Centrum voor Wiskunde en Informatica
REPORT*RAPPORT*

Exploring the Space of Emotional Faces of Subjects without Acting Experience

J. Hendrix, Zs.M. Ruttkay

Information Systems (INS)

INS-R0013 June 30, 2000

Report INS-R0013
ISSN 1386-3681

CWI
P.O. Box 94079
1090 GB Amsterdam
The Netherlands

Exploring the Space of Emotional Faces of Subjects without Acting Experience

J. Hendrix, Zs. Ruttkay

CWI

P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

Email: {jeroen.hendrix, zsofia.ruttkay}@cwi.nl

ABSTRACT

The current state of semi-automated facial animation demands better understanding of how facial expressions are produced by real people. As a first step of a series of empirical studies, we investigated snapshots of facial expressions of the six basic emotions, produced by 18 subjects in 'out of context' sessions. The MPEG-4 coded vectors were analysed on principal components and canonical variates, which served as a basis to draw conclusions about the (for the negative emotions: lack of) generic characteristics of the six emotions and their mutual distances. Our results are compared to ones obtained by facial image analysis and human perception researchers.

1998 ACM Computing Classification System: H.5.2, I.5.3, J.4

Keywords and Phrases: facial expressions, classification, multi dimensional scaling, principal component analysis.

Note: An on-line version of this report with figures in colours is available from `ftp://ftp.cwi.nl/pub/CWIreports/INS/INS-R0013.ps.Z`.

1. Introduction

In the framework of the Facial Analysis and Synthesis of Expressions (FASE) project [FASE 1998] we have been aiming at (re)producing 2D and 3D synthetic faces with expressions. We have adopted and improved a 3D physically-based facial model, Persona, which can be deformed by pulling facial muscles. We also developed the CharToon environment [Noot, H., Ruttkay, Zs. (2000)][Ten Hagen, P., Noot, H., Ruttkay, Zs. (1999)] to design 2D cartoonish faces with deformable features. Whatever model one uses, when one wishes to show emotional expressions, it is essential to know 'what makes a face look e.g. sad', and what is the correct timing of an expression. In order to be able to generate convincing synthetic emotional faces, one has to answer the following two questions:

- What are the general and person-specific characteristics of (dynamical) facial expressions on real faces?
- How can/should these characteristics be mapped on simpler synthetic faces, especially 2D cartoon faces, if one would like to achieve the similar emotional expression?

In this paper we deal with the first question, the second is going to be addressed in a coming study [Ten Hagen at al. 2000]. The first, in itself interesting question should be answered by analysing emotional expressions on real faces. Our analysis is based on tracked facial data gained by the point tracking system developed by our partner in the project at the Technical University Delft [Veenman, C.J., Hendriks, E.A., Reinders, M.J.T. (1998)]. Below we report on the first pilot investigation, in which we restricted ourselves to snapshots of 6 basic expressions, coded according to MPEG-4 [ISO (1998)]. In this framework, each snapshot is represented by a vector of FAPs. Further on we refer to the multi-dimensional space of the parameters as the *expression space*. Our goal was, naturally, first of all to find answers to questions like:

1) What does the expression space look like?
   - Can the space be reproduced in lower dimensions?
   - What are the decisive components of facial expressions?
   - What part of the expression space is perceived as emotional expressions?
2) Can the 6 emotional expressions be identified in the space?
   - Is our set of parameters (or even a subset of it) sufficient?
   - Can 'prototypes' of all 6 basic expressions be defined?
   - Are there essentially different alternatives for the same expressions?
   - What are the distances between the 6 basic expressions?
3) Are our findings comparable to characterisations given by others?
   - Are the typical characteristics of the 6 basic emotions similar to descriptions used by psychologists?
   - Can the 2D space, suggested by perception analists [Schlosberg, H. (1952)][Russell, J. A. (1980)], be detected in our approach?
   - How do distances of expressions correspond to conclusions by other facial expression analysis researchers?

We knew from the beginning that our single experiment cannot be used as decisive to answer the questions above. Hence we only considered this case as a first 'pilot study', to develop the appropriate tools and gain experience to fine-tune further experiments.

In this paper we first sum up the conclusions of previous research on a computational framework of facial expressions. Then in Section 3 we discuss our data collection method. In Section 4 we explore the data. First we introduce the expression space and the data analysis methods used to explore it. Then we draw conclusions about the characteristics of expressions and their distances. In Section 5 we further investigate the data set, looking at the size of the clusters and the correlated parameters. We make assumptions to explain the mistakenly perceived expressions and show that the negative emotions in our experiments cannot be better separated by taking more detail into account. We end the paper by giving (sometimes negative or partial) answer to the above listed questions and by outlining further work.

2. Related work

The analysis of facial expressions has challenged many researchers, both from the field of psychology and of image processing. Below we list those who have been trying to provide a computational model to classify facial expressions.

One of the earliest frameworks to describe and classify facial expressions is the FACS system [Ekman, P., Friesen, W. (1978)]. Though initially developed for the psychologists to hand-code facial expressions, it has become popular in software systems. On the one hand, many image processing systems have adapted the underlying principles to analyse facial expressions, and a system has been developed which automatically codes facial expressions with higher accuracy than human coders do [Barlett, M. S., Hager, J. C., Ekman, P., Sejnowski, T. J. (1999)][Donato, G., Bartlett, M. S., Hager, J. C., Ekman, P., Sejnowski, J. (1999)]. (It was assumed that the trained performers produced correct expressions, and the recognition rate was taken as indication of accuracy.) On the other hand, it has been common practice to use the '6 basic emotional expressions' which Ekman and his collegues claim to be universal. In the community of psychologists, however, there has been criticism of the categorial approach of Ekman [Russell, J. A. (1994)]. Based on early works by Schlosberg [Schlosberg, H. (1952)], Russell places the 6 basic and many other facial expressions in a 2D space, in a circular form [Russell, J. A. (1980)]. In his approach emotions are defined by 2 coordinates of pleasure and arousal in the continuus emotion space, in contrast to the discrete categories of Ekman. In a recent paper [Schiano, D. J., Ehrlich, S. M., Rahardja, K., Sheridan, K. (2000)] not only the (mostly methodological) criticism by Russell has been proven to be uncorrect, but it was also shown that the circular arrangement could not be reproduced when visualizing the 'perceptual closeness' of the 6 basic emotions in a 2D space, by using multidimensional scaling.

Pilowsky and Katsikitis [Pilowsky, I., Katsikitis, M (1993)] used snapshots of 'peaks' of emotions in videorecordings of the 6 basic emotions posed by 23 drama students. The facial expressions on the snaphots were described by 12 characteristic standardised distances on the face, including data about the eyes. They classified the data by means of specific software which also produced the best number of classes automatically. The result was 5 classes, two of them containing a majority of a single expression, namely happiness and surprise. The other 3 classes each contained a mixture of expressions. Though not homogeneous, each class could be characterised by typical deformed features and distances. The authors concluded that their computational investigation served as justification for the existence of 3 fundamental emotions: surprise, smile and 'negative'. They also raise the issue that the existence of the three mixture classes might be caused by the lack of clear unique prototypes for the negative emotions.

Yamada and his collegues [Yamada, H., Watari, C., Suenaga, T. (1993)] did investigations similar to ours: they used canonical discriminant analysis to visualize the 6 basic expressions performed by 12 females and coded in the form of MPEG-4 like parameters. They found three major canonical variables, the first one for lifting the eyebrow and opening the mouth, the second (roughly) for pulling up the cornes of the mouth and the third one for the position of the eyelids and eye corners.

Essa [Essa, I. (1994)] used naive performers to pose the 6 basic emotions 'out of context'. With optical flow analysis he characterised the expressions in terms of time-curves of muscle contractions. He used the peak of the curves (which he found unique for muscle actuations) as representative snapshots of expressions. He reported that subjects had difficulty with producing fear and sadeness, hence his database contained holes, and fear was not present at all. He used dot products of the muscle contraction vectors as an indication of closeness of expressions. He found that anger and disgust were close to each other, and suprisingly, anger and smile too. For the latter observation he refered to Minsky [Minsky, M. (1985)] claiming that in the case of these expressions which have similar snapshots at the peak, the time behaviour is an important differentiating factor.

Yacoob and Davis [Yacoob, Y., Davis, L. (1994)] used a rule-based system to recognise facial expressions based on temporal and spacial analysis. We quote them especially, as they also lacked fear, disgust and sadness samples due to poor performance. Their system had difficulty with differentiating between fear and surprise, sadness and disgust and sadeness and anger for certain faces.

3. Collecting data of expressions

3.1. The recording setting

The 18 experimental subjects were all students or young co-workers at an electronic engineering department, 17 males and 1 female, 11 of Dutch origin. After an introduction about the goal of the project, blue dots of about 0.5 cm diameter were put on the subjects' face according to a fixed scheme (see Figure 1).

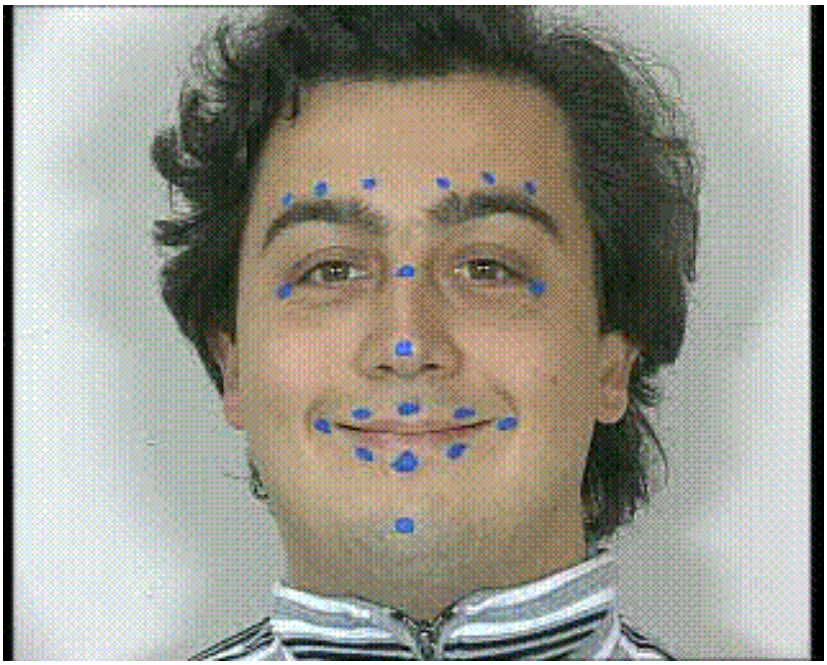

Figure 1. Performer's face with blue dot markers to be traced

Subjects were asked in the framework of individual sessions to make the 6 basic expressions written on a blackboard in a predefined order (smile - surprise - anger - disgust - fear - sadness), each twice. We were aware of the fact that the experiment was done 'out of context', that is: subjects were ordered to pose the emotional expressions without stimuli, which is known to have (at least two) deficiencies:

- for the average person, it is difficult to produce expressions 'on demand',
- for some expressions there is a difference in appearance between 'false' and 'really motivated' facial expressions.

The sessions were videorecorded with two cameras, and based on the 3D (time dependent) positiont of the blue dots, MPEG-4 FAPs were computed for each frame. Most of the blue points were moving ones, which were used to compute the corresponding FAPs. The non-moving blue dots were used as spatial reference points, and also to compute the 3D position of the head. Hence the movement of the face was expressed in terms of 15 FAPs, each telling the vertical or horizontal movement of specific feature points (see Figure 2). Finally what we got was the time curve of the 15 FAPs for each person (see Figure 3).

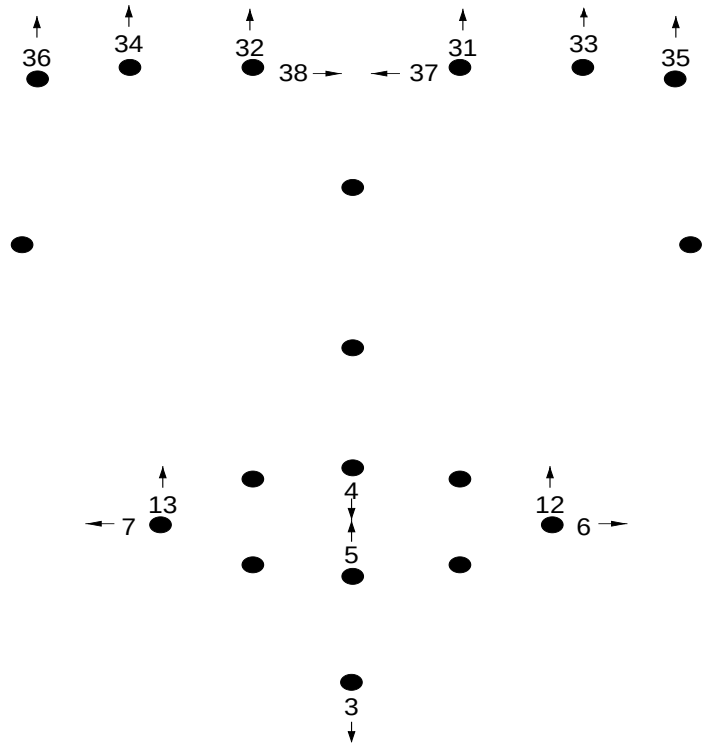

Figure 2. The tracked FAPs with directions

3.2. Getting snapshots of expressions from the recordings

From the complete recording we took 'the most extreme' snapshot for each expression. This was done essentially by choosing the snapshot at the 'peak' of most of the curves. Note that in many cases the 15 curves did not all have the extreme at the same time. In controversial cases (e.g. when it was not clear where to find the extreme of 'most of the curves'), the corresponding videoframes were looked at and the 'best' frame was identified, always by the first author. As a result, we ended up with two sets of 6 expression snapshots for each person, containing the first and second trials. We kept going on with the set of second trials, as in general we got the impression that for the second time subjects could show expressions better.

3.3. Pruning erroneous data

We noticed, however, both from the shape of the curves and from the video that several subjects had difficulties with producing especially the negative emotions: they were laughing, or produced faces which did not show the intended expression. In order to prune such 'erroneous' recordings, we asked 56 volunteer collegues at our institute to re-label the 108 snapshots, arranged in an identical random sequence and shown one after the other on a Web page. They had to give their first impression of the snapshots, which could be one of the 6 basic expressions or the 'none of them' answer. The rate of identical labelling for expressions was considered as an indication of the 'success of producing the emotion'. (We did not take into account the possibly disturbing effect of the 'blue dots' on the face.) A performed expression was *correct* if at least 50% of the evaluators perceived it as intended, providing the GOOD data set. A performed expression was considered as *mismatch* if at least 50% of the evaluators agreed on preceiving it as an expression different of the intended one, providing the MISS data set. Correct and mismatched expressions together form the *accepted* expressions (ACC data set). The rest of the cases were rejected. All the data is referred to as the ALL data set.

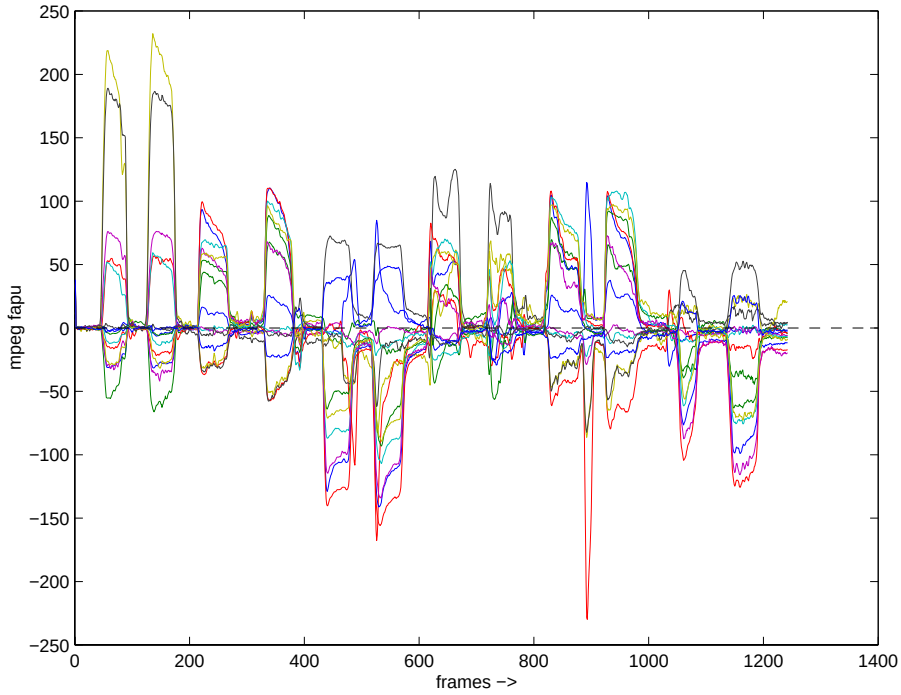

Figure 3. The time curves for a recording session. One can identify the patterns of expressions, each twice.

The outcome of re-labeling is shown in Table 1 below. The results suggest that happy and surprise are easy to produce (100% and 86% correct for these two expressions), while when producing the 4 negative emotions, the result is often a rejected expression (26%) or a mismatch (62 %). Particularly, none of the 18 performed 'fears' was correct, all 11 of them which got accepted were mismatches. Though the results are in line with results reported by others, the very low success rate for negative emotions was an unpleasant surprise to us. We dwell upon this issue further in Section 4 and Section 5.

For the rest of out investigations, we used only the 47 correct expressions to draw conclusions. Then we looked at the 18 mismatches to find out if in our computational framework the mismatch in perception could be explained.

Table 1: Result of relabeling expressions of the data set

intended expression	total relabelled	smile	surprise	anger	disgust	fear	sadness	none
smile	17	17	0	0	0	0	0	0
surprise	15	1	12	0	0	0	0	2
anger	13	0	0	6	2	0	1	4
disgust	14	1	1	2	6	0	2	2
fear	11	0	4	1	0	0	1	5
sadness	9	1	0	1	0	0	6	1

4. Visualising the data

4.1. The expression space

Each recorded expression was represented in the dataset by a vector of 15 normalised FAP values. By normalising the data we expressed displacements relative to the extremes of a person. E.g. one person never pulls his mouth wider than 1.5 times the unit distance used in MPEG-4, and for another person this extreme is 2.5. By normalising the displacements relative to the person's maximum, we compensated for the differences due to differences in dynamical ranges, and tried to focus on patterns. Hence each normalised FAP value was between -1 and 1.

Our data defined points in a box of the 15 dimensional space. In order to get a 'picture' of the different data sets in the 15 dimensional space, both visually and in an abstract sense, we performed different data analysis. For our investigations, we used Matlab.

4.2. Visualization of the expression space

Multi Dimensional Scaling (MDS) is a common tool to approximate data in a high-dimensional space with data in a lower dimensional space. Particularly, if the space used for approximation is 2 or 3 dimensional, it can be visualised.

We used Principle Component Analysis (PCA), a special kind of MDS. The basic idea of PCA is the following: A new orthogonal basis of the original (in our case 15 dimensional) space is constructed in such a way that the original data vectors can be well approximated by the (unique) linear combination of a low number of the basis vectors. The components in this combination are the projections of the original vector on the basis vectors. On the other hand, the basis vectors can be expressed as linear combination of (some of the) the original vectors, giving an insight into the 'semantics' of the new basis vectors. That is, the first new basis vector now becomes a linear combination of the 15 original vectors for which the variance of the projections of all vectors on this coordinate is maximal. A comprehensive explanation of PCA can be found in [Krzanowski, W. J. (1988)].

4.3. The results of PCA

First we performed PCA for the ALL dataset. When expressing the new basis vectors as linear combinations of the original ones (corresponding to FAP parameters in the data set), the coefficients give an insight into the nature of the components. For the first four components the coefficients are given in Table 2.

As we can see in Table 2, the first two components make up for about 73% of the total variation. If we examine the values in these two components, we can get an insight into the FAPs that play the biggest role in differentiating the expressions.

The large values of FAPs 31 to 36 in component 1 suggest that component 1 is dominated by the raising and lowering of the eyebrows. The other FAPs do not have zero values in the first component, so of course the raising of the eyebrows is not the only factor in component 1.

In component 2 we can find large values at FAPs 6,7,12 and 13, which all are concerned with the 'smiling' movements of the mouth.

Component 3 can be seen as dealing with the opening of the mouth. If we plot only the first two components, we will lose a large part of the information about the 'openness' of the mouth. However, in Section 5.4 we will show that the third component does not add much to distinguishing between the expressions in this experiment.

Table 2: The first four principal components expressed in terms of the original FAPs

FAP variables	component 1 47.6%	component 2 26.3%	component 3 9.41%	component 4 5.53%
3	-0.17	0.22	0.63	0.13
4	-0.20	0.25	-0.31	0.33
5	0.17	-0.26	-0.59	-0.13
6	0.12	-0.34	-0.04	0.57
7	0.14	-0.39	0.13	0.38
12	0.15	-0.41	0.19	-0.06
13	0.15	-0.41	0.20	-0.13
31	-0.34	-0.12	-0.04	0.22
32	-0.34	-0.09	-0.02	0.29
33	-0.35	-0.14	-0.05	0.01
34	-0.35	-0.14	-0.02	-0.01
35	-0.35	-0.07	-0.13	-0.03
36	-0.35	-0.10	-0.06	-0.09
37	0.22	0.30	-0.19	0.25
38	0.23	0.23	0.02	0.40

In Figure 4 the first two PCA components of the GOOD dataset are plotted. For each expression the convex hull of the cluster of points is plotted. Notice the clearly distinct clouds of smiles and surprises. The negative emotions are all in the top-right corner (eyebrows down, sad mouth), but are rather mixed. It is evident that the the first two components of the tracked FAPs are not sufficient to differentiate between negative emotions. We'll ponder further about the overlap of the negative emotion clusters in Section 5.3 and Section 5.4.

The surprises are divided amongst two subclusters: one raising the eyebrows very much and the other mainly lowering the mouth corners. Lowering of the mouth corners is largely due to the opening of the mouth, not visible in these two components. Thus we can conclude that a surprised face was made in two different ways: with closed and with open mouth.

In Figure 5 the first two components of the ACC dataset are plotted. The resulting picture is largely the same as the previous one of GOOD. The added mismatched expressions of MISS fall nicely into clusters formed by corresponding GOOD data. This suggests that concerning the tracked FAPs, there is not much difference between an expression which was produced as intended and an expression which was produced against intention.

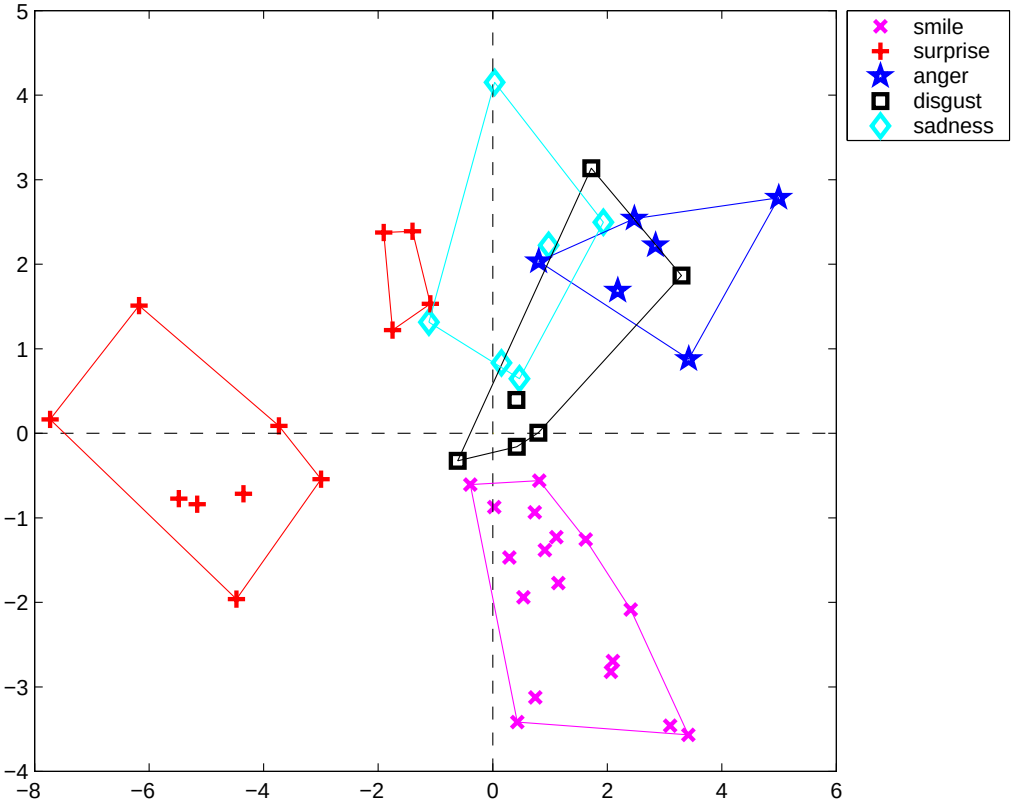

Figure 4. The GOOD dataset

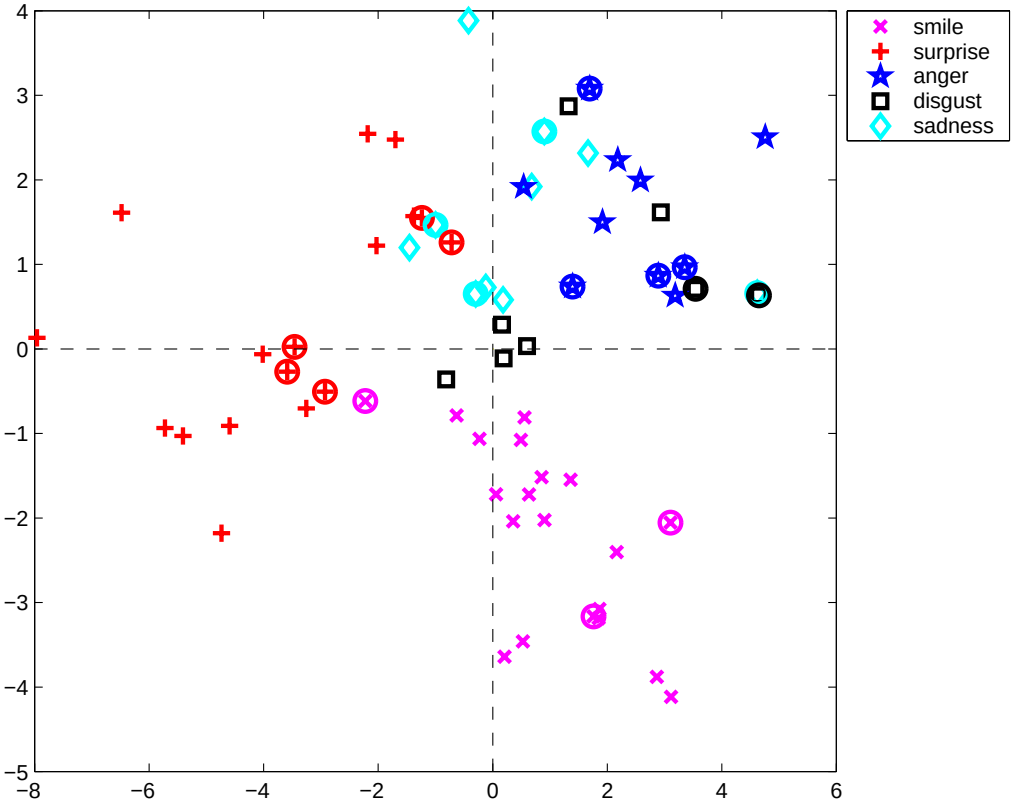

Figure 5. The ACC dataset, labelled to how the expressions were perceived

4.4. The generic expression

Having for each expression a bundle of vectors in 15 dimensional space, we can compute an average vector for each expression. How much such an average can be considered as a generic expression of the emotion in question is hard to say. An expression which has two possible configurations (surprise with mouth open or with raised eyebrows) will probably yield an average expression which is neither.

Notwithstanding these concerns, the average expressions can perhaps give some insight into the general structure of an expression. We calculated the averages for the GOOD dataset and plotted them in the PCA graph (Figure 6). The averages thus found we used as a representation of the 'generic performer'.

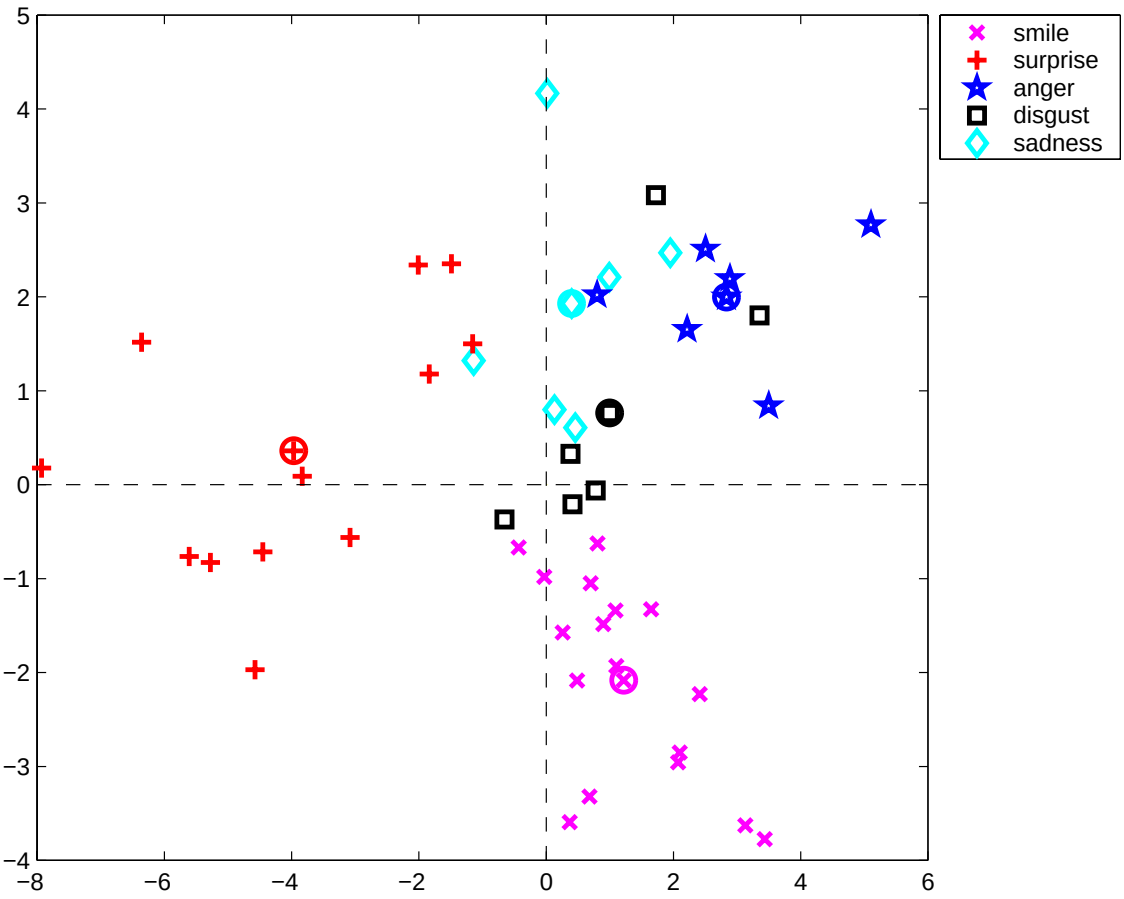

Figure 6. The average expressions of the GOOD dataset

4.5. Distances between expressions

In [Schlosberg, H. (1952)] an empirical analysis of six expressions is performed, of course not by computer, but based on human perception of photos of emotional faces. They provided a two dimensional representation of (a part of) the expression space (see Figure 7), based on the observation that people made very confined mistakes when identifying expressions from stills. Each expression was only mistaken for two others, and in such a way that easy to mistake 'neighbouring' expressions result in a circular graph (see Figure 7). The neighbouring expressions can be said to be close to each other with respect to perception.

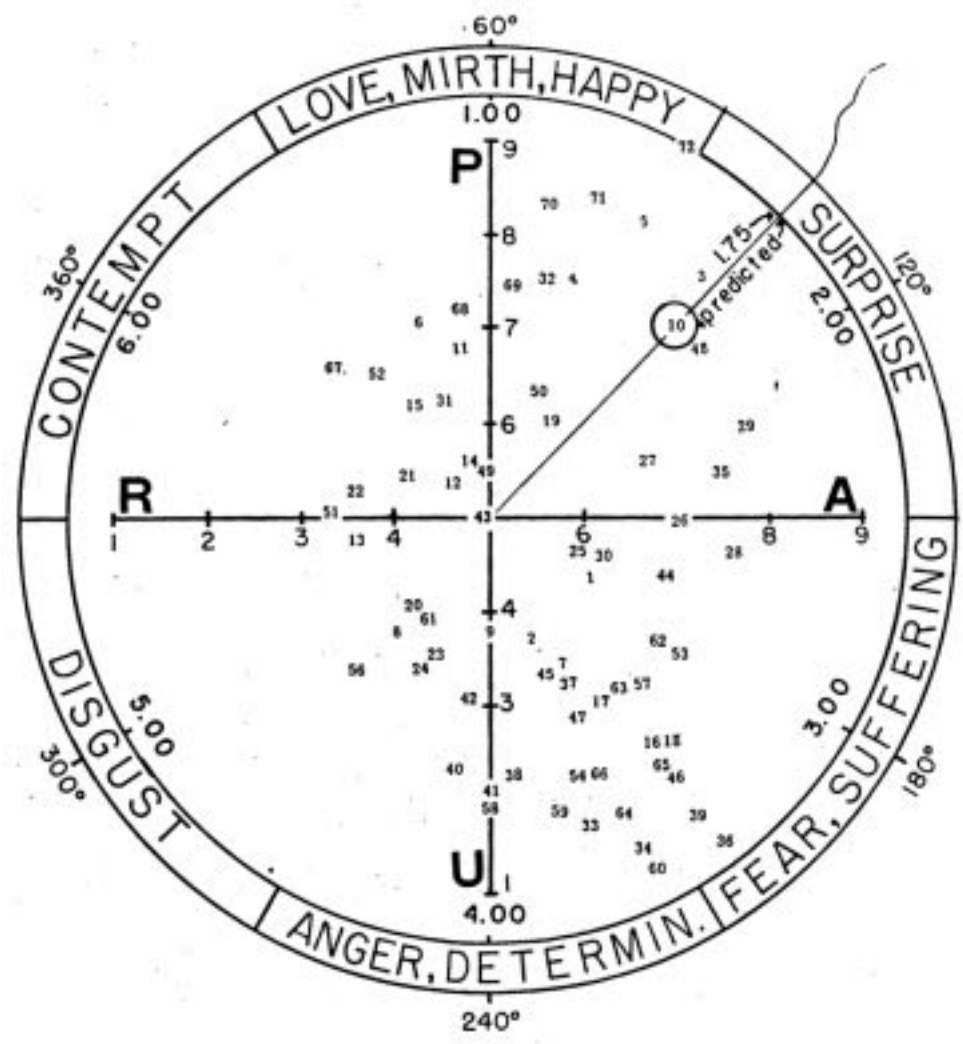

Figure 7. The 'emotion circle' of Schlosberg (reproduced from the referred article)

We wondered if the above described, perception-based closeness matches some kind of closeness of the FAP vectors in our expression space. A closeness in our 15 dimensional space can be defined by a distance function in this space. We chose the Euclidean distance.

We used the single 'generic emotions' mentioned in Chapter 4.4 to test this conception. Table 3 shows the distances between the generic expressions. Notice that we have no information about fear, because no fear was included in the GOOD dataset.

Table 3: Distances between the generic expressions

	smile	surprise	anger	disgust	sadness
smile	0	3.195	2.637	1.926	2.554
surprise		0	3.436	2.298	2.084
anger			0	1.506	1.645
disgust				0	1.040
sadness					0

To visualize these distances in a clear way the second author wrote an application, which shows the six basic expressions as points in a circular arrangement and connects only those that have a distance smaller than some treshold. The user can change this treshold interactively. The circular arrangement of the six points makes it easy to test the hypotheses mentioned earlier. The picture in Figure 8 show that the distances between FAP-vectors do not conform to the psycological theory of close neighbours in a circle reported in [Schlosberg, H. (1952)]. Particularly, smile and surprise are far apart. On the other hand, disgust and anger are very close to each other, which was also suggested by Essa [Essa, I. (1994)]. The closeness of sadness to anger and to disgust are in line with the findings of [Yacoob, Y., Davis, L. (1994)].

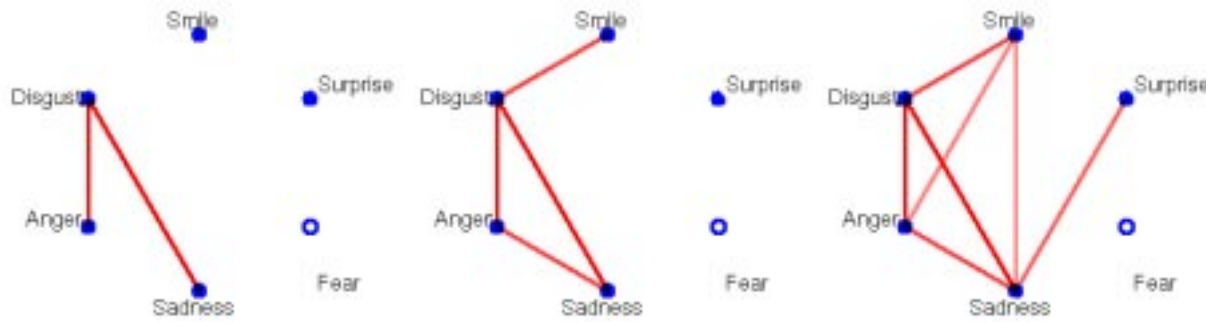

Figure 8. The 2, the 4 and the 7 smallest distances

5. Further analysis of expressions

5.1. Similarities within expression clusters

In Table 4 for each vector in the GOOD dataset a closest neighbour (Euclidean distance) is assigned. We can see that all smiles have a closest neighbour which is a smile as well, 8.33% of all surprises are closer to a sadness than to any other surprise etc.

Table 4: The closest neighbours in the GOOD dataset.

original	closest neighbour				
	smile	surprise	anger	disgust	sadness
smile	100%	0%	0%	0%	0%
surprise	0%	91.7%	0%	0%	8.33%
anger	0%	0%	100%	0%	0%
disgust	0%	0%	16.7%	33.3%	50%
sadness	0%	0%	0%	33.3%	66.7%

In Table 5 the maxima of the Euclidean distances between pairs of expressions from two clusters is shown. Notice that smile and surprise have the largest diameter (but these clusters have more points than the others). Also, disgust has, relative to it's diameter, small distances to anger and sadness, suggesting that there is no real coherence in this cluster.

Table 5: The maximum distances within and between clusters.

	smile	surprise	anger	disgust	sadness
smile	3.3798	6.4425	4.9192	5.3948	5.5451
surprise	6.4425	3.1227	5.0611	4.4295	3.7811
anger	4.9192	5.0611	1.7488	2.9630	3.2903
disgust	5.3948	4.4295	2.9630	2.8907	2.6678
sadness	5.5451	3.7811	3.2903	2.6678	2.1357

5.2. The correlations in the data

We statistically analized our ALL dataset, to gain some insight into the correlation between the movement of the different points on a face. By just looking at the correlations for all 15 parameters, we can immediately make some intuitively apparent remarks (see Table 6).

All vertical displacements of points on the eyebrows are strongly correlated. When the left part of the left eyebrow goes up, the right part of the same eyebrow and all points of the other eyebrow almost always go up too. The vertical displacement of the cornerpoints of the mouth are heavily correlated.

The construction of the principal components is such that highly correlated variables will be comprised into one component. Hence, we did not have to bother with minimalising the numbers of considered points in our analysis. But, on the other hand, by using a single representative of the correlated FAPs, we would not have lost relevant information about the expressions. In other words, it would have been sufficient to track one of the pairs of the FAPs, and only the mid-eyebrow, to get similarly accurate characterisation of the expressions. One always has to keep in mind, though, that these correlation are valid for the expressions in this experiment. It doesn't imply that for instance the cornerpoints of the mouth always move symmetrically.

Table 6: Top seventeen highest correlation coefficients. (The six points on the eyebrows, seen from the front are given with o and x marks, where. x's stand for the points of considered pair.)

FAP1	FAP2	correlation	description
34	36	0.96692	(xxo ooo) right eyebrow
3	5	-0.96575	open jaw - close lower lip
12	13	0.96383	raising left cornerpoint mouth - right cornerpoint mouth
33	35	0.96107	(ooo oxx) left
31	32	0.96020	(oox xoo) right left symmetrical
31	33	0.95249	(ooo xxo) left
33	34	0.94833	(oxo oxo) right left symmetrical
33	36	0.92108	((xoo oxo) right left
35	36	0.91355	(xoo oox) right left symmetrical
32	34	0.90003	(oxx ooo) right
31	34	0.89912	(oxo xoo) right left
34	35	0.89654	(oxo oox) right left
32	33	0.89439	(oox oxo) right left
31	35	0.89161	(ooo xox) left
32	35	0.84358	(oox oox) right left
31	36	0.83849	(xoo xoo) right left
32	36	0.82849	(xox ooo) right

5.3. Explanation for mismatches

In the ACC dataset there are 18 entries more than in the GOOD dataset. These 18 expressions are perceived as different from the intended expression.

For the smiles there is a simple explanation, namely that the performer didn't succeed in making the intended expression and laughed because of the failed attempt. Thus the different labeling of these expressions is correct, and it is thus also justified that the data of these expressions 'falls in' the cluster of smiles in the GOOD data set.

Many negative emotions were precieved as a different negative emotion or surprise. Though the FAP vector of the mismatched expressions also fall into the cluster of the correctly perceived expressions (see Figure 5), one should not conclude that these cases were labeled only by their FAP values. E.g. when making a sad face, many performers look down or partly close the eyelids. These factors were not measured with the FAPs in this experiment.

Probably the eye region and the head orientation play an essential role in judging the negative emotions. This assumption is supported by the example of the expression depicted in Figure 9.

On the left is a graph of the ACC dataset plus those expressions that were perceived as 'None'. These cases are indicated by circles.

The still on the right is the snapshot of an intended fear. As probably everyone agrees, it doesn't look much like fear. More than 50% of the evaluators labelled this expression as 'None of the above'. Yet, the first two components of the expression space suggest that this expression resembles very much a surprise (the circled green triangle on the left)! Probably, the half closed eyes is what makes the evaluators think otherwise.

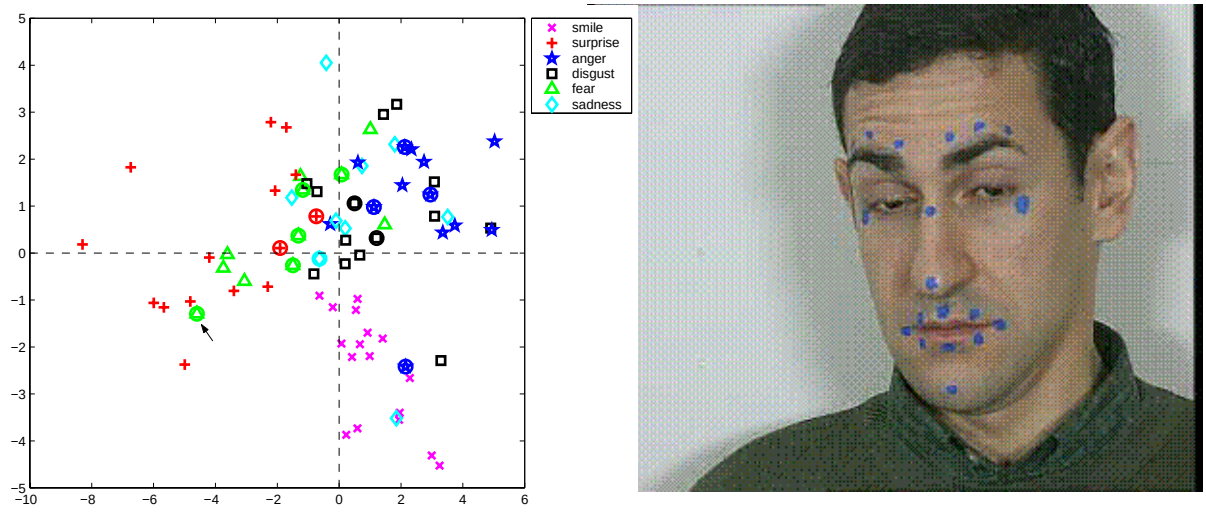

Figure 9. Intended as fear, perceived as none, but in the 'surprise area'

5.4. What is the contribution of higher components?

The first two components include 73% percent of the total variation. This still leaves 27% not accounted for. Can the information in the remaining component be used to make a clearer distinction between the negative emotions? The answer is no, which we explain on the basis of a concrete case.

Let's try to find a criterium to decide between anger and disgust. For this we use only the expressions labelled anger or disgust from the GOOD dataset and performing PCA on these few expressions. This way, the components are optimally constructed to capture the variation in this subset.

If the clouds of anger and disgust were distinguishable in 15 dimensional space, then it would be possible to see this in the projections on at least one of the principal components. Figure 10 shows no clearly separable clouds in the first three components. With a willing eye, one could see a cloud of angers and two small clouds of disgust in the left picture, but this is hard to ascertain on such a small set.

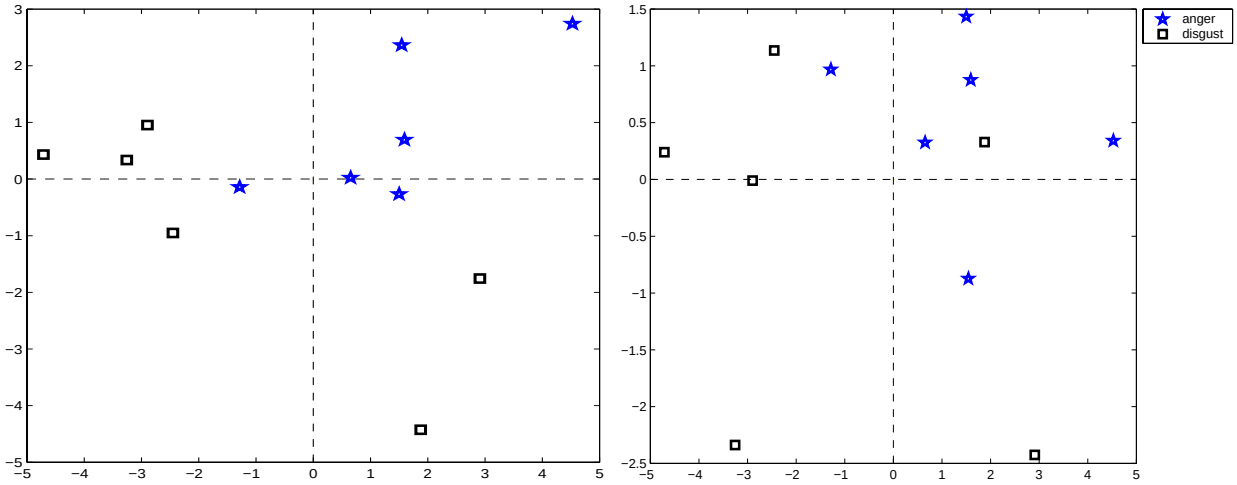

Figure 10. Anger and disgust, components (1,2) and (1,3)

The previous analysis was repeated for all combinations of two negative expressions and for all principal components, and resulted in similar negative results. Hence again it is underlined that the tracked data is not sufficient to differentiate between negative emotions.

5.5. Canonical variate analysis

When applying PCA, we completely disregarded the fact that we are already aware of a certain structure in the dataset. We knew beforehand that the expressions originated from a set of ‘families’: smile, surprise, anger etc. This might help us to get maximal information regarding the dissimmilarities between expressions into a few-dimensional picture.

Canonical variate analysis (see [Krzanowski, W. J. (1988)]) can be used for this purpose. It creates projections like PCA, but in such a way that it maximizes the mean differences between the families. Applying this technique yields Figure 11.

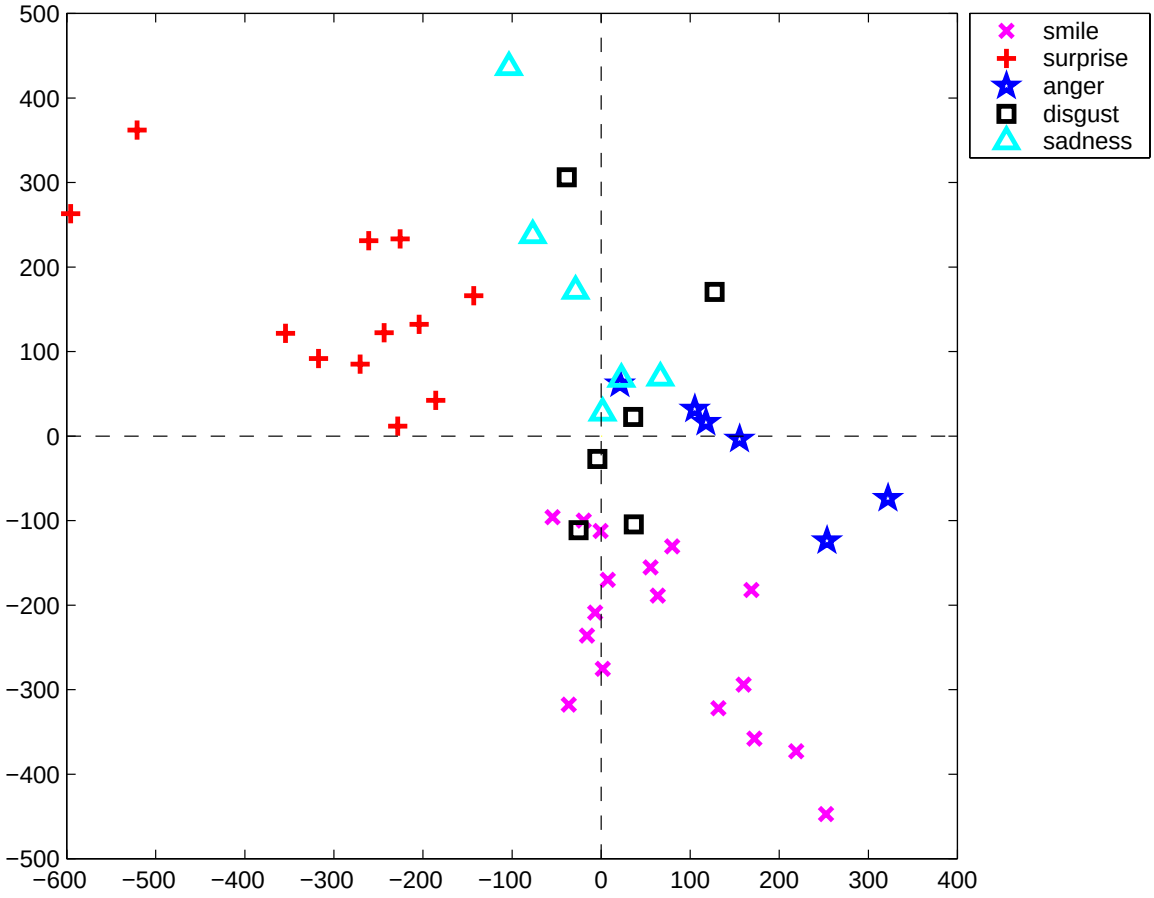

Figure 11. CVA on the GOOD dataset

A small improvement to the PCA picture can be found, namely that the clusters of anger and sadness are more distinguisable. The ‘meanings’ of these first two canonical variates (just like in PCA) are:

1) eyebrows down and/or mouth close

2) mouth corners down

Disgust is still hard to characterize, but anger and sadness are different in the way that anger has clearly more of the ‘eyebrow down’ property and sadness has more ‘mouth corner down’.

6. Discussion: answered, unanswered and new questions

Our experiment, though small, gives a computational framework to compare facial expressions in still images. To evaluate the results, we return to the questions posed in Section 1.

6.1. The expression space

The FAP description of a face leads to a box in a multidimensional space of real numbers, bounded by the maximal and minimal values of each separate FAP. The results of PCA showed to some extent how the six basic emotional expressions are distributed in this hard to visualise 15 dimensional space. The heavily correllated FAPs of the eyebrows, the mouthcorners and the opening of the mouth seem to form a three dimensional subspace in which most of the actions can be well approximated. Though the 'meaning' of the components is different than reported in other works [Yamada, H., Watari, C., Suenaga, T. (1993)], due to the different set of tracked feature points, our experiments support the idea that emotional expressions can be expressed in a 2 or 3 dimensional continous space. Of course this doesn't imply that other expressions, not considered in this experiment, wouldn't need a different or larger subspace.

An interesting and still open question is to explore what part of the expression space is perceived as an emotional expression. From our experiments it is possible to conclude that no (basic) emotion is in the corners of the 2D box, and some rectangles in the left bottom and top of the box are also empty. Regarding the interpretations of the two displayed components, we can conclude that none of the six basic expressions is characterised by raising the eyebrows while at the same time having a smiling mouth position (left bottom corner).

It requires further investigations to find out which part of the expression space corresponds to 'meaningful' or physically feasible expressions. With more subtle analysis of the changes of expressions, one could get a picture about the 'transition pathses' between expressions.

6.2. Characterisation of expressions

Due to the small size of the dataset used in this experiment, it is not possible to draw very firm conclusions respecting statistical laws. In the ACC dataset most expressions had about 10 to 15 samples. These clusters of expressions, i.e. 15 dimensional vectors, are quite close to each other relative to their size. In the case of the three negative emotions, the clusters even overlap one another. As the flocks of points are so thin it is hard to assign a unique characterisation to each of them. As we saw in Section 4.3, surprise is produced in two different ways, leading to smaller subclusters. In Section 4.4 we neglected these properties and assigned a unique vector to each emotion by means of calculating the average of the clusters. This lead to a representation of five of the six emotions.

As we expected, the subset of FAPs we considered was not sufficient to always be able to distinguish two emotions from each other. Two different emotions sometimes were very close in terms of the distance their FAP vectors (Section 4.5), yet people could differentiate between the original expressions easily. Other factors must be of influence when perceiving expressions: the region of the eyes, the orientation of the head. The importance of the eye region is clear from the work by Yamada [Yamada, H., Watari, C., Suenaga, T. (1993)]: in contrast to us, they found a distinctive and important third component, which was containing movements of the eye lids and corners. Using FAPs for additional properties of the face (eye gaze, head orientation) could perhaps lead to much better results. Also, some of the FAPs now used, can be discarded with neglectible loss because of the strong correlations between some FAPs.

6.3. Comparison to psychological findings

In [Schlosberg, H. (1952)] psychologists propose a two dimensional representation of the space of emotional faces where the six basic emotions are situated on a circle in a plane with perpendicular axis pleasantness-unpleasantness and attention-rejection. The willing eye can see this arrangement in the graphs of the first two PCA components: smile and sadness are opposed on the vertical mouth corners axis, similar to pleasantness-unpleasantness, surprise and anger/disgust are opposed on the eyebrows axis, a little like attention-rejection.

The article also mentions some sort of similarity in perception in pairs that are next to each other on the circle. In this respect smile and surprise should be similar in our computationally based framework. This assumption was tested with the generic expressions calculated by averaging. In Section 4.5 is shown that euclidean distances between neighbouring expressions are often larger than other distances and hence do not validate the assumption.

The components of the expressions, expressed in FAP parameters, are in line with the description given by Ekman [Ekman, P., Friesen, W. (1978)]. Our technique of analysis could be applied to components only (e.g. eye brow

vertical/horizontal movement, mouth corners, mouth middle), in order to justify the 'componential approach' of facial expressions [Smith, C., Scott, S. H. (1997)].

6.4. Getting better performer data

The low success rate of producing negative expressions has casted (again) the light on how careful and critical one must be when collecting empirical data. We believe that the low success rate was basically due to the 'bad performance' of the subjects. In general it has been known, that it is difficult to produce faithful expressions in 'out of context' experiments. However, it could be identified that some of the 'bad performers', i.e. subjects for whom at most 2 expressions were accepted, had very lazy facial mimick: all the produced 'expressions' were close to each other. Hence no wonder, that the evaluators could not differentiate between the produced expressions.

Ideally, one would like to have facial expressions coded from 'real-life' videos (e.g. ones shown at news reports). As for the time being no powerful enough image processing tools are available, one has to rely on data gained in experimental settings. One can inprove the setting in three ways:

- using 'better performers' (selected and/or trained people);
- using stimuli to induce facial expressions;
- using more performers.

We have repeated the experiments with drama students, who are supposed to be better performers. The similar analysis of these cases is going on, and they will be compared to the cases in this experiment. Though one would probably get higher succes rate with performance, such a special set of people cannot be used to make subtle conclusions about facial expressions in general.

One could try to improve the success rate by using stimuli. However, recordings of expressions as responses to stimuli are still far from 'realistic' expressions, and such settings would require more powerful facial data tracking technology than what we have access to.

Another approach would be to accept the low success rate, make experiments with a huge number of 'ordinary' subjects, and apply some objective criteria to prune the data. Ideally, a joint pool of facial expressions, may be taken in different circumstances and by different image-processing means, but coded in some standard way. At least access to individual databases would facilitate the work of individual research groups, all of them suffering from similar problems concerning data collection [Physta Home Page].

6.5. Getting more data on facial expressions

In our analysis of facial expressions eye gaze and eye lids were not taken into account, due to the limitations of the available tracking system. It is appearant for the casual observer too that for the negative expressions (especially sadness, fear) these characteristics are significant, and may be even decisive. To investigate the role of these factors, we plan two further experiments:

- To process the accepted snapshots of negative emotions by replacing the performed gaze with 'neutral' eyes, and testing the effect n perception of the expressions,
- To extend the parameters with ones on gaze and eye lids, and repeating the classification based on the extended set of parameters.

One should not forget, however, that 'more data' about the face does not necessarily mean significant and useful data from the point of view of classifying facial expressions. One would like to identify exactly those characteristics which differentiate the basic expressions. Our correlation analysis showed that several FAPs were redundant.

6.6. The time factor

We noticed the difficulty of differentiating between snapshots of negative emotions. For these emotions, the process of making the expression may be relevant for recognition. Moreover, when making animations for synthetic heads, it is of vital importance to also know about the dynamical aspects of the expression: how it emerges, how long it is maintained and how it is released. A semi-automatically produced animation will be very unsatisfactory if these issues are not considered.

To be able to analyse the temporal aspects of expressions, we have to find a way of representing time curves of FAPs. Possibilities to do this include stacking the FAP vectors of multiple frames to obtain one big vector describing the curves by some sort of parameterisation. This is the next target of our research.

References

Barlett, M. S., Hager, J. C., Ekman, P., Sejnowski, T. J. (1999)
Measuring facial expressions by computer image analysis, Psychology, 36, pp 253-263.

Donato, G., Bartlett, M. S., Hager, J. C., Ekman, P., Sejnowski, J. (1999)
Calssifying facial actions, IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(10), pp 974-989.

Ekman, P., Friesen, W. (1978)
Facial Action Coding System. Consulting Psychology Press Inc. Palo Alto, California

Essa, I. (1994)
Analysis, Interpretation, and Synthesis of Facial Expressions. PhD thesis, MIT Medial Laboratory, available as MIT Media Lab Perceptual Computing Techreport #272 from http://www-white.media.mit.edu/vismod/

FASE Project Home Page (1998)
http://www.cwi.nl/FASE/Project/

ISO (1998)
Information Technology – Generic coding of audio-visual objects – Part 2: visual, ISO/IEC 14496-2 Final Draft International Standard, Atlantic City

Krzanowski, W. J. (1988)
Principles of multivariate Analysis, a users perpective. Clarendon Press, Oxford

Minsky, M. (1985)
The Society of Mind, Simon and Schuster Inc., New York

Noot, H., Ruttkay, Zs. (2000)
CharToon 2.0 Manual, Report INS-R0004, CWI, Amsterdam

Physta Home Page
http://manolito.image.ece.ntua.gr/physta/, Institute of Communications and Computer Systems, National Technical University of Athens

Pilowsky, I., Katsikitis, M (1993)
The classification of facial emotions: a computer-based taxonomic approach, Journal of Affective Disorders, 30, pp 61-71.

Russell, J. A. (1980)
A circumplex model of affect. Journal of Personality and Social Psychiology, 39(6), pp 1161-1178.

Russell, J. A. (1994)
Is there an universal recognition of emotion from facial expressions? A review of the cross-cultural studies. Psychology Bulletin, 115, pp 102-141.

Schiano, D. J., Ehrlich, S. M., Rahardja, K., Sheridan, K. (2000)
Face to InetrFace: facial

Schlosberg, H. (1952)
The description of facial expressions in terms of two dimensions, Journal of Experimental Psychology, 44, No. 4.

Smith, C., Scott, S. H. (1997)
A componentional approach to the meaning of facial expressions, in: Russell, J. A., Fernandez-Dols, J. M. (eds): The psychology of facial expressions, Cambridge University Press, New York. pp 229-254.

Ten Hagen, P., Noot, H., Ruttkay, Zs. (1999)
CharToon: a system to animate 2D cartoon faces, Short Papers Proceedings of Eurographics'99

Ten Hagen, P. (2000)
The CharToon Repertoire, to appear as INS Report at CWI, Amsterdam

Veenman, C.J., Hendriks, E.A., Reinders, M.J.T. (1998)
A Fast and Robust Point Tracking Algorithm, Proceedings of the Fifth IEEE International Conference on Image Processing, pp. 653-657. Chicago, USA,

Yamada, H., Watari, C., Suenaga, T. (1993)
Dimensions of visual information for categorizing facial expressionbs of emotion, Japanese Psychological Research, 35(4), 172-181.

Yacoob, Y., Davis, L. (1994)
Recognizing Facial Expressions by Spatio-Temporal Analysis, 12th International Conference on Pattern Recognition, Jerusalem, Israel, October, 1994, 747-749.