PERGAMON

Vision Research 42 (2002) 2345–2355

Vision Research

www.elsevier.com/locate/visres

Estimating the efficiency of recognizing gender and affect from biological motion

Frank E. Pollick [a,*], Vaia Lestou [a], Jungwon Ryu [b], Sung-Bae Cho [b]

[a] Department of Psychology, Glasgow University, 58 Hillhead Street, G12 8QB Glasgow, UK
[b] Department of Computer Science, Yonsei University, 134 Shinchon-dong, Sudaemoon-ku, Seoul 120-749, South Korea

Received 8 February 2002; received in revised form 16 May 2002

Abstract

It is often claimed that point-light displays provide sufficient information to easily recognize properties of the actor and action being performed. We examined this claim by obtaining estimates of human efficiency in the categorization of movement. We began by recording a database of three-dimensional human arm movements from 13 males and 13 females that contained multiple repetitions of knocking, waving and lifting movements done both in an angry and a neutral style. Point-light displays of each individual for all of the six different combinations were presented to participants who were asked to judge the gender of the model in Experiment 1 and the affect in Experiment 2. To obtain estimates of efficiency, results of human performance were compared to the output of automatic pattern classifiers based on artificial neural networks designed and trained to perform the same classification task on the same movements. Efficiency was expressed as the squared ratio of human sensitivity (d') to neural network sensitivity (d'). Average results for gender recognition showed a proportion correct of 0.51 and an efficiency of 0.27%. Results for affect recognition showed a proportion correct of 0.71 and an efficiency of 32.5%. These results are discussed in the context of how different cues inform the recognition of movement style.

1. Introduction

The ability to perceptually organize point-light displays (Johansson, 1973) into the percept of a specific human action has long served as a demonstration that humans are adept at recognizing the actions of their conspecifics. While much research has been directed at gaining an understanding of the processes involved in the perception of human movement, there has been little work directed towards quantifying the general limits of our abilities to recognize different styles of human movement. This is surprising given that our ability to recognize stylistic gender differences from point-light walkers often appears as a cornerstone to support claims of exquisite abilities in the recognition of biological motion. In the present research we compare human recognition of movement style to that of an automatic pattern classifier in order to estimate how well humans extract the available information from biological motion displays.

There have been numerous investigations of the ability to recognize different styles of human movement from human movement, including recognizing gender (Barclay, Cutting, & Kozlowski, 1978; Kozlowski & Cutting, 1977; Troje, 2001), running versus walking (Hoenkamp, 1978; Todd, 1983), identity (Beardsworth & Buckner, 1981; Cutting & Kozlowski, 1977; Hill & Pollick, 2000; Stevenage, Nixon, & Vince, 1999) and emotion (Dittrich, Troscianko, Lea, & Morgan, 1996; Pollick, Paterson, Bruderlin, & Sanford, 2001b; Walk & Homan, 1984) from point-light displays, as well as style of tennis serve from solid body animations (Pollick, Fidopiastis, & Braden, 2001a). These studies have essentially used three approaches to examine the basis of our ability to recognize different styles of movement. These have included: (1) parametric manipulation of biomechanical models to study how manipulating particular parameters influence perception, (2) using recordings of natural movements to correlate movement kinematics with movement recognition, and (3) creation of movement spaces and examination of how interpolating and

* Corresponding author. Tel.: +44-141-330-3945.
E-mail address: frank@psy.gla.ac.uk (F.E. Pollick).

0042-6989/02/$ - see front matter © 2002 Published by Elsevier Science Ltd.
PII: S0042-6989(02)00196-7

extrapolating between movements influence the recognizability of movement style. In the following we first review these three approaches and then discuss how they relate to our proposal for measuring the efficiency of movement recognition.

1.1. Biomechanical models

The study of how we recognize different styles of gait has extensively used parametric biomechanical models. By isolating the appropriate parameters of the model and by systematically varying these parameters to generate synthetic movement one can study how recognition of gait covaries with the manipulated parameters (Hoenkamp, 1978). Examples of such an approach include the influence of center of moment between the hips and shoulders in the perception of gender (Cutting, 1978), the influence of lower leg angle on the perception of running versus walking (Todd, 1983), and the influence of lateral body sway in the recognition of gender (Mather & Murdoch, 1994).

1.2. Correlating movement recognition to movement kinematics

Another way to explore the recognition of movement style is by relating kinematic properties of measured movements to the recognition of displays of these same movements. Early examples of this approach are provided by description of how kinematic properties relate to the perception of lifted weight (Bingham, 1987; Runeson & Frykholm, 1981) and are embodied in the principle of kinematic specification of dynamics (Runeson, 1994; Runeson & Frykholm, 1983). A more recent example of this approach is provided by investigation of the recognition of affect from arm movements (Pollick et al., 2001b). The first step of this research was to record three-dimensional (3D) arm movements of actors performing drinking and knocking actions with different affective states. These movements were subsequently shown as point-light displays to observers who categorized the displays. From these judgments multidimensional scaling (Kruskal & Wish, 1978) was used to construct a two-dimensional (2D) psychological space of the perceived movements and the dimensions of this psychological space was correlated to measurements of the movement kinematics. It was shown that one dimension of the psychological space was highly correlated to the measured speed of the movement, the other dimension appeared related to the phase relations among the limb segments. This correlation between movement speed and one dimension of the psychological space was consistent with cognitive models of affect that represent affect in a 2D space of activation and valence (Russell, 1980; Yik, Russell, & Barrett, 1999).

1.3. Exaggerating movement style

The recognition of human movement style is an example of recognition at the subordinate level where recognition is achieved among movements that share the same basic features and differ only in subtle aspects of the movement. One can consider a hypothetical space of movements, similar to the face space proposed in the related domain of subordinate recognition of faces which act as a representation of the movements (Valentine, 1991; for a review see Rhodes, 1996). It has been shown that in this space of movements if one travels from the point representing the grand average of all movements in the direction of the point representing a stylistic movement then interpolated points (movements) along this line are less readily identified as the particular movement style. However, if one continues to move in the same direction and extrapolates past the stylistic movement then one finds points (movements) which are more readily identified as that particular movement style. The effectiveness of these movement exaggerations to enhance recognition has been shown for both temporal (Hill & Pollick, 2000) and spatial (Pollick et al., 2001a) exaggerations of movement. In both cases the space of possible movements was defined upon a database of 3D measurements of several joint locations at a high sample rate. One possible interpretation of the success of both spatial and temporal exaggerations is that spatiotemporal interactions such as velocity (which were not controlled for in the construction of either spatial or temporal exaggerations) are crucial for the recognition of movement style.

1.4. Comparison among techniques

Each of the approaches described above has its own particular advantages and disadvantages. For example, the biomechanical model approach is well suited for cases where a well-developed model is available such as gait. However, for arbitrary movements where it might not be possible to find an appropriate model then this approach is not practical. The correlation technique is well suited for relating movement properties to perception. However, since movement kinematics are the consequences of both motor planning and execution, and many kinematic properties are correlated to one another, the correlation of perception to kinematics does not necessarily reveal the feature crucial for recognition. Finally, the exaggeration technique is well suited to understanding the representation of movement. However, since it constructs an arbitrarily defined movement space and treats each movement as a point in this space the relation between movement space and physical properties of the movement is not necessarily well defined. Thus there can be difficulty in providing a

simple physical explanation of how the changes in movement perception are coming about. The property common to all the approaches is that they quantify how some movement property (a parameter of a biomechanical model, a measured kinematic feature, or a displacement in a constructed stimulus space) influences the recognition of movement style.

The limitation of all the approaches is that none of them indicate whether the effectiveness of the movement property in modulating perception is due to large or small variation in the amount of information available with which to base the recognition judgement. In other words, when recognition of movement style becomes easy it is not known whether this is due to large amounts of information becoming available, or great sensitivity to the available information. In addition, for cases when recognition is not possible, there is no guarantee that the information is provided in the image but is not being used. In the present research we attempt to address this issue by using the concept of ideal observers (Barlow, 1978) to obtain estimates of the efficiency of style recognition.

1.5. Estimates of human efficiency in style recognition

As discussed above, previous research into recognition of style has indicated styles that can be recognized and properties of movement that can be manipulated to influence the recognition of movement style. However, these studies have not addressed the issue of how effective we are at utilizing the information available in a point-light display. This question can be addressed by comparing human performance to that of an ideal observer (Barlow, 1978; Liu, Knill, & Kersten, 1995), with efficiency expressed as the squared ratio of human sensitivity (d') to that of the ideal observer. The ideal observer is an optimal algorithm that uses all the relevant information available in the stimulus and thus efficiency provides a means of normalizing human performance to a standard absolute level of performance. Efficiencies of approximately 10% and above are generally thought to indicate high levels of human performance. At the heart of ideal observer analysis is the assumption that the algorithm to which humans are compared is optimal. In many cases it is possible to rigorously define a model in which optimal performance can be defined in a strict sense, however, for the case of recognizing human movement this is not possible. The reason for this is that human movements are tremendously complex and to generally define the space of natural human movements in precise analytic terms is not possible (cf. Bowden, 2000; Faraway, 2002; Sidenbladh, Black, & Sigal, 2002). Probably the best alternative at present, which we adopt in this current research, is to take a considerable sample of natural movements, devise an optimal algorithm for categorizing them, and use this as an approximation which approaches optimal performance. Since efficiency is measured as the ratio of human sensitivity to algorithm sensitivity, this approximate solution would provide a useful upper bound on efficiency.

In the present set of experiments we examined human recognition of gender and affect (angry or neutral) from displays of point-light arm movements of knocking, lifting and waving, and compared these human results to those of automatic pattern classifiers based on artificial neural networks. Arm movements were chosen since they are more compact stimuli than whole-body motion and have been shown to be able to carry visual information such as identity and affect (Hill & Pollick, 2000; Pollick et al., 2001b). While no evidence exists about visually perceived gender differences from arm movements there is evidence that, relative to body height, males have a longer forearm (Geary, 1998). Thus it would not appear unreasonable to conjecture that visually perceived differences might exist. Artificial neural networks were chosen from a variety of other automatic pattern classifiers such as self-organizing maps, decision trees and various statistical techniques since they are reliable and straightforward to implement. Neural networks have been used before in related applications of human action recognition (Howell & Buxton, 1997; Lakany & Hayes, 1997) and are generally regarded as effective means to obtain pattern classifiers for complex systems where it is difficult to obtain parametric representations of the underlying processes. We consider arm movements to fit this criterion since although a parametric model of arm biomechanics is certainly possible, accurate modelling of the way these parameters naturally vary across the population, across different affective states and across different neural control signals is at present beyond our abilities.

To achieve our examination of the efficiency of recognizing gender and affect from arm movements we began with the collection of a large 3D database of movements. For each of six combinations of two affects (neutral, angry) and three actions (knocking, lifting, waving) we obtained 10 repetitions of a movement for each of 26 actors (13 males, 13 females). This resulted in a database of 1560 $(13 \times 2 \times 6 \times 10)$ movements. Of these 1560 movements 1248 were used to train the artificial neural networks on the categorization task that human participants were asked to perform. A small minority of 312 movements were excluded from use in training the networks and instead were used both in the testing of human and artificial neural network performance. In Experiment 1 we examined human recognition of gender and compared it to artificial neural networks trained to recognize gender. Similarly in Experiment 2 we examined human recognition of affect and compared it to artificial neural networks trained to recognize affect.

2. Methods

2.1. Collection and preprocessing of movement data

The movement data was obtained using a 3D position measurement system (Optotrak, Northern Digital). Positions of the head, right shoulder, elbow, wrist, and the first and fourth metacarpal joints were recorded at a rate of 60 Hz while an actor performed a movement. These arm movements were recorded from 13 males and 13 females that were instructed to perform knocking, waving and lifting movements in both a neutral and an angry style. For each of the six combinations of action and affect 10 repetitions were recorded for every actor; resulting in a database of 1560 movements.

For all arm movement recordings actors used their right arm and started and ended the movement with their arm to their right side. For knocking they struck the surface directly in front of them approximately three times. For waving they waved their hand for several cycles as towards an individual facing them. For lifting they raised a plastic bottle from its location on a desk directly in front of them to a small elevated platform located at the back of the desk.

Each 3D movement data record was processed to obtain the start and end of the movement and the tangential velocity of the wrist. The start of the movement was defined as the instant the tangential velocity of the wrist rose above a criterion value, and the end by the instant after the velocity passed below the criterion. This start/end velocity criterion was defined as 5% of the peak tangential velocity of the wrist. In addition, short segments of missing data, resulting from a marker going out of view of the cameras, were interpolated to remove this artifact. Finally, to eliminate the structural cue to gender that the males were on average taller than the females, the size of all the stimuli were standardized. This standardization was to set the distance from the head to the first metacarpal joint, in the first frame of the movement, identical for all the actors. This standard height was the average height of all the actors.

2.2. Automatic recognition of human movement

In this section we provide only an overview of the techniques used in the automatic recognition of human movement, further details are provided in Appendix A. Our procedure was to first randomly divide the entire set of collected movements into a training set and a testing set. The training set was used to train a neural network to make classifications of the gender or affect of a movement and the testing set was used to test both the trained network and human participants. Estimates of efficiency came from comparison of the sensitivity of human observers to the sensitivity of the network.

There are two important considerations in the development of the neural networks; these are the particular design of the neural network architecture and the representation of the data input to the networks. Additional factors which will influence the final performance include the training procedure (gradient descent, stochastic annealing, etc.) and the number of training examples. Since it is impossible to know in advance either the optimal architecture or representation for a particular problem we examined several basic architectures and several different data representations. The multilayer perceptron (MLP) was ultimately chosen because of its simplicity and its consistently high performance in categorizing the movements. While it is possible that better performance could be found by a more sophisticated architecture, this result would still serve as a bound on efficiency and we felt it beyond the scope of the present work to emphasize network architecture. The representations we examined included the 3D position, 3D velocity, 2D position and 2D velocity of the recorded point lights. To select the best representation we trained a MLP network with each of the input representations and selected the one with the best average performance over all combinations of action and affect. For the case of gender recognition the best representation was 2D position and for affect the best representation was 3D position.

2.3. Presentation of movements

The movements were presented as moving point-light displays using orthographic projection. The knocking and lifting movements were presented at a side view while the waving movements were shown at a front view. The point-light animations were displayed on a Silicon Graphics, Octane computer with MXI graphics. The frame rate achieved by the computer system was approximately 30 Hz. Subjects were seated approximately 1 m away from the computer monitor and viewed the display under binocular conditions. The stimuli covered approximately 8° of the visual field. Examples of the displays are presented in Fig. 1.

3. Experiment 1

In Experiment 1 we examined the ability of participants to judge the gender of point-light displays performing movements at the six combinations of action and affect.

3.1. Participants

A total of 20 participants volunteered to participate in the study. All were naïve to the purpose of the study and were paid for their participation.

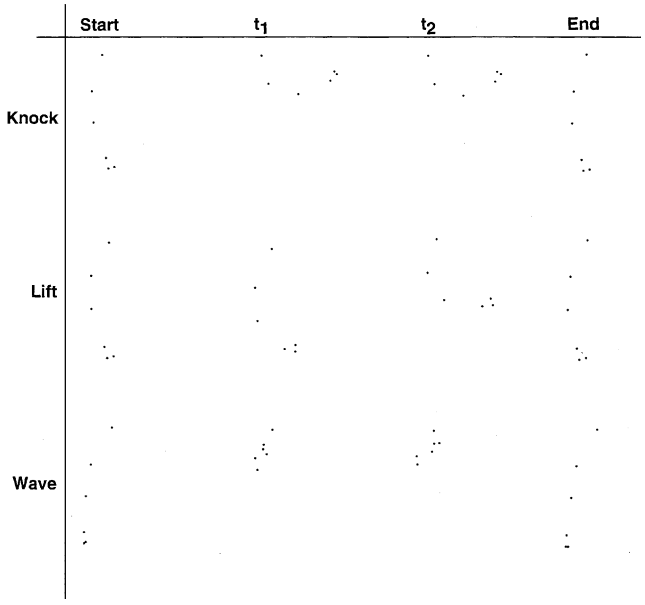

Fig. 1. Frames from the knocking, lifting and waving movement sequences are shown at the start, end and two intermediate points (t_1 and t_2) of the movement. For knocking a side view was presented, t_1 corresponds to the fist hitting the surface while frame t_2 corresponds to the fist pulling back before striking the surface again. For lifting a side view was also shown, t_1 corresponds to the hand first coming into contact with the object to be lifted and t_2 shows the hand position as the object is being placed at the back of the table. For waving a front view was shown, t_1 corresponds to the hand position at one maximum of lateral displacement and t_2 shows the other maximum.

3.2. Design

The two variables of action and affect were explored. There were three levels of the variable action (knocking, waving and lifting) and two levels of the variable affect (neutral and angry). Stimuli were blocked by the six possible combinations of action and affect. In each block participants viewed the two test trials for each of the 13 male and 13 female models, resulting in a total of 52 trials per block. The order of the six blocks was randomized for each participant.

3.3. Procedure

Participants were instructed that they would see an arm movement performing either a knocking, waving or lifting movement and that at the completion of the movement display they were to judge the gender of the actor performing the movement. Several practice trials were given at the beginning of the experiment to familiarize participants with the procedure. Each movement display was viewed on the computer monitor and at its completion a graphical user interface would appear in the corner of the screen where, by a mouse click, participants could enter their response of the gender of the actor. After responding the next trial would commence and there was no feedback provided on whether a response was correct. Each trial took a few seconds to complete and thus an entire block of 52 trials was typically performed in less than 5 min, and the duration of the entire experiment was less than 30 min.

3.4. Results

The average proportion correct for judgments of the gender of point-light actors was 0.51 (standard error of means, SEM—0.01). In addition to proportion correct we examined sensitivity (d') of the judgments. This was accomplished by defining hits to be accurately judging a male as male and false alarms to be misclassifying a female as male. The resulting average value of d' obtained was 0.05 (SEM 0.07). Further examination of the results was provided by examining the sensitivity by the six individual conditions. These results are shown in Fig. 2A where it can be seen that performance differed between the neutral and angry movements. The variation in d' between experimental conditions was examined through a within subjects analysis of variance (ANOVA) using the factors of action (knock, lift, wave) and affect (angry, neutral). Results of the ANOVA showed a significant effect of affect, $F(1, 19) = 12.3$, $p < 0.01$ with no other significant effects or interactions. Tukey's post hoc comparison revealed a significant

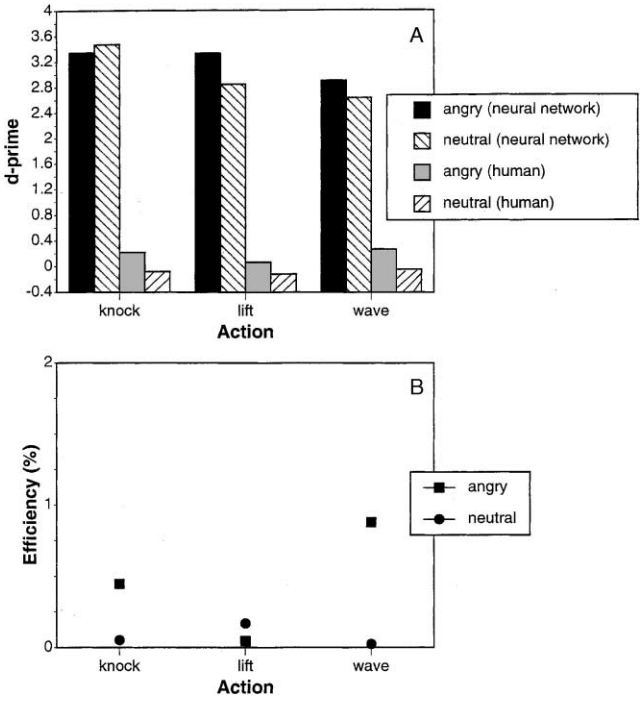

Fig. 2. Results of Experiment 1 on the recognition of gender. Panel A shows the recognition results expressed as d'. Panel B shows the comparison of human and neural network performance expressed as efficiency. Results are shown for the two different cases of neutral (●) and angry (■) movements.

difference ($p < 0.01$) between the angry ($d' = 0.19$) and neutral ($d' = -0.08$) levels of d'.

The neural network for gender recognition performed best using the input representation of 2D position and these results were used for comparison with human performance. The average value of d' obtained by the network for the six experimental conditions is plotted in Fig. 2A and are similar to the human results for the actions of lifting and waving in showing better performance for angry movements. An ANOVA of resulting d' values was conducted using the factors of action (knock, lift, wave) and affect (angry, neutral). Results of the ANOVA showed significant effects of action, $F(2,98) = 53.9$, $p < 0.01$, affect, $F(1,49) = 19.4$, $p < 0.01$ and their interaction $F(2,98) = 13.0$, $p < 0.01$. Results of Tukey's post hoc analysis revealed that at a 0.05 significance level that all levels of action were different for neutral affects but that for angry affect knocks were not different from lifts. In addition neutral movements were different from angry ones for the lifts and waves but not for knocks.

A final result to examine is the estimate of human efficiency expressed as the ratio of d' squared between human and neural network performance. This result is shown in Fig. 2B where it can be seen that efficiency was below 1% for all the experimental conditions. The average value of efficiency obtained was 0.27%.

3.5. Discussion of Experiment 1

The overall percentage correct and d' indicate that viewers could not reliably recognise the gender of the actor, however, the neural network was able to perform the discrimination with reasonably high accuracy. Thus, it would appear that although the information to discriminate gender is available in the displays it could not be used by participants to provide reliable recognition. The low level of performance achieved by the viewers is consistent with some of the comments of participants during debriefing that indicated that they found the task extremely difficult. However, the significant difference found between the neutral and angry movements, that was also visible in the performance of the neural network for lifting and waving, suggests that participants' responses were not purely random. We examined the relation between movement kinematics and participants' patterns of responses for evidence of any heuristic that could have explained human performance. To do this we took the motion of the wrist marker and derived kinematic landmarks such as the peak velocity, peak acceleration, peak deceleration as well as global properties such as average velocity and duration. Next we examined the correlation between these kinematic characteristics and judged maleness as defined by the proportion of times a particular actor was judged male. The hypothesis being that participants were consistently basing their judgments on some kinematic property of the stimulus which did not reliably indicate gender. These examinations were largely inconclusive, although there was evidence that the male actors had a greater chance of being judged male the faster they performed the movement ($R^2 = 0.59$ between proportion of male responses and average velocity, the equivalent R^2 value for females was 0.29).

4. Experiment 2

In Experiment 2 we examined the ability of participants to judge the affect of point-light displays performing movements at the six combinations of action and gender.

4.1. Participants

A total of 18 participants were run, all were naïve to the purpose of the experiment and were paid for their participation.

4.2. Design

The experiment was a 3(action) × 2(gender) within subjects factorial design. The three levels of action used were knocking, lifting and waving. The two levels of gender were male and female.

4.3. Procedure

The procedure was identical to that of Experiment 1 except that participants judged whether a movement exhibited a neutral or an angry affect.

4.4. Results

The average proportion correct for judging the affect of point-light actors was 0.71 (standard error of means, SEM—0.02). In addition to proportion correct we examined sensitivity (d') of the judgments. This was accomplished by defining hits to be accurately judging an angry movement as angry and false alarms to be misclassifying a neutral movement as angry. The resulting average value of d' obtained was 1.43 (SEM 0.22). Further examination of the results was provided by examining the sensitivity by the six individual conditions. These results are shown in Fig. 3A where it can be seen that performance appeared to differ among the different conditions. The variation in d' between experimental conditions was examined through a within subjects ANOVA using the factors of action (knock, lift, wave) and gender (male, female). Results of the ANOVA showed a significant effect of gender, $F(1,17) = 9.7$,

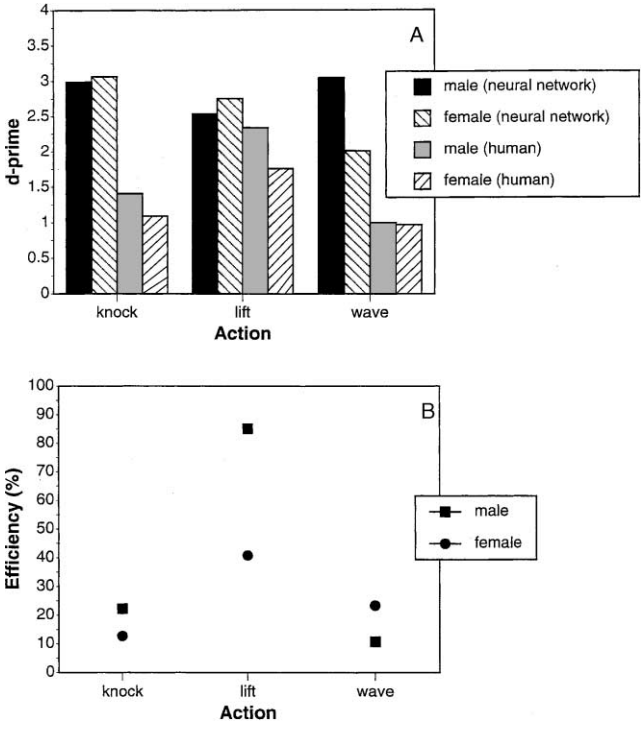

Fig. 3. Results of Experiment 2 on the recognition of neutral and angry affect. Panel A shows the recognition results expressed as d'. Panel B shows the comparison of human and neural network performance expressed as efficiency. Results are shown for the two different cases of male (■) and female (●) movements.

$p < 0.01$ and a significant effect of action $F(2, 34) = 15.0$, $p < 0.01$ but no significant interaction. Tukey's post hoc comparison on the effect of gender revealed a significant difference ($p < 0.01$) between the male ($d' = 1.58$) and female ($d' = 1.27$) levels of d'. Tukey's post hoc comparison on the effect of action revealed a significant difference ($p < 0.01$) between lifting ($d' = 2.05$) and knocking ($d' = 1.25$) as well as a significant difference ($p < 0.001$) between lifting and waving ($d' = 0.98$).

The neural network for affect recognition performed best using the input representation of 3D position and these results were used for comparison with human performance. The average value of d' obtained by the network for the six experimental conditions is plotted in Fig. 3A. An ANOVA examined the performance of the neural network using the factors of action (knock, lift, wave) and gender (male, female). Results of the ANOVA showed significant effects of action, $F(2, 78) = 19.6$, $p < 0.01$, gender, $F(1, 39) = 13.3$, $p < 0.01$ and their interaction $F(2, 78) = 38.9$, $p < 0.01$. Results of Tukey's post hoc analysis revealed that at a 0.05 significance level, all levels of action were different for female movements but that for male movements knocks were not different from waves. In addition male movements were different from female ones for waves but not for knocks or lifts.

A final result to examine is the estimate of human efficiency expressed as the ratio of d' squared between human and neural network performance. This result is shown in Fig. 3B where it can be seen that efficiency varied between around 10% and 90% for the various conditions, and had an average of 32.5%. These values are considerably greater than those obtained for gender recognition.

4.5. Discussion

Results showed that both the human observers and the neural networks obtained reliable performance at discriminating affect from the point-light displays. In addition it was found that the estimates of efficiency were generally high suggesting either that observers were using a large proportion of the available information or that the neural network was very poor at performing the task. If we assume that an optimal algorithm would achieve near perfect performance then we can examine the two possibilities. Near perfect performance could be considered to be a hit rate of 99% and a false alarm rate of 1%, resulting in a d' value of 4.65. The efficiency resulting from this assumption of near perfect performance would be 9.4%, thus it would appear that participants were indeed using a substantial proportion of the information available to them.

5. General discussion

Consistent with claims that humans are adept at perceiving biological motion, the results of Experiment 2 showed high levels of efficiency in the recognition of affect from arm movements. However, the results of Experiment 1 told a different story. Here low efficiency was found, with the neural network able to discriminate gender while human participants performed at chance. This result is unique in the domain of biological motion research in demonstrating that a failure to recognize biological motion was due to an inability to use the available information.

Comparing across the two experiments reveals that although the performance of the neural network was approximately the same across both experiments, human performance varied dramatically. Unfortunately we do not know what features the neural network was using to obtain this consistent performance, or indeed if the same features were used by the neural network across the experiments. However, we can conjecture what information human observers might have used for discriminating between neutral and angry affect. Evidence from Pollick et al. (2001b) indicates that two separate stimulus dimensions are used to encode affect—one incorporating the speed of the movement and the other likely incorporating the phase among the limb

segments. Neutral and angry movements fall along the velocity dimension and thus velocity would have been diagnostic for the recognition task. These results of Pollick et al. (2001b) also indicated that the velocity/activation dimension accounted for more of the variability of participants' responses than the dimension apparently related to phase, or form of the movement. Thus, it is possible that gender information was carried in the form of the movement and human observers were simply not adequately sensitive to it. That information about gender is carried in form information is consistent with previous claims about gender recognition from gait (Cutting, 1978) and taken together with the current results suggest that the encoding of gender in human movement varies across actions and limb segments.

Previous results using ideal observers to explore human recognition have indicated differences between the input representations such as 2D or 3D position representations (Liu et al., 1995; Tjan & Legge, 1998). However, the current study found no consistent pattern between input representation and recognition accuracy. One possible explanation for this is that for human movement position and velocity information is tightly correlated. For example, Pollick and Sapiro (1997) showed that the 1/3 power law relationship between speed and shape (radius of curvature) in planar movement production implies motion at constant affine velocity—a property that would facilitate recognition of the shape from arbitrary viewing directions.

One potential issue with the current ideal observer approach is that it deals with the problem of style recognition in a bottom up fashion. Numerous researchers have provided evidence to support the view that recognition of human movement relies on recruiting higher level cognitive processes in the interpretation of the motion signal (Dittrich, 1991; Shiffrar & Freyd, 1993; Thornton, Pinto, & Shiffrar, 1998). An additional issue is that one can model human movement recognition to be the result of converging form and motion information rather than from a single source of position or velocity information as done in the present research (Giese, 2001). We would not debate the computational advantages of these approaches. However, our primary concern was to compare human performance to a standard model, and since these more sophisticated models have no advantage in providing a more rigorous definition of optimality we did not think the increased detail of a more sophisticated model would have facilitated this comparison. What the current results do provide is a clear comparison of human performance against a single representation of the input information, thus allowing various recognition tasks to be compared to a standard.

The current results do allow us to compare human performance to a standard, however, there are important shortcomings to use this quantity for the calculation of efficiency. The use of artificial neural networks carries with it several heuristically determined components including the network architecture, input representation and training paradigm. Although efforts were made to maximize performance of the neural networks there is no guarantee that optimal performance was obtained. Indeed, given the complexity of the stimulus and the largely unknown nature of how it is represented in human perception and cognition it is difficult to propose an alternative that could guarantee the optimal performance required to define a true ideal observer. Without an ideal observer we must consider the efficiencies obtained as an upper bound on efficiency. In this sense obtaining low efficiency where the neural network was able to discriminate the input while humans were not is more significant than cases of high efficiency. This is because high efficiency can always be accounted for by the possibility of greatly suboptimal performance of the neural network, while low efficiency indicates that human participants failed to exploit a regularity in the movement data that allowed discrimination.

The present results can be compared to investigations of the recognition of identity from gait. Psychophysical studies have revealed that subsequent to training, individuals could recognize identity from gait even when displays were presented as impoverished point-light displays (Stevenage et al., 1999). These studies of human performance can be compared to computational approaches to the recognition of identity from gait (Huang, Harris, & Nixon, 1998; Little & Boyd, 1998). The work of Huang et al. (1998) showed that both spatial and flow templates could be used to identify identity from video sequences of gait. These templates were constructed by eigenspace and canonical space transformations upon the input image data as a means of data reduction in obtaining a compact representation of gait. The work of Little and Boyd (1998) found moments of the video input image data and identified phase relations among the motion of these moments that were suitable features to indicate identity. While no direct comparisons of human and machine performance for the recognition of identity from gait have been performed, the current results suggest that psychophysical and computational results are consistent in indicating the ability to recognize identity from gait. Moreover, the computational studies indicate a variety of features which are sufficient for recognizing identity from gait.

The current experiments were successful in showing that recognition of biological motion varies between tasks and that a failure to recognize movement style could not be accounted for by the lack of available information to perform the task. What this highlights is that subordinate class recognition of human movement likely relies on particular forms of information being available. Biological motion research has focused on

describing basic level recognition (i.e. whether a display is human motion or not) and has provided examples of subordinate level classification to bolster claims of high levels of competence in perceiving human movement. However, a clear view of subordinate level recognition of human movement style is not fully developed. The development of an understanding of subordinate level recognition of human movement style is essential to understanding the mechanisms behind the perception and recognition of human movement. For example, current findings in neuropsychology indicate that a particular brain region in the STS is activated when human movement is observed (Grezes et al., 2001; Grossman et al., 2000; Stevens, Fonlupt, Shiffrar, & Decety, 2000). Whether this brain region should be interpreted as simply a locus for biological motion detection, or a region involved in more sophisticated processes of recognizing subordinate styles of movements will be facilitated by a clear behavioural understanding of human abilities at movement style recognition.

Acknowledgements

We would like to thank the Royal Society for support of this research through a UK–Korea Joint Project Grant as well as the Wellcome Trust and Nuffield Foundation.

Appendix A. The neural network classifier

The MLP is commonly used in many fields of pattern recognition due to the power and stability of its learning algorithms (Lippman, 1987). Common among these is the backpropagation algorithm based on the delta learning rule (Haykin, 1999; Werbos, 1994, Chap. 11). The power of the backpropagation algorithm lies in two main points: it takes place locally to update the synaptic weights and biases of MLP, and it is easy to compute all the partial derivatives of the cost function with respect to free parameters of network (Beale, 1996). We used a MLP with one hidden layer, given by an input vector X, hidden neuron Z and output neuron Y and activated by following equations:

$$Z = f[\mathrm{Net}(w_1 X)], \quad Y = f[\mathrm{Net}(w_2 Z)]$$

where w_1 and w_2 are the weights between first and second, and second and third layers. $\mathrm{Net}(\cdot)$ is the weighted sum of nodes connected from the prior layer and $f[\cdot]$ is a sigmoid activation function. The errors on the output and hidden layers are expressed as:

$$\delta_2 = (d - Y) f'[\mathrm{Net}(w_2 Z)],$$
$$\delta_1 = \left(\sum \delta_2 w_2 \right) f'[\mathrm{Net}(w_1 X)]$$

where d is the desired output. The weights are updated by:

$$w_2(t+1) = w_2(t) + \alpha \delta_2 Z, \quad w_1(t+1) = w_1(t) + \alpha \delta_1 X$$

where α is a learning constant.

The final values of the weights were obtained using gradient descent until the network reached a recognition rate of 98% on the training data. The dimensions of the input vector X depended upon the representation being examined and ranged from 900 to 2700. There were 50 hidden nodes and two output nodes. Learning rates and momentum were chosen experimentally for each network to find the values that provided best performance.

The four input representations examined were 2D position, 2D velocity, 3D position and 3D velocity. Each of these representations incorporated all six of the points measured on the body. The 2D representation was based on the two coordinates of the movement that were derived from the image information shown to human participants and the 3D representation used all three recorded coordinate data. For velocity we used the magnitude of the tangential velocity in the image plane (2D) or in 3D. This resulted in an input size of 900 for 2D or 3D velocity (6×150), 1800 for 2D position ($6 \times 2 \times 150$) and 2700 for 3D position ($6 \times 3 \times 150$). All movements were normalized in time to have a constant time of 150 time steps from start to finish. The amplitude of each input vector was also normalized so that all vectors varied between the maximum and minimum range of values accepted by the neural network.

For both gender and affect recognition, the MLP was trained with 1248 training samples and then tested on 312 test samples. To obtain a stable estimate of the performance of the network training and testing was conducted 40 times for each condition with the network beginning at a randomly selected seed value. To find performance of the algorithm, average sensitivity d' was calculated for each individual condition. These d' results for gender recognition are provided in Fig. 4 and correspond to average values of 3.1, 2.7, 2.5 and 2.5 for 2D position, 3D position, 2D velocity and 3D velocity respectively. The d' results for affect recognition are provided in Fig. 5 and correspond to average values 2.3, 2.7, 2.4 and 2.3 for 2D position, 3D position, 2D velocity and 3D velocity respectively. Subsequent comparison to human performance for gender were obtained using the values of the MLP network with a 2D position representation; and for affect were obtained using the values of the MLP network with a 3D position representation.

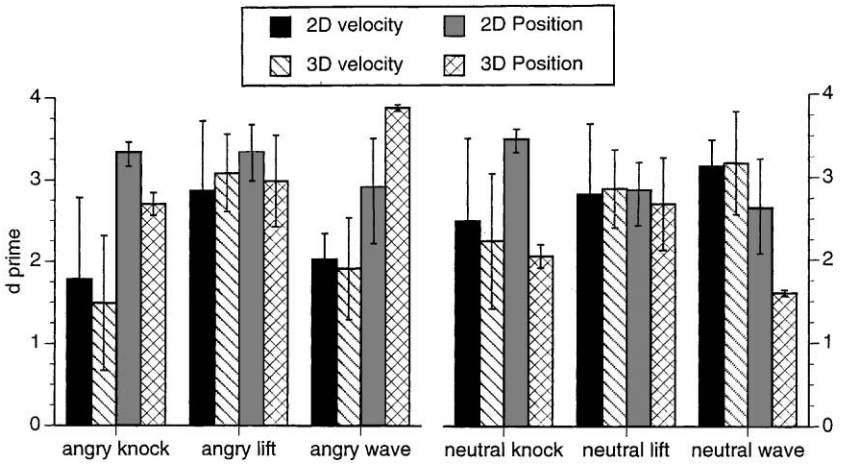

Fig. 4. Results of the neural network classification of gender are shown as d' for the four different input representations examined (2D velocity, 3D velocity, 2D position, 3D position). Bars indicate 95% confidence intervals. The results are plotted for the six different experimental conditions examined. The best overall results for gender recognition were produced by the 2D position representation and these values are used in Fig. 2.

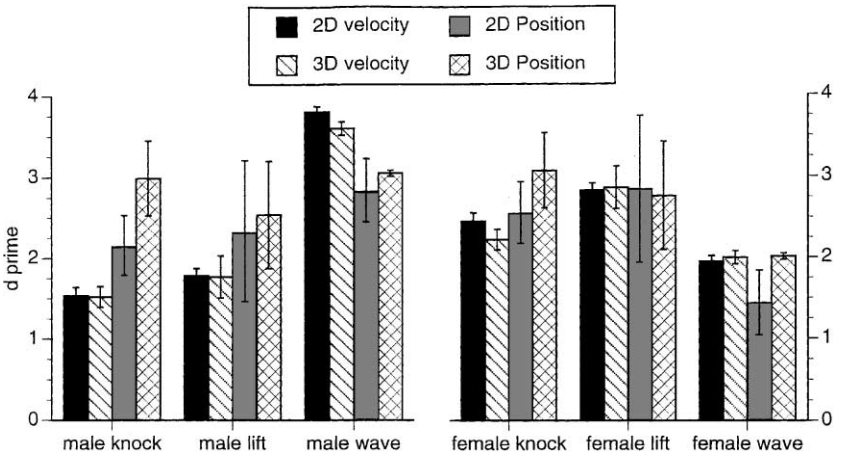

Fig. 5. Results of the neural network classification of affect shown as d' for the four different input representations examined (2D velocity, 3D velocity, 2D position, 3D position). Bars indicate 95% confidence intervals. The results are plotted for the six different experimental conditions examined. The best overall results for affect recognition were produced by the 3D position representation and these values are used in Fig. 3.

References

Barclay, C. D., Cutting, J. E., & Kozlowski, L. T. (1978). Temporal and spatial factors in gait perception that influence gender recognition. *Perception & Psychophysics, 23*, 145–152.

Barlow, H. B. (1978). The efficiency of detecting changes in random dot patterns. *Vision Research, 18*, 637–650.

Beale, H. D. (1996). *Neural network design*. PWS Publish Company.

Beardsworth, T., & Buckner, T. (1981). The ability to recognize oneself from a video recording of one's movements without seeing one's body. *Bulletin of the Psychonomic Society, 18*, 19–22.

Bingham, G. P. (1987). Kinematic form and scaling: Further investigations on the visual perception of lifted weight. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 155–177.

Bowden, R. (2000). Learning statistical models of human motion. In *IEEE Workshop on Human Modelling, Analysis and Synthesis, CVPR 2000*.

Cutting, E. J. (1978). Generation of synthetic male and female walkers through manipulation of a biomechanical invariant. *Perception, 7*, 393–405.

Cutting, J. E., & Kozlowski, L. T. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society, 9*, 353–356.

Dittrich, W. H. (1991). Action categories and the perception of biological motion. *Perception, 22*, 15–22.

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.

Faraway, J. (2002). *Human animation using nonparametric regression*. Technical Report #382, Department of Statistics, University of Michigan.

Geary, D. C. (1998). *Male, female the evolution of human sex differences*. Washington, DC: American Psychological Association.

Giese, M. A. (2001). Hierarchical neural model for the recognition of biological motion. *Journal of Vision, 1*(3), abstract 356.

Grezes, J., Fonlupt, P., Bertenthal, B., Delon-Martin, C., Segebarth, C., & Decety, J. (2001). Does perception of biological motion rely on specific brain regions? *Neuroimage, 13*(5), 775–785.

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., & Blake, R. (2000). Brain areas involved in

perception of biological motion. *Journal of Cognitive Neuroscience, 12*(5), 711–720.

Haykin, S. (1999). *Neural networks* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Hill, H., & Pollick, F. E. (2000). Exaggerating temporal differences enhances recognition of individuals from point light displays. *Psychological Science, 11*, 223–228.

Hoenkamp, E. (1978). Perceptual cues that determine the labeling of human gait. *Journal of Human Movement Studies, 4*, 59–69.

Howell, A. J. & Buxton, H. (1997). Recognising simple behaviours using time-delay RBF networks. *Neural Processing Letters*, *5*, 97–104.

Huang, P. S., Harris, C. J., & Nixon, M. S. (1998). Comparing different template features for recognizing people by their gait. In *British Machine Vision Conference (BMVC)*.

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

Kozlowski, L. T., & Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. *Perception & Psychophysics, 21*(6), 575–580.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling, sage university paper series on quantitative applications in the social sciences* (pp. 7–11). Beverly Hills and London: Sage Publications.

Lakany, H. & Hayes, G. (1997). An algorithm for recognising walkers. In J. Bigun, G. Chollet, & G. Borgefors (Eds.), *Audio- and video-based biometric person authentication* (pp. 112–118). IAPR, Springer.

Lippman, R. P. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine*, 4–22.

Little, J., & Boyd, J. (1998). Describing motion for recognition. *Videre, 1*, 1–32.

Liu, Z., Knill, D. C., & Kersten, D. (1995). Object classification for human and ideal observers. *Vision Research, 35*, 549–568.

Mather, G., & Murdoch, L. (1994). Gender discrimination in biological motion displays based on dynamic cues. *Proceedings of the Royal Society of London, 258*, 273–279.

Pollick, F. E., & Sapiro, G. (1997). Constant affine velocity predicts the 1/3 power law of drawing and planar motion perception. *Vision Research, 37*, 347–353.

Pollick, F. E., Fidopiastis, C. M., & Braden, V. (2001a). Recognizing 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. (1996). *Superportraits: Caricatures and recognition*. Hove, England: Psychology Press.

Runeson, S. (1994). Perception of biological motion: The KSD-principle. In G. Jansson, S. S. Bergstrom, & W. Epstein (Eds.), *Perceiving events and objects* (pp. 383–405).

Runeson, S., & Frykholm, G. (1981). Visual perception of lifted weight. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 733–740.

Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General, 112*, 585–615.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*, 1161–1178.

Shiffrar, M., & Freyd, J. J. (1993). Timing and apparent motion path choice with human body photographs. *Psychological Science, 4*, 379–384.

Sidenbladh, H., Black, M. J., & Sigal, L. (2002). Implicit probabilistic models of human motion for synthesis and tracking. In *European Conference on Computer Vision (ECCV)*.

Stevenage, S. V., Nixon, M. S., & Vince, K. (1999). Visual analysis of gait as a cue to identity. *Applied Cognitive Psychology, 13*, 51–526.

Stevens, J. A., Fonlupt, P., Shiffrar, M., & Decety, J. (2000). New aspects of motion perception: Selective neural encoding of apparent human movements. *Neuroreport, 11*, 109–115.

Thornton, I. M., Pinto, J., & Shiffrar, M. (1998). The visual perception of human locomotion. *Cognitive Neuropsychology, 15*, 535–552.

Tjan, B., & Legge, G. E. (1998). The viewpoint complexity of an object recognition task. *Vision Research, 38*, 2335–2350.

Todd, J. T. (1983). Perception of gait. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 31–42.

Troje, N. F. (2001). Decomposing biological motion: A linear model for analysis and synthesis of human gait patterns. *Journal of Vision, 1*(3), abstract 353.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race upon face recognition. *Quarterly Journal of Experimental Psychology, 43A*, 161–204.

Walk, R. D., & Homan, C. P. (1984). Emotion and dance in dynamic light displays. *Bulletin of the Psychonomic Society, 22*, 437–440.

Werbos, P. J. (1994). *Beyond regression: New tools for prediction and analysis in the behavioral sciences. The roots of backpropagation*. New York: John Wiley & Sons.

Yik, M. S. M., Russell, J. A., & Barrett, L. F. (1999). Structure of self-reported current affect: Integration and beyond. *Journal of Personality and Social Psychology, 77*(3), 600–619.