Simplified facial animation control utilizing novel input devices: a comparative study

Nikolaus Bee, Bernhard Falk, Elisabeth André

Angaben zur Veröffentlichung / Publication details:

Bee, Nikolaus, Bernhard Falk, and Elisabeth André. 2008. "Simplified facial animation control utilizing novel input devices: a comparative study." In *Proceedingsc of the 13th international conference on intelligent user interfaces - IUI '09, Sanibel Island, Florida, USA, February 08 - 11, 2009*, edited by Cristina Conati, Mathias Bauer, Nuria Oliver, and Dan Weld, 197–206. New York, NY: ACM Press. https://doi.org/10.1145/1502650.1502680.

Simplified Facial Animation Control Utilizing Novel Input Devices: A Comparative Study

Nikolaus Bee, Bernhard Falk and Elisabeth André
Institute of Computer Science
University of Augsburg
Eichleitnerstr. 30, 86135 Augsburg, Germany
{bee, andre}@informatik.uni-augsburg.de

ABSTRACT

Editing facial expressions of virtual characters is quite a complex task. The face is made up of many muscles, which are partly activated concurrently. Virtual faces with human expressiveness are usually designed with a limited amount of facial regulators. Such regulators are derived from the facial muscle parts that are concurrently activated. Common tools for editing such facial expressions use slider-based interfaces where only a single input at a time is possible. Novel input devices, such as gamepads or data gloves, which allow parallel editing, could not only speed up editing, but also simplify the composition of new facial expressions. We created a virtual face with 23 facial controls and connected it with a slider-based GUI, a gamepad, and a data glove. We first conducted a survey with professional graphics designers to find out how the latter two new input devices would be received in a commercial context. A second comparative study with 17 subjects was conducted to analyze the performance and quality of these two new input devices using subjective and objective measurements.

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: Input devices and strategies

Author Keywords

Input Device, Gamepad, Data Glove, Facial Expression, Virtual Avatar

INTRODUCTION

Virtual worlds, such as Second Life, Lively by Google, or World of Warcraft, provide a rich platform for embodied interaction between people all over the world through the internet. The social component of such platforms is a fundamental part of their success. When it comes to close interaction, facial emotional expressions play an important role as non-verbal behavior to underline the written words during a chat. Especially in game-based multi player platforms, such as World of Warcraft, where users conduct quests with 2-40 companions, expressing emotions, for instance, becomes essential after you succeed or fail to accomplish a cooperative quest.

While the visual capabilities for displaying expressions through virtual characters has advanced quickly in the last few years, the control of facial expressions still remains a challenge. Common tools to adjust facial expressions use slider-based graphical interfaces that allow users to edit one facial parameter after the other. However, facial expressions involve several facial muscles in parallel. As a consequence, new input devices which support intelligent parallel control should not only speed up the editing of facial expressions, but also simplify the editing for inexperienced users.

In the next section, we first discuss related work on facial animation control systems. After that, we present a virtual character we created whose facial animation system [21] is based on Ekman's FACS (facial action coding system) [8]. Basically, it enables the creation of an infinite number of facial expressions. We connected this facial animation system to three controllers: (1) a slider-based GUI, which is the current standard interface for such a task; (2) a gamepad, which allows parallel control and is widely used in computer games; (3) and a data glove, which enables continuous control while editing five facial parameters in parallel. We first conducted a survey with professional graphics designers to find out how the latter two new input devices would be received in a commercial context. A second study with 17 subjects was conducted to analyze the performance and quality of these two new input devices.

RELATED WORK

Slider interfaces bear the advantage that they are both easy and quick to implement. Furthermore, most users are familiar with this kind of interface. Nevertheless, there are some serious pitfalls to be considered. First of all, users may only manipulate one parameter at a time. Yet, the interplay of different parameters is crucial in generating high quality facial animations. As a consequence, users need to switch back and forth between different sliders to adjust the parameters for the desired facial expression. Furthermore, the use of sliders is hardly intuitive since there is no obvious mapping between the manipulation of a slider and the movement of the corresponding mimic muscle. In order to know what effect a particular slider achieves, the user needs to interpret

IUI'09, February 8 - 11, 2009, Sanibel Island, Florida, USA.

the description of the slider correctly. Uncommon anatomic technical terms may further hinder the user's understanding.

Approaches to facial animation control based on the manipulation of images also offer alternative solutions, for example see [26]. So-called sketch-based interfaces as introduced by Chang and colleagues [5] or Natanelli and colleagues [17] go a step further and generate facial expressions from sketches drawn by a user. Jacquemin developed a 3D interface of editing facial expressions. The tangible interface named Pogany maps a real model of a human face to a computer generated 3D face [13]. Depending on which region is touched on the physical model, the virtual match is activated and enables one to compose a facial expression. Jacquemin could show that such a novel interface is easily accepted and engages users in a pleasant way.

While these interfaces need special mapping, including pattern recognition or even special hardware to control facial expression of virtual faces, Thalmann [27] analyzed a variety of more common hardware devices for animation control, including position and orientation trackers, data gloves, data suits, 6D-devices and midi keyboards. Particular emphasis was given to data gloves and midi keyboards as promising control devices for facial animation. The computer game "Indigo Prophecy" by Quantic Dream [4] provides evidence of the practical use of data gloves. The facial expressions of this game were produced by translating the finger bends of the gloves an animator was wearing into the corresponding morph target animation parameters. The facial animations included emotional expressions as well as lip syncing.

The aforementioned mapping approaches use a direct mapping to facial muscles or regions to control the facial expression of a virtual agent. They could be described as *direct* mapping. Another, *indirect*, approach to map emotions to facial expressions is to use a descriptive or a model representation of emotions. Ruttkay et al. [23] developed EmotionDisc, where discrete emotions are arranged in a circular way. The distance from the center of the disc is equivalent to the intensity of the current emotion dependent on the current angle. Albrecht et al. [2] describe the usage of an emotion dimension model recommended by Cowie et al. [7] to control the facial expressions of a virtual character. This model describes emotions in an activation-evaluation space. Depending on spatial position, the respective facial expression is displayed (e.g. the center displays neutral, the upper right area display happy or excited, ...). Courgeon et al. [6] use a 3D model to describe emotions. They place a discrete emotion on every corner of a cube. Users control the 3D representation of it with a joystick and, depending on the position in the 3D space, an appropriate blended facial expression will be generated.

This way of controlling emotional facial expressions does not require the understanding of how to design or model facial expression. Thus, it makes it easily usable for inexperienced users.

FACS-BASED FACIAL EXPRESSION GENERATION

"Alfred" (see Fig. 1) is a butler-like character used to display facial expressions. Alfred's facial animations are based on the Facial Action Coding System (FACS) by Ekman and Friesen [8]. Although FACS was originally designed to analyze natural facial expressions, it turned out to be usable as a standard for production purposes too. That is why FACS based coding systems are used with the generation of facial expressions displayed by virtual characters, like Gollum in the movie trilogy The Lord of the Rings or Kong in Peter Jackson's King Kong [24]. But the usage of FACS is not only limited to virtual characters in movies. The gaming industry with Half-Life 2 by Valve, also utilizes FACS to produce the facial expressions of their characters [1].

FACS defines 32 so called "Action Units" (AU) which are motivated through the human facial muscle system (e.g. inner brow raiser, upper lid raiser, or lip corner depressor). The action units describe the movement of a facial part of one or several muscles. FACS consists of 32 action units and additionally of 26 action descriptors, which describe more complex movements outside the mimic muscles, e.g. the rotation of the head or the eyes.

To implement FACS [11] [25] in Alfred, morph targets (also known as blend shapes) were used. They describe the translation of a set of vertices to a defined new position in the 3D space. In our implementation, each morph target represents one of the action units. As not all action units are necessary – some of the action units overlap and are not needed for generation – we could limit them to 23.

Another system, mostly used in academia, to generate facial expressions is the MPEG-4 standard [3], [22]. It defines 66 facial action parameters (FAP) which control specific regions of the face (e.g. shift tongue tip, raise left middle eyebrow, stretch left corner lip). The basic principles of controlling facial expressions with the MPEG-4 standard and FACS are the same [20].

We chose the FACS-based approach for our facial animation system, because of the Facial Expression Repertoire (FER) [10], which maps over 150 emotional expressions to the action units of FACS. Not only does it explain in detail, which action unit must be activated for certain facial expressions, it further provides a rich dataset of videos which show how the action units ought to be designed.

Alfred's mesh has a resolution of about 21.000 triangles. For displaying more detailed wrinkles in the face, normal maps baked from a high-resolution mesh are used [18]. The morph targets for the action units are modeled using the actor's templates from the FER. For rendering the character and its animations the Horde3D [12] graphics engine is used.

To control Alfred's facial expressions (i.e. action units), we use the UDP network protocol. This allows us to easily connect new interfaces to control the virtual face. Any controller can send the desired expression in terms of a string array with the values of all action units to the Alfred application.

Figure 1. The virtual character Alfred is designed utilizing FACS to compose facial expressions.

Figure 2. The XBox 360 controller with two analog sticks, two analog shoulder buttons, one digital stick and several buttons.

DESIGN AND IMPLEMENTATION OF THE INTERFACE

We identified a number of serious disadvantages regarding wide spread slider-based user interfaces for facial expression generation. Hardware controllers represent a promising alternative which has not yet been explored in depth. First we analyzed the capabilities of such input devices and then defined an intuitive mapping between the input devices and the facial expression control.

Use of Novel Input Devices

Slider-based GUIs limit the composition of facial expressions to sequential control and thus lack transparency. Since users can just edit a single facial unit at a time, it is hard for them to imagine for what the final result of the composing might look like. Novel input devices, such as gamepads or data gloves, allow users to modify several facial units at once.

Gamepad

The first type of hardware controller we studied in our work was the gamepad. Gamepads are today's standard controller for gaming consoles like the XBox 360 or the Playstation 3. They have the major advantage of being widely available, cheap, and many users are familiar with them. Gamepads were originally designed for long hours of computer gaming and thus their design takes into account many ergonomic aspects. We focused on the XBox 360 game controller (see Fig. 2), which can also be easily connected to any Windows compatible PC. Since most of today's gamepads are constructed in a similar manner, our analysis and results can be easily transferred to other gamepads.

To control facial expression, it is important that a controller returns a continuous data stream. In this way, the intensities of the action units or morph targets can be controlled in real-time. The XBox 360 gamepad provides a variety of analog and digital controls: two analog sticks, one four-way digital cross, six buttons on the top of the gamepad and four buttons, two of which are analog, on the front of it (see Fig. 2). Each of them can be controlled independently and in parallel by moving a finger or thumb. The analog buttons provide a one dimensional signal similar to sliders and the analog sticks provide a two dimensional signal. Two basic approaches should be considered to interpret this two-dimensional signal and to transfer it into the one-dimensional "action unit"-space:

1. The signals from the analog sticks consist of two dimensions: an x- and a y-dimension. Each dimension of an analog stick can be mapped to one parameter of an action unit. In this way, two action units can be controlled simultaneously. But, contrary to sliders or analog buttons, analog sticks provide positive and negative values. This allowed us to map negative values into a positive space and thus control four different parameters with one analog stick (i.e. moving the analog stick forward to control one action unit and moving it backward to control a second action unit – the same when moving the stick sidewards).
2. Since analog sticks can be moved circularly, signals can also be interpreted as polar coordinates. In that way, the angular coordinates can be used to select an action unit and the radial coordinate can be used as its weight.

The first approach controls two action units at once, since the horizontal and vertical activations are independent. With the second approach, only one action unit can be controlled at once, as every angular activation selects an action unit. In addition to the analog controls, the XBox 360 gamepad has a couple of digital buttons and a directional pad, which can be used for further control functions (e.g. switching the current setting of action unit mapping).

Data Glove

Data gloves (see Fig. 3) measure the bends of the fingers and, often too, the orientation and the position of the hand wearing the data glove. While the position of the hand can be very useful for performing a selection task (e.g. selecting the setting for a certain region of the face), the posture of the hand can be used for expression control. The human hand consists of five fingers, which can be bent relatively independently. Since a finger bend is a one-dimensional signal, it is an ideal candidate to replace slider-based interfaces.

The “P5 Glove” was originally developed for gamers, and its low cost makes it ideal for developing prototypical applications. The “P5 Glove” provides the following data:

- absolute position (x,y,z), relative position (x,y,z), and rotation (yaw, pitch, roll)
- finger bend
- three additional digital buttons

Figure 3. P5 data glove with five analog controllers and three buttons.

Mapping Models

The question arose as to how the single signals of a controller could be projected onto the FACS model. Both the gamepad and the data glove offer just a limited number of controls which do not suffice to cover our 23 action units. In this section, we present three different mapping models to solve this problem.

Direct Mapping

The basic idea of direct mapping is to transfer the structure and layout of the human face onto the hardware controller. To this end, the face is decomposed into logical groups (e.g. eyes, nose, mouth etc.). Ekman et al. [9] already defined logical groups for facial regions. They decomposed the face into an upper part with 7 action units and a lower part with 16 action units.

Since there are more action units than may be controlled at a time, it was necessary to assign multiple settings to one control. Suitable controls are those that are not required for manipulating facial action units. Here, principles of ergonomics should be applied. Controls which are easy to operate should be reserved for the more important facial action units. The importance of an action unit is defined by its frequency of occurrence and its influence on the facial expression.

Mapping for Gamepad

Based on the considerations above, we defined the following settings for the gamepad. The upper face with 7 action units could be directly mapped to a gamepad setting with two action units mapped to the two front buttons, two action units mapped directly to the right analog stick and three action units mapped to the left analog stick using circular mapping. The lower face had to be split into two settings, since 16 action units represented too many regulators for the gamepad to be able to deal with them all at once. The second setting controlled parts of the lower face, excluding the action units for the inner lips. We again used the analog front buttons to directly map two action units. The left analog stick was used with circular mapping to control the lip corners and the right stick was mapped to the raising and lowering of the chin. The third setting controlled the inner lips. The front buttons were used to control two action units, the left stick to control one action unit and the right stick to control two action units. Figure 4 illustrates the gamepad settings for the upper face, the lower face, and the inner lips. The four-way digital cross used to switch between the three settings.

Mapping for Data Glove

The data glove provides five analog controls, one for each finger. That means that only five action units can be controlled simultaneously. The 23 action units could be mapped to five different settings for the data glove. To keep from Ekman et al. [9]’s distinction between upper and lower face, we opted to use six logical settings for the data glove: brows (3 AUs), lids (3 AUs), cheek and nose (3 AUs), corners of the mouth (5 AUs), chin and inner lips (4 AUs), and lips (3 AUs). More important action units were mapped to more prominent fingers. The user could select one of the six settings by moving the data glove horizontally.

Context-Sensitive Mapping

One disadvantage of direct mapping is the necessity to assign multiple functions to a single control. Permanently switching between different settings increases the complexity of the interface and thus the time required to generate a facial expression. To avoid this problem, we investigated whether it would be possible to have the user only manipulate action units that are relevant in a specific context. An example of such a context would be an emotion the user wishes to express. In such a situation, it might be helpful to provide the user with just the action units that are necessary to adjust the corresponding emotional expression as desired. We conducted an analysis of the facial expressions stored in the FER database in order to find out which action units were mostly involved in a particular facial expression considering their frequency of occurrence as well as the variance of occurrence. In addition, we performed a correlation analysis in order to identify action units occurring together.

The FER database contains variations of Ekman’s basic emotions: joy, anger, fear, sadness, disgust, and surprise. We use these six emotional expressions as the context to be modified. The action units to define these basic expressions were selected by collating the listed action units from the FER database with Ekman et al. [9] and with Kätsyri [14]. The overlapping action units were used to define a basic emotion. The action units that influence a basic emotion (e.g. surprise → puzzled) were selected by calculating the mean

Figure 4. Settings for the gamepad to control the action units for the upper face (left), lower face without inner lips (middle), and the inner lips (right).

and the variance. Action units with a high mean value were considered important for varying the basic emotion and were automatically mapped to an analog control on the gamepad. Action units with a medium mean value but a high variance were considered important for a broader range of different facial expressions and thus were mapped to an analog stick using the circular approach. Action units with a low mean value and variance were checked manually to see whether they played an important tole in influence the facial expression of a basic emotion and, where applicable, were omitted.

Using the gamepad to modify a basic emotion, the user could select the desired context by pressing the four-way digital cross. The respective basic facial expression was presented and the user could manipulate this expression using a setting which was adjusted to this particular context. For instance, if a user selected the context "joy", the setting contained, among other modifiers, the action units "cheek raiser" and "lip corner puller", whereas the action unit "lip corner depressor" was not allocated to a control in this setting, as it was not needed.

Mapping Based on Basic Emotion Categories

Ekman and Friesen found that a large portion of emotional expressions could be generated by blending the six basic emotions [8]. McCloud seized on this idea and illustrated how comic characters can express emotions by blending two or more of the six basic emotions [16]. We implemented two mappings for blending the basic emotions with the gamepad.

One uses the six independent analog controls and maps the intensity of all action units for one basic emotion to one control of the gamepad. The user simultaneously controls all six basic emotions with four fingers, two emotions for each analog stick and one emotion for each analog front button. This approach might be challenging for the user, as all six emotions can be blended at once. McCloud [16] mostly blends two basic emotions, thus it could be sufficient to limit the controls to blending two to four basic emotions at once.

The second mapping uses the two analog sticks of the gamepad based on the circular approach. This allows the user to blend two to four basic emotions using two controls simultaneously.

Although the blending on face level produces a variety of different emotional facial expressions, blending on face regions would not only increase the variations, but also improve the quality of such blended expressions. Especially, if emotions overlap (e.g. you feel sad but want to show joy), the way on how treat the blending on the different facial regions is challenging [8] [19]. In this paper, we did not consider the blending at the level of facial regions, as we wanted to keep the blending of facial expressions as simple as possible in this first approach.

Anatomical Constraint Model

One problem with morph targets are the interferences that might occur when simultaneously activating several morph targets. Lewis et al. [15] describes this phenomenon and offers a solution to avoid such interferences. Since the FACS model was originally defined to analyze, and not generate facial expressions, it is possible to simultaneously activate certain action units, which anatomically speaking would be impossible, and the result of this is an unnatural facial expression (see Fig. 5).

Our constraint model, to prevent such unnatural facial expressions, is based on facial regions. Action units that are anatomically impossible within one facial region are reduced to a realistic value.

$$Totalforce_{region} = \sum_i AU_i * weight_{i,region}$$

When the total force of a region exceeds 1.5, all action units within this region are reduced by the factor $Totalforce/1.5$:

$$AU_{i,reduced} = \frac{1.5 * AU_i}{Totalforce_{region}}$$

The face is divided into four regions, which are not dependent on each other: eye brows, eye lids, inner lips, and corners of the mouth. The single weights for each action unit were derived from video clips in the FER database and man-

Figure 5. Facial expression without (left) and with (right) constraint model.

ually adjusted. Although the constraint model is in principle independent from the facial model as it is based on the activation of the AUs, it might depend on how the single morph targets for the AUs are designed.

STUDIES WITH PROFESSIONAL GRAPHICS DESIGNERS

To find out how the new interfaces would be received in a commercial context, we recruited two professional graphics designers from the computer game developer "Chimera Entertainment". The study served to clarify a number of questions that came up during the design of the interface. In particular, we hoped to get useful hints regarding the assignment of functions to hardware controls. The introduction to the system followed the Coaching Method in order to obtain additional information on the users' behavior during the learning phase. During the actual test, the users performed different tasks following the "Thinking Aloud" method.

Our users appreciated the direct gamepad control interface. In particular, they found that this interface had a clearer structure and layout than the slider-based interface. Regarding the context-sensitive gamepad control interface, a number of concerns were uttered. Firstly, the continuous switch between different settings required the users to re-orient themselves again and again and made it difficult for them to get familiar with the interface. Secondly, the users had the feeling that they had less control over the system as a whole since it was not always obvious to them which action units were to be manipulated.

The basic emotion composition approach was positively received. This approach was described as intuitive and easy to use. In particular, the participants appreciated the fact that this approach could speed up the production process. The participants, however, had some doubts as tow whether it would be possible to adjust the settings in such a way that all desired emotional expressions could be generated. Nevertheless, they regarded this approach as a solution to come up with fast and creative pre-settings. In particular, the potential of the composition approach in combination with direct mapping was emphasized. A designer could, for instance, first create a rough pre-model and then refine it using the direct mapping approach.

The data glove profited in particular from the novelty effect. As graphics designers, our users were familiar with gamepads, but not with data gloves, which are less common in the game industry. Yet, they found that the data glove was not accurate enough. Furthermore, it was perceived as physically tiring after some time.

The users emphasized the importance of comprehensive functions for storing and changing authored expressions for day-to-day production. Moreover, they mentioned the noisy signal from the data glove regarding the tracking of the hand in space. They found it quite difficult to select a setting, which was mapped to the horizontal movement of the glove. To improve the selection process a noise reduction filter was applied to the signal from the glove.

At first, there were two versions of the gamepad visualization. One was with text that labeled the controller with the controllable action unit (see Fig. 4) and the other was with small icons of the controllable action unit. The users preferred the one with text and therefore the icons were omitted.

EVALUATION OF THE HARDWARE CONTROLLERS

To compare the novel hardware devices against the traditional sliders, we conducted a formal user study. We were particularly interested in finding out (1) how users would get along with the novel input devices in comparison to the sliders, (2) whether they would enjoy using them and (3) how they assessed their technical features. Besides subjective user ratings, we also aimed at objective performance measurements. In particular, we assessed the quality of the users' creations and recorded how much time it took them to accomplish a task.

Based on the preliminary user study, we expected the novel input devices to be positively received. We assumed that both the data glove and the gamepad would contribute to an enjoyable interaction experience. In particular, we believed that the gamepad would successfully compete against the sliders thanks to better usability and performance. Due to the feedback we got from the professionals, we did, however, expect some usability and performance issues with the data glove.

The formal study was structured as follows: Each input device was tested by each participant in random order. After a short training phase, we presented the participants with concrete modeling tasks that they had to accomplish using a particular input device and measured task completion in terms of quality and time. Before the participants were given the tasks for the next input device to be tested, they were asked for their subjective assessment of the input device they had just used.

Users and Experimental Method

We recruited 17 subjects aged 20 to 40 (13 males and 4 females). Most of the subjects (76 %) were students. To assess participants' prior experience, we used a 5-point rating scale: "none", "little", "medium", "high" and "extremely high". Most participants were familiar with the use of gamepads (mean value: 2.82), but had in general little experience with 3D modeling (mean value: 1.88), facial expression animation (mean value: 1.47) or emotional models (mean value: 1.88). Due to the large number of users required for a statistical analysis, it was not possible to rely on professional graphics designers for the formal user study. The different user groups, however, gave us the opportunity to investigate whether the comments of the professionals could confirmed by non-professionals.

At the beginning of the experiment, the participants had to input the required demographic data. After that, they tested each input device in random order to avoid any bias due to habituation effects. To compare the single input devices, we decided to rely on the direct mapping approach. First of all, the interviews with the professionals had revealed a preference for the direct model approach. Secondly, it would have been more difficult to identify the factors responsible for a particular effect using the other two mapping models since they heavily differed for the single input devices. For each input device, the participants had to go through the following phases:

- *Training Phase*
  The single participants were given an individual introduction to the input device to be used next while they were holding it in their hands. After that, they got one minute to test the input device themselves.
- *Modeling Phase*
  The participants were asked to create three facial expressions based on photos of an actor. The photos were taken from the FER database which provides a list of the relevant action units for each facial expression. To test the single input devices with different photos, we collected nine photos that were distinguished by different levels of complexity:
  - Facial expressions with *complex eye area* and *simple mouth area*
  - Facial expressions with *simple eye area* and *complex mouth area*
  - Facial expressions with *complex eye area* and *complex mouth area*

  The complexity was defined by the number of action units that had to be manipulated in order to create a particular facial expression. The facial expressions were randomly assigned to the three input devices. Using a particular input device, the participants had to generate one facial expression per category. They were allowed to spend as much time as they wished on a specific modeling task. While they were interacting with the system, the time was logged. After each task, participants were asked to indicate how satisfied they were with their result using a questionnaire with a 5-point scale attitude statement (disagree, somewhat disagree, neutral, somewhat agree, agree) for each task.
- *Questionnaire*
  After accomplishing all three tasks with a particular controller, the participant was asked to fill in a post-task questionnaire. The post-task questionnaire used eight attitude statements with a 5-ary scale to evaluate how the participants perceived the interaction with the system when using a particular device. The question referred to the usability of the device (four questions: U1, U2, U3, U4), the user's subjective perception of the interaction experience (two questions: E1, E2) and the technical features of the device (two questions: T1, T2).

Finally, the participants were asked which controller they would choose if they had to repeat the test again with all nine photos.

Results of the Experiment

To evaluate the gamepad and the data glove, we compared them with the sliders as a reference interface. In particular, we applied two-tailed t-tests to each of the two novel input devices and the sliders.

First we analyzed how the participants had assessed the results of their own work. Overall, the participants were most satisfied with the facial expressions they created using the sliders with a mean value of 3.84, followed by the facial expressions they created using the gamepad with a mean value of 3.63. The data glove scored worst with a mean value of 3.30. Significant differences were only found for the data glove and the sliders ($p = 0.018$).

Having participants assess their own results is, however, a subjective quality measurement. Since it is unclear which criteria the participants used and what factors influenced their ratings, such data should be interpreted with caution. We therefore decided to complement the partipants' subjective ratings by objective quality measurements. In particular, we computed to what extent a facial expression created by the subjects deviated from a reference facial expression. To this end, we created for each of the nine tasks a standard expression based on the action units that were listed for the corresponding photo in the FER database. We then calculated the deviation of the facial expressions created by the participants from the corresponding standard reference expression using the following formula

$$Deviation = \sum_i |AU_{ref,i} - AU_{user,i}|$$

where AU is a floating value between 0 and 1 and i is the index of all action units. Using this formula, we obtained the following mean values for the deviation of user-generated facial expressions from the corresponding standard expressions: 4.26 for the gamepad, 4.53 for the sliders and 4.94 for the data glove. Neither the value for the gamepad nor for the data glove was significantly different from the value for the sliders.

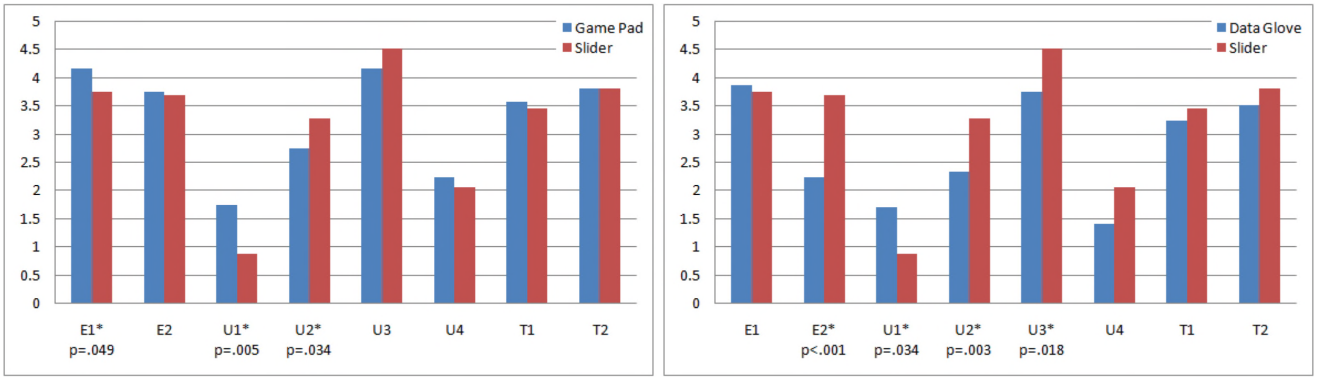

Figure 6. Subjective User Ratings. The slider-based interface compared with the gamepad (left) and the slider-based interface compared with the data glove (right).

When analyzing the time the participants spent on the creation of facial expressions, we found that they needed significantly less time with the gamepad (148.06 s) than with the sliders (168.29 s), while they needed significantly more time with the data glove (263.31 s). The time advantage for the gamepad was of about 12 % averaged over all values. For six out of seventeen users, the time advantage for the gamepad was even above 30 %, while just one user got a time advantage of above 30 % for the sliders. However, only the difference between the values for the data glove and the values for the sliders were significant ($p < 0.0005$), while the difference between the values for the gamepad and the values for the sliders were just tendentially significant ($p = 0.056$).

In addition to evaluating the performance of the single devices in terms of quality and time, we were interested in the participants' subjective impression. Overall, the gamepad achieved the best mean scores for most attitude statements. Figure 6 shows the results of two-tailed t-tests for each of the two hardware devices and the sliders. Results that were statistically significant are marked by a star. The findings of the experiment may be summarized as follows:

- *Interaction Experience:* The participants found it more enjoyable to use the gamepad and the data glove than to use the sliders (E1). However, they found the data glove physically more tiring than the sliders (E2).
- *Usability:* A major advantage of the novel input devices in comparison to the traditional sliders is that they allow the users to keep their eyes on their work. In the experiment, the participants had the feeling that they had to shift their gaze more often between the input devices and the characters when using the sliders than when using the gamepad or the data glove (U1). The sliders, however, scored best regarding the predictability (U2) and the plausibility (U3) of the devices' behavior. Compared to the sliders and the data glove, the gamepad enabled better tuning. The participants found it less difficult to adjust parameters with the gamepad than with the other two devices (U4).
- *Technical Features:* The participants had the impression that the gamepad offered them more options to adjust parameters than the other two devices (T1) and were more satisfied with the accuracy of the gamepad than with the accuracy of the data glove (T2).

Discussion

Overall, the use of a gamepad for facial expression generation can be regarded as promising. It reduced the production time without causing a loss of quality. This result is all the more remarkable as the gamepad hardware is obviously not adjusted to the specific requirements of facial animation design. Thus it came as no surprise that a large proportion of our users (49 %) expressed a preference for the gamepad. The sliders only scored better regarding predictability and plausibility, which could be explained by the fact that the sliders were labeled with the action units the user could manipulate.

The bad score of the data glove deserves further discussion. Even though data gloves were recommended as an input device for facial animation by Thalmann [27] and already used in production by Quantic Dreams, they obtained a significantly lower rating on almost all attitude statements. Furthermore, it took our users significantly longer to come up with a result than with any of the other devices, and the quality of the result was significantly lower. The only advantage found over traditional input devices was that data gloves allow the users to direct their gaze fully onto their work and do not require them to permanently shift their gaze between the interface device and the graphical display. Tthe low wearing comfort and insufficient accuracy of the very low-priced hardware may explain the poor ratings. Furthermore, moving five fingers in parallel might have been too difficult for unexperienced users. Finally, pressing buttons with the left hand was most likely too complicated and caused an interruption of the work flow. Nevertheless, the data glove should not be discarded as a completely inoperative input device. After all, 24% of the users preferred it as a controller – nearly as many as those who chose the definitely superior sliders. One of the users indicated that the data glove offered him the maximum amount of parallel control over the facial action units.

CONCLUSION

In this work, we investigated three different interfaces to a FACS-based animation system. Based on ergonomic principles, we defined three mapping strategies to assign facial actions to controls and showed how they could be applied to gamepads and data gloves. The appropriateness of the mapping strategies was investigated by conducting interviews with professional graphics designers. Based on these studies, we tested the most promising mapping strategies for the gamepad and the data glove in an experiment with non-experienced users. The users had to accomplish various tasks which were evaluated based on time and quality. The users were not only satisfied with the facial expressions they created. In addition, there was a high congruence between the users' creations and the corresponding standard reference expressions. A comparison of the novel hardware devices with the conventional sliders revealed that the gamepad scored best on most dimensions. It helped reduce production time without loss of quality.

Our work was based on existing hardware controllers. An interesting idea would be to create new hardware controllers that are especially adapted to the specific requirements of facial animation generation. In this case, we even expect a greater advantage of the gamepad over the other two interfaces. The future lies most likely in a combination of different interfaces. The wish to combine various interfaces was also uttered by our users when testing the mapping approach that was based on basic emotion categories.

ACKNOWLEDGMENTS

This work has been funded in part by the European Commission via the CALLAS Integrated Project. (ref. 034800, `http://www.callas-newmedia.eu/`).

REFERENCES

1. Facial Expressions Primer from Half-Life 2 by Valve. `http://developer.valvesoftware.com/wiki/Facial_Expressions_Primer`.
2. I. Albrecht, M. Schröder, J. Haber, and H.-P. Seidel. Mixed feelings: expression of non-basic emotions in a muscle-based talking head. *Virtual Reality*, 8(4):201–212, 2005.
3. K. Balci. Xface: MPEG-4 based open source toolkit for 3D Facial Animation. In *AVI '04: Proceedings of the working conference on Advanced Visual Interfaces*, pages 399–402. ACM, 2004.
4. D. Cage. Postmortem: Indigo Prophecy. `http://gamasutra.com/features/20060620/cage_01.shtml`.
5. E. Chang and O. C. Jenkins. Sketching articulation and pose for facial animation. In *SCA '06: Proceedings of the 2006 ACM SIGGRAPH / Eurographics symposium on Computer animation*, pages 271–280, 2006.
6. M. Courgeon, J.-C. Martin, and C. Jacquemin. User's gestural exploration of different virtual agents' expressive profiles. In *AAMAS '08: Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems*, pages 1237–1240, 2008.
7. R. Cowie, E. Douglas-Cowie, S. Savvidou, E. Mcmahon, M. Sawey, and M. Schröder. Feeltrace: an instrument for recording perceived emotion in real time, 2000.
8. P. Ekman and W. Friesen. *Unmasking the Face*. Prentice Hall, 1975.
9. P. Ekman, W. Friesen, and J. C. Hager. *Facial Action Coding System*. Research Nexus eBook, 2. edition, 2002.
10. Facial Expression Repertoire (FER) `http://research.animationsinstitut.de/`.
11. M. Frydrych, M. Dobsik, J. Kätsyri, and M. Sams. Toolkit for animation of finnish talking head. In *ISCA Tutorial and Research Workshop on Audio Visual Speech Processing (AVSP'03)*, pages 199–204. International Speech Communication Association, 2003.
12. Horde3D Graphics Engine `http://horde3d.org/`.
13. C. Jacquemin. Pogany: A tangible cephalomorphic interface for expressive facial animation. In *ACII '07: Affective Computing and Intelligent Interaction*, pages 558–569, 2007.
14. J. Kätsyri. *Human Recognition of Basic Emotions from Posed and Animated Dynamic Facial Expressions*. PhD thesis, Helsinki University of Technology, 2006.
15. J. P. Lewis, J. Mooser, Z. Deng, and U. Neumann. Reducing blendshape interference by selected motion attenuation. In *I3D '05: Proceedings of the 2005 symposium on Interactive 3D graphics and games*, pages 25–29, New York, NY, USA, 2005. ACM.
16. S. McCloud. *Making Comics: Storytelling Secrets of Comics, Manga and Graphic Novels*. Harper Paperbacks, 2006.
17. G. Nataneli and P. Faloutsos. Sketching facial expressions. In *International Conference on Computer Graphics and Interactive Techniques*, page 60. ACM, 2007.
18. C. Oat. Animated wrinkle maps. In *SIGGRAPH '07: ACM SIGGRAPH 2007 courses*, pages 33–37, New York, NY, USA, 2007. ACM.
19. M. Ochs, R. Niewiadomski, C. Pelachaud, and D. Sadek. Intelligent expressions of emotions. In *Affective Computing and Intelligent Interaction (ACII)*, 2005.
20. I. S. Pandzic and R. Forchheimer. *MPEG-4 Facial Animation: The Standard, Implementation and Applications*. Wiley, 2002.

21. F. Parke. Parameterized models for facial animation. *IEEE Computer Graphics and Applications*, 2(9):61–68, 1982.

22. S. Pasquariello and C. Pelachaud. Greta: A simple facial animation engine. In *6th Online World Conference on Soft Computing in Industrial Applications*, 2001.

23. Z. Ruttkay, H. Noot, and P. Hagen. Emotion disc and emotion squares: Tools to explore the facial expression space. *Computer Graphics Forum*, 22, 2003.

24. M. Sagar. Facial performance capture and expressive translation for king kong. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Sketches*, page 26, New York, NY, USA, 2006. ACM.

25. J. Spencer-Smith, H. Wild, A. Innes-Ker, J. Townsend, C. Duffy, C. Edwards, K. Ervin, N. Merritt, and J. Paik. Making faces: Creating three-dimensional parameterized models of facial expression. *Behavior Research Methods, Instruments & Computers*, 33:115–123, 2001.

26. T. Sucontphunt, Z. Mo, U. Neumann, and Z. Deng. Interactive 3D facial expression posing through 2D portrait manipulation. In *GI '08: Proceedings of graphics interface 2008*, pages 177–184, Toronto, Ont., Canada, Canada, 2008. Canadian Information Processing Society.

27. D. Thalmann. Using Virtual Reality Techniques in the Animation Process. *Virtual Reality Systems (Earnshaw R, Gigante M, Jones H eds), Academic Press*, pages 143–159, 1993.