Mouthbrush: Drawing and Painting by Hand and Mouth

Chi-ho Chan
ATR Media Information Science Labs
2-2-2 Hikaridai, Seika-cho,
Soraku-gun, Kyoto, Japan 619-0288

Michael J. Lyons
ATR MIS Labs
2-2-2 Hikaridai, Seika-cho,
Soraku-gun, Kyoto, Japan 619-0288

Nobuji Tetsutani
ATR MIS Labs
2-2-2 Hikaridai, Seika-cho,
Soraku-gun, Kyoto, Japan 619-0288

E-mail: mlyons@atr.co.jp

ABSTRACT
We present a novel multimodal interface which permits users to draw or paint using coordinated gestures of hand and mouth. A headworn camera captures an image of the mouth and the mouth cavity region is extracted by Fisher discriminant analysis of the pixel colour information. A normalized area parameter is read by a drawing or painting program to allow read-time gestural control of pen/brush parameters by mouth gesture while sketching with a digital pen/tablet. A new performance task, the Radius Control Task, is proposed as a means of systematic evaluation of performance of the interface. Data from preliminary experiments show that with some practice users can achieve single pixel radius control with ease. A trial of the system by a professional artist shows that it is ready for use as a novel tool for creative artistic expression.

Categories and Subject Descriptors
I.3.1 [**Computer Graphics**]: Hardware Architecture – *Input devices.* I.5.5 [**Pattern Recognition**]: Implementation – *Interactive systems.*

General Terms
Performance, Design, Experimentation, Human Factors

Keywords
Vision-based interface; mouth controller; alternative input devices.

1. INTRODUCTION
Drawing and painting are expressive forms of human communication in which hand gestures create a lasting impression on media such as paper, canvas, or an electronic image by means of a pen, brush, or digital tablet. We are exploring the extension of this concept to allow the creation of expressive digital images using coordinated gestures of hand and mouth. The mouth is involved in many human activities and the associated sensorimotor control systems allow fine motor control. The central role of the mouth in facial expression and speech production, especially, inspired us to consider its use for artistic expression.

Our desire for a non-invasive, comfortable, and low-cost solution led us to select computer vision as a promising candidate for real-time capture of mouth gestures. A simple vision algorithm is used to measure the area of the open mouth. This single parameter may be mapped to control the size, hardness, opacity, colour or other property of the pen or brush in real-time as the artist paints or draws.

An outline of this paper is as follows: in section 2 we briefly describe related work; in section 3 we discuss design considerations and give details of the hardware and software implementation; section 4 introduces a novel task for evaluating performance of the human/machine interface, the Radius Control Task, along with preliminary results of user studies; in section 5 we describe how the mouth controller is used together with the GNU Image Manipulation Program (GIMP) and a pen/tablet device for creative drawing/painting. The paper concludes with suggestions for future work.

2. RELATED WORK
Several prior works have described devices for capture of mouth gestures for use as computer input.

Salem and Zhai [5] developed *Tonguepoint,* a Tongue pointing device in which a small pressure sensitive isometric joystick is mounted on a mouthpiece held with the teeth. This allows the user to control a cursor with their tongue.

Orio [3] described a system which monitors the internal shape of the oral cavity by analyzing the response to stimulation with an acoustic source. It was found that users could easily control two independent parameters by varying their mouth shape.

Lyons [2] developed the *Mouthesizer*, a vision-based system for extracting shape parameters of the external appearance of the mouth cavity such as height, width, and aspect ratio and mapping these to musical control parameters.

Vogt et al. [7], with their *Tongue 'n' Groove* system, used ultrasound imaging to measure the tongue position and movement in real time for real-time sound synthesis control.

3. IMPLEMENTATION
3.1 Hardware
Figure 1 shows the set up. A head mounted camera is used to acquire an image of the mouth region. This reduces the complexity of the vision task by eliminating the need to detect

ICMI'03, NOVEMBER 5–7, 2003, VANCOUVER, BRITISH COLUMBIA, CANADA.

and track the head. Moreover it affords the user greater mobility and comfort than if they had to face a fixed remote camera.

The headset is a modified Shure SM10A with microphone and beam assembly replaced by a USB web camera on a flexible arm. The flexible arm lets users easily adjust the camera position. The FlexiC@m B100 USB camera used in this work is very small: the 1/5” CMOS imaging sensor and the USB interface are built into the same circuit board which has dimensions 45×18 mm. The camera is powered by the USB port.

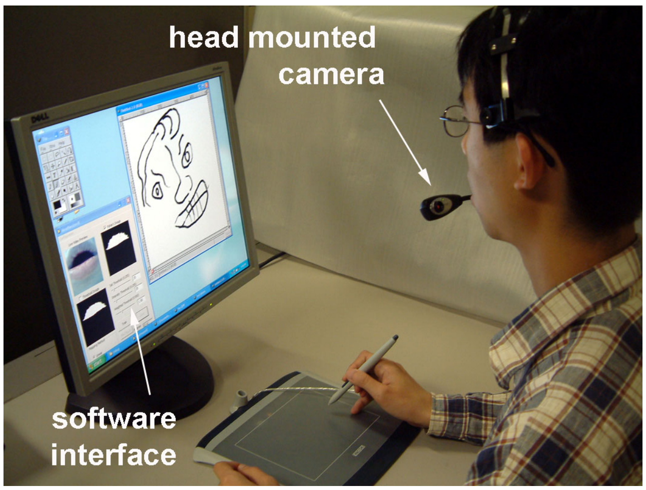


Figure 1. Drawing with the mouth controller and a digital tablet.

3.2 Software

The input image is reflected about the vertical axis to allow for intuitive adjustment of camera position by the user. The captured image is converted to a binary array corresponding to pixels belonging to the mouth cavity or other regions. The area of the open mouth is computed by summing the mouth cavity pixels. This area is normalized and sent via TCP/IP to an application such as a drawing program. The system block diagram is shown in Figure 2.

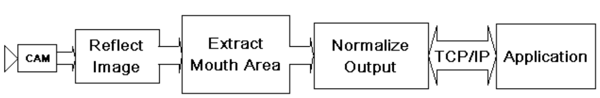


Figure 2. Block diagram of the Mouthbrush system.

The software was developed in Visual C++, under Windows XP, using DirectShow. The system runs as a self-contained application under Windows XP or 2000.

3.2.1 Colour Segmentation of the Mouth Cavity

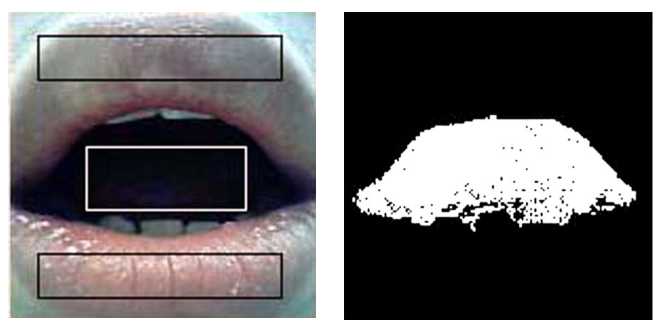
There is a significant body of work on algorithms for recognizing facial action [4]. Here, it is important to realize that the task is not to recognize the shape of the mouth, but to devise a system which provides a smooth, repeatable, and intuitive function of mouth shape. Here we choose to extract the apparent area of the mouth cavity in the image using a segmentation algorithm based on colour.

Colour is known to be a useful feature for segmenting human skin, as for example in detecting the face [6]. Our approach is to use Fisher Discriminant analysis [1] in a two dimensional colour space to determine a decision boundary mouth cavity and other regions of the input image. We first empirically determined that the red component (R) and total intensity ($I = R+G+B$) provided a representation of the input colour information suitable for detecting mouth cavity pixels. Intensity is useful because the mouth cavity appears as a shadow under usual lighting conditions. The inside of the mouth is also characterized by a relatively high red component.

The system has two modes: a training mode in which the discriminant analysis is initialized and run and output normalization facts determined, and an operating mode during which the system acts as a TCP/IP server outputting mouth cavity area in real-time.

When training mode is entered, the user positions their mouth cavity region in a central rectangular area such that skin areas occupy two flanking rectangles (see Figure 3a). (R,I) values from pixels in the two classes (cavity and non-cavity) are used to calculate a discriminant vector using Fisher analysis [1]. The position of the decision boundary between cavity and non-cavity pixels is decided using a threshold determined from the mean μ_m and standard deviation σ_m of projection of the mouth cavity *(R,I)* values onto the discriminant vector and the corresponding mean μ_s for non-cavity pixels.

In operating mode, *(R,I)* values of pixels in the input image are classified as cavity or non-cavity using the Fisher discriminant classifier. Each pixel in the image is then labeled as cavity or non-cavity. The labeled mouth cavity image is shown in Figure 3b.



(a) **(b)**

Figure 3. a) Mouth training image, b) Segmented image.

3.2.2 Output Normalization

The area of the mouth cavity is the number of pixels, M, in the image labeled as belonging to the mouth cavity. This parameter is invariant with regards to small translations or rotations of the camera, making it insensitive to small movements or vibration of the camera. The area is normalized according to the maximum and minimum mouth cavity areas, M_{max} and M_{min}:

$$A = \begin{cases} 0 & M < M_{min} \\ \dfrac{M - M_{min}}{M_{\max} - M_{min}} & \\ 1 & M_{max} < M \end{cases}$$

so that the output parameter $A \in (0,1)$. The user inputs maximum and minimum mouths in training mode.

3.2.3 Gain

The normalized area, A, is mapped to a pen parameter, P, which can be used to control size, hardness, opacity, colour etc…

$$P/P_{\max} = gA$$

where $P_{\max}$ is the maximum pen parameter. ($P/P_{\max}$), A, and the gain, g, are dimensionless parameters taking values between 0 and 1. With a larger value for g, the user need not open the mouth as much to input a given pen parameter, P and the system should require less effort to operate. However, if g is too large we expect it will be difficult to control P because involuntary movements of the mouth will be amplified.

4. EVALUATION

To investigate the expected trade-off between gain and accuracy of control we devised an evaluation task: the radius control task. This section reports results of our preliminary experiments with the mouth controller using this task.

4.1 Radius Control Task

In this task, the experimental subject uses the mouth controller and opens/closes their mouth to adjust the radius of circle drawn on the computer display until it is exactly tangent to a target square (Figure 4).

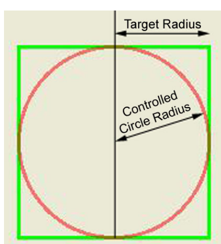


Figure 4. The Radius Control Task.

When the target size is reached, the user clicks the space bar and attempts to hold the same position for 3 seconds, whence user is prompted to begin the next trial.

4.2 Experimental Procedure

The independent variables of the experiment are the gain, g, and the size of the target square. The half side of the square (= target radius) took values (5, 23, 42, 61, 80) pixels and g took values (0.25, 0.5, 0.8). Each combination of gain and target size was repeated three times for a total of 45 trials. The order of presentation of trials was randomized for each experimental subject. Users were free to rest between trials. Subjects sat comfortably at a distance of approximately 60 cm from a 19" LCD having a resolution of 1280x1024 pixels. Completion of the experiment took about 10 minutes.

Four volunteers participated in the experiment. Two had some prior experience using the mouth controller and the other two were naïve subjects.

For each trial, pressing the space bar causes the computer to record the radius of the controlled circle at 67 Hz for the subsequent 3 seconds. The standard deviation, σ, from the mean position was calculated as a measure of precision of user control. The mean absolute error, E, (difference between target and actual radius) was calculated as a measure of the accuracy of control.

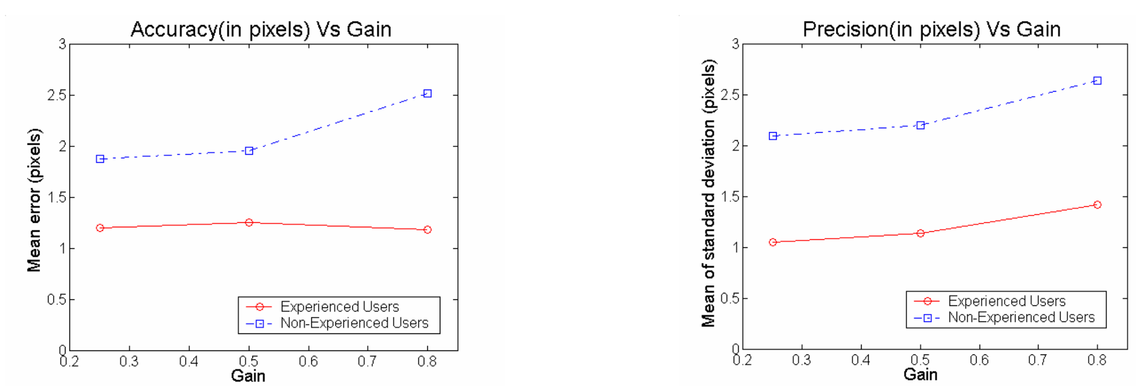
$$E(x,y) = \sum_{i=0}^{n} \left| C(x,y)_i - T(x,y) \right| \Big/ n$$

$$\sigma(x,y) = \sum_{i=0}^{n} \left(C(x,y)_i - \overline{C}(x,y) \right)^2 \Big/ n(n-1)$$

where x and y are indices of gain and target parameters, n is the total sample number in each target test. T is the target radius and C is the input circle radius. $\overline{C}$ is the mean of input circle radius.

5. Results

Figure 5a plots the mean absolute error, averaged over all target sizes for each value of the gain, while figure 5b plots the standard deviation averaged over target sizes for each g. Averages for experienced and inexperienced users are shown separately. Both the mean absolute error and standard deviation were greater for inexperienced users demonstrating performance improvement as a result of practice with the mouth controller. Accuracy was not strongly dependent on gain for experienced users whereas for inexperienced users lower gain resulted in higher accuracy.



(a) (b)

Figure 5. Graphs of user accuracy and precision for the radius control task.

Precision, a measure of how well the user can keep the input radius at a fixed value was higher at lower gain for both experienced and inexperienced users.

These results suggest that, overall, the lower gain setting, g = 0.25, resulted in the greatest accuracy and precision of control of the input radius.

6. APPLICATION

We studied an example application in which the mouth controller was used to control pen and brush parameters for creative drawing and painting.

Communications between the mouth controller and applications were implemented using Windows sockets and the TCP/IP protocol, with the mouth controller acting as a server providing the product gA. The GNU Image Manipulation Program (GIMP) was modified to act as a TCP/IP client able to connect to the server and read the control parameter, gA, using it to set pen or brush parameters such as size, hardness, opacity, and colour. A Wacom Graphire drawing tablet was used with the system as illustrated in figure 1.

A professional artist used the system to sketch and paint for approximately two hours. One of the pictures created during the trial period is shown in figure 6. In this picture mouth gestures were used to vary brush size and hardness while the hand controlled the direction of the line. The artist's overall impression after using the interface was that it allowed an interesting and challenging new technique for expressive drawing.

Figure 6. A digital painting created using the Mouthbrush. (Artist: Mao Makino)

7. CONCLUSIONS

In this work we have developed a new interface, the *Mouthbrush*, which allows drawing and painting using coordinated mouth and hand gestures. A novel method was introduced for automatically setting the colour thresholds by training a Fisher discriminant vector on mouth cavity and skin regions. We also introduced the radius control task for systematic evaluation of the performance of the human-machine interface. Results of user experiments with the task suggested that precision and accuracy of performance improve with practice using the mouth controller and also that lower gain settings are preferable. Users with some practice using the mouth controller were able to obtain single-pixel control of input radius with ease.

Trials of the system by an experienced artist showed that the system is sufficiently developed to be used comfortably as a novel means of creative artistic expression.

In future work we plan to extend the Mouthbrush to output 2 or more mouth shape parameters. We also wish to study the mapping issues more extensively and conduct further user studies with both experienced and beginning artists.

Updated information about the progress of the Mouthbrush project will be made available at the web site:

http://www.mis.atr.co.jp/~mlyons/mouthbrush.html

8. ACKNOWLEDGMENTS

We thank Mao Makino for contributing her artistic skill, Nicolas Fay and Mei-gi Lee for participating in the experiment. This work was supported by the Telecommunications Advancement Organization of Japan.

9. REFERENCES

[1] Duda, R. and Hart, P. Pattern Classification and Scene Analysis. New York: Wiley, 1973.

[2] Lyons, M. J. and Tetsutani, N. Facing the Music: A Facial Action Controlled Musical Interface. In CHI 2001 Extended Abstracts, ACM Press, 309-310.

[3] Orio, N. A Gesture Interface Controlled by the Oral Cavity, in Proc. of International Computer Music Conference, 1997, pp. 141-144.

[4] Pantic, M. and Rothkrantz, L.J.M. Automatic Analysis of Facial Expressions: The state of the Art. IEEE PAMI 22: 1425-1445, 2000.

[5] Salem, C., Zhai, S. An Isometric Tongue Pointing Device, Proceedings of CHI'97, 538-539.

[6] Terrillon, J.-C., Shirazi, M.N., Fukamachi, H. and Akamatsu, S. Comparative Performance of Different Skin Chrominance Models and Chrominance Space for the Automatic Detection of Human Faces in Color Images. Proc. Of Automatic Face and Gesture Recognition, 54 -61, 2000.

[7] Vogt, F., McCaig, G., Ali, M.A. and Fels, S. Tongue 'n' Groove: An Ultrasound based Music Controller. In Proc of the 2002 Conference on New Interfaces for Musical Expression (NIME-02), 60-64.