either to changes in the pattern of the velocity field or to static cues which became available. If the manipulation of movement duration had had a strong effect on perceived emotion, then this would have provided clearer evidence that temporal factors, independent of

spatial configuration, were important for perceived emotion. However, this result was not obtained. In the following we discuss these issues in greater detail.

Previous studies of the caricature of a single facial image provide a backdrop to viewing the results of our spatial exaggerations. Calder et al (2000) have suggested that static facial expressions are coded as vectors in a multi-dimensional space, and that exaggerating the expression along the trajectory of the vector enhances the displayed emotion. Within this context we can pose the question whether, for moving faces, an additional dimension represents some essential property of motion or whether time simply serves as an index for multiple static frames—perhaps with a particular frame, such as the apex, having a privileged status. Previous work by Dittrich (1991) and Bassili (1978, 1979) with randomly placed point-lights on the face supports the notion that motion itself is informative to decode facial emotion. Thus, these results support the view that properties intrinsic to motion would contribute to any additional dimension of representation. Moreover, the fact that uniform velocity scaling had some, although minimal, effect on judgments of facial emotion is consistent with a role for motion in the decoding of emotional expressions.

One issue with the results of the temporal manipulation is that the general advantage for longer durations did not conform to the pattern predicted by Kamachi et al (2001) who found that happiness is associated with fast movements and sadness with slow movements. These findings of Kamachi et al (2001) are also supported by recent work by Pollick et al (2001b) which showed that, for arm movement, velocity is an important cue for signalling the difference between happy and sad knocking and drinking movements. However, one possible difference for arm movements is that for movements such as knocking there are only small spatial differences among the potential paths of sad and happy arm movements and thus movement velocity predominates in signalling affect. Conversely, for faces the different paths for different expressions might dominate over the speed along which the path is traversed. This would suggest that a property largely invariant to movement speed is critical. However, a crucial temporal property might yet be revealed for more complex facial movements where relative timing among segments of the movement could possibly encode affect. For example, in the recognition of identity from arm movements it has been found that relative timing of movement segments is critical.

In conclusion, the results of the current experiments are consistent with the view that spatial differences in the velocity field predominate over uniform scaling of the velocity field in conveying facial emotion, although this does not preclude uniform scaling from being diagnostic under particular situations, or temporal information becoming more influential for more complex, multiphasic facial movements (Knappmeyer et al 2003). The current results, showing the effectiveness of spatial exaggeration of facial expression relative to a neutral expression, provide an additional method for the investigation of facial emotion.

Acknowledgments. We wish to thank Graeme Seawright and the reviewers for many helpful suggestions. This work was supported by grants from the Engineering and Physical Sciences Research Council and the Economic and Social Research Council.

References

- Bassili J N, 1978 "Facial motion in the perception of faces and of emotional expression" *Journal of Experimental Psychology: Human Perception and Performance* **4** 373–379
- 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
- Benson P J, Perrett D I, 1991 "Perception and recognition of photographic quality facial caricatures: implications for the recognition of natural images" *European Journal of Cognitive Psychology* **3** 105–135
- Calder A J, Rowland D, Young A W, Nimmo-Smith I, Keane J, Perrett D I, 2000 "Caricaturing facial expressions" *Cognition* **76** 105–146

-
- Calder A J, Young A W, Benson P J, Perrett D I, 1996 "Self priming from distinctive and caricatured faces" *British Journal of Psychology* **87** 141–162
- Calder A J, Young A W, Rowland D, Perrett D I, 1997 "Computer-enhanced emotion in facial expressions" *Proceedings of the Royal Society of London, Series B* **264** 919–925
- Dittrich W, 1991 "Das Erkennen von Emotionen aus Ausdrucksbewegungen des Gesichts" *Psychologische Beiträge* **33** 366–377
- Dittrich W H, Troscianko T, Lea S E G, Morgan D, 1996 "Perception of emotion from dynamic point-light displays represented in dance" *Perception* **25** 727–738
- Ekman P, Friesen W V, 1976 *Pictures of Facial Affect* (Palo Alto, CA: Consulting Psychologists Press)
- Hill H, Pollick F E, 2000 "Exaggerating temporal differences enhances recognition of individuals from point-light displays" *Psychological Science* **11** 223–228
- Johansson G, 1973 "Visual perception of biological motion and a model for its analysis" *Perception & Psychophysics* **14** 201–211
- Kamachi M, Bruce V, Mukaida S, Gyoba J, Yoshikawa S, Akamatsu S, 2001 "Dynamic properties influence the perception of facial expressions" *Perception* **30** 875–887
- Knappmeyer B, Thornton I M, Bühlhoff H H, 2003 "The use of facial motion and facial form during the processing of identity" *Vision Research* in press
- Pollick F E, Fidopiastis C M, Braden V, 2001a "Recognising the style of spatially exaggerated tennis serves" *Perception* **30** 323–338
- Pollick F E, Paterson H, Bruderlin A, Sanford A J, 2001b "Perceiving affect from arm movement" *Cognition* **82** B51–B61
- Rhodes G, Brennan S, Carey S, 1987 "Identification and ratings of caricatures: Implications for learning the mental representations of faces" *Cognitive Psychology* **19** 473–497
- Roth P L, 1994 "Missing data: A conceptual review for applied psychologists" *Personnel Psychology* **47** 537–560
- Walk R D, Homan C P, 1984 "Emotion and dance in dynamic light displays" *Bulletin of the Psychonomic Society* **22** 437–440