Chapter III
Facial Expression Analysis, Modeling and Synthesis: Overcoming the Limitations of Artificial Intelligence with the Art of the Soluble

Christoph Bartneck
Eindhoven University of Technology, The Netherlands

Michael J. Lyons
Ritsumeikan University, Japan

ABSTRACT

The human face plays a central role in most forms of natural human interaction so we may expect that computational methods for analysis of facial information, modeling of internal emotional states, and methods for graphical synthesis of faces and facial expressions will play a growing role in human-computer and human-robot interaction. However, certain areas of face-based HCI, such as facial expression recognition and robotic facial display have lagged others, such as eye-gaze tracking, facial recognition, and conversational characters. Our goal in this paper is to review the situation in HCI with regards to the human face, and to discuss strategies, which could bring more slowly developing areas up to speed. In particular, we are proposing the “The Art of the Soluble” as a strategy forward and provide examples that successfully applied this strategy.

INTRODUCTION

The human face is used in many aspects of verbal and non-verbal communication: speech, the facial expression of emotions, gestures such as nods, winks, and other human communicative acts. Subfields of neuroscience, cognitive science, and psychology are devoted to study of this information. Computer scientists and engineers have worked on the face in graphics, animation, computer vision, and pattern recognition. A widely stated motivation for this work is to improve hu-
man computer interaction. However, relatively few HCI technologies employ face processing (FP). At first sight this seems to reflect technical limitations to the development of practical, viable applications of FP technologies.

This paper has two aims: (a) to introduce current research on HCI applications of FP, identifying both successes and outstanding issues, and (b) to propose that an efficient strategy for progress could be to identify and approach soluble problems rather than aim for unrealistically difficult applications. While some of the outstanding issues in FP may indeed be as difficult as many unsolved problems in artificial intelligence, we will argue that skillful framing of a research problem can allow HCI researchers to pursue interesting, soluble, and productive research.

For concreteness, this article will focus on the analysis of facial expressions from video input, as well as their synthesis with animated characters or robots. Techniques for automatic facial expression processing have been studied intensively in the pattern recognition community and the findings are highly relevant to HCI (2004; Lyons, Budynek, & Akamatsu, 1999). Work on animated avatars may be considered to be mature (Cassell, Sullivan, Prevost, & Churchill, 2000), while the younger field of social robotics is expanding rapidly (Bartneck & Okada, 2001; Bartneck & Suzuki, 2005; Fong, Nourbakhsh, & Dautenhahn, 2003). FP is a central concern in both of these fields, and HCI researchers can contribute to and benefit from the results.

However, an examination of the HCI research literature indicates that activity is restricted to a relatively narrow selection of these areas. Eye gaze has occupied the greatest share of HCI research on the human face (e.g. (Zhai, Morimoto, & Ihde, 1999)). Eye gaze tracking technology is now sufficiently advanced that several commercial solutions are available (e.g. Tobii Technology (2007)). Gaze tracking is a widely used technique in interface usability, machine-mediated human communication, and alternative input devices. This area can be viewed as a successful sub-field related to face-based HCI.

Numerous studies have emphasized the neglect of human affect in interface design and argued this could have major impact on the human aspects of computing (Picard, 1997). Accordingly, there has been much effort in the pattern recognition, AI, and robotics communities towards the analysis, understanding, and synthesis of emotion and expression. In the following sections we briefly introduce the areas related to analysis, modeling and synthesis of facial expressions. Next, we report on insights on these areas gained during a workshop we organized on the topic. A gap between the available FP technology and its envisioned applications was identified, and based on this insight, we propose the “Art of the Soluble” strategy for FP. Last, we provide successful examples in the field of FP that took the Art of the Soluble approach.

**ANALYSIS: FACIAL EXPRESSION CLASSIFICATION**

The attractive prospect of being able to gain insight into a user’s affective state may be considered one of the key unsolved problems in HCI. It is known that it is difficult to measure the “valence” component of affective state, as compared to “arousal”, which may be gauged using biosensors. However, a smile, or frown, provides a clue that goes beyond physiological measurements. It is also attractive that expressions can be gauged non-invasively with inexpensive video cameras.

Automatic analysis of video data displaying facial expressions has become a topic of active area of computer vision and pattern recognition research (for reviews see (Fasel & Luettin, 2003; Pantic & Rothkrantz, 2000)). The scope of the problem statement has, however, been relatively narrow (Ellis & Bryson, 2005; Hara & Kobayashi, 1996; Shugrina, Betke, & Collomosse, 2006). Typically one measures the performance of a