Expression Glasses: A wearable device for facial expression recognition

Jocelyn Scheirer, Raul Fernandez, Rosalind W. Picard
MIT Media Laboratory
Room E15-383, 20 Ames Street, Cambridge, MA 02139-4307
{rise, galt, picard}@media.mit.edu

ABSTRACT
Expression Glasses provide a wearable “appliance-based” alternative to general-purpose machine vision face recognition systems. The glasses sense facial muscle movements, and use pattern recognition to identify meaningful expressions such as confusion or interest. A prototype of the glasses has been built and evaluated. The prototype uses piezoelectric sensors hidden in a visor extension to a pair of glasses, providing for compactness, user control, and anonymity. On users who received no training or feedback, the glasses initially performed at 94% accuracy in detecting an expression, and at 74% accuracy in recognizing whether the expression was confusion or interest. Significant improvement beyond these numbers appears to be possible with extended use, and with a small amount of feedback (letting the user see the output of the system).

Keywords
Wearable computing, affective computing, input devices, human-computer interaction, facial expressions

INTRODUCTION
Human facial expression is a vital and efficient means of exchanging information in conversation, communicating messages such as interest or confusion, approval or disapproval, and a variety of other so-called “basic emotions,” all while operating in parallel with language. There have been a number of efforts to give computers the ability to recognize facial expressions and other expressions of human affective state [1]. Facial expression recognition by computer has been dominated by a computer vision approach whereby a video camera, with a general-purpose computer, records and analyzes a sequence of images of a face. These systems perform in the 80-98% range when choosing among a set of six exaggerated “basic emotion” expressions, but do not run in real time, and have not been tested on their ability to detect expressions such as confusion (C) and interest (I).

A wearable appliance-style recognizer is limited in certain ways, but also has several advantages over a general purpose computer-based recognition system. Currently, the Expression Glasses are not able to image the whole face, although future sensing technology could enable this. Instead, the glasses discriminate signals involving motion around the eyes. One might think that an “off-board” system is preferable because the user doesn’t have to wear anything. However, when considering privacy and control, glasses offer an important advantage. A user can easily remove eyeglasses or disable their sensor, whereas it is virtually impossible for a user to disable sensing done in most “smart environments” by cameras and computers hidden behind walls. Also, the fact that the glasses only sense certain muscle movements, and cannot sense identity or other appearance characteristics is an advantage in many situations. Because glasses are a personal item, like jewelry or clothing accessories, they offer a fundamentally more comfortable, adaptable, and controllable interface. Another advantage is that glasses can be used anywhere, especially in ambulatory wearable systems; they are not restricted to installations with fixed cameras and lighting.

APPARATUS

Ordinary glasses have been modified to include a 3-panel vinyl extension with vinyl piping and reinforcement, attached to the frame with stitching and heavy-duty glue. An optional “privacy” visor is illustrated above. The sensors are two small pieces of piezoelectric film developed by AMP, connected via a 30mm electrode plug to a standard ribbon wire. The sensors are connected to a Dell PC running Windows 95 via a multichannel digital I/O board (ComputerBoards, Inc.) The piezo eye sensors are easily snapped on and off of the glasses as needed, and the connecting wires are tucked behind the user’s ears. A wireless version is a future possibility.

The system software is implemented in LabView. The software is trained on each user by having the user make 5 expressions from each class (C and I). Following median filtering of each channel to reduce noise, and fully rectifying the signal (taking its absolute value), the 5 highest peaks in the training data are found by fitting quadratic polynomials. The heights of these peaks from the two channels provide five 2D feature vectors x = col(x1, x2) for each class. The system fits a Gaussian to
each class by estimating the sample mean and covariance of the class features. In the real-time testing stage, the system, using a moving 3 sec. window, applies an equal-
prior likelihood ratio test [2], and, if the detected peaks

![Figure 1: Recognition of expressions (left bar shows confusion level, right bar shows interest level)](image)
exceeds a preset threshold (150 in this implementation),
the data vector \(x\) is classified to the class \(k\) that
maximizes the class conditional distribution.

**USER TESTING**

Eight subjects put on the glasses and made a sequence of
expressions of C and I. The first 10 of these were used
to train the system, and the second 12 of these were used
to test the recognition system. The order of the
expressions was varied randomly across users during
test, and contained 6 of each class. During testing, users
were given no feedback about how well they were
making the expressions or how well the system worked
on them.

One of the issues we confronted was that of sensor
placement because of variability across subjects' facial
expressions. We explored two fixed settings: one in
which the sensors were placed parallel to each other on
top of the corrugator and frontalis muscles [3], and one in
which they were offset to cover a wider area on the
forehead. Holding the setting constant, the recognition
accuracy on the subset of detected expressions was 62% and
70% respectively. However, when the experimenter
was allowed to select the setting that gave best subject-
dependent performance, this recognition rate increased to
74%. The tables below summarize the results of the
detection and recognition system for the latter case.

During the experimental session, the system gathered a
total of 190 3 sec. frames. For each frame, when an
expression was detected, the system classified it as either
a C or an I expression (right-hand table). However,
because the system doesn't detect the expressions only
when they're made, the left-hand table includes figures
for the number of times an expression was made \(E\) but
none \(N\) was recognized, or alternatively no expression
was made, but the system recognized one. If one
considers the subset of detected expressions, the
recognition rate of the system is 74%. Taking into
account all frames to evaluate the detection performance,
the system performs at 94%.

From experience with the glasses, we expected
performance to improve if the user saw feedback from
the system. In one case, we took a user who initially had

<table>
<thead>
<tr>
<th>Made</th>
<th>Detected</th>
<th>Recognized</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E 74</td>
<td>C 20</td>
</tr>
<tr>
<td></td>
<td>N 105</td>
<td>I 35</td>
</tr>
</tbody>
</table>

slightly above random recognition accuracy (57%), and
exposed him to a minute of feedback, during which he
made expressions and saw the system's response. Then,
his performance was re-measured (without his getting to
see feedback from the system.) Accuracy jumped to 81%.
It is reasonable to expect that individuals will make
expressions in different ways, and that the best
performance will be attained as the eyeglasses learn an
individual's pattern of expression over time, including
how that pattern may vary with context.

**IMPLICATIONS FOR USE**

A wearable expression-sensing appliance has many
applications. One example is feedback on one's own
emotions; for example, a practice session for certain
professions (such as counseling), where individuals are
trained specifically to refrain from expressing negativity.
For human-to-human communication, a device like this
allows a video lecturer access to the confusion and
interest levels of her students in a remote location,
providing a "barometer" of collective emotional
expression. Use of a device like the glasses gives
students an opportunity to communicate low-bandwidth,
but key information about their experience in a
nondistracting way, while concentrating on the lecture. The
anonymity provided by the visor-option may be
particularly useful in classrooms, focus groups, or other
situations where individuals might otherwise feel
inhibited about communicating negative emotions.

**CONCLUSIONS AND FUTURE WORK**

Expression glasses are a new, wearable, special-purpose
device designed to detect and recognize certain facial
expressions and to communicate these to a computer,
software agent, or to other people via networked
technology. An initial prototype has been built and
tested, and has attained significantly better than random
recognition accuracy, especially as users are given a
small amount of feedback about how the device works.
Future work includes improving upon the sensing
technology and pattern recognition, visualization of the
results for larger groups, and long term evaluation of use.

**REFERENCES**

Press, Cambridge, MA
Classification*, John Wiley & Sons.
The Skeletomotor System. in Cacioppo, J. &
Tassinary, L. (Eds.) *Principles of Psychophysiology:*
Physical, social and inferential elements. Cambridge