The Emotional Hearing Aid: An assistive tool for Autism

Rana El Kaliouby and Peter Robinson

University of Cambridge Computer Lab
William Gates Building
JJ Thomson Avenue, Cambridge CB3 0FD, U.K.
rana.el-kaliouby@cl.cam.ac.uk and peter.robinson@cl.cam.ac.uk

Abstract

The ability to read the mind from the face is a cornerstone for the development of social functions and emotional intelligence in humans. People diagnosed with autism are thought to lack or have an impairment in those representational set of abilities. As a result they have difficulties operating in our highly complex social environment, and are for the most part, unable to understand other people's emotions. In this paper, we present the *emotional hearing aid*, an assistive tool designed for people diagnosed with a mild form of autism – Asperger's Syndrome. The application is implemented as an affective facial-expression recognition system, and is designed to operate on spontaneous human expression. Finally, we propose a tentative hardware and interface design, and discuss how the *emotional hearing aid* would be used in a typical social scenario.

1 Motivation

Social intelligence or "mind reading" encompasses our abilities to interpret others' behaviours in terms of mental states (thoughts, intentions, desires, beliefs), which we need to interact in a highly complex social world (Baron-Cohen et al., 2001; O' Connell, 1998). Those representational set of abilities, referred to as "Theory of Mind" allow an individual to attribute mental states to others, and understand their actions within an intentional or goal-directed framework.

While subtle and somewhat elusive, the ability to mind read is essential to the social functions we take for granted. In his book "Mindblindness: an essay on autism and theory of mind", Baron-Cohen (1995) explains how people mind read all the time, effortlessly, and mostly unconsciously. He contrasts that with the abilities of individuals with autism, who have difficulties operating in the highly complex social environment in which we live and are, for the most part, unable to understand other people's emotions. This lack of or impairment in a theory of mind cause many symptoms, which include insensitivity to other people's feelings, inability to take into account what other people know, and the inability to read and respond to people's facial expressions.

While a number of therapeutic and educational technologies exist for autism, there is an insufficiency of tools that assist people diagnosed with autism in social and emotional situations. We draw on the need for such assistive technologies to develop the *emotional hearing aid*. The application is designed to help people diagnosed with a mild form of autism (Asperger's Syndrome) with reading emotions from facial expression. We believe that introducing such a tool can enable people with emotion understanding difficulties to participate in more social

interactions. In a sense, the application is analogous to a hearing aid, which allows people with hearing problems to communicate with the rest of the world.

In the following section, we identify key educational and therapeutic technologies developed for autism, as those are the closest to the *emotional hearing aid* from an application point of view. In addition, we survey major approaches to facial expression recognition, as they lie closest to our research from a technical point of view.

2 Related Work

Much of our thinking about the *emotional hearing aid* is motivated by Picard's vision of affective computing (Picard, 1997; Picard and Wexelblat, 2002), and research on affective systems such as StartleCam (Healey and Picard, 1998). The AURORA project (Dautenhahn, 1999; Werry et al., 2001) utilizes socially intelligent agents in a number of different environments including educational and therapeutic ones for autism. Closely related are a number of researchers building theory of mind into humanoid robots as a tool to test and evaluate social developmental theories (Scassellati, 2000; Deak et al., 2001).

From a technical point of view though, facial expression recognition systems lie closest to our work. Automated facial expression systems are mostly concerned with detecting and recognizing facial action units. The action unit based classifiers use the Facial Action Coding System (Ekman and Friesan, 1978), a coding system that enumerates all possible facial movements. The majority of attempts to recognize facial expressions automatically are based on computer vision techniques. We advocate this non-intrusive vision-based approach, which favours usability over accuracy. In particular, we use dynamic facial action analysis, similar to the work implemented by Tian et al. (2001) and Bartlett et al. (2001). In this approach, facial expression sequences are analyzed to classify the feature motion into one of a possible set of action units.

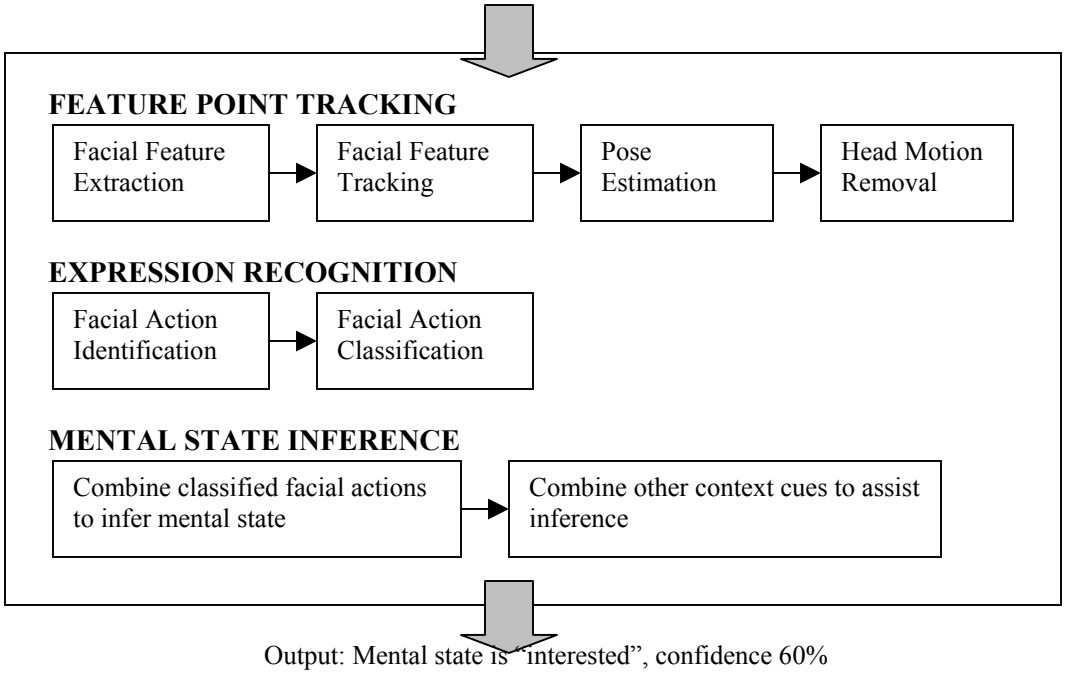
In contrast to the majority of facial action recognition systems, our goal in designing the *emotional hearing aid* is to span a wider range of spontaneous emotions, including "complex" ones referred to as cognitive mental states (Baron-Cohen, 1995). Because humans provide the best example we have of social intelligence, we examine the models from social and experimental psychology on the role of facial expressions in emotion understanding, and apply those in our design.

3 On Reading Emotions from the Face

Inferring other people's mental state can be thought of as a two-step process (Baron-Cohen et al., 2001). Until recently, only "basic" emotions (happiness, sadness, fear, surprise, disgust, and anger) were thought to be recognised reliably from facial expression. On the contrary, complex emotions such as interest, boredom and confusion, which involve attribution of a belief or intention – a cognitive mental state – to the person (Baron-Cohen et al., 2001) were held to be private, unobservable, and therefore not available from facial expression. In "Reading the Mind in the Eyes", Baron-Cohen et al. (2001), demonstrate through experimentation that humans are indeed capable of inferring a large array of mental states from visual cues.

The second step in mind reading is about inferring the contents of the mental state, a process that involves integrating contextual and temporal information. Bruce and Young (1998) explain how humans make considerable use of the contexts in which expressions occur to assist interpretation. Similarly, Ratner (1989) and Edwards (1998) show that the reliability of facial expressions as indicators of emotion is significantly improved when they are perceived in relation to contextual events instead of in isolation.

Input: sequence of spontaneous facial expression



Output: Mental state is "interested", confidence 60%

Figure 1: Overall architecture of the Emotional Hearing Aid.

Although the role of context is very much emphasized in models of how humans perceive emotions, to our knowledge, none of the prevalent facial affect recognition systems make use of contextual cues in the process of inferring emotions. In designing the *emotional hearing aid* we follow the two-step process of inferring emotions.

4 The Emotional Hearing Aid

The *emotional hearing aid* combines facial expression information with temporal and contextual clues to make inferences about a person's mental state. Figure 1 shows the overall architecture of the system. The system runs in real time, where the person is facing a video camera, and the frames (at 25-30fps) are fed on the fly to the tracker. While most systems operate on short clips (1-3 seconds), we place no constraints on the duration of the input, allowing us to consider longer sequences of expressions. This is of particular importance if temporal cues and expression transitions are going to be considered.

4.1 Feature Point Tracking

Twenty-two points identified on the face using prior knowledge of face shapes. The points include pose estimation ones (such as nose tips, nose root, and nostrils), and feature points used for emotion classification (such as inner and outer eyebrow, upper and lower lips, and pupils). Once the points are extracted, feature point tracking is then performed using optical flow (Lucas, 1984). The tracker accounts for the motion characteristics of individual feature points across frames of an image, and can deal with rigid head motion. Although experiments on gaze direction indicate that head direction plays a significant role in determining other peoples' mental states (Baron-Cohen 1995), it is important that we separate the smaller feature movement from the larger head motion before any facial action classification can be done.

4.2 Expression Recognition

The displacements of each of the twenty-two points across the sequence are fed to a classifier. We implement Support Vector Machines (SVM), which were first introduced by Vapnik (1995), as our classifying methodology. SVMs perform well with high dimensional data, and have been successfully applied to several applications including face detection.

4.3 Mental State Inference

Once individual facial actions are identified, the temporal relationship between them is examined to further identify the underlying emotion. While prevalent classification methodologies treat every facial action in isolation of preceding ones, we implement a classification system in which the temporal relationship between facial actions is utilized in recognizing the current expression (El Kaliouby and Robinson, 2003).

Other context cues that could be integrated in the inference process include information on whether the person is looking at (or away from) the camera, and if the person is moving towards (or away from) the camera.

5 Hardware and Interface Design

Rapid advances in key computing technologies have made the development of non-obtrusive wearable computers feasible. We propose that the *emotional hearing aid* be designed as a wearable computer system, similar to that of StartleCam (Healey and Picard, 1998). The system consists of a digital video camera, some form of audio feedback interface, and a wireless connection to a server.

In a typical scenario, the wearer is engaged with a person in some social situation. Depending on the capabilities of the digital video camera, raw video frames or extracted feature points are sent back to the server for classification. The server is responsible for keeping a rotating buffer of the incoming frames, and running the classification application. If the classifier detects an expression of interest, the results are sent back to the wearer in real time; continuous feedback is avoided so as not to distract the wearer from the interaction. Non-obtrusive audio messages using ambient sound rather than voice synthesis can indicate different mental states, and will hopefully make the social interaction more meaningful to the wearer.

6 Conclusion

In this paper we present the *emotional hearing aid*, a facial-expression based emotion recognition application currently under development. It is designed to assist people diagnosed with Asperger Syndrome with emotion understanding in natural social situations. We believe that such a tool will

open new possibilities for people who have difficulties recognizing emotional expression in others, such as those diagnosed with Asperger Syndrome.

References

Baron-Cohen, S. (1995) Mindblindness: an essay on autism and theory of mind. MIT Press.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., and Plumb, I. (2001) 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): 241±251.

Bartlett, M.S., Braathen, B., Littlewort-Ford, G., Hershey, J., Fasel, I., Marks, T., Smith, E., and Movellan, J.R. (2001) Automatic Analysis of Spontaneous Facial Behavior. *Tech Report, UCSD MPLab 2001.08*; see ftp://markov.ucsd.edu/pub/MPLABTR/mplab2001.08.pdf

Bruce, V. and Young, A. (1998) *In the Eye of the Beholder: The Science of Face Perception*. Oxford University Press.

Dautenhahn, K. (1999) Robots as Social Actors: AURORA and the Case of Autism. *Proceedings of The Third International Cognitive Technology Conference (CT99), August, San Francisco*.

Deak, G., Fasel, I., and Movellan, J.R. (2001) The Emergence of Shared Attention: Using Robots to Test Developmental Theories. *First International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*.

Edwards, K. (1998) The face of time: Temporal cues in facial expression of emotion. *Psychological Science*, 9, 270-276.

Ekman, P., and Friesen, W. (1978) *Facial Action Coding System: A technique for the measurement of facial movement*. Consulting Psychologists Press.

El Kaliouby, R., and Robinson, P. (2003) Temporal Context and the Recognition of Emotion from Facial Expressions. *To Appear in Proceedings of The HCII2003 International conference on Human-Computer Interaction.* Lawrence Erlbaum Associates, Inc.

Healey, J. and Picard, R (1998) StartleCam: A Cybernetic Wearable Camera. *In Proceedings of the International Symposium on Wearable Computers.*

Lucas, B. (1984) Generalized Image matching by the method of Differences. Tech. Report, CMU-CS-85-160, PhD Dissertation, Carnegie Mellon University School of Computer Science; see http://reports-archive.adm.cs.cmu.edu/anon/1985/abstracts/85-160.html

O' Connell, S. (1998) *Mindreading: How we learn to love and lie*. Arrow Books.

Picard, R.W. (1997) *Affective Computing*. MIT Press.

Picard, R.W., & Wexelblat, A. (2002) Future Interfaces: Social and Emotional. *Extended Abstracts of The CHI 2002 Conference on Human Factors in Computing Systems*, ACM Press, 698-699.

Ratner, C (1989) A Social Constructionist Critique of Naturalistic Theories of Emotion. *Journal of Mind and Behavior*, 10, 211-230.

Scassellati, B. (2000) Models of Social Development Using a Humanoid Robot. *Biorobotics*. Webb, B. and Consi, T. Ed. MIT Press.

Tian, Y., Kanade, T. and Cohn, J. (2001) Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23 (2), 97-115.

Vapnik, V. 1995. The Nature of Statistical Learning Theory. Springer, New York.

Werry, I., Dautenhahn, K., Ogden, B., Harwin, W. (2001) Can Social Interaction Skills Be Taught by a Social Agent? The Role of a Robotic Mediator in Autism Therapy. *Proceedings CT2001, The Fourth International Conference on Cognitive Technology: Springer Verlag, Lecture Notes in Computer Science, sub series Lecture Notes in Artificial Intelligence*.