RECOGNITION OF PAIN EXPRESSIONS

P.J.B. HANCOCK[A]   C.D. FROWD[A]   E. BRODIE[B]   C.A. NIVEN[C]

[A] *Department of Psychology, University of Stirling,*
Stirling, FK9 4LA, UK
pjbh1@stir.ac.uk

[B] *Department of Psychology, Glasgow Caledonian University, George Moore Building,*
Cowcaddens Road, Glasgow G4 0BA, UK

[C] *Department of Nursing and Midwifery, University of Stirling,*
Stirling, FK9 4LA, UK

The automatic recognition of facial expressions of pain has potential medical significance: some patients are unable or unwilling to ask for analgesia, but it has been rather little studied. We report some preliminary work. The expressions were produced by actors. Human participants were able to distinguish the expressions intended to be pain from others with an average accuracy of around 80%, with disgust the most difficult to distinguish. A simple computer model consisting of a Gabor filtering front end and a linear discriminator network performed similarly well, suggesting that quite a simple model is able to emulate human performance.

1. Introduction

Much work has been done on classifying facial expressions by computer, but this has mostly been confined to Ekman's six basic emotions: anger, happiness, fear, disgust, sadness and surprise. Performance on these six expressions has been quite good (e.g. Dailey et al. (2000) report average performance of 86%). Pain, however, appears to have been largely neglected. It has potential medical significance: there are patients who either cannot or do not wish to report their pain (e.g stroke victims, post operative patients, premature babies or just not wanting to make a fuss, e.g. LeResche and Dworkin (1988)). There are some patients that medical staff are supposed to monitor, to check they are not in pain, but they may either forget to do it or fail to recognise it. In addition, clinicians regard patients' verbal reports of pain as unreliable and may therefore rely on non-verbal cues (Hill and Craig 2002). Unfortunately, medical staff tend to be worse than lay public at detecting pain expressions (Prkachin, 1997). Indeed, it has even been suggested that cancer patients be taught how to express pain in ways that medical practitioners will recognise (Keefe and Dunsmore, 1992). An alternative approach to helping

2

with this problem would be to build a computer system capable of recognising when someone is in pain by looking at them. Other possibilities for automation exist, for example attempting to classify the cry made by babies, to detect when they are in pain (Petroni et al., 1995).

Ekman developed the Facial Action Coding Scheme, that identifies the individual facial muscle movements, termed Action Units (AUs), which combine to make up expressions (Ekman and Friesen, 1978). Each expression can therefore be categorized in terms of the AUs that are displayed. For pain, there are 11, including AU4, brow lowering, AU12, lip corner puller and AU 43, eyes closed (Williams, 2002). However, it is to be expected that the combination of AUs displayed will vary with the type of pain: a sudden, sharp pain differing from more chronic suffering. They will also differ in terms of dynamics, with sharp pains producing a rapid change in expression while chronic pain results in a more fixed, drawn appearance.

There appears to be little previous work on computer classification of pain expressions (e.g. Mori et al., 2000), though Pantic and Rothkrantz (2002) speculate about the feasibility of doing so. However, there have been a number of attempts to build general expression recognition systems, reviewed by Pantic and Rothkrantz (2000). One method of doing this is to attempt to identify the individual AUs: if this can be done, then the appropriate combinations can be used to identify, in principle, any expressions using Ekman's classification. Pantic and Rothkrantz (2004) use a very rule-based approach, first locating key fiducial points on a face and then classifying the individual AUs in terms of relationships between them. For example, if the position of the eyebrow "arcade" viewed from profile moves up or down by more than some threshold amount, AU1 is indicated. The individual AUs can then be combined to identify expressions. Bartlett et al (2004) use a more holistic approach. Following an automatic face finding and alignment stage, the faces are filtered by multi resolution and orientation Gabor filters. Individual features are selected by an Adaboost procedure, before individual support vector machines classify each AU. However, Dailey et al (2000) reported quite good performance on the six basic expressions by using the same Gabor filtering stage, followed by a simple network-based classifier. This simple approach is used in the pilot experiments reported here.

2. Human Experiments

2.1.*Stimuli*

Our stimulus set derived from 23 actors from a local drama school, who were asked to simulate expressions for the six basic emotions (anger, disgust, fear, happiness, sadness and surprise), plus pain. A neutral face was added to make eight classes. The actors were videoed, and ten different examples of pain and two of each other expression were captured to give still frames. These were 720x576 pixels, of which the face occupied approximately the central 300x400 pixels.

2.2.*Study 1*

The task here was two alternative forced choice (2AFC): participants had to decide which of two displayed faces depicted pain. One example of each of the non-pain expressions was selected from each of eight actors, four male and four female. Each expression was paired with a different example of pain from the same actor, so no images were repeated. This gave a total of 56 pairs, with pain appearing on the left and right equally often for all the other expressions. The participants were first given four practice trials, using images from an actor unused in the main experiment. They were asked to press the 'z' key if pain was on the left and 'm' if it was on the right. For the practice trials, participants received feedback on their answer; for the main experiment they received none.

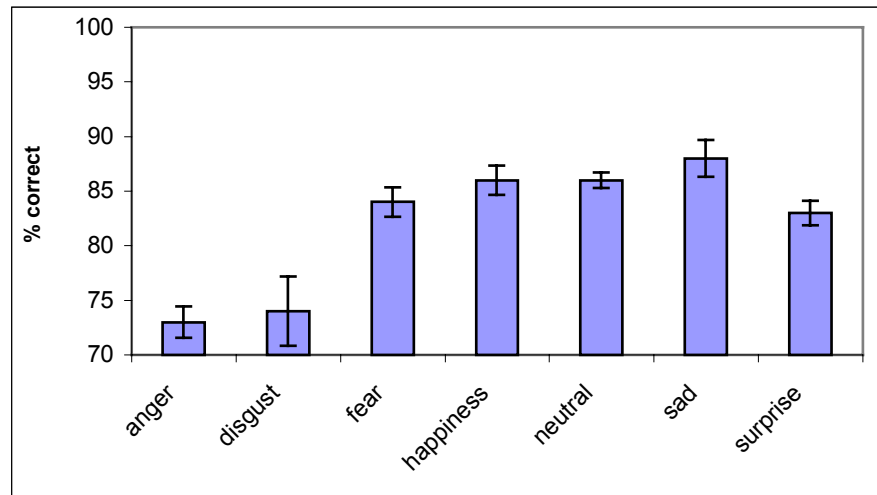Participants were 24 students from the university, mean age 24, who took part voluntarily.

Figure 1 Accuracy with which pain is selected for each alternative expression on the 2AFC task in study 1. Chance is at 50%

2.3. *Results*

Overall average accuracy was 82%. A breakdown of accuracy for each expression is shown in Figure 1. Each bar shows the accuracy with which pain was differentiated from the particular expression, since it is a 2AFC task, chance is at 50%. Anger appears to be the hardest to distinguish, with the pain expression being chosen 73% of the time, implying that the Anger expression was incorrectly chosen as being pain on 27% of occasions. A repeated measures Anova confirmed an effect of expression ($F_{6,132}$=13.4, p< 0.001), i.e. there is a significant difference between the accuracy rates for the different expressions. Testing these differences with pairwise comparisons showed that anger and disgust produced the lowest accuracy rates, significantly lower than the others, but still significantly above chance. Anger and disgust are therefore most confused with pain, but the pain expression is still correctly chosen about three times out of four.

2.4. *Study 2*

The task for this study was 4AFC, with a pain expression placed among fear, disgust and sadness from the same actor. These three were chosen as being three similarly negative emotions. The faces were positioned in a 2x2 array on

the screen, with a number, 1-4, next to each. Participants were asked to respond with the number of the face they considered to be displaying pain. An initial four practice trials were displayed, using images not used in the main experiment. Feedback was given for practice, but not for the main set of trials. In the main set, each actor featured twice, using different examples of each of the expressions. Actors were displayed in a randomised order.

Participants were 20 university students, mean age 26, who took part voluntarily.

2.5. *Results*

Overall accuracy, i.e. correctly choosing pain, was 79%. Figure 2 shows the distribution of choices. Since this is a 4AFC task, with each of the four expressions being presented on each trial, the bars represent the percentage of trials on which that expression was picked as being the one depicting pain. Most of the errors were caused by selecting the disgust expression, 13.5% of the time. Paired t-tests indicated that disgust was chosen significantly more often that the other two, which did not differ significantly from each other (e.g. disgust-fear t_{19}=5.14, p< 0.001).
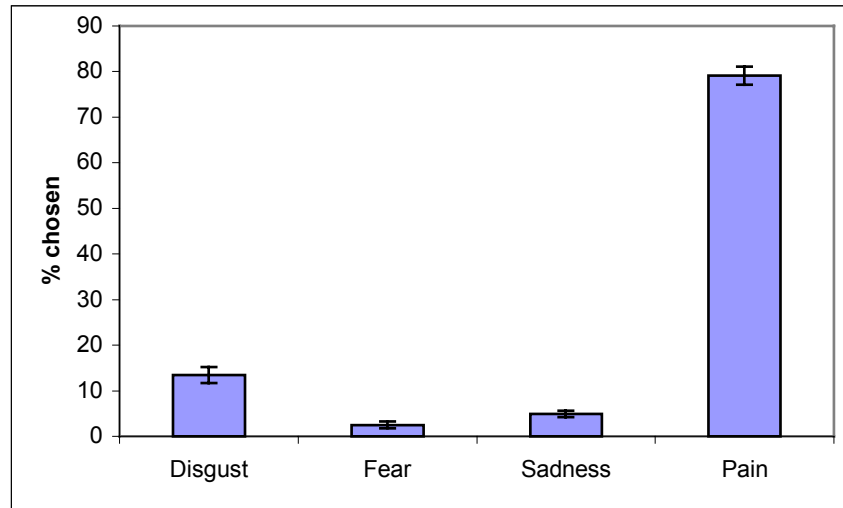
Figure 2 Results for the 4AFC task: percentage of trials on which each expression was chosen as being the one best depicting pain. Chance performance would be 25%

2.6. *Discussion*

The first aim of these studies was to provide some verification of the stimuli. Since they were performed by actors, it was possible that they would not be very clear. In the event, our participants recognized the pain expression about 80% of the time for both tasks. Since the second experiment always included a comparison with disgust, this suggests that they are quite recognisable as pain, even if perhaps not quite the same as real pain expressions. Pain was most commonly mistaken for disgust in a study using health professionals as participants (Kapesser & Williams, 2002). They reported an error rate of 18% for pain misclassified as disgust, though only 3% the other way round, disgust being more often mistaken for anger in their experiment.

3. Computer simulation

The simulation followed the design of Dailey et al. (2000), with images processed by multi-resolution Gabor filtering, followed by a linear network classification stage. We used 10 male and 10 female faces each with 2 examples of the 8 expressions for 320 images in total – 2 (gender) x 10 (individuals) x 8 (emotions) x 2 (examples). These were formed into a training and a test set, each with one of the examples from each face. The faces were

rotated and scaled to put the centre of the eyes in a common location, then cropped to 180x240 pixels, in monochrome. Figure 3 shows some examples of the edited images.

Each image was convolved with Gabor filters, at up to five different spatial scales, each at seven orientations (0, 30, 60, 90, 120, 150 degrees). Initial tests used five scales, of sd from 1-5 pixels, this was varied to obtain the best results. The filtered images were sampled at eight pixel intervals to give a 21 x 29 array of intensity values for each scale and orientation. All the samples for each original image were concatenated to give a vector of size 21315 (7 x 5 x 21 x 29). These long vectors were then compressed by a principal components analysis, taking the first 160 components.

Figure 3 Four examples of pain expressions, as cropped for image processing.

The PCA reduced vectors were then classified using a linear neural net, with 160 inputs and 8 outputs. One set of images was used as a training set, the other for testing and the two sets then reversed for cross validation. Training used a learning rate of 0.0005 for 10000 epochs (although correct classification on the training set was generally obtained after 500). We experimented with the number of inputs used and found that training was unsuccessful with fewer, for example 90 inputs gave only 70% correct classification.

Testing indicated that the initial rather fine scale filtering used (a sd range from 1-5) gave rather poor test performance, but that this improved with larger spatial scales in the range of 10-15. Marginally the best overall performance was given by the use of three spatial scales 10, 11 and 12, at 71.6%, averaged across both test sets. However, performance varied rather little with the precise choice of spatial scales and the average of four different runs, using a variety of combinations of scale from 10-15, averaged 69.5% overall. However, pain is relatively poorly classified, averaging only 46.3% overall.
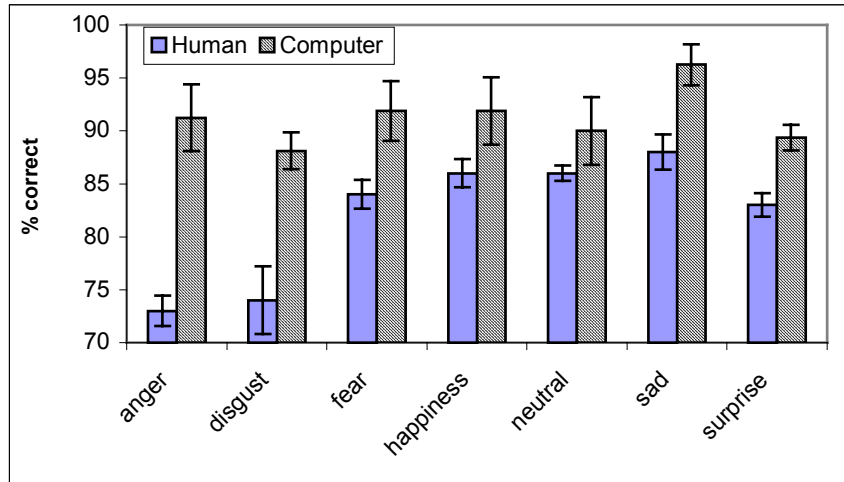
Figure 4 Accuracy of pain identification for each competing expression on the 2AFC task for humans in study 1 and for the computer simulation.

3.1. *Comparison between computer and human performance*

While the computer classification performance is not especially good, it is being asked to do a different task from the human participants. They were asked only to decide which of 2 or 4 images was pain, not to decide what expression a given image portrayed. We therefore asked the same of the computer simulation: for each identity in the set, compare the pain image with the others to see which gave the higher signal on the pain output unit. Note that while the images used in the computer and human studies are from the same set, for a variety of reasons they are not identical, e.g. in the 2AFC task, 8 actors were used for the human study to limit the length of the task, while 20 were used in simulation.

Average performance on the 2AFC task, across the four different spatial scale sets, is shown in Figure 4. The scores are much higher than for the overall classification task: the pain expression only has to produces a higher signal on the pain output than the one alternative image, rather than a higher output on pain than for any other output. That is, the system may classify the pain image as anger, for example, but still more like pain than the alternative image, which would therefore count as a correct answer. Performance is also higher than

humans on the same task, though it is interesting that both find sadness the easiest to discriminate from pain and disgust one of the hardest.

Average performance on the 4AFC task is shown in Figure 5. On this task the performance of the computer system is almost identical to that of the human participants. Both correctly pick the pain expression about 80% of the time, while making most errors for disgust.
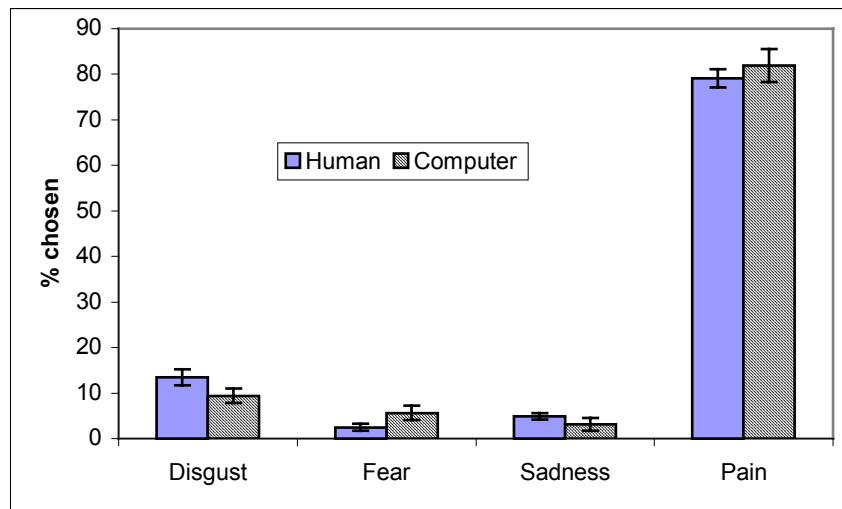

Figure 5 Results for the 4AFC task: the second experiment for human participants, and for the computer simulation.

4. Discussion

The work described here is very much pilot work, intended as a proof of concept, both to get an indication of the validity of the posed expression set and to see how easy it might be to classify pain. The computer model is very simple, though based on one that successfully classified the six basic expressions (Dailey et al., 2000). The classification method, though implemented as a neural net, is linear, and therefore equivalent to discriminant analysis. It is assumed that a more sophisticated classifier might perform better, though a larger training set might be required to constrain it adequately. The net was only asked to generalize over different example images from individual actors that were also in the training set: it was not asked to generalize across identities. This would obviously be needed in a useful system and is one focus for future development. This system also required entirely manual intervention

to identify and normalize the position and size of each face. Accurately locating a face is a surprisingly difficult task to automate, even with quite consistent images such as these, though good progress is being made (e.g. Bartlett et al 2004).

The match between computer and human participants on the 4AFC task is striking but should not be over-interpreted. Although the early spatial filtering stage of the computer system is loosely modelled on human visual processing, the rest of it is very simple and the match may well be coincidental: certainly the 2AFC results are not so obviously similar. In any case, emulating human performance is not the target of this work: ideally we need to improve upon it. However, comparison with human performance provides a good benchmark for any computerized system, and more extensive similarities might provide some circumstantial evidence as to how humans perform the task, which would be of more general interest.

Acknowledgments

References

Bartlett M., Littlewort G., Lainscsek C., Fasel I., Movellan J., (2004). Machine learning methods for fully automatic recognition of facial expressions and facial actions. In *IEEE Conference on Systems, Man & Cybernetics*.

Dailey M.N., Cottrell G.W., Adolphs R. (2000). A six-unit network is all you need to discover happiness. *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society.*

Ekman P., Friesen W.V. (1978). *Facial Action Coding System* Consulting Psychologists Press.

Hill M.L., Craig D.M. (2002). Detecting deception in pain expressions: the structure of genuine and deceptive facial displays. *Pain*, 99, 135-144.

Kapesser J., Williams A.C.de C. (2002). Pain and negative emotions in the face: judgements by healthcare professionals. *Pain*, 99, 197-206.

Keefe F.J., Dunsmore J. (1992). Pain behavior: concepts and controversies. *American Pain Society Journal*, 1, 92-100.

LeResche L., Dworkin S.F. (1988). Facial expressions of pain and emotions in chronic TMD patients. *Pain*, 35, 71-78.

Mori T., Okazaki Y., Kawai T., Sato T. (2000). Objective estimation of pain based on facial expression, vital signs and body movements *Advanced Robotics* 14, 413-414.

Pantic M., Rothkrantz L.J.M. (2000). Automatic analysis of facial expressions: the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 22, 1424-1445.

Pantic M., Rothkrantz L.J.M. (2002). Machine understanding of facial expressions of pain. *Behavioral and Brain Sciences*, 25, 469-470.

Petroni M., Malowany A.S., Johnston C.C., Stevens B.J. (1995). Classification of infant cry vocalizations using artificial neural networks (ANNs). *Proceedings of IEEE International Conference on Acoustics and Speech Signal Processing*, Vol 5, 3475-3478.

Prkachin K.M. (1997). Afterword: the consistency of facial expressions of pain. In Ekman P. and Rosenberg E. (Eds.), *What the face reveals.* Oxford: Oxford University Press.

Williams A.C.de C. (2002). Facial expression of pain: An evolutionary account. *Behavioral and Brain Sciences*, 25, 439-488.