FAIM: Integrating Automated Facial Affect Analysis in Instant Messaging

Rana El Kaliouby and Peter Robinson
Computer Laboratory
University of Cambridge
Cambridge, CB3 0FD UK
+44 1223 763669
{rana.elkaliouby, peter.robinson}@cl.cam.ac.uk

ABSTRACT
One of the limitations in traditional instant messaging platforms is that they predominantly rely on text messages as the primary form of expression. This paper presents FAIM, an instant messaging application that analyzes a person's facial affect in real time and augments the dialogue with an emotive character representing them. Throughout the paper, we identify a number of design challenges that arise from integrating facial affect into instant messaging, and discuss how each of these issues is addressed in the design of FAIM. We also present a use case scenario of how FAIM works.

Categories and Subject Descriptors
H.1.2 [**User/Machine Systems**]: Human factors. H.5.3 [**Group and Organization Interfaces**]: Synchronous interaction. I.2.10 [**Vision and Scene Understanding**]: Perceptual reasoning, video analysis.

General Terms
Design, Human Factors

Keywords
Affective Interfaces, facial affect analysis, facial expression, emotionally intelligent applications, instant messaging

INTRODUCTION
Instant messaging is quickly becoming one of the most prevalent tools for communicating with peers over the Internet. It is being used to support collaborative work, to coordinate social meetings, and to keep in touch with friends and family [5,6]. Essentially text-based, IM dialogues are commonly augmented by the use of "emoticons" as a means of expressing emotions. Although emoticons are quite useful, they provide a very limited means of expressing emotion. They do not capture the dynamics underlying the emotional processes, and have to be explicitly inserted in-line by the user, removing the spontaneity of affective interactions.

A number of research efforts are underway with the aim of integrating various affective channels into IM dialogues. Conductive chat [3] incorporates users' fluctuating skin conductivity levels into the dialogue interface. Bodine and Pignol [2] develop a kinetic typography-based instant messenger. EmpathyBuddy [7] gives the user automatic affective feedback by displaying different emotion faces to match the affective context of the story being told through the user's email. Although EmpathyBuddy is an email application, the very encouraging results reported in their user studies are equally applicable to IM. In addition, extra functions are being built into IM applications in order to allow users to express their identity. In rear view mirror [5], the identity of other users is shown by means of a thumbnail picture, and the presence of each user is indicated by the border color around the picture. In the latest release of Microsoft's MSN Messenger®, users represent themselves by displaying a picture of their choice that appears in the conversation window.

Facial expressions provide an important spontaneous channel for the communication of both emotional and social displays [1]. They are used to communicate feelings, show empathy and acknowledge the actions of other people[1]. In this paper, we apply our previous work on automated real time facial affect inference [4], in instant messaging. The objective of FAIM is to provide a natural yet powerful means of communicating affect by enabling users to represent their affective state through an emotive virtual character. Our hypothesis is that since facial expressions are to a large extent communicative, including them in computer-mediated communication would enhance

IUI'04, Jan. 13–16, 2004, Madeira, Funchal, Portugal.
ACM 1-58113-815-6/04/0001.

[1] Throughout this paper, we will use the term "affective" state to cover all the possible functions of facial displays.

the communication process, and ultimately increase its effectiveness.

OVERVIEW OF FAIM

FAIM is an IM application, based on Microsoft's MSN Messenger®. In this IM platform, users communicate through a conversation window, where each user can display a picture representing them (e.g. a personal photo or a virtual character) in their personal "tile". That picture however is static and does not change irrespective of the user's affective state.

FAIM provides a means to communicate the user's current affective state to people on his/her contact list, by automatically changing the emotive content of that display area using information from the user's facial expressions.

The following section identifies several issues that arise when attempting to integrate facial affect into IM applications. We then present how FAIM addresses each of those.

DESIGN CHALLENGES

Functions of Facial Expression

Facial expressions have a number of functions; they convey inner feelings, are used to show empathy, are used to acknowledge actions of other people, and are communicative (e.g. turn-taking). FAIM must be able to function in situations when multiple facial gestures and actions are layered on top of each other (e.g. nodding in agreement to something just said, amidst a conversation where that person was mostly feeling upset).

Timing and Synchronization

Design issues associated with the timing and synchronization of an animation, such as when to display one, how long should it persist for, and how frequently should it be updated, have direct implications on usability. An affective display usually lasts around 6-8 seconds on average. It would only distract (and indeed frustrate!) both parties in a conversation if new animations are displayed with every inference. Similarly important, is how soon after a different emotional state is detected should a change be made, and when is it too late to react to a recognized emotional state.

Multitasking and Multiple Conversations

Studies by Isaac *et al.* [6] have shown that in 85% of conversations, people multitask while using IM, performing other computing tasks. They have also shown that around 23% were engaged in multiple simultaneous IM conversations. In order to ensure that the displayed emotive character is relevant to an active conversation, synchronizing between the facial affect analyzer and the different applications is an important design consideration.

FUNCTIONALITY OF FAIM

FAIM consists of three main modules. The facial affect analyzer is responsible for inferring affective information from the face. The classifier filters the inferences made by the analyzer, and is responsible for taking decisions related to timing and synchronization of affective states. Finally, the animator updates the user's display to reflect the current emotive state.

The Facial Affect Analyzer

Video sequences of the user are input to the facial affect analyzer in real time. The analyzer identifies facial feature points on the face using prior knowledge about faces. Shape parameters (e.g. mouth width), color information (e.g. presence of teeth), and head rotation parameters are extracted and used to infer facial actions (e.g. frown) and head gestures (e.g. head nod). A screenshot showing feature landmarks, color information and lip contours is shown in Figure 1. Concurrent facial and head displays are combined into micro expressions, which typically last between two to four seconds. Temporal sequences of micro expressions portray an affective state. Our system currently supports the following states: happy, surprised, agreeing, disagreeing, confused, indecisive, and neutral.

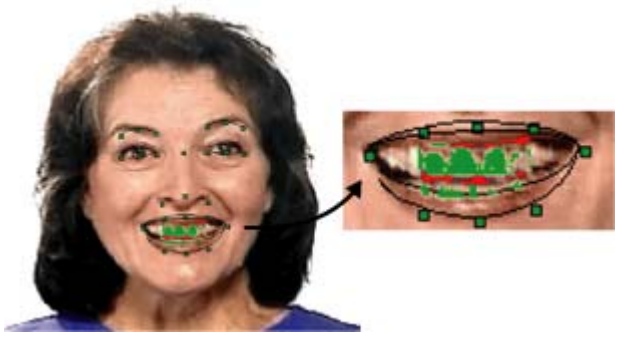

Figure 1: A screenshot of the facial affect analyzer (left). A close up of the mouth feature points, the inner and outer lip contours, and teeth / aperture color information (right).

Affective State Manager

While the analyzer produces an inference every six to eight seconds on average, updating an emotive display with this frequency would distract users. Hence, this module manages affective states over time to choose which emotion to display. This includes smoothing transitions from one emotional state to another and dealing with state decay.

The Animator

The affective state broadcasted by the manager, along with strength parameters are input into the animator. The animator superimposes the emotion onto the user's virtual character. Currently this is simply done by choosing the instance of the display picture or character with the matching emotion. The new character is then shown in the user's display area. Other possibilities involve using various morphing and animations techniques.

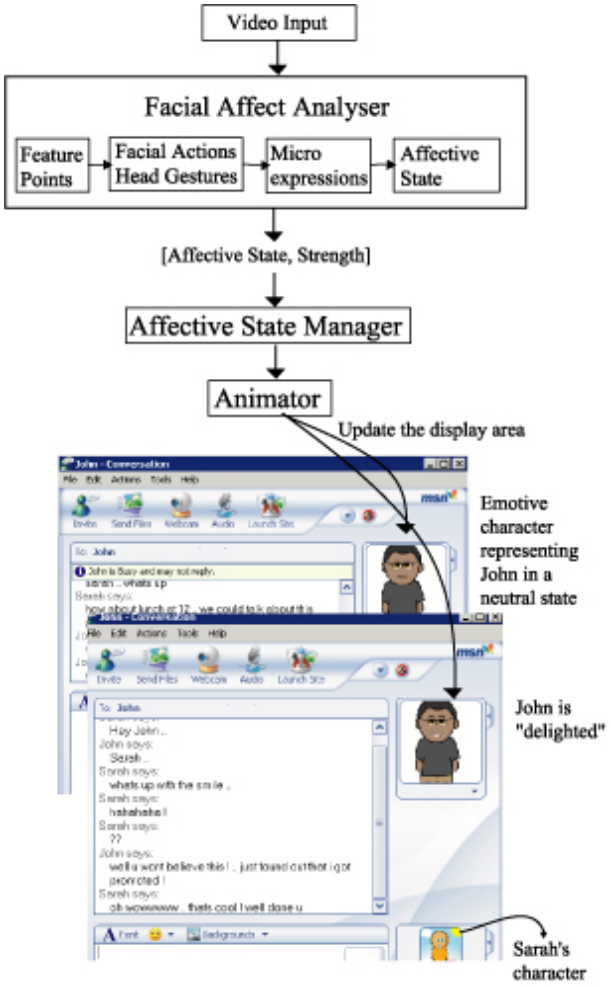

Figure 2: Example scenario showing how FAIM works.

Example Interaction Scenario

A schematic diagram of how FAIM works in a typical scenario is shown in Figure 2. We follow Sarah and John through three different stages of a conversation. John starts with his character displaying a neutral state (screenshot in the background). During the conversation, John gets an email from his manager at work informing him that he just got promoted. John is delighted with the news. The facial affect analyzer picks on this expression and sends it to the affective state manager, which records a change in affect and informs the animator module. The animator in turn picks a smiley version of John's character. John's display area gets updated (screenshot in the foreground). Sarah immediately notices that "John" suddenly looks much happier, and asks him about that. John is happy to share the news with Sarah. A few days later, John's delight fades away as he becomes more stressed at work, and his character is back to the neutral look once more.

DISCUSSION AND CONCLUSION

This paper reported work in progress on FAIM, an IM application that utilizes a real time automated facial affect analyzer to animate a virtual character representing the user. The emotive character is displayed in conversation windows of active instances of the IM application. The paper also presented a number of design challenges, and discussed how FAIM addresses each of those. Facial expressions play an important role in every day-to-day interaction. We hypothesize that integrating facial affect in IM would enhance the spontaneity of the communication and ultimately improve the overall effectiveness and satisfaction with the interaction. Our immediate next step is to perform user studies aimed at quantifying the impact of FAIM on empathy and social presence.

REFERENCES

1. Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., and Plumb, I. The "Reading the Mind in the Eyes" Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. *Journal of Child Psychology and Psychiatry,* 42:2 (2001), 241-251.
2. Bodine, K. and Pignol, M. Kinetic Typography-Based Instant Messaging, in *CHI'03 Extended Abstracts on Human Factors in Computer Systems* (2003), ACM Press, 914-915.
3. DiMicco, J.M., Lakshmipathy, V. and Fiore, A.T. Conductive Chat: Instant Messaging With a Skin Conductivity Channel. Poster Presentation in *ACM Conference on Computer Supported Cooperative Work* (2002), ACM Press.
4. El Kaliouby, R. and Robinson, P. Real time head gesture recognition in affective interfaces. In M. Rauterberg, M. Menozzi and J. Wesson (Eds.) *Human Computer Interaction Interact'03* (2003), IOS Press, 950-953.
5. Handel, M. and Herbsleb, J.D. What is Chat doing in the workplace? In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (2002), ACM Press, 1-10.
6. Isaacs, E., Walendowski, A., Whittaker, S., Schiano, D.J. and Kamm, C. The Character, Functions, and Styles of Instant Messaging in the Workplace, in *Proceedings of the ACM Conference on Computer-Supported Cooperative Work* (2002), ACM Press, 11-20.
7. Liu, H., Lieberman, H. and Selker, T. Automated Affective Feedback in an Affective Browser. MIT Media Laboratory, Software Agents Group, Technical Report (2002).