Head-shaped Tangible Interface for Affective Expression

Christian Jacquemin
LIMSI-CNRS & Univ. Paris 11
BP 133, Batiment 508
91403 ORSAY, France
+33 1 69 85 81 11

FirstName.Name@limsi.fr

ABSTRACT

A head shaped resin with several holes and equipped with a camera is used for facial expression synthesis through intuitive multiple finger contacts and gestures. The calibration of the interface is presented together with an evaluation of its accuracy under different light and equipment conditions. Three experiments on using the interface for facial expression synthesis are described: a synthesis of emoticons and two modes of expressive animation by associating face zones with expressions of a 3D face. Evaluations confirm the usability of the interface and show that subjects have appreciated its nuanced and sensitive interaction modes.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces – *Input devices and strategies*.

General Terms

Design, Human Factor.

Keywords

Tangible interfaces, Affective computing, Anthropomorphic interfaces.

1. INTRODUCTION

The use of anthropomorphic shapes for output interfaces is quite common (e.g. Embodied Conversational Agents), but it is more rare for tangible input interfaces. When compared with non human-shaped interfaces, anthropomorphic interfaces have the double advantage to be quite intuitive to handle, and to allow for immediate expressive and emotional communication. Intuitiveness is exploited in the doll's head interface of Hinckley et al. [1] for neurosurgical planning through brain visualization. The head shape of the interface provides the appropriate affordances for intuitive and efficient brain view selection. SenToy [2] takes advantage of our natural affective relationship with human body. It is a full body doll equipped with sensors so that players can engage an affective communication with the characters of a game by manipulating the toy.

The work presented here focuses on the head as a natural and intuitive tool for the capture of expressive gestures. We report

on the design of a low-cost input interface to affective communication through multi-point finger contacts. It is made of a head-shaped resin equipped with 42 holes and an internal video camera (Figure 1). The interface is used to control the facial animation of a virtual 3D model through finger contacts. The evaluations made through 45 minutes sessions with a panel of 22 subjects aged between 15 and 56 show that such an interface can offer an intuitive and efficient device for the expression of emotions.

Figure 1. Cross-section of the Interface

2. MULTI-TOUCH FINGER CAPTURE

The purpose of the interface is to use the tangible affordances of a human head to control the expressions of an animated head. Real-time facial animation raises difficult issues because an expression is the result of the association of several local and fine actions called Action Units (AUs) by Friesen and Eckman [3]. We want the users to acquire quickly the production of AUs so that they can focus on their combinations. For this purpose, we have associated face parts with AUs such as *Lip Corner Depressor* or *Brow Lowerer*. The users can combine these elementary expressions through progressive multi-touch interaction with the interface. Through the interface we want users to be able to act simultaneously on several parts of the head, and to produce complex expressions by combining partially expressed elementary AUs..

The multi-touch capture of finger positions on the interface is made by a video camera inside the resin molding. Because of the small dimensions of the interface (approximately the height of a joystick), it is necessary to use a device for wide-angle video capture. The use of a concave mirror has been preferred to lenses because it makes the visual distortion independent of

the camera, because it does not result in any loss of brightness and does not augment the camera size, and because it increases the distance between face and camera through reflection (figure 1).

For each of the holes used to capture finger positions, finger contact is recognized by computing the difference between the image brightness at calibration time and the brightness at capture time. Such an image processing technique allows for the simultaneous acquisition of all the finger positions (without recognition of the obturating fingers). The partial obturation of a hole by the fingers can result in the progressive expression of the associated emotion. Thus the tangible interface is a contact sensor that has the two expected properties: it is both gradual and multi-touch.

In order to accept slight changes in camera position and however preserve an accurate capture of light on head holes, the image capture can be calibrated. The camera image is presented to the user together with four anchors that must be placed on four specific holes (Figure 2.a). With the mouse, the user can drag the anchors so that they cover the corresponding holes (Figure 2.b). The correspondences between screen coordinates and hole locations are recomputed and the capture zones around holes (squares) are shown to the user so that she can evaluate the quality of the calibration and possibly makes additional adjustments (Figure 2.c).

Figure 2. Interface Calibration.

The computation of image and hole center transformations relies on an homography: the geometric transformation that associates the initial view of the face, with the current view of the video camera. An homography is appropriate because camera motions are small rotations close to the center of view (they are due to the imperfect construction of the interface).

The matrix M for a 3D homography in homogenous coordinates is given by 8 floats. The values of the 4 initial and final locations A_i and A'_i of the calibration anchors result in a system of 8 equations with 8 unknowns that is solved by a Gaussian elimination:

$$\forall i \in 1..4 \begin{pmatrix} a'_{ix} \\ a'_{iy} \\ 1 \end{pmatrix} = \begin{pmatrix} m_{1,1} & m_{1,2} & m_{1,3} \\ m_{2,1} & m_{2,2} & m_{2,3} \\ m_{3,1} & m_{3,2} & 1 \end{pmatrix} \begin{pmatrix} a_{ix} \\ a_{iy} \\ 1 \end{pmatrix}$$

At calibration time, the gray level in each hole is recorded. During interaction, the new values of brightness are computed and each hole activation is proportional to the ratio between the current brightness, the one captured initially, and full darkness. Thus hole activation is a function of the opacity of a finger and of the strength with which a hole is obturated.

Figure 3 shows logs of the activation values (darknesses) of a hole and its neighbors at different light levels and with two equipments (black glove for the figure 3.a or bare hand for 3.b). The captured gesture is a finger that points at a hole, reaches its target and obturates it, and finally moves away. The time is the horizontal coordinate and the vertical coordinate is the level of obturation of each hole. The activation of the targeted hole (#3.4) is represented by a solid line. The concurrent activations of neighboring holes are represented by dotted lines, they result from shadowing effects of the fingers and hand palm. Hole #3.4 is the bottom center of the right eye, the coactivated holes (#3.6, #3.2, #3.12) correspond respectively to eye center, upper center, and right corner of the same eye.

Because of the finger translucency, the level of obturation with bare hand and strong light is weaker than with lower lighting conditions (figure 3.b), while it is constant with black gloves whatever the light intensity (figure 3.a). Whereas it may seem desirable to have a better obturation by using black gloves, wearing gloves has the drawback to increase the shadowing effect on neighboring holes: coactivated holes have closer values to #3.4 in Figure 3.a than in Figure 3.b. Bare hand manipulation has been preferred to gloves because it does not require an additional equipment, and because it reduces the level of undesirable shadows. In the case of bare hand manipulation, the results of Figure 3.b show that the better discrimination is obtained for strong light even though the obturation by the fingers is not very high (~0.2). At all light levels, the targeted hole is the most activated, but strong light yields better discrimination.

Obturation of point 3.4 and its neighbors

Figure 3. Hole activations with finger pointing at hole #3.4.

3. EXPERIMENTS

Among the multiple applications of this interface, we have chosen to focus on its use for facial animation. Animating a 3D face is a skilled activity, and we wanted to provide the users with a tool that would enable them to control easily the expressions of a synthetic face. The presence of undesirable coactivations of neighboring holes has oriented towards the use of face zones instead of individual holes for the mapping between finger contacts and expressions. For this purpose, we have divided the tangible interface into 6 zones and associated each zone with a predefined expression (an Action Unit of [3]): the forehead is associated with *raised brows*, the nose with *lower brows*, the eyes with *closed eyes*, the mouth with smile (*lip corner puller*), the jaws with *jaw drop*, and the chin with *lip corner depressor* (Figure 4.b). A similar mapping has been established between the face zones and emoticons (Figure 4.a).

Figure 4. Mapping between face zones and base expressions

We have designed two modes of facial animation: a *boolean exclusive mode*: only one basic expression can be activated at a time, and a *blended mode*: expressions can be gradually activated (from neutral face to full expression) and can be combined. Emoticons cannot be used in blended mode, this mode only concerns 3D facial expressions.

In order to evaluate the usability of the interface together with the user's pleasure, 22 volunteer subjects aged between 15 and 58 (average 29.1) have made a 45' experiment with on interactive facial animation. Figure 5 shows the experimental setup: the subject sits in front of the interface, manipulates it with her fingers, and tries to reproduce a model (top left of the screen) on an animated head or emoticon (center of the screen). The subject has two sheets that correspond to Figures 4.a and 4.b to help her memorize the mappings between face zones and emoticons or Action Units.

Each evaluation session consists of a presentation of the purpose of the interface by the experimenter and three

experiments. The first two experiments, are in the boolean exclusive mode (emoticons for experiment 1 and facial expressions for experiment 2). In these experiments, the subject is presented successively 15 random target emoticons or expressions and must reproduce them by touching the associated zone of the interface and holding the expression for half a second. Every time a target expression is reached, the next one is displayed until the last one.

Figure 5. Experimental setup

The third experiment is in the blended mode. Blended expressions can be produced by touching simultaneously and more or less strongly several zones. The resulting expression is the linear combination of the elementary expressions with weights computed from the zone activation values [4] (the activation of a zone is the level of the most occluded hole). For instance, the simultaneous activations of the jaw and mouth zones result in a combination of smile and lip protrusion. In this experiment, the subjects are asked to reproduce as accurately as possible each target blended expression and to decide by themselves when they are satisfied before turning to the next target.

Before starting each experiment, the subjects are asked to feel comfortable, to take as much time as they need to practice the interface in the current mode. They can hold the interface either facing them or oriented the other way.

4. EVALUATIONS

Quantitative evaluations rely on the time needed to reproduce a target expression. For the third experiment, we also use the error between the target and the produced expression. Tables 1 and 2 report the percentage of subjects at each expertise level by the average time taken to reach a target. These tables show that expert users tend to be quicker in the first experiment (emoticon). But the advantage of expert subjects tends to disappear in the second experiment, either because non expert subjects take longer to learn how to control the interface or because the additional difficulty due to the need to identify a target 3D expression before reproducing it (emoticons are easier to recognize) tends to leverage the completion speeds between expert and non expert users.

Table 1. Distribution of the subjects by average time and expertise (1st task)

Expertise:	Low	Average	High
0-5 sec.	33%	30%	67%
5-10 sec.	56%	60%	33%
> 10 sec.	11%	10%	0%

Table 2. Distribution of the subjects by average time and expertise (2nd task)

Expertise:	Low	Average	High
0-5 sec.	0%	20%	33%
5-10 sec.	89%	70%	67%
> 10 sec.	11%	10%	0%

The third experiment brings an additional difficulty: the subjects must combine zone activations to reach a blended expression. In order to take into account the quality of the produced expression, Table 3 reports, for each duration and for each level of expertise an average measure of the error between the achieved expression and the target one. The error is the sum of the distances between the coordinates of the face made by the user and the target face in the 6D space of blended expressions.

Due to the difficulty of the task, the time taken to reach the targets are longer than in the first two experiments. Table 3 shows that expert subjects get better quality, but they do not perform significantly faster than non expert subjects: low, average, and high experts have the respective average durations 27.2", 26.9", and 25.9".

Table 3. Error by average time and expertise (3rd task)

Expertise:	Low	Average	High
10-20 sec.	1.74	1.14	1.06
20-30 sec.	1.52	1.66	1.0
30-40 sec.	1.53	1.34	1.38
> 40 sec.	-	1.91	0.89

The qualitative evaluation is based on an anonymous questionnaire that contained questions about the expertise of the subjects, about the subjective level of difficulty and the pleasure of use, and about general thoughts on the interface and its possible applications. All the subjects but one have appreciated using the interface and have been deeply engaged in the experiments that were proposed.

Among the positive aspects of the interface, 12 subjects out of 22 appreciate the soft and sensitive tactile interaction with such arguments as *the possibility of nuanced control, it is touch friendly and natural based or the tactile interface and its reactivity*. The negative points concern the difficulties in the manipulation of the interface: the shadowing, the small distances between holes, and a lack of relief on the mouth.

In response to the question whether they could envision some applications to this interface, the subjects have been very positive. They have made many propositions such as graphical modeling, emoticons for chat or email, plastic surgery, telepresence and affective communication, interface for the blind and visually impaired people, theater, animation, embodied conversational agents...

Last, subjects were asked to comment freely on the interface. We just report one comment that is certainly representative of the impression given by such a device: *The contact of fingers on a face is a particular gesture that we neither often nor easily make (You do not let people easily touch your face). Luckily, this uncomfortable impression does not last very long. After a few trials, you feel like a sculptor working with clay. . .*

5. CONCLUSION

The interface presented in this paper was based on the idea that giving access to a tactile interaction with a head-shaped interface could engage users in a deep and sensitive experience. The results of the experiments show that usability is satisfactory even though some of the technical aspects could be improved. More importantly, the observation of the subjects and the analysis of the questionnaire confirm that users have experienced the interface as an appealing device for affective communication. Some of the users have even start talking with the interface *Come on! Close your mouth...* in situation where they could not make it look as they wished.

More applications should be conducted in the near future to evaluate the interface for interpersonal communication and telepresence. The first experiments reported here show that the characteristics of the device are a good starting point for further investigations in its use for affective communication.

6. ACKNOWLEDGEMENTS

Many thanks to Clarisse Beau, Vincent Bourdin, Laurent Pointal and Sébastien Rieublanc (LIMSI-CNRS) for their help in the design of the interface; Jean-Noël Montagné (Centre de Ressources Art Sensitif), Francis Bras, and Sandrine Chiri (Interface Z) for their help on sensitive interfaces; Catherine Pelachaud (Univ. Paris 8) for her help on ECAs. This work is supported by LIMSI-CNRS *Talking Head* action coordinated by Jean-Claude Martin.

REFERENCES

- [1] K. Hinckley, R. Pausch, J. C. Goble, and N. F. Kassell. Passive real-world interface props for neurosurgical visualization. In *CHI '94: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 452–458, New York, NY, USA, 1994. ACM Press.
- [2] A. Paiva, G. Andersson, K. Höök, D. Mourao, M. Costa, and C. Martinho. Sentoy in Fantasya. Designing an affective sympathetic interface to a computer game. *Personal Ubiquitous Comput.*, 6(5-6):378–389, 2002.
- [3] P. Ekman, and W.V. Friesen. Facial action coding system: *A technique for the measurement of facial movement*. Consulting Psychologists Press, Palo Alto, CA, USA, 1978.
- [4] N. Tsapatsoulis, A. Raouzaïou, S. Kollias, R. Crowie, E. Douglas-Cowie. Emotion recognition and synthesis based on MPEG-4 FAPs. In I.S. Pandzic, R. Forchheimer, eds.: *MPEG-4 Facial Animation*. Wiley, Chichester, UK, 141–167, 2002.