A Facial Repertoire for Avatars

Zsófia Ruttkay, Jeroen Hendrix, Paul ten Hagen, Alban Lelièvre, Han Noot, Behr de Ruiter
Centre for Mathematics and Computer Science
Kruislaan 413, 1090 GB Amsterdam, The Netherlands
{Zsofia.Ruttkay | Jeroen.Hendrix | Paul.ten.Hagen | Alban.Lelievre | Han.Noot | Behr.de.Ruiter }@cwi.nl

Abstract

Facial expressions are becoming more and more important in today's computer systems with humanoid user interfaces. Avatars have become popular, however their facial communication is usually limited. This is partly due to the fact that many questions, especially on the dynamics of expressions, are still open. Moreover, the few commercial facial animation tools have limited facilities, and are not aimed at lightweight Web applications.

In this article we discuss the empirical basis and a software tool to produce faces with emotional expressions and lip-sync. In order to elicit the characteristics of expressions on real human faces and to map them on synthetic, non-realistic ones, we analysed expressions on real and artist-drawn cartoon faces. We developed CharToon, a software tool that allows the construction of faces and to animate them from scratch or by re-using components of the facial feature and expression repertoire. CharToon faces can be animated real time. The list of applications includes 3D faces of avatars in a VRML environment.

Keywords: Avatar, animation, facial expressions, facial analysis, non-verbal communication.

1 INTRODUCTION

Human faces convey a multitude of information in every-day communication. Some information (e.g. emotional state) can be read mostly, sometimes exclusively, from the face, in other cases the face provides auxiliary input to interpret communication via other channels (e.g. watching the mouth improves understanding speech). Faces serve also as primary basis for identification of people. These aspects are becoming relevant for today's computer systems.
Present day systems share one or more of the following characteristics:

- there are many occasional users;
- there are multiple users who need to be identified and recognisable for each other;
- the application domain is such that there are one or more human participants with a special role (tutor, shop-assistant, game-players).

Avatars have become the solution to provide the user with a humanoid representation of himself, other users and/or system-related assistants. Though an avatar need not look, per se, realistic, it must be capable of communicational modalities that can be easily recognised by the users. Facial expressions (of emotions, of cognitive states, of speech) are of major importance, both to improve efficiency of the interaction and to make the user feel at ease when using a system. However, as humans are very trained and critical with reading real faces, it is a big challenge to endow avatars with 'right' facial expressions.

In this paper we give an account of how we took the challenge. First we specify our motivations and constraints, and give an overview of work on related issues by others. In Chapter 2 we outline what we have learnt by analysing still and dynamical expressions on human faces, and compare those to expressions on drawn, non-realistic faces. In Chapter 3 we introduce CharToon, a facial animation system, and explain the constituents of the facial repertoire serving as a basis to make a big variety of expressive faces. In Chapter 4 we address the issues of exploring the facial expression repertoire and a mechanism to define dynamical facial expressions. Finally, in Chapter 5 we enumerate some application such as putting 2D cartoon faces on 3D avatars, using 2D CharToon faces to convey emotions and to make talking heads. We close the paper by talking about ongoing work and further research issues.

1.1 MOTIVATIONS AND RESEARCH ISSUES

In the recent years, avatars have been a popular research topic. The first results can be seen in commercial avatar-making software packages and a range of applications. (For a query on 'avatar', AltaVista provides more than 160,000 hits, more than 35,000 of the pages dating from this year.) However, concerning facial expressions, avatars are usually poor. (For the query 'avatar' + 'facial expression', there are only 198 hits.)

The use of expressions on avatar faces is hampered by two factors:

- the lack of sufficient knowledge about facial expressions, especially dynamical behaviour;
- the lack of effective methods of creating and presenting expressive faces with appropriate emotional and cognitive behaviour.

In the framework of the FASE project [11], our main goal was to produce expressive synthetic faces to be used in humanoid user interfaces. The applications we envisioned are interactive and web-based. As 3D physically-based facial models [25] [36] are still too slow and complex for such applications, we turned our attention to cartoons, 2D faces which can be controlled easily and fast.

As a first step, we had to know what should be reproduced, when trying to make expressive cartoon faces. As we were interested in making non-realistic 2D faces, we compared the characteristics of 'realistic' expressions to ones on faces drawn by an artist. We also had to check if cartoon faces can convey the same expressions as real human faces.

We needed a handy, easy to use and flexible tool, which, first of all, allowed us to make experimental expressive 2D cartoon faces, and secondly can be used by HCI experts, artists and ordinary users to make expressive faces for different applications. As none of the available facial animation tools fulfilled these expectations, we had to develop one ourselves. The result is CharToon, a platform-independent, Java-based tool.

To design a face and even more, to animate it, is a very difficult (if not impossible) task for an average user. Higher-level ready-to-use (but adaptable) building blocks are needed to help him in making a big variety of expressive faces. As part of the CharToon, we provide a ready-to-use facial feature and expression repertoire. Parallel to the repertoires, we have been developing a declarative framework to design and re-use (static and dynamic) facial expressions.

Once ready-made expressions are (technically) available, the user has to have a tool to explore them. To this end we extended CharToon with Emotion Disc, which represents an emotion space in 2D as a disc, and allows the generation of variations of the 6 basic expressions.

1.2 RELATED WORK

As of analysing facial expressions, the well-known description of the 6 basic static expressions by Ekman [6] has been available. Though initially developed for the psychologists to hand-code facial expressions, it has become popular in software systems. In the community of psychologists, however, there has been criticism [28] of the categorical approach of Ekman. Based on early works by Schlosberg [33], Russell places the 6 basic and many other facial expressions in a 2D space, in a circular form [27]. In his approach emotions are defined by 2 co-ordinates of pleasure and arousal in the continuous expression space, in contrast to the discrete categories of Ekman. In a recent paper [32] not only the (mostly methodological) criticism by Russell has been proven to be incorrect, but it was also shown that the circular arrangement could not be reproduced when visualising the 'perceptual closeness' of the 6 basic emotions in a 2D space, by using multidimensional scaling.

Pilowsky and Katsikitis [26] classified snapshots of 'peaks' of emotions in video recordings of the 6 basic emotions posed by 23 drama students. The result was 5 classes, two of them containing a majority of a single expression, namely happiness and surprise. 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 colleagues [42] did investigations similar to ours: they used canonical discriminant analysis to visualise 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 corners of the mouth and the third one for the position of the eyelids and eye corners.

Essa [8] used naive performers to pose the 6 basic emotions 'out of context'. He reported, similarly to Yacoob and Davis [43], that subjects had difficulty with producing fear and sadness, 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 surprisingly, anger and smile too. For the latter observation he referred to Minsky [22] claiming that in the case of these expressions which have similar snapshots at the peak, the time behaviour is an important differentiating factor. He himself produced time functions of muscle actuations and made qualitative observation on the profile, such as the existence of a ‘second peak’ in the relax phase for smile.

There has been earlier work on **2D cartoon face** [3][31][37] and general **light-weight 2D animation systems** [12][20][21][24][38]. The first two facial animation systems do not allow the design of dynamical expressions: in [3] cartoon faces can be animated by image morphing, in ComicChat [31] stills are used. Our system is especially equipped for facial animation, and in this application field is superior to the listed general ones, serving a wider domain of 2D animations. Comparing CharToon to Inkwell [20], there are similarities in the main objectives (light and easy to use, flexible animation system) and the technical solutions (exploiting layers, allowing the manipulation of motion functions, grouping/hierarchy of components). While Inkwell has several nice features which CharToon lacks, CharToon offers extras which are especially useful for facial animation: special skeleton-driven components and an extensive set of building blocks to design faces; the support to re-use components and pieces of animations, a separate graphical editor to design and manipulate animations and real-time performance. Similar arguments hold for MoHo [21], a recent general, light and vector-based 2D animation system. While skeleton-based motion (with inverse kinematics) is supported in MoHo, it is not possible to manipulate time-curves of parameters. Also, there is no player to generate real-time animation from ASCII files.

Editing animations with CharToon can be seen as extension to **parametric keyframing**, supported by all commercial animation packages. In CharToon, editing operations are allowed on pieces of parameter curves. Moreover, CharToon is being extended with constraint mechanisms, which will provide a basis for manipulating animations on a higher level and in a descriptive way.

Current commercial facial animation packages all assume a 3D facial model, which can be animated either by re-using a set of predefined expressions without the possibility of fine-tuning them [9], or by tracking the facial motion of a performer [10]. In the latter case, the editing operations are performed as Bezier curve operations.

This also applies for most of the general **motion warping** [41] and signal processing based motion curve transformations [4] techniques. An exception is the work on **constraint-based motion adaptation** [13], which uses the combination of motion signal processing methods and constraint-based direct manipulation, in order to be able to modify an existing motion to meet certain requirements. There is a lot of literature of **motion synthesis** and **motion control** systems based on some general constraints and principles of (realistic) physical motion [15][18]. CharToon is more general in the sense that any object, with non-realistic dynamical characteristics can be animated.

From the technical point of view, by using vector-based graphics to achieve real-time performance and possibilities for Web applications, CharToon is in line with the current research in the W3C to incorporate real-time vector-based animation into Web pages [35].

2 ANALYSIS OF FACIAL EXPRESSIONS

In order to provide the right tools and repertoire elements for facial animation, we had to gain insight in the following issues:

- What are the generic and specific characteristics of expressions on human faces?
- What are the dynamical properties of expressions?
- How do non-realistic cartoon faces convey expressions?

In this chapter we give an account on our (partly still ongoing) empirical investigations on the above questions. The expressions (both on real human faces and on drawn cartoon ones) were expressed as MPEG FAPs [16]. Further on we refer to the multidimensional space of the parameters as the expression space.

We have analysed three bodies of data:

- Stills of tracked facial data.
- Stills of facial cartoon data.
- Time curves of tracked facial data.

Tracked data was gained by the point tracking system developed by our partner in the project at the Technical University Delft [40], while the cartoon drawings were produced by an experienced animator in out team [29].

2.1 ANALYSIS OF TRACKED EXPRESSIONS

2.1.1 Collecting data

The 18 subjects, 17 males and 1 female, were asked in the framework of individual sessions to make the 6 basic expressions (smile - surprise - anger - disgust - fear - sadness), each twice [14]. Blue markers on their faces were tracked, producing time curves of the 15 FAPs for each person (see Figure 1).

We took 'the most extreme' snapshot for each expression, by choosing the snapshot at the 'peak' of most of the curves.

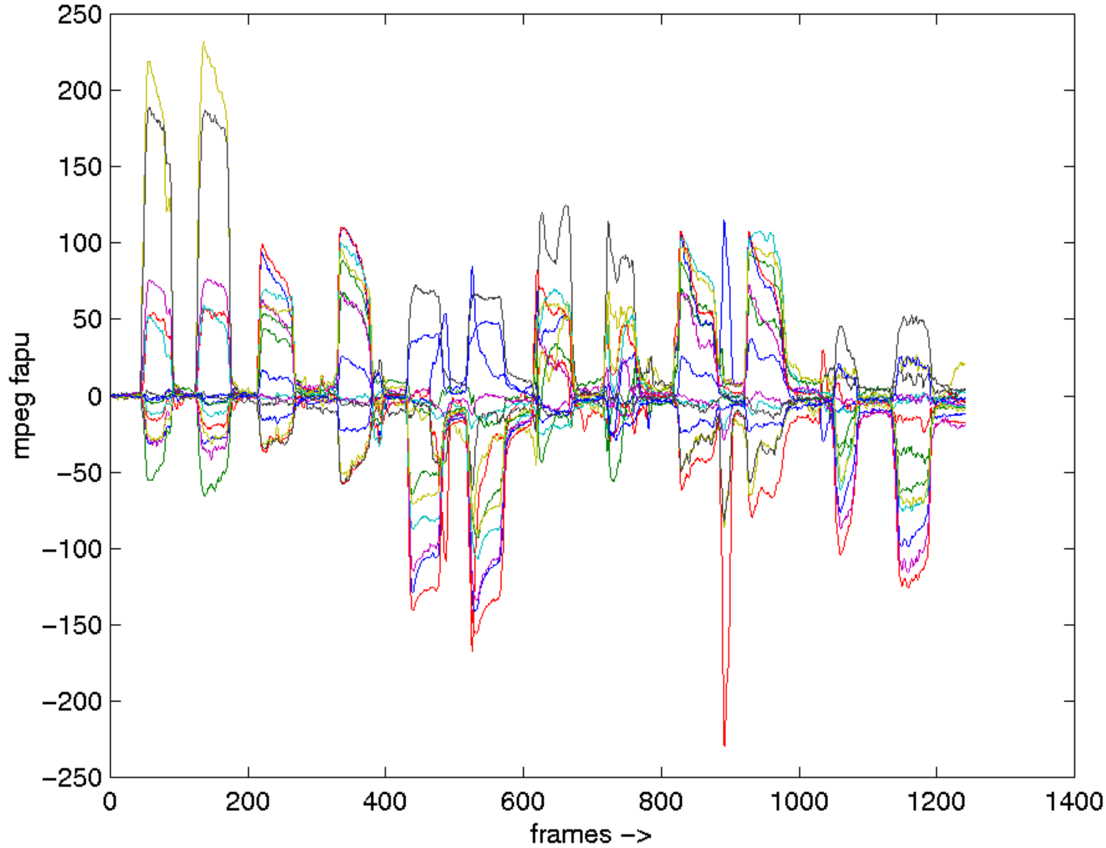


***Figure 1**: The time curves of a recording session.*

Not all performers succeeded in producing all six expressions. In order to prune 'erroneous' recordings, we asked 56 volunteer colleagues at our institute to re-label the 108 snapshots. A performed expression was **good** 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 perceiving it as an expression different from 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.

Each recorded expression was represented in the data set by a vector of 15 normalised FAP values. By normalising the data we expressed displacements relative to the extremes of a person. Our data defined points in a box of the 15 dimensional so-called **expression space**.

2.1.2 Principle component analysis of the data

We performed Principle Component Analysis (PCA) [19] for the ALL data set. 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 three components the coefficients are given in Table 1.

As we can see in Table 1, the first two components make up for about 73% of the total variation. 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.

FAP variables	component 1 47.6%	component 2 26.3%	component 3 9.41%
3 open mouth	-0.17	0.22	0.63
4 lower middle upper lip	-0.20	0.25	-0.31
5 raise middle lower lip	0.17	-0.26	-0.59
6 raise left corner point mouth	0.12	-0.34	-0.04
7 raise right corner point mouth	0.14	-0.39	0.13
12 stretch left corner point mouth	0.15	-0.41	0.19
13 stretch right corner point mouth	0.15	-0.41	0.20
31 raise left inner eyebrow	-0.34	-0.12	-0.04
32 raise right inner eyebrow	-0.34	-0.09	-0.02
33 raise left middle eyebrow	-0.35	-0.14	-0.05
34 raise right middle eyebrow	-0.35	-0.14	-0.02
35 raise left outer eyebrow	-0.35	-0.07	-0.13
36 raise right outer eyebrow	-0.35	-0.10	-0.06
37 squeeze left eyebrow	0.22	0.30	-0.19
38 squeeze right eyebrow	0.23	0.23	0.02

***Table 1:** The first three principal components expressed in terms of the original FAPs*

In Figure 2 the first two PCA components of the ACC data set 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 first two components of the tracked FAPs are not sufficient to differentiate negative emotions.

The surprises are divided amongst two sub-clusters: 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.

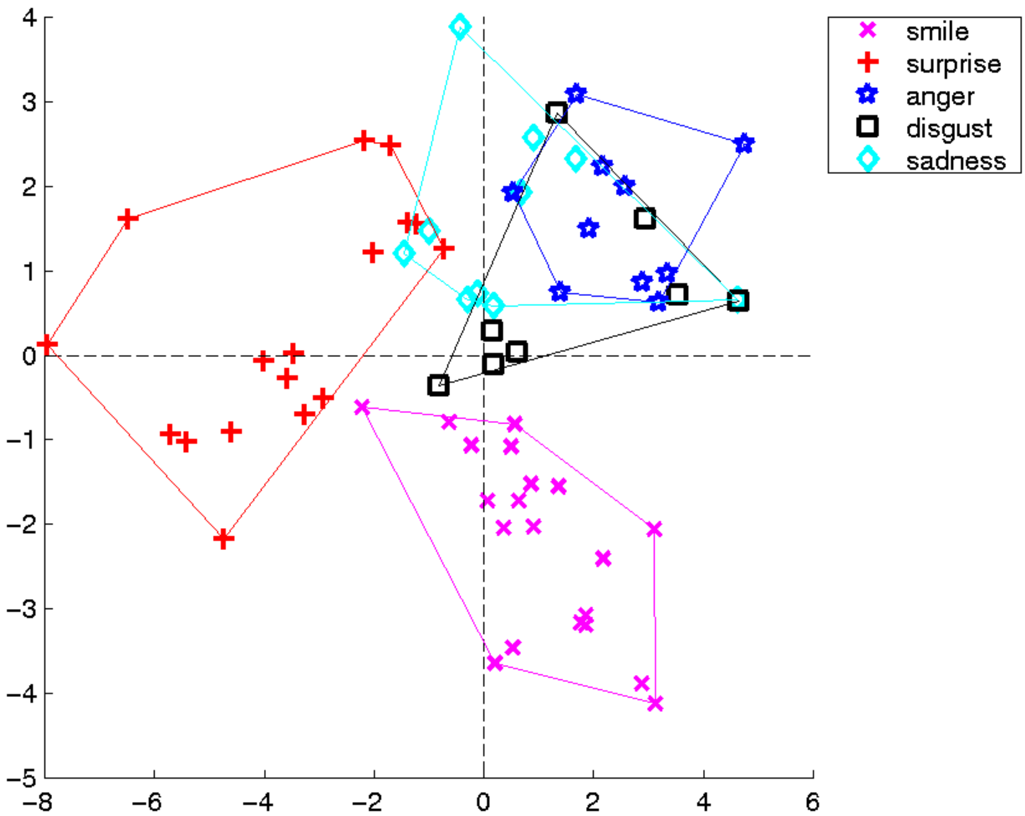


Figure 2: PCA on the ACC data set.

2.1.3 Further analysis

Correlation in the data

We statistically analysed our ALL data set, to gain some insight into the correlation between the different points on a face. All vertical displacements of points on the eyebrows are strongly correlated (>0.82). The vertical displacements of the corner points of the mouth are heavily correlated (0.96). By using a single representative of the correlated FAPs, we would not have lost relevant information about the expressions.

Canonical variate analysis

When applying PCA, we completely disregarded the fact that we are already aware of a certain structure in the data set. 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 dissimilarities between expressions into a few-dimensional picture.

Canonical variate analysis [19] can be used for this purpose. Applying this technique yields Figure 3.

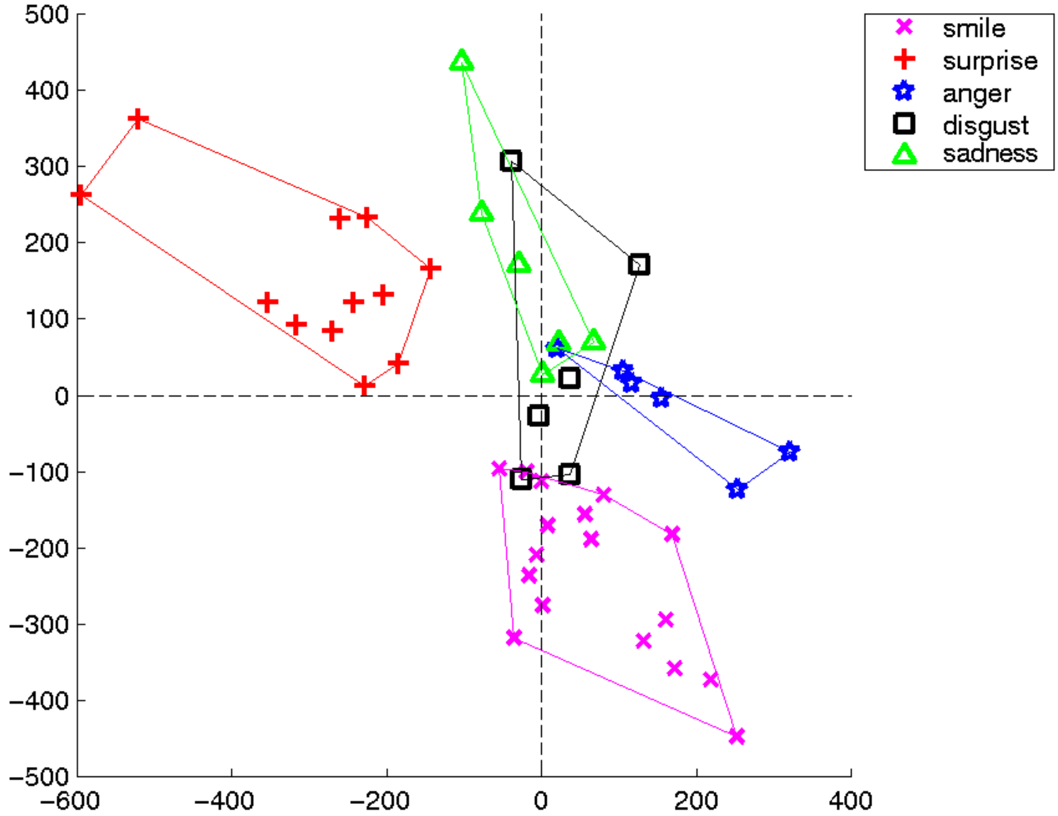


Figure 3*: CVA on the GOOD data set.*

A small improvement to the PCA picture can be found, namely that the clusters of anger and sadness are more distinguishable. 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 characterise, 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'.

2.1.4 Conclusions

The characteristics of the expressions, expressed in FAP parameters, are in line with the description given by Ekman [6]. Our technique of analysis could be applied to these main features only (e.g. eyebrow vertical/horizontal movement, mouth corners, mouth middle), in order to justify the 'componential approach' of facial expressions [34].

As we expected, the subset of FAPs that was made available by the face tracker, was not always sufficient to distinguish two emotions (especially the negative ones) from each other. Two different emotions sometimes were very close in terms of the distance of their FAP vectors, 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 [42].

2.2 Analysis of Expressions on CharToon Faces

2.2.1 Collecting data

An experienced animator, educated in drawing expressive faces, produced 59 expression stills. He made his designs in such a way that they were applicable to cartoon faces constructed from elements of the facial repertoire (see Chapter 3.2). The animator himself categorised the expressions either as variants of the six basic expressions or as belonging to a group of 'other' expressions.

We treated this data set in the same way as the tracked data. We used the MPEG coding of the expressions applied to a face made of components with the same degrees of freedom as the tracked data has, i.e. the same FAPs. Because the animator also supplied gaze and eye-openness, we also investigated whether these contributed to the characterisation of the basic expressions.

2.2.2 Principal component analysis of the data

We performed principal component analysis exactly the same way as on the tracked data. The main difference in the result was that the expression space of the CharToon expressions is of higher dimensionality (see Table 2). This is not surprising, as an animator can do much more with a cartoon face than most people can do with their own face. E.g.: for most persons raising only one eyebrow can be very hard, but not for a cartoon face.

FAP variables	Comp 1	Comp 2	Comp 3	Comp 4	Comp 5
	41.1%	18.3%	16.8%	8.6%	5.1%
4 lower middle upper lip	0.19075	-0.23138	0.15325	-0.46144	0.34207
5 raise middle lower lip	0.33587	0.13540	0.02133	-0.24099	0.16361
6 raise left corner point mouth	-0.11463	-0.33613	-0.32187	0.18391	0.59554
7 raise right corner point mouth	-0.03057	-0.44140	-0.36246	0.06872	0.24405
12 stretch left corner point mouth	-0.08300	0.09852	-0.56841	-0.19100	-0.24003
13 stretch right corner point mouth	-0.05897	0.07415	-0.58285	-0.23294	-0.26824
31 raise left inner eyebrow	-0.25529	-0.30880	0.16546	-0.46024	-0.17439
32 raise right inner eyebrow	-0.27385	-0.27600	0.16553	-0.42159	-0.07878
33 raise left middle eyebrow	-0.38813	-0.09145	0.03220	-0.00472	-0.12113
34 raise right middle eyebrow	-0.39951	-0.01192	0.07258	0.08047	0.08511
35 raise left outer eyebrow	-0.37324	0.01444	0.06658	0.16504	-0.06684
36 raise right outer eyebrow	-0.38028	0.09061	0.08733	0.22483	0.09709
37 squeeze left eyebrow	-0.21955	0.44916	-0.06142	-0.28107	0.30800
38 squeeze right eyebrow	0.20930	-0.46358	0.03686	0.21649	-0.38095

Table 2: *Components of PCA on CharToon, basic expressions.*

Table 2 displays the first 5 components of the PCA analysis done on all basic CharToon expressions. Note that component 3 and even component 4 contribute significantly to the total variation in the CharToon data set.

If we assign meaning to the most significant components, it can be something like:

Component	'Meaning'
1	Lowering of the eyebrows
2	Stretching the mouth, raising the upper lip
3	Raising the corner points of the mouth, asymmetric squeezing of the brows
4	Among others, raising the inner brows

Many principal components are asymmetric with respect to the face. Many facial features are part of multiple components. This makes it harder to unambiguously interpret the essential components.

For the graphical visualisation of the PCA data, plotting only the first two components yields Figure 4, losing the information in component 3 (dealing mainly with the smiling shape of the mouth!).

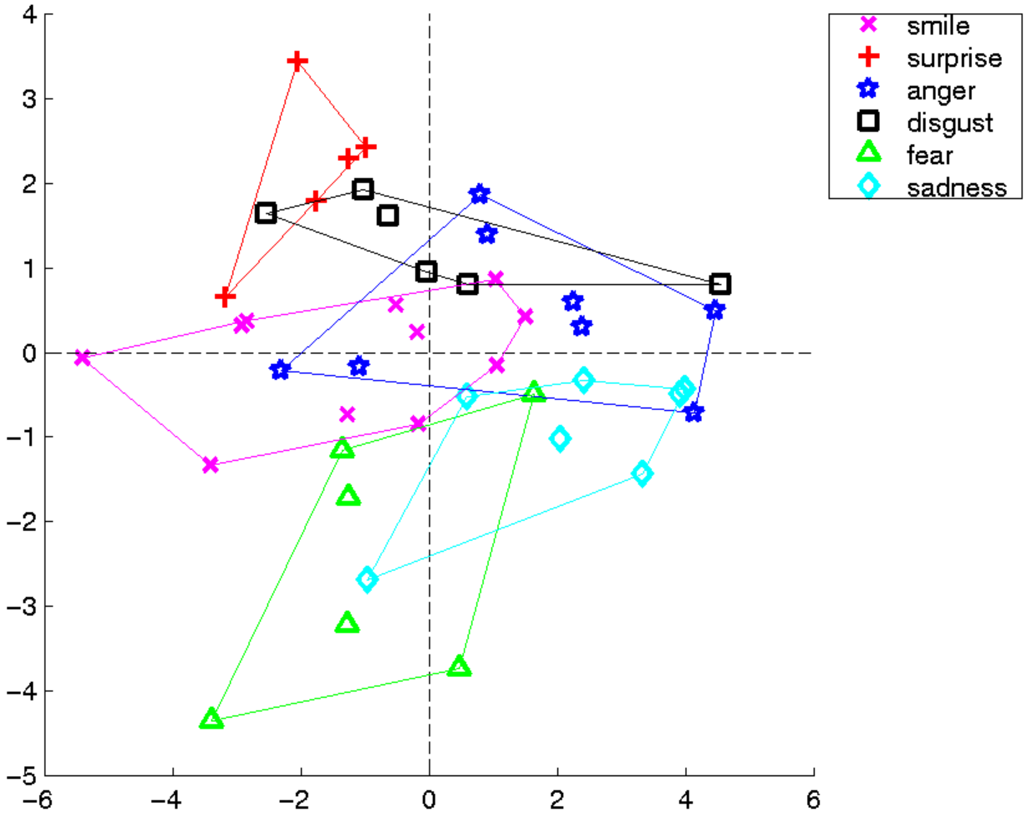


Figure 4: PCA on CharToon expressions.

2.2.3 Further Analysis

In the PCA graph the variation on the complete set of points is maximised. Again, PCA is not the most suitable tool to find variations between groups, CVA is. As the structure of the expression space is quite complex, applying CVA to all groups at once is not optimal. We applied CVA on each separate pair of expression groups (smile versus anger, fear versus sadness etc.). This way we were able to find separate clusters for almost every pair of expressions.

We tried using the information supplied by the animator about the eyes to achieve even better results, but this did not contribute much. The animator used the eyes mainly to make different expressions within groups (cf. anger-annoyed: eyes half-closed and anger-furious: eyes wide open), and hardly to distinguish between groups.

We plotted the CharToon expressions into the PCA graph of the convex hulls of the tracked data (see Figure 5). As we can see, there is not much difference between the CharToon smiles and the tracked smiles. The CharToon 'Smiling-for-the-camera' can be seen as an extreme case of the tracked smiles.

The CharToon sadness is close to the origin of the axes, this is because the artist mainly used the eyes to convey sadness, which isn't visible in this graph.

Each of the artist-drawn versions of the basic negative expressions is very near or inside the convex hull of the tracked versions.

It is interesting to see that the artist-drawn surprise is outside of the convex hull of the tracked data. As surprise was easy to produce for the performers, here we have an example of an expression that has different characteristics on an artist-drawn cartoon face than on real human faces.

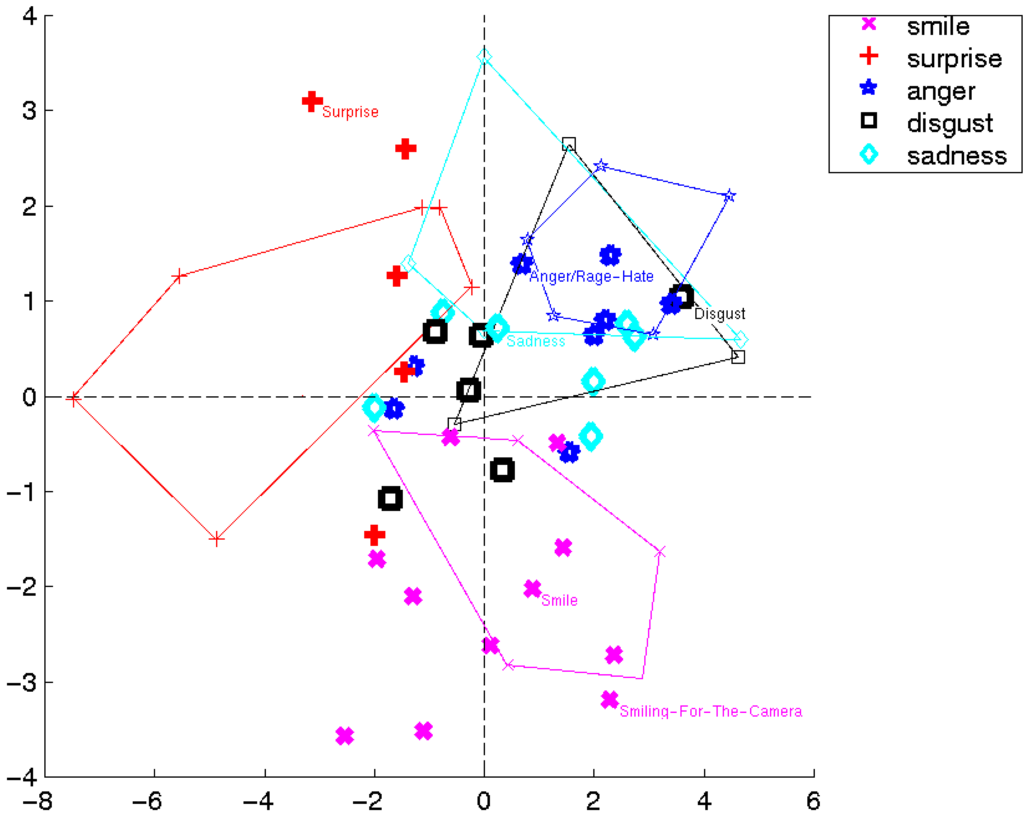


Figure 5: *CharToon expressions plotted in PCA graph of ACC data.*

2.2.4 Conclusion

As mentioned already, the artists used the facial features in a more varied way to exhibit expressions, than the human subjects did.

Are the CharToon expressions close to the tracked ones? An indication to an answer to that can be found by searching for each of the CharToon expressions for the closest tracked expression. 18 out of the 38 CharToon expressions (not counting ones in categories 'other' and 'fear') had a neighbouring tracked expression of the same kind.

2.3 ANALYSIS OF DYNAMICS OF EXPRESSIONS

As mentioned before, having information about temporal aspects of facial expressions is highly valuable. We aim at several goals:

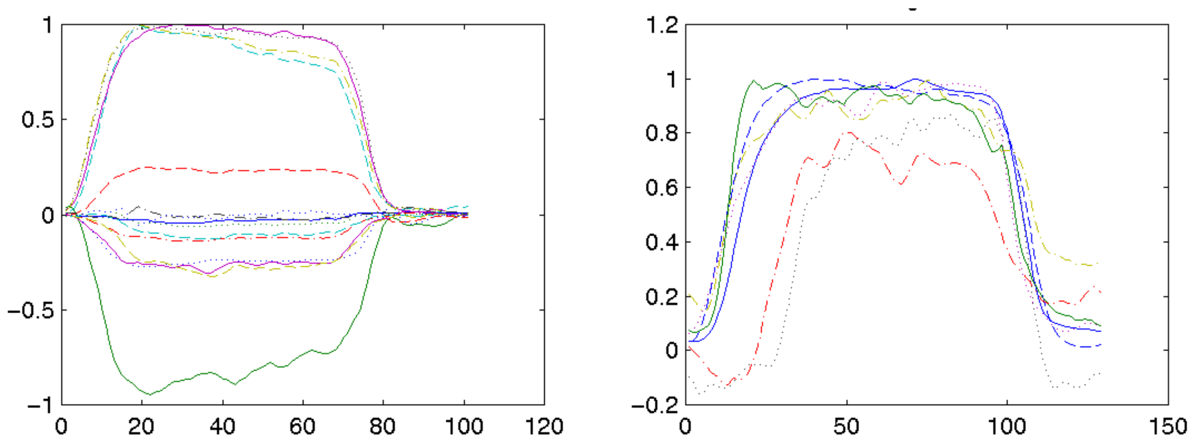
1. Finding constraints describing expressions, describing time curves of FAPs that are e.g. a 'smile'.
2. Finding a generic set of time curves of FAPs for every expression, e.g. the 'generic' smile.

And, as an interesting by-product, not strongly related to our project:

3. Finding a way to recognise expressions from time curves of FAPs.

Currently we are busy with answering these questions, based on tracked data. We separated the expressions from each recording and scaled all time curves of the expressions to a common length. Each expression is represented as a graph of time lines for the FAP values (see Figure 6).

We use wavelet decomposition as a tool to characterise individual time curves.



a) b)

Figure 6*: Time curves for all FAPs in smile (a), Time curves for 'raise mouth corner' FAP in all smiles (b).*

2.3.1 Using wavelets to characterise individual FAP curves

We currently are applying the discrete wavelet transform [17] to each individual time curve. This gives a time-frequency spectrum of each curve, a first raw characterisation of what the curve looks like.

When choosing for wavelet analysis, one also has to select what wavelet base to use. This highly determines which characteristics of the curve are visible in the wavelet coefficients. The characteristics interesting for our purpose include:

- Duration of the three stages (application, sustain, release) of actuation of the expression (finding the lowest frequency).
- Steepness of ascend and descend of the activation part (finding the main frequency or frequencies at the start and end time).
- Smoothness (finding the highest frequency).
- Overall shape (presence and location of appropriately chosen frequencies).

For all this, some not negligible, post-processing has to be done on the wavelet output to really find these characteristics. This is currently being investigated upon.

Up to now, we have obtained results matching visual observations using a Coiflet of order 1 to find the starting and ending points and the duration of an expression curve. Further analysis has yet to be conducted.

2.3.2 Using the characterisations to analyse complete expressions

As soon as we have a detailed wavelet characterisation of all FAP curves, we can use these to achieve the goals mentioned before.

Using all characterisations of one FAP of all sequences in the data set, we can extract one generic curve, or perhaps a few, by averaging the curves that are sufficiently similar. Doing this for all FAPs will give a generic expression.

Calculating the extremes for the characteristics for each FAP at every expression will give boundaries to the deformation of the base expressions. We will also analyse the characteristics of FAP time curves of a single expression, to elicit constraints on critical characteristics (like co-articulation) of several FAPs.

When confronted with a new set of FAP curves, one can calculate the characteristics of it and compare them with the constraints on each expression to find out which expression matches the new set. Thus being able to recognise it.

3 FACIAL ANIMATION WITH CHARTOON

3.1 THE CHARTOON SYSTEM

CharToon is a system we have developed to design $2\frac{1}{2}$D faces (and other objects) that can be animated, to compose animations for such faces, and to play animations. The corresponding components of the system are Face Editor, Animation Editor and Face Player (see Figure 7).

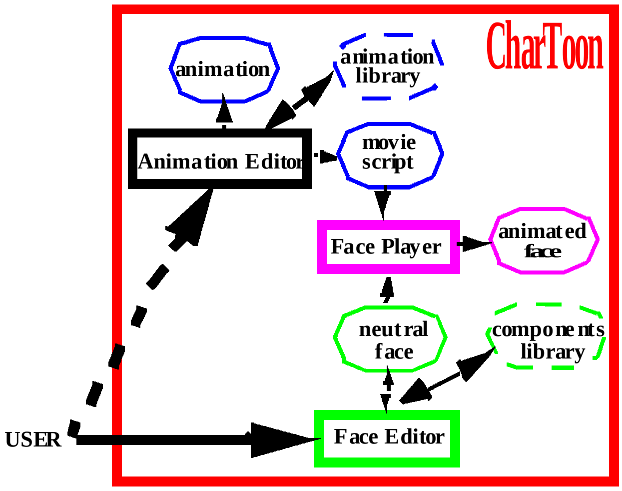


Figure 7: The components of CharToon.

Face Editor is a 2½D drawing program with which one can define the structure, the geometry, the colours and the potential motions of the face. A face is built up of a layered arrangement of vector-graphics components. Different components may be animated in different ways by changing the location of so-called control points (see Figure 8). A collection of extensible building blocks facilitates the construction of faces.

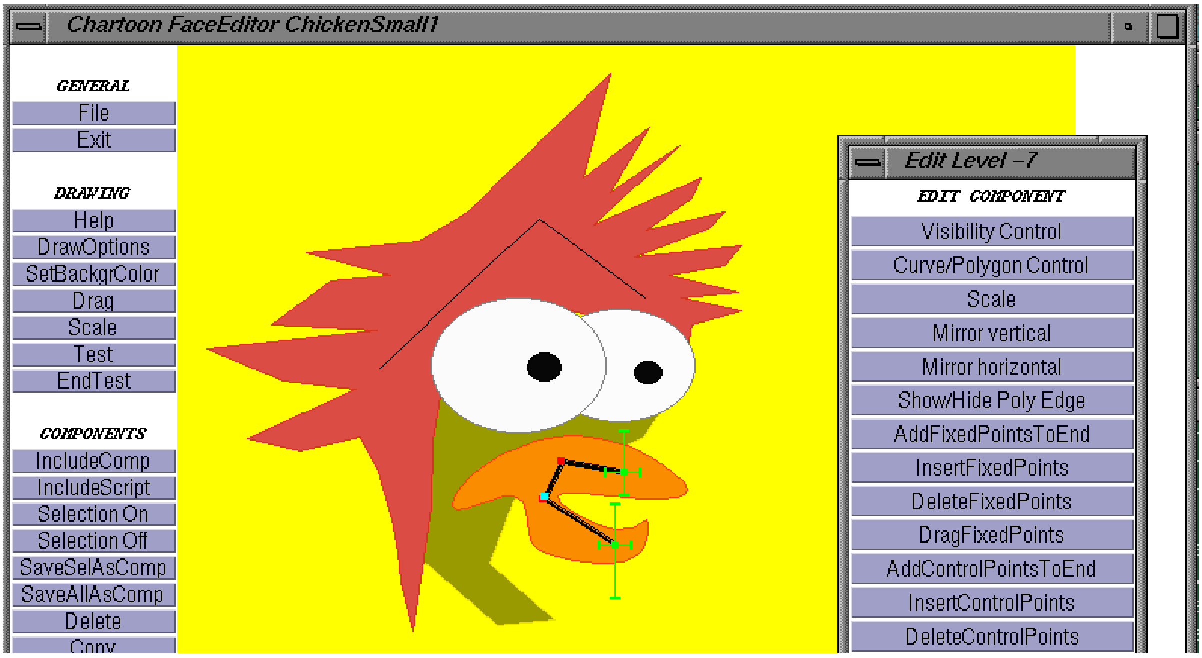


Figure 8: *The Face Editor window. The beak is selected, and its skeleton is shown with its two control points.*

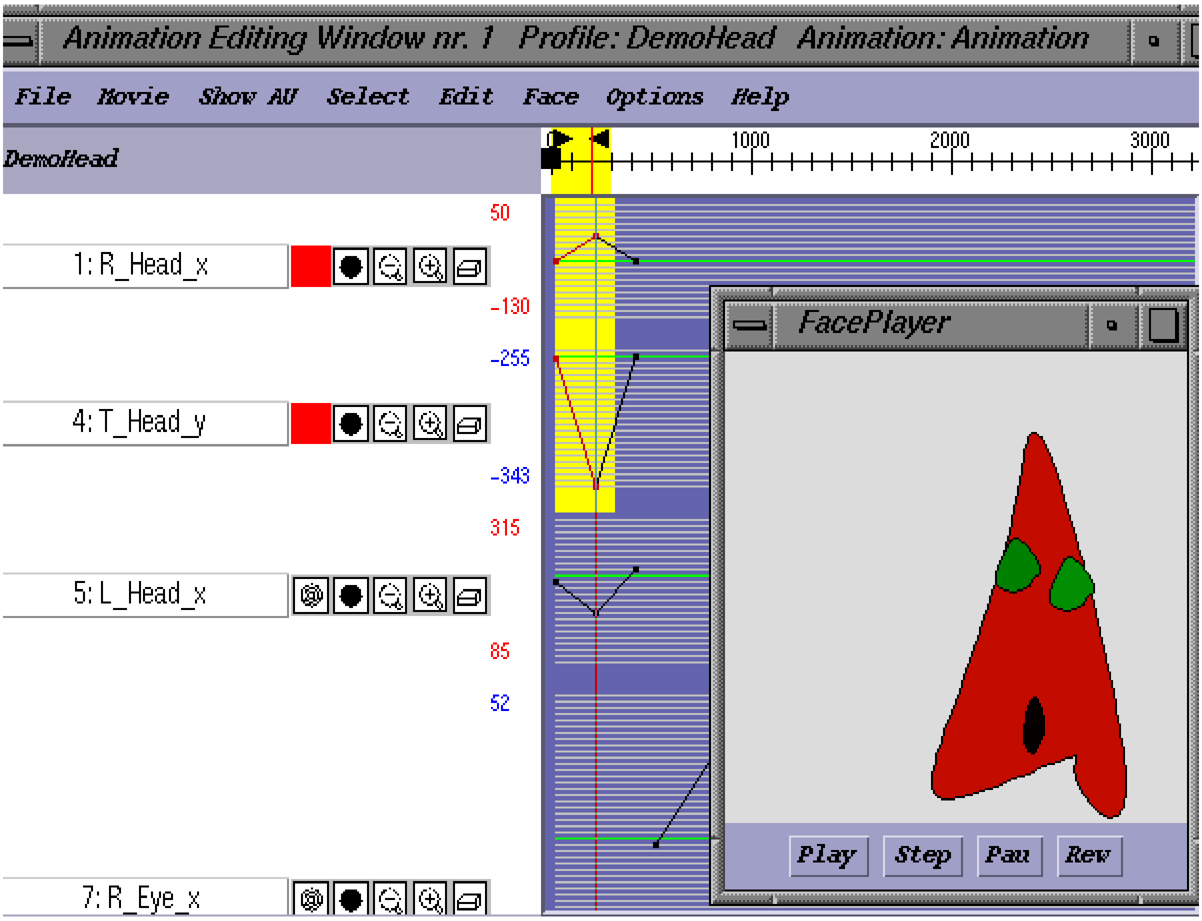


Figure 9: *The components of CharToon: Animation Editor, with a FacePlayer window showing the expression corresponding to the cursor position.*

Animation Editor is an interactive 'animation composing' program, to define the time-behaviour of a drawing's animation parameters, provided by Face Editor (see Figure 9). Animations can be saved as a script (for later re-use). **Face Player** actually generates the frames of an animation, on the basis of the animation parameter values in the movie script file provided by Animation Editor and the face description file provided by Face Editor. When playing a script-driven animation, it is possible to generate image dumps to make movies of them later by using commercial movie making software, or to generate Flash output.

A complete technical description of the CharToon system can be found in [23]. The programs exchange data with each other and possibly with other applications via ASCII files. All components are written in Java, which makes Web-based applications possible. (See [5] for an applet demo.)

CharToon separates the **appearance**, the **dynamism** (possible deformations) and the **behaviour** of a face. The first two aspects are incorporated in the definition of the face, while the latter is defined as an animation. CharToon technically supports the re-use of facial components and pieces of animations as building blocks.

3.2 Facial Features

Based on careful analysis of deformation of specific facial features of the basic expressions – happiness, surprise, fear, sadness, anger and disgust –, for each feature (eye, mouth, eyebrow...) different alternative designs were produced, forming together the **facial feature repertoire**. One can easily compile a face by selecting an element for each feature from the repertoire. Each element of it is editable, that is, the creator of a character can adjust any component included. The user has full freedom concerning the appearance and dynamism of the face being created.

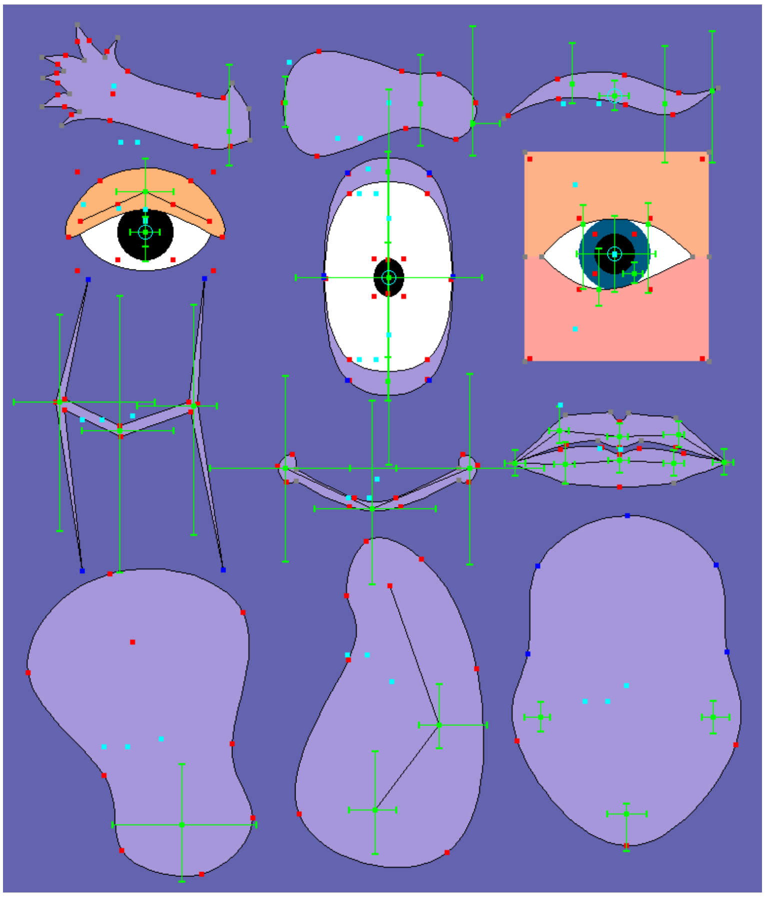


Figure 10: *Facial feature repertoire elements.*

3.3 FACIAL EXPRESSIONS AND VISEMES

For each feature, the deformations resulting in a certain set of expressions were given (in terms of animation parameters), forming the **expression repertoire**. As part of CharToon, 59 expressions are provided, containing the widely used 6 basic expressions and subtle variants of them. An expression from the repertoire can be applied to any face constructed from elements of the facial feature repertoire.

A **viseme repertoire** consists of mouth shapes (for each mouth element in the facial feature repertoire), defined as snapshots. The viseme repertoire which is supplied with CharToon is the so-called Extended English Visemes, consisting of 47 visemes appropriate for lip-sync for English.

3.4 LEVELS OF QUALITY

The alternatives available in the repertoire for a facial feature differ concerning deformation control mechanism and/or structure. E.g. the functionally simplest eyebrows are the ones which do not change shape but may be moved up/down, and the most complex ones have 4 control points, with which one can produce subtly deformed eyebrow shapes. In general, one can use the repertoire on four levels of quality: High, Medium, Low and Primitive. Higher level quality is computationally demanding and requires more effort from the designer to deal with all the details, but produces very expressive, subtle faces. Hence the level of quality should be chosen according to required expressiveness, expected operation environment, the technical circumstances and designer's expertise. In Figure 11 four faces of different levels of quality exhibit different expressions from the expression repertoire. All the faces were made of facial repertoire elements. In case of the female face, the used elements from the feature repertoire were adapted.

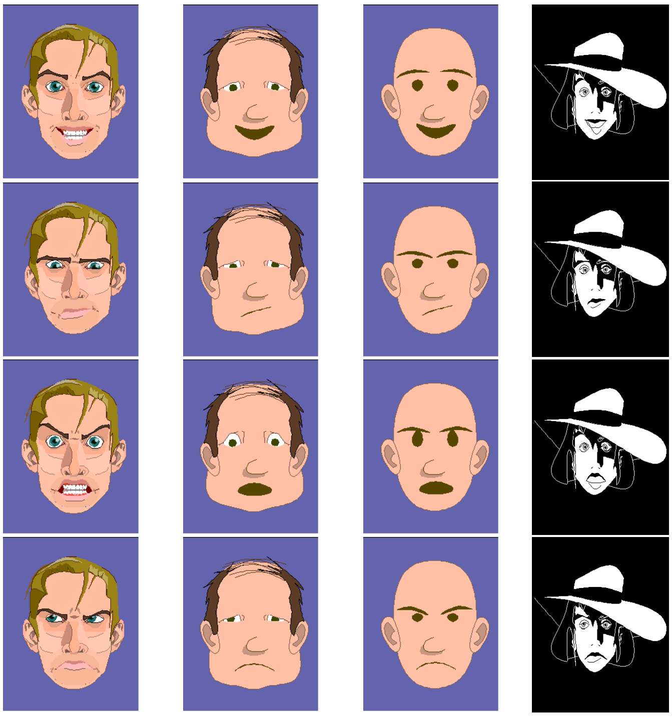


Figure 11: *Four faces made up from facial feature repertoire elements, showing the same expression in each row.*

4 EXPLORING THE REPERTOIRE

4.1 EMOTION DISC

The emotion disc (see Figure 12) allows the user to explore the space of emotions, assuming that the 6 basic expressions are defined (as snapshots) for the face in question. The elements of the emotion space are generated by blending two of the given emotions in a certain way. The space is mapped on a disc, which serves as a handy user interface. The emotion disc is based on the following properties:

- Each facial component has, in addition to its basic neutral shape, information defining the shape variations corresponding to the six basic emotional expressions for joy, surprise, fear, sadness, anger and disgust.
- According to Schlosberg [5], the six basic emotional expressions for joy, surprise, fear, sadness, anger and disgust are perceptually related in such away that they can be arranged in a two dimensional space as a visual continuum. The space is arranged as a round disc showing a neutral face in the centre and a maximal expression on the perimeter. Each position in the so-called emotion disc corresponds to an expression obtained by interpolation between the known expressions positioned on the disc.

The emotion disc can be used in all stages of the animators' work, to judge the expressiveness of the character, or to support expression transition planning or key frame selection. It can be used as a direct controller of the expression of an avatar's face. Though the empirical evidence for the arrangement of the expressions has been debated, the device has proven to be very useful and popular among users.

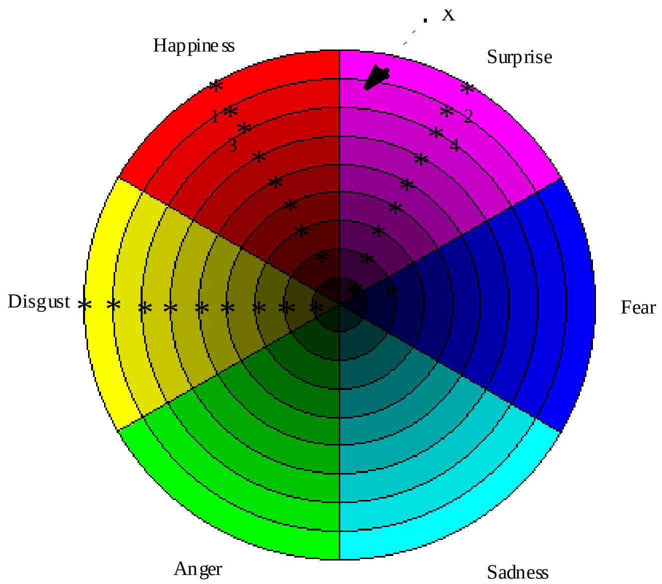


Figure 12: *An annotated Emotion Disc example, showing positions of expression samples and possible assignments of emotions. Expression sample positions are indicated by asterisks in the Surprise, Happiness and Disgust segments. There are analogous sample positions in the other sectors, which are not shown.*

4.2 SCULPTURING FACIAL EXPRESSIONS

A designer needs high-level building blocks that can be reused and adapted when compiling new animations for a face. The facial expression repertoire as a collection of animations allows only the low, control-parameter-level re-use and modification of pieces of animations. Higher-level building blocks should be defined in terms of general characteristics of the expression (e.g. symmetrical motion of the face, synchrony of the motion of certain features, duration). Moreover, the designer should have tools to modify these characteristics within allowed limits as well as to add and modify requirements when editing an animation.

In a new version of CharToon, the above facilities are available [30]. The underlying mechanism is the manipulation of **interval constraints** [2]. The characteristics of high-level building blocks are expressed in terms of constraints. E.g. in case of a smile, both mouth-corners should be pulled up for some time, and then after a short while the expression should be released. The durations and final location of the mouth corners are not set to a specific value, but some limits are prescribed. Moreover, if one wishes to have a perfectly symmetrical smile, the motion of the two mouth corners should be perfectly 'mirrored'. Otherwise some degree of asynchrony is allowed.

When making an animation, the actual parameter values must satisfy the prescribed constraints. The extensions and refinement of the high-level building blocks is done by adding new sets of constraints as the definition of a new expression, and by tightening constraints of existing building blocks. When working on an animation, the user may prescribe further constraints, expressing requirements to be met for the given animation. In this way the constraint-based animation editing tool has two usages:

- to sculpture the mimic and expression repertoire of a face to be animated;
- to make animations for a face with a given mimic repertoire, meeting certain further requirements set for the particular animation.

These characteristics will be automatically enforced, and predefined building blocks (e.g. a smile) can be re-used in the course of editing an animation for the face. Moreover several, non-identical expressions of the same kind can be generated, avoiding the unpleasant effect of using identical pieces of animations whenever an expression is to be produced.

The underlying constraint mechanism requires appropriate extension of the user interface of Animation Editor too. The new version has been implemented and being tested. In the first implementation, 7 types of constraints can be used [30]. For constraint handling, the OpaC solver [1] is used. Figure 13 gives an impression of the new user interface, adapted for constraint visualisation and handling.

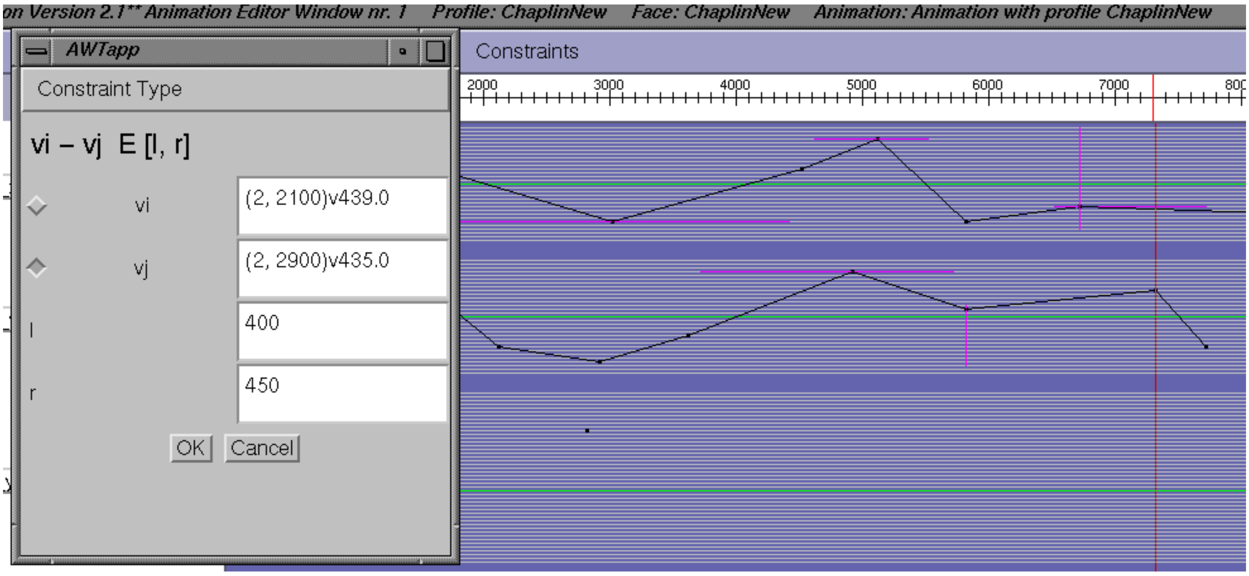


Figure 13: *Snapshot of the new Animation Editor with constraint handling facilities..*

5 APPLICATIONS

5.1 3D AVATARS WITH FACIAL EXPRESSIONS

CharToon has been used to supply avatars with facial expressions in the RIF virtual environment [7], where avatar-embodied agents offer help to users. In order to have a lightweight solution to generate the facial expressions, the 3D facial features were arranged on a plane, which was placed in front of the simple, cylindrical head of the avatar. Because of the small size and the simple facial features, this solution gave a satisfactory appearance. Moreover, the expression repertoire and control mechanism of CharToon could be used in a straightforward way. The Emotion Disc was re-implemented in VRML. The user could set the facial expression by using the disc (see Figure 14). Also, automatic facial expression generation was supported, by exploiting the Blaxxun Technology's facility of linking gestures to textual phases.

5.2 2D FACES IN WEB APPLICATIONS

CharToon has been developed to make expressive cartoon-like faces. One of the motivations for this choice was technical: we wanted to have lightweight synthetic faces that can be animated real-time, with a special eye on potential interactive web-based application. A natural and essential question is whether simplified, cartoon faces can convey expressions in similar detail and clarity as they appear on real human faces or realistic 3D models. To answer this question, a test has been conducted. Human ergonomists have tested the expressive effect of CharToon faces [39], and found that the experimental subjects could recognise as well as reconstruct emotions on different non-realistic faces. Hence the choice for cartoon faces does not constraint the expressiveness. On the other hand, the fact that the user is confronted with a face, which does not pretend to look real, makes him to adjust his expectations. Instead of being frustrated by imperfections of 3D realistic faces, he can enjoy the special aesthetics and extra expression methods common to the world of cartoon characters.

CharToon makes web-based applications possible, either by using a Face Player applet, or by producing output in web-confirmed formats like Flash. Currently we are investigating the possibility to an interface of CharToon to the SVG format [35], a recent recommendation by W3C for vector graphics on the web.

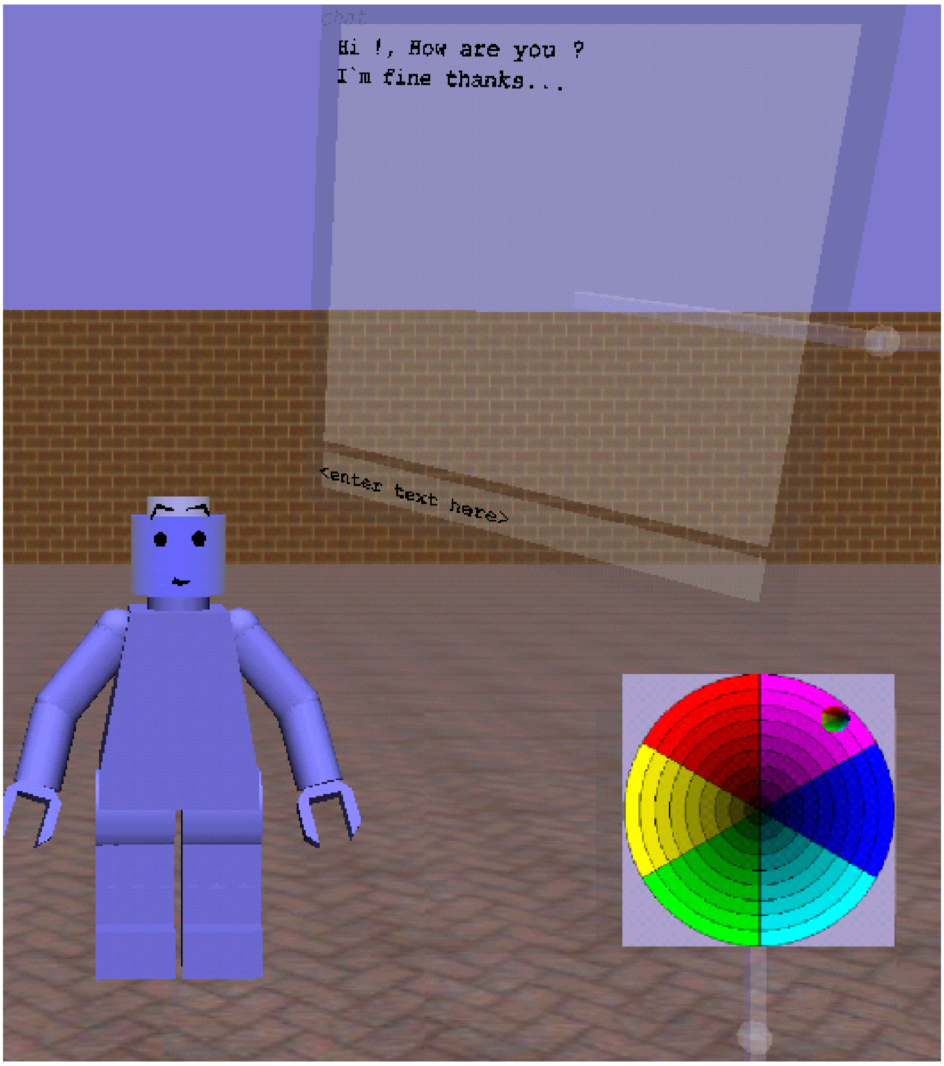


Figure 14: *An avatar in a VRML space. The facial expression has been generated according to typed text the avatar is supposed to say. The user can set the expression on the face with the Emotion Disc shown in the right corner.*

5.3 TALKING HEADS

By using a viseme repertoire, one can produce lip-sync with CharToon. It is possible to generate the mouth movements automatically, assuming that a script of the viseme sequence to be shown is available. Face Player can play the generated talking head with the corresponding audio.

By using different viseme repertoires, one can easily generate lip-sync of the same head for different languages or different types of users, such as the hearing impaired.

6 FUTURE WORK

The concept of repertoire has proven to be very useful for UI designers as well as for novice users. In the near future, we will extend the repertoire in two ways. On the one hand, we will provide further sets of facial elements (e.g. for woman and child faces). On the other hand, we will extend the expression repertoire with cognitive and communicational expressions.

It requires further investigations to find out which part of the expression space corresponds to 'meaningful' (realistic or cartoon) expressions. With more subtle analysis of changes of expressions, one could get a picture about the 'transition paths' between expressions. Such knowledge could be the basis for expression blending and concatenation mechanisms.

The fact that 4 principle components contribute to the subtle expressions on cartoon faces, makes it necessary to design a device (consisting of two 2-dimensional emotion discs) to provide full control over the expressions.

The introduction of higher-level animation building blocks allows higher-level script-driven generation of animations. Relying on the extensive research on multi-modal communication and prosody, we would like to investigate the possibility of generating complete facial expressions (lip-sync, emotional, prosodic, cognitive) based on high-level scripts in a semi-automatic way.

We have several partners who wish to use the basic mechanisms of CharToon in specific application domains such as telecommunication and education. In the framework of a new project, CharToon will be used in distributed, multi-agent systems to generate the proper appearance of an avatar on the fly. The appearance and gesture would adapt to the information to be presented, to the user's profile and to the resources available.

While CharToon was developed with facial animation as the envisioned application domain, experiments by an artist user have proven that CharToon is appropriate to design body parts and animate hand and body gestures. We will investigate if it is possible to extend the facial repertoire with a body repertoire.

Acknowledgements
We thank Frederic Benhamou and Frederic Goualard for making the OpaC solver available for us. The work has been carried out as part of the ongoing FASE project (CharToon 1999) sponsored by STW under no. CWI 66.4088.

REFERENCES

[1] Benhamou, F., Goualard, F. (1999) The OpaC solver for interval constraints, Personal communication.

[2] Benhamou, F., Older, W. (1997) Applying interval arithmetic to real, integer and boolean constraints, *The Journal of Logic Programming*, Vol 32 No 1. pp. 1-24.

[3] Brennan, S. (1985) Caricature Generator: The dynamic exaggeration of faces by computer, *LEONARDO*, 18(3), pp. 170-178.

[4] Bruderlin, A., Williams, L. (1995) Motion signal processing, Motion warping, *Proc. of Siggraph'95*, 97-104.

[5] CharToon Home Page, (1999) http://www.cwi.nl/CharToon

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

[7] Elians, A., Van Ballegooij, A. (2000) Avatars in VRML worlds with expressions, *Proc. of Face2Face Symposium*, Amsterdam, http://blaxxun.cwi.nl:4499/ VRML_Experiments/FASE/

[8] 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/

[9] FaceWorks (1998) DIGITAL *FaceWorks Animation Creation Guide*, Digital.

[10] FAMOUS Home Page (1989) http://www.famoustech.com/

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

[12] Fekete, J. D., Bizouarn, E., Cournaire, E., Galas, T., Taillefer, F. (1995) TicTacToon: A paperless system for professional 2D animation, *Proc. of Siggraph'95*, pp. 79-90.

[13] Gleicher, M., Litwinowicz, P. (1996) Constraint-based motion adaptation, *Apple TR 96-153*.

[14] Hendrix, J., Ruttkay, Zs.M. (2000) Exploring the space of emotional faces of subjects without acting experience, *Report INS-R0013*, CWI, Amsterdam.

[15] Hodgins, J., Wooten, W. L., Borgan, D. C., O'Brien, J. F. (1995) Animating human athletics, *Proc. of Siggraph'95*, pp. 71-78.

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

[17] Kaiser, G. (1994) *A Friendly Guide to Wavelets*, Birkhauser.

[18] Kokkevis, E., Metaxas, D., Badler, N. (1996) User-controlled physics-based animation for articulated figures, *Proc. of Computer Animation'96*, pp. 16-25.

[19] Krzanowski, W. J. (1988) *Principles of Multivariate Analysis, a Users Perpective*. Clarendon Press, Oxford.

[20] Litwinowicz, P.C. (1991) Inkwell: a 2/1/2-D animation system, *Computer Graphic*s, Vol. 25/4, pp. 113-122.

[21] Lost Marble (1999) Moho, http://www.lostmarble.com/aboutmoho.html

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

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

[24] Owen, M., Willis, P. (1994) Modelling and interpolating cartoon characters, *Proc. of Computer Animation'94*, pp. 148-155.

[25] Parke,F., Waters, K. (1996) *Computer Facial Animation*, A. K. Peters.

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

[27] Russell, J. A. (1980) A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), pp. 1161-1178.

[28] 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.

[29] Ruttkay, Zs.M., Lelièvre, A.D.F. (2000) CharToon 2.1 extensions: Expression repertoire and lip sync, *CWI Report INS-R0016*, CWI Amsterdam.

[30] Ruttkay, Zs. (1999) Constraint-based facial animation, *CWI Report INS R9907*, 1999. Also available from ftp://ftp.cwi.nl/pub/CWIreports/INS/INS-R9907.ps.Z.

[31] Salesin, D., Kurlander, D. Skelly, T. (1996) Comic Chat, *Proc. of Siggraph'96*, pp. 225-236.

[32] Schiano, D. J., Ehrlich, S. M., Rahardja, K., Sheridan, K. (2000) Face to InterFace: Facial Affect in (Hu)man and Machine, *Proc. of CHI'2000*.

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

[34] 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.

[35] SVG (1999) http://www.w3.org/1999/07/30/WD-SVG-19990730/

[36] Takacs, B. (1999) Digital cloning system, *Abstracts and Applications Proc of Siggraph'99*, 188.

[37] Thórisson, K. (1996) ToonFace: A system for creating and animating interactive cartoon faces, M.I.T. Media Laboratory Technical Report, pp 96-01.

[38] Van Reeth, F. (1996) Integrating $2^1/_2$-D computer animation techniques for supporting traditional animation, *Proc. of Computer Animation'96*, pp. 118-125.

[39] Van Veen, H., Smeele, P., Werkhoven, P. (2000) Report on the MCCW (Mediated Communication in Collaborative Work) Project of the Telematica Institute, TNO.

[40] 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.

[41] Witkin, A., Popovic, Z. (1995) Motion warping, Proc. of Siggraph'95, pp. 105-108.

[42] Yamada, H., Watari, C., Suenaga, T. (1993) Dimensions of visual information for categorizing facial expressions of emotion, *Japanese Psychological Research*, 35(4), pp. 172-181.

[43] Yacoob, Y., Davis, L. (1994) Recognizing Facial Expressions by Spatio-Temporal Analysis, *Proc. of 12th International Conference on Pattern Recognition*, Jerusalem, Israel, pp. 747-749.