A Text Based Talking Face

L.J.M. Rothkrantz and A. Wojdel

Delft University of technology, Knowledge Based Systems
Zuidplantsoen 4, 2628 BZ Delft, The Netherlands
Phone +312787504 Fax +312787141
L.J.M.Rothkrantz@cs.tudelft.nl

Abstract. Facial expressions and speech are means to convey information. They can be used to reinforce speech or even complementary to speech. The main goal of our research is to investigate how facial expressions can be associated to text based speech in an automated way. As a first step we studied how people attach *smileys* to text in chat sessions and facial expressions to text balloons in cartoons. We developed an expert system with a set of rules that describe dependencies between text and facial expressions. The specific facial expressions are stored in a nonverbal dictionary and we developed a search engine for that dictionary. Finally we present a tool to generate 3D facial animations.

1 Introduction

Human communication is based on verbal and nonverbal behavior. It is commonly assumed that natural language is used to communicate objective information to other people and nonverbal behavior is used to convey subjective and affective information. Speech has a verbal and a nonverbal aspect. It is more appropriate to speak about the denotative and connotative aspect of multimodal communication. The denotative aspect is based on the grammar and the denotative aspect is based on the rules of communication.

In [1] P. Ekman introduced the concepts emblems and emotional emblems. The last ones are expressed by employing parts of the corresponding affect they refer to, while the first ones are used to replace and repeat verbal elements. Most of the time both are intentional, deliberate actions used to communicate. In general they are produced consciously and are driven by the semantics of the utterance. They are conventionalized. Since they are discourse driven, the user enters their appearance. What is needed is a library of possible emblems. Efron gave a large list of them [2] and Ekman proposes a set of words, which have a corresponding emblem. Nevertheless the user can build his/her own emblem and add them to the library [3].

In this paper we mainly investigate which facial expressions people show, reading a text. This knowledge can be used to create a text based talking face. This talking face can be used as an affective user-friendly communicative human computer interface. Human computer-interaction is mainly based on text/picture and mouse, keyboard and screen input/output. But one of the most user-friendly interface is a talking face.

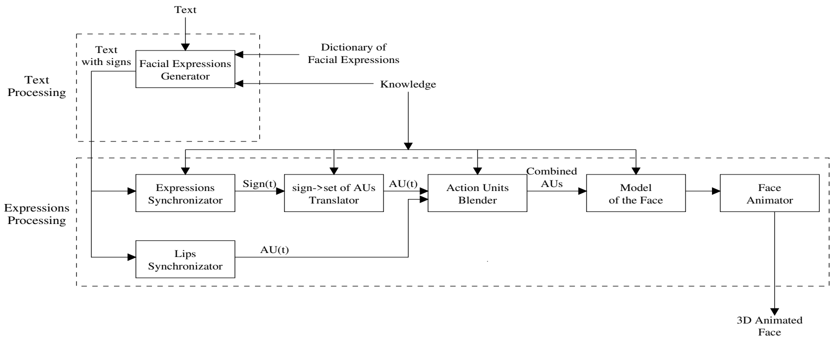


Fig. 1. Design of the system for generating facial expressions

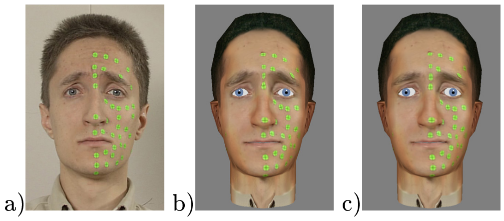


Fig. 2. Activation of AU1. (a) original photo, (b) manually adapted model, (c) parametrized deformation of the model

2 Animated 3D Faces

The whole idea of the system for generating facial animation is based on a "facial script language". In 1970 P. Ekman and F.W. Friesen [4] developed an universal Facial Action Coding System (FACS) based on Action Units (AUs), where each facial expression can be described as a combination of AUs. Because most of all we want to represent real human behavior, not only artistic imagination, we decided to base facial movements on AUs as defined by Ekman. That means that we want to design and to develop a script language of facial expressions, where basic variables are Action Units. We consider AUs as words in a "normal" language. When we have AUs as characters we can define words of our script language: facial expressions. Facial expressions, the same as words in a "normal" language have their own syntax and semantics. We discuss this in more detail in section 3.

In order to have a full definition of the language we need also to define a grammar of the script language. That means we have to define rules of how facial expressions can be combined together. Grammatical rules will be implemented in different modules of the system in such a way that they will support the user while creating animation.

Our design of the system for facial animation is as follows: it has a modular structure, where each module is dedicated to a given task and each module has his own knowledge about dependencies between facial expressions for its

Fig. 3. Examples of different facial expressions (a) disgust (b) happiness (c) weirdness (d) fear (e) sadness

level. The schematic design of the system is presented in Fig. 1. We developed a first prototype of a synthetic 3D face showing facial expressions. The facial expressions can be generated by activating the corresponding AUs. In Fig. 2 we give an example. More details about the underling model for facial animation is given in paper [5].

3 Nonverbal dictionary

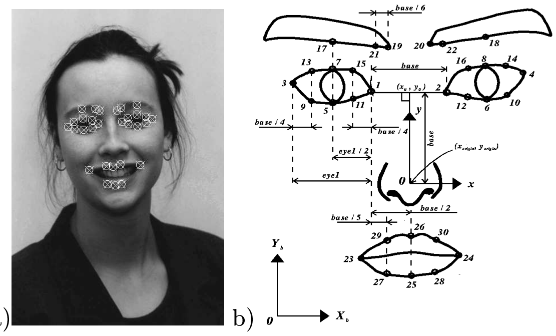
In a common dictionary of some language words are presented in alphabetic order. We can find the spelling of the words, sometimes the phoneme presentation, the meanings in different contexts and rules of transformations of the words. Usually a dictionary will be used to check up the spelling of words or to find different meanings of words. Our nonverbal dictionary will have the same functionality. We developed a prototype of dictionary with 200 nonverbal words.

Nonverbal dictionary is composed of nonverbal words i.e. facial expressions. Every facial expression can be described in terms of Action Units (AUs). We use these AUs and their intensity as the spelling components. So the nonverbal alphabet is the set of AUs in numerical order AU1 to AU43 and in order of intensity. We associate with every expression one characteristic verbal label, namely the name of the emblem. The facial expressions displayed in Fig. 3 are labeled as disgust, happiness, weirdness, fear and sadness. And we give different meanings of the corresponding expression in different contexts.

Finally we give a geometric description of facial expressions. The characteristic shape of the eyes, mouth, eyebrows, the appearance of wrinkles, and the color, light-intensity of different parts of the skin characterizes every facial expression. To have a uniform description of this geometric features, we use a fixed list of items and every expression is scored according to this list (i.e. eyeballs in a central, upwards, downwards, left, right position). Some features are scored on an ordinal 3-points scale corresponding to the intensity.

4 Search Modalities

We implemented different search modalities and we discuss them in more details.



a) b)

Fig. 4. Facial characteristic points (a) on a real face (b) on a generic model

Pictures:
The most natural way to find the meaning of a given facial expression is to look it up in the dictionary and try to find the best pattern match. We automated this procedure and designed and implemented a system named ISFER for the automatic recognition of facial expressions. From a picture of a facial expression we extract the position of some characteristic facial points (see Fig. 4). We feed the coordinates of those points into a fuzzy rule expert system [6]. The output of the system is a list of activated AUs and their intensity. With use of activated AUs we can look up the corresponding facial expression in our dictionary. If the facial expression is out of vocabulary we find the best similar expression using S*-algorithm.

Line Drawings:
It is well known that schematic drawings of facial expressions can convey emotional meanings very well. If we want to lookup a facial expression and there is no picture available the user can generate a line drawing of the corresponding facial expression. To support the user we designed a special tool to generate facial expressions. With use of sliders we can change the shape of the mouth, eyes and eyebrows. Next the user can ask for the N best matching faces from the nonverbal dictionary.

Genetic Algorithm :
If the user has a specific facial expression in his mind he can look it up using a tool based on GA. The system comes up with four facial expressions as representatives of 4 clusters covering the whole nonverbal dictionary. The user selects the best fitting picture. Based on this user feedback the system comes up with another 4 pictures and again the user is requested the best fitting one. The system uses GA to select appropriate representative facial expressions and the user feedback is used as fitness function.

5 Text generated emotions

During on line chatting *emoticons/smileys* can be attached to a text. These *smileys* are composed of "keyboard characters". There are software tools available, which transform these characters to corresponding pictures/line drawings of facial expressions. We use these symbols to generate appropriate 3D animated faces as background or next to the text. In an experiment we asked 28 post-graduate students from the Department of Computer Science to take part in 5 chat-sessions. To support the user we designed a nonverbal keyboard with 50 characteristic facial expressions as buttons. These buttons are available in a window and can be added to the text by simple clicking by the mouse. Students were requested to chat with fellow students and use the *emoticons* as much as possible. It was even allowed to send *emoticons* without any text.

All the chat sessions were recorded in a logfile. The question is whether it is possible to generate the *smileys* automatically. To put it in another way: *is it possible to define knowledge rules, which associate smileys with a text in an automated way*? From the corpus of the logged chat-sessions we extracted more than 300 production rules. Some examples are displayed in Table 1. With these rules we developed an expert system which generates the *emoticons* in an automated way. The input of the system is one line of text. We developed a robust chart parser to parse and to extract the relevant features from the text. It proved that chatters use simple language. Unfortunately the text doesn't satisfy the rules of the Dutch grammars. Chatters use their own words and their own grammar.

But to associate the right *emoticon* it is important to know if the chatter is speaking about his emotions or the emotions of other people. Further it is important to know that he is stating that he wants to convey a specific emotions or that does not want to convey a specific emotion.

Chatting is a way of interaction. A single utterance is related to the utterance of the last speaker. We found out that chatters adapt to some role playing. They can play different roles depending of the context, their moods etc. In case of ambiguity, a safe heuristic is to reflect the emotions of the last speaker.

It proved that in case the *emoticon* was not related to emotional features in the text (punctuators, special words, and onomatopoeia) we couldn't generate them in an automated way. In that case the affective/emotion is not enclosed in the text. We need information from the context and the history of the dialogue or information with respect to the prosody or intonation of spoken language related to that text.

But in case *emoticons* are used to stress some text features with an emotional loading we were able to generate *emoticons* in an automated way, that is to say we generate "default options" or common used options. In many cases there are many options possible. But not every option has the same user appreciation.

6 Text generated facial expressions

In many cartoons we have a lot of facial expressions. Some of these facial expressions are closely related to text balloons. We assume that there is a high correlation between the text in the balloon and the corresponding facial expression. In some cartoons this is stressed by underlining some words. To test this

Table 1. Examples of *emoticons*

<table>
<tr><th colspan="2">Punctuation</th><th colspan="2">Emblems</th><th colspan="2">Emotions</th><th colspan="2">Onomatopoeia</th></tr>
<tr><td>?</td><td>:-Q</td><td>wink</td><td>;-)</td><td>happy</td><td>:-)</td><td>Oops</td><td>;-*</td></tr>
<tr><td>!</td><td>:-o</td><td>woman</td><td>>-</td><td>laughing</td><td>:-D</td><td>Ha ha</td><td>;-D</td></tr>
<tr><td>, ;</td><td>':-|</td><td>unclear</td><td>:-$</td><td>excited</td><td>8-)</td><td>Hmmm</td><td>:-I</td></tr>
<tr><td>.</td><td>:-|</td><td>Lincoln</td><td>=|:-)=</td><td>sad</td><td>:-(</td><td>Hi hi</td><td>:->></td></tr>
</table>

hypothesis we removed the facial expressions in the picures of a cartoon magazine. We presented the facial expressions on a different sheets. In an experiment students were requested to select the most appropriate facial expression related to a text balloon. To be sure that no context information was used we mixed the text balloons in a random way. Again we found that facial expressions were selected on the basis of emotional features related to the text.

7 Conclusion

In this paper we described a general model for a talking face. A prototype for facial animation is described. The main problem was how to choose appropriate facial expressions reading a text or listning to speech. As a first step we created a nonverbal dictionary with facial expressions and special designed search facilities. Emoticons were used as a facial expression script language. In an experiment students were requested to add emoticons to their text in chat-sessions and to add facial expressions to predefined text balloons from cartoons.We were able to define some rules and heuristics how to associate facial expressions to text. It was possible to associate facial expressions to specific keywords in an automated way. But in general there are any choices of facial expressions. The appropriateness of the choice depends of features not included in the text such as prosody and context. Next future we will investigate how to associate facial expressions to spoken text, i.e text with prosidic information used in speech synthesis.

References

1. P. Ekman, Movements with precise meanings. *Journal of Communication, 26*, pp. 14-26, 1976
2. D. Efron, *Gesture, Race and Culture*, The Hague, Mouton & Company,1972
3. C.Pelachaud, N.I. Badler, and M.Steedman, Generating facial expressions for speech, *Cognitive Science*, vol 20, no. 1, pp. 1-46, 1996
4. P. Ekman and W.F. Friesen, *Unmasking the Face.* Englewood Cliffs, New Jersey, USA: Prentice-Hall, Inc., 1975
5. A. Wojdel and L.J.M. Rothkrantz, A Performance Based Parametric Model for Facial Animation, to be published in *Proceedings of IEEE International Conference on Multimedia and Expo*, New York City, NY, USA, August 2000
6. M. Pantic and L.J.M. Rothkrantz, Expert system for automatic analysis of facial expressions, to appear in *Image and Vision*, 2000