33

# 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 human 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 guaged 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 novel classification algorithm on recognition of the basic expression classes proposed by Ekman and Friesen (1975). Expression data often consists of a segmented headshot taken under relatively controlled conditions and classification accuracy is based on comparison with emotion labels provided by human experts.

This bird's eye caricature of the methodology used by the pattern recognition community given above is necessarily simplistic, however it underlines two general reflections. First, pattern recognition has successfully framed the essentials of the facial expression problem to allow for effective comparison of algorithms. This narrowing of focus has led to impressive developments of the techniques for facial expression analysis and substantial understanding. Second, the narrow framing of the FP problem typical in the computer vision and pattern recognition may not be appropriate for HCI problems. This observation is a main theme of this paper, and we suggest that progress on use of FP in HCI may require re-framing the problem.

To do so, we have to overcome several controversies that are associated with the most fundamental issues of facial expression research, and it has been suggested (Bartneck & Lyons, 2007; Schiano, Ehrlich, Rahardja, & Sheridan, 2000), that these unresolved issues may be significantly impeding progress in the development of workable HCI systems. The nature of the method used to represent facial expressions is seen as a key issue in this regard. One school of thought, famously affiliated with Ekman (1999) but dating back at least to Charles Darwin (1872), holds that a discrete set of facial expression categories serves to communicate affective, categorical states, which, likewise, can be represented using a set of emotion categories. Another view with a long history, which was articulated clearly by Harold Schlosberg (1952, 1954), but again with roots in older work, holds that emotional facial expressions are better suited to representation in a continuous multi-dimensional space. Common interpretations for the affective dimensions are valence (pleasure/displeasure), arousal and intensity. Differences between categorical and dimensional models have sometimes been a source of controversy in the study of facial expressions (Schiano, Ehrlich, & Sheridan, 2004).

Choice of an appropriate representation scheme is no doubt of paramount importance for the success of any facial expression system, however categorical and dimensional views are by no means incompatible in the context of their application to HCI technologies. One of our earliest studies of dimensional facial expression representation conducted with my colleagues Miyuki Kamachi and Jiro Gyoba and reported in Lyons et al. (1998), was the result of a larger project to build a facial expression categorization system. While studying classification methods for images of facial expressions, we explored the dimensional structure of the facial expression image data and discovered that a nonlinear two-dimensional projection of the data, captured a large proportion of the variance in our data. A slightly greater proportion of the variance was accounted for with addition of a third dimension. Interestingly, the two dimensional projection closely resembled the well-known "circumplex" model of facial expressions, itself a low-dimensional projection of empirical data from semantic differential ratings of facial expression images. The correlation between the image-processing derived and semantic-rating derived spaces was unexpectedly high and provided support for our image-filter derived representation of facial expressions, as well as for the possibly utility of a dimensional representation in classifying facial expressions. At the same time, we observed a natural clustering of facial expression images within our low-dimensional affect space into basic emotional categories of happiness, anger,

surprise, and so on. This finding suggested that the concept of facial expression categories could also be a viable component of our facial expression classification system.

The findings reported in Lyons et al. (1998) and briefly summarized above showed that both categorical and dimensional representations could be used at different stages of a facial expression classification system and guided a subsequent project to build a facial expression classification system as reported by Lyons et al. (1999). The basic idea of the classification system to first process facial images with filters modeled on complex cells of primary visual cortex (area V1), then project the filter outputs into a low dimensional space learned from an ensemble of facial expression images and finally categorize expressions on the basis clusters. This system embodies dimensional and categorical approaches to facial expression representation and combines the power of both: an outcome of the project was the development of one of the early successful facial expression classifiers. Subsequent studies (see, for example, (Dailey, Cottrell, Padgett, & Adolphs, 2002)) have provided further support for the general approach of combining V1-like image filtering, dimensionality reduction followed by categorization.

In addition to utility of this approach for classifying images of facial expression, the schema discussed above is helpful in thinking about how dimensional and categorical facial expression representations might relate to what happens in the brain. For example, dimensional and categorical aspects of processing may be different facets of a single neural scheme for processing facial expressions. Loosely speaking, dimensionality reduction might take places at an earlier stage of processing, to reduce the complexity, and increase the robustness of a facial expression recognition system. Independently of how emotions are actually processed in the brain, artificial characters and robots also require a model to be able to process the external world into emotional states that can then be expressed. In the next section, we will discuss the modeling of emotions.

#### SYNTHESIS: EMOTION MODELING

Emotions are an essential part of the believability of embodied characters that interact with humans (Elliott, 1992; Koda, 1996; O'Reilly, 1996). Characters need an emotion model to synthesize emotions and express them. The emotion model should enable the character to argue about emotions the way humans do. An event that upsets humans, for example the loss of money, should also upset the character. The emotion model must be able to evaluate all situations that the character might encounter and must also provide a structure for variables influencing the intensity of an emotion. Such an emotion model enables the character to show the right emotion with the right intensity at the right time, which is necessary for the convincingness of its emotional expressions (Bartneck, 2001). Creating such an emotion model is a daring task and in this section we will outline some of its problems. In particular, we will argue for the importance of the context in which the emotion model operates.

Emotions are particularly important for conversational embodied characters, because they are an essential part of the self-revelation feature of messages. The messages of human communication consist of four features: facts, relationship, appeal and self-revelation (Schulz, 1981) The inability of a conversational character to reveal its emotional state would possibly be interpreted by the user as missing sympathy. It would sound strange if the character, for example, opened the front door of the house for the user to enter and spoke with an absolute monotonous voice: "Welcome home".

From a practical point of view, the developer of a screen character of robot is wise to build upon existing models to avoid reinvent the wheel. Several emotion models are available (Roseman, Antoniou, & Jose, 1996; Sloman, 1999). However, Ortony, Clore and Collins (1988) developed a computational emotion model, that is often referred to as the OCC model, which has established itself as the standard model for emotion synthesis.

A large number of studies employed the OCC model to generate emotions humans (Bondarev, 2002; Elliott, 1992; Koda, 1996; O'Reilly, 1996; Studdard, 1995). This model specifies 22 emotion categories based on valenced reactions to situations constructed either as being goal relevant events, as acts of an accountable agent (including itself), or as attractive or unattractive objects (see Figure 1). It also offers a structure for the variables, such as likelihood of an event or the familiarity of an object, which determines the intensity of the emotion types. It contains a sufficient level of complexity and detail to cover most situations an emotional interface character might have to deal with.

When confronted with the complexity of the OCC model many developers of characters believe that this model will be all they ever need to add emotions to their character. Only during the development process the missing features of the model and the problem of the context become apparent. These missing features and the context in which emotions arise are often underestimated and have the potential to turn the character into an unconvincing clown. We will point out what the OCC model is able to do for an embodied emotional character and what it does not.

The OCC model is complex and this paper discusses its features in terms of the process that characters follow from the initial categorization of an event to the resulting behaviour of the character. The process can be split into five phases:

- 1. **Categorization:** In the categorization phase the character evaluates an event, action or object, resulting in information on what emotional categories are affected.
- 2. **Quantification:** In the quantification phase, the character calculates the intensities of the affected emotional categories.
- 3. **Interaction:** The classification and quantification define the emotional value of a certain event, action or object. This emotional value will interact with the current emotional categories of the character.
- 4. **Mapping:** The OCC model distinguishes 22 emotional categories. These need to be mapped to a possibly lower number of different emotional expressions.

Figure 1. The OCC model of emotions. It contains a classification schema and variables to calculate the intensity of emotions.



#### Categorization

In the categorization phase an event, action or object is evaluated by the character, which results in information on what emotional categories are affected. This categorization requires the character to know the relation of a particular object, for example, to its attitudes. Depending on this evaluation either the "love" or "hate" emotional category will be affected by the object.

Consider this example: a character likes bananas and the user gives him a whole bunch. The character will evaluate the consequences of the event for the user, which results in pity, since the user has a whole bunch of bananas less. It will also evaluate the consequences of the event for itself, which results in satisfaction because it received a bunch of bananas. Next, it evaluates the action of the user, which results in admiration and finally the aspect of the object, which results in love. It appears that ironic that the category "love" is being used in the OCC model only for objects, since the more important usage for this word is certainly found in human-human relationships.

To do this classification the character needs an extensive amount of knowledge. First, it needs to know its relationship to the user, which was assumed to be good. Hence, pity is triggered and not resentment. Moreover, it needs to know what this event means to the user. Otherwise the character's happy-for category might be triggered (User Model). Second, it needs to have a goal "staying alive" to which the bananas contribute (Goals). Third, it needs to know what to expect from the user. Only knowing that the user does not have to hand out bananas every other minute the character will feel admiration (Standards). Last, it needs to know that it likes bananas (Attitudes).

The standards, goals and attitudes of the character that the OCC model requires need to be specified, organized and stored by the designer of the character. A new character knows even less than a newborn baby. It does not even have basic instincts. One way to store this knowledge could be an exhaustive table in which all possible events, actions and objects that the character might

encounter are listed together with information on which emotional categories they affect and how their intensity may be calculated. This approach is well suited for characters that act in a limited world. However, it would be rather difficult, for example, to create such an exhaustive list for all the events, actions and objects that the character might encounter at the home of the user. With an increasing number of events, actions and objects, it becomes necessary to define abstractions. The bananas could be abstracted to food, to which also bread and coconuts belong. The categorization for the event of receiving food will be the same for all types of food. Only their intensity might be different, since a certain food could be more nutritious or tasty. However, even this approach is inherently limited. The world is highly complex and this approach can only function in very limited "cube" worlds.

This world model is not only necessary for the emotion model, but also for other components of the character. If, for example, the character uses the popular Belief, Desires and Intention (BDI) architecture (Bratman, Israel, & Pollack, 1988), then the desires correspond to the goals of the emotion model. The structure of the goals is shared knowledge. So are the standards and attitudes. The complexity of the OCC model has a direct influence on the size of the required world model. However, the AI community has long given up the hope to be able to create extensive world models, such as Cyc (Cycorp, 2007). The amount of information and its organization appears overwhelming. Only within the tight constraints of limited worlds was it possible so far to create operational world models.

As mentioned above, the OCC model distinguishes 22 emotional categories (see Figure 1). This rather cumbersome and to some degree arbitrary model appears to be too complex for the development of believable characters (Ortony, 2003). The OCC model was created to model human emotions. However, it is not necessary to model a precise human emotion system to develop a believable character. A "Black Box" approach (Wehrle, 1998) appears to be sufficient. The

purpose of this approach is to produce outcomes or decisions that are similar to those resulting from humans, disregarding both the processes whereby these outcomes are attained as well as the structures involved. Such a "Black Box" approach is more suitable, particularly since the sensory, motoric and cognitive abilities of artificial characters are still far behind the ones of humans. The characters emotion system should be in balance with its abilities. Several reason speak for a simplification of the OCC model.

First, only those emotional categories of the OCC model should be used that the character can actually use. If a character uses the emotional model only to change its facial expression then its emotion categories should be limited to the ones it can express. Elliot (1992) implemented all 22 emotional categories in his agents because they were able to communicate each and every one to each other. This is of course only possible for character-character interaction in a virtual world. It would be impossible for characters that interact with humans, since characters are not able to express 22 different emotional categories on their face. Ekman (1972) proposed six basic emotions that can be communicated efficiently and across cultures through facial expressions.

Second, some emotional categories of the OCC model appear to be very closely related to others, such as gratitude and gratification, even thought the conditions that trigger them are different. Gratification results from a praiseworthy action the character did itself and gratitude from an action another character did. It is not clear if such a fine grained distinction has any practical advantages for the believability of characters.

Last, if the character does not have a user model then it will by definition not be able to evaluate the consequences of an event for the user. In this case, the "fortunes of others" emotional categories would need to be excluded.

Ortony acknowledged that the OCC model might be too complex for the development of believable characters (Ortony, 2003). He proposed to use five positive categories (joy, hope, relief, pride, gratitude and love) and five negative categories (distress, fear, disappointment remorse, anger and hate). Interestingly, he excluded the emotional categories that require a user model. These ten emotional categories might still be too much for a character that only uses facial expressions. Several studies simplified the emotional model even further to allow a one-to-one mapping of the emotion model to the expressions of the character (Bartneck, 2002; Koda, 1996).

#### Quantification

The intensity of an emotional category is defined separately for events, actions and objects. The intensity of the emotional categories resulting from an event is defined as the desirability and for actions and objects praiseworthiness and appealingness respectively (see Figure 1). One of the variables that is necessary to calculate desirability is the hierarchy of the character's goals. A certain goal, such as downloading a certain music album from the internet, would have several sub goals, such as download a specific song of that album. The completed goal of downloading of a whole album will evoke a higher desirability than the completed goal of downloading of a certain song, because it is positioned higher in the hierarchy. However, events might also happen outside of the character's current goal structure. The character needs to be able to evaluate such events as well. Besides the goal hierarchy, the emotion model also needs to keep a history of events, actions and objects. If the user, for example, gives the character one banana after the other in a short interval then the desirability of each of these events must decrease over time. The character needs to be less and less enthusiastic about each new banana. This history function is not described in the original OCC model, but plays an important role for the believability of the character. The history function has another important advantage. According to the OCC model, the likelihood of an event needs to be considered to calculate its desirability. The history function can help calculating this likeli-

hood. Lets use the banana example again: The first time the character receives a banana, it will use its default likelihood to calculate the desirability of the event. When the character receives the next banana, it will look at the history and calculate how often it received a banana in the last moments. The more often it received a banana in the past the higher is the likelihood of this event and hence the lower is its desirability. After a certain period of not receiving any bananas the likelihood will fall back to its original default value. This value should not be decreased below its default value, because otherwise the character might experience an overdose of desirability the next time it receives a banana. Another benefit of the history function is the possibility to monitor the progress the character makes trying to achieve a certain goal. According to the OCC model, the effort and realization of an event needs to be considered to calculate its desirability. The history function can keep track of what the character has done and hence be the base for the calculation of effort and realization

#### Interaction

The OCC model does not describe another important aspect of an emotion model: the interaction of the different emotional categories. Lets assume that the character was not able to download a certain song from the internet and is therefore angry. Next, the user gives it a banana. This event should not suddenly make it happy, but make it less angry. The emotional value of a certain event interacts with the current emotional state of the character. Little is known how this interaction might work, but a very simple approach could be to counter effect of the positive and negative categories.

#### Mapping

If the emotion model has more categories than the character has abilities to express them, the emotional categories need to be mapped to the available expressions. If the character, for example, uses only facial expression then it may focus on the six basic emotions of happiness, sadness, anger, disgust, fear and surprise (Ekman, Friesen, & Ellsworth, 1972). Interestingly, there is only one positive facial expression to which all 11 positive OCC categories need to be mapped to: the smile. Ekman (1985) identified several different types of smiles but their mapping to the positive OCC categories remains unclear. The 11 negative OCC categories need to be mapped to four negative expressions: Anger, Sadness, Disgust and Fear. The facial expression of surprise cannot be linked to any OCC categories, since surprise is not considered to be an emotion in the OCC model. Even though the character might only be able to show six emotional expressions on its face, the user might very well be able to distinguish between the expression of love and pride with the help of context information. Each expression appears in a certain context that provides further information to the viewer. The user might interpret the smile of a mother next to her son receiving an academic degree as pride, but exactly the same smile towards her husband as love.

#### Reflection

The main limitation of the OCC model is its reliance on world model. Such models have only been successfully used in very limited worlds, such as pure virtual worlds in which only virtual characters operate. Furthermore, the OCC model will most likely only be one part of a larger system architecture that controls the character or robot. The emotional states of the OCC model must interact with the other states. Not only the face of the character is influenced by the emotional state of the character, but also its actions. It would be unbelievable if the character showed an angry expression on its face, but acted cooperatively. The mapping of the emotional state should be based on strong theoretical foundations. Such theoreti-

cal foundations might not be available for every action that a character might be able to execute and thus force the developer of the character to invent these mappings. This procedure has the intrinsic disadvantage that the developer might introduce an uncontrolled bias based on his or her own experiences and opinions.

Besides the actions of the character, the emotional state may also influence the attention and evaluation of events, actions and objects. In stress situations, for example, humans tend to focus their attention on the problem up to the point of "tunnel vision". Ortony (2003) categorized the behavioural changes of the character through its emotional state in self-regulation (such as calming down), other-modulation (punish the other to feel better) and problem solving (try to avoid repetition). The latter will require the history function mentioned above. The emotional state of the character might even create new goals, such as calming down, which would result in actions like meditation.

#### FACIAL EXPRESSION SYNTHESIS

There is a long tradition within the HCI community of investigating and building screen based characters that communicate with users (Cassell, Sullivan, Prevost, & Churchill, 2000). Recently, robots have also been introduced to communicate with the users and this area has progressed sufficiently that some review articles are available (Bartneck & Okada, 2001; Fong, Nourbakhsh, & Dautenhahn, 2003). The main advantage that robots have over screen based agents is that they are able to directly manipulate the world. They not only converse with users, but also perform embodied physical actions.

Nevertheless, screen based characters and robots share an overlap in motivations for and problems with communicating with users. Bartneck et al. (Bartneck, Reichenbach, & Breemen, 2004) has shown, for example, that there is no significant difference in the users' perception of emotions as expressed by a robot or a screen based character and that subtle emotional expressions have been neglected (Bartneck & Reichenbach, 2005). The main motivation for using facial expressions to communicate with a user is that it is, in fact, impossible not to communicate. If the face of a character or robot remains inert, it communicates indifference. To put it another way, since humans are trained to recognize and interpret facial expressions it would be wasteful to ignore this rich communication channel.

Compared to the state of the art in screen-based characters, such as Embodied Conversational Agents (Cassell, Sullivan, Prevost, & Churchill, 2000), however, the field of robot's facial expressions is underdeveloped. Much attention has been paid to robot motor skills, such as locomotion and gesturing, but relatively little work has been done on their facial expression. Two main approaches can be observed in the field of robotics and screen based characters. In one camp are researchers and engineers who work on the generation of highly realistic faces. A recent example of a highly realistic robot is the Geminoid H1 which has 13 degrees of freedom (DOF) in its face alone. The annual Miss Digital award (Cerami, 2006) may be thought of as a benchmark for the development of this kind of realistic computer generated face. While significant progress has been made in these areas, we have not yet reached human-like detail and realism, and this is acutely true for the animation of facial expressions. Hence, many highly realistic robots and character currently struggle with the phenomena of the "Uncanny Valley" (Mori, 1970), with users experiencing these artificial beings to be spooky or unnerving. Even the Replice Q1Expo is only able to convince humans of the naturalness of its expressions for at best a few seconds (Ishiguro, 2005). In summary, natural robotic expressions remain in their infancy (Fong, Nourbakhsh, & Dautenhahn, 2003).

Major obstacles to the development of realistic robots lie with the actuators and the skin. At least 25 muscles are involved in the expression in the human face. These muscles are flexible, small and can be activated very quickly. Electric

motors emit noise while pneumatic actuators are difficult to control. These problems often result in robotic heads that either have a small number of actuators or a somewhat larger-than-normal head. The Geminoid H1 robot, for example, is approximately five percent larger than its human counterpart. It also remains difficult to attach skin, which is often made of latex, to the head. This results in unnatural and non-human looking wrinkles and folds in the face.

At the other end of the spectrum, there are many researchers who are developing more iconic faces. Bartneck (2003) showed that a robot with only two DOF in the face can produce a considerable repertoire of emotional expressions that make the interaction with the robot more enjoyable. Many popular robots, such as Asimo (Honda, 2002), Aibo (Sony, 1999) and PaPeRo (NEC, 2001) have only a schematic face with few or no actuators. Some of these only feature LEDs for creating facial expressions. The recently developed iCat robot is a good example of an iconic robot that has a simple physically-animated face (Breemen, Yan, & Meerbeek, 2005). The eyebrows and lips of this robot move and this allows synthesis of a wide range of expressions.

While there is progress in the facial expressions of robot faces, we are sill facing several conceptional problems that stem from the field of Artificial Intelligence. Lets take the example of emotions that we discussed in detailed above. The emotional state of the character is defined through values for each of its emotional categories. This emotional state needs to be expressed through all available channels. A conversational embodied character, for example, needs to express its emotional state through its speech and facial expressions. It would be unconvincing if the character would smile, but speak with a monotonous voice. However, the systematic manipulation of speech to express emotions remains a challenge for the research community. Emotional facial expressions are understood better, but a fundamental questions remains. Shall the character only express the most dominant emotional category, or shall it express every category at the same time and hence show a blend of emotions. The blending of emotional

expression requires a sophisticated face, such as Baldi from the CSLU Toolkit. Cartoon like characters, such as eMuu (Bartneck, 2002) or Koda's Poker Playing Agent (Koda, 1996) are not able to show blends and therefore they can only express the most dominant emotional category.

Another important issue that needs to be considered when designing the facial expression of the character is that they need to be convincing and distinct at low intensity levels. Most events that a character encounters will not trigger an ecstatic state of happiness. The evaluation of a certain event should be roughly the same as could be expected of a human and most events that humans encounter in everyday life do unfortunately not result in ecstasy. If the character managed to download a complete album of music it still did not save the world from global warming. Hence, it should only show an appropriate level of happiness.

It becomes obvious that the problems inherited by HRI researchers from the field of AI can be severe. Even if we neglect philosophical aspects of the AI problem and are satisfied with a computer that passes the Turing test, independently of how it achieves this, we will still encounter many practical problems. This leads us to the so-called "weak AI" position, namely claims of achieving human cognitive abilities are abandoned. Instead, this approach focuses on specific problem solving or reasoning tasks.

There has certainly been progress in weak AI, but this has not yet matured sufficiently to support artificial entities. Indeed, at present, developers of artificial entities must to resort to scripting behaviors. Clearly, the scripting approach has its limits and even the most advanced common sense database, Cyc (Cycorp, 2007), is largely incomplete. FP should therefore not bet on the arrival of strong AI solutions, but focus on what weak AI solutions can offer today. Of course there is still hope that eventually also strong AI applications will become possible, but this may take a long time.

When we look at what types of HRI solutions are currently being built, we see that a large number of them do barely have any facial features at all. Qrio, Asimo and Hoap-2, for example, are only

able to turn their heads with 2 degrees of freedom (DOF). Other robots, such as Aibo, are able to move their head, but have only LEDs to express their inner states in an abstract way. While these robots are intended to interact with humans, they certainly avoid facial expression synthesis. When we look at robots that have truly animated faces, we can distinguish between two dimensions: DOF and iconic/realistic appearance (see Figure 2).

Robots in the High DOF/Realistic quadrant not only have to fight with the uncannieness (Bartneck, Kanda, Ishiguro, & Hagita, 2007; MacDorman, 2006) they also may raise user expectations of a strong AI which they are not able to fulfill. By contrast, the low DOF/Iconic quadrant includes robots that are extremely simple and perform well in their limited application domain. These robots lie well within the domain of the soluble in FP. The most interesting quadrant is the High DOF/Iconic quadrant. These robots have rich facial expressions but avoid evoking associations with a strong AI through their iconic appearance. We propose that research on such robots has the greatest potential for significant advances in the use of FP in HRI.

# WORKSHOP ON "HCI AND THE FACE"

As part of our effort to examine the state of the field of FP in HCI, we organized a day-long workshop the ACM CHI'2006 conference (see: http:// www.bartneck.de/2006/04/22/hci-and-the-face/

Figure 2. Robots with animated faces



for details). The workshop included research reports, focus groups, and general discussions. This has informed our perspective on the role of FP in HCI, as presented in the current paper.

One focus group summarized the state of the art in facial expression analysis and synthesis, while another brainstormed HCI applications. The idea was to examine whether current technology sufficient advanced to support HCI applications. The proposed applications were organized with regards to the factors "Application domain" and "Intention" (see Table 1). Group discussion seemed to naturally focus on applications that involve some type of agent, avatar or robot. It is nearly impossible to provide an exhaustive list of applications for each field in the matrix. The ones listed in the table should therefore be only considered as representative examples.

These examples well illustrate a fundamental problem of this research field. The workshop participants can be considered experts in the field and all the proposed example applications were related to artificial characters, such as robots, conversational agents and avatars. Yet not one of these applications has become a lasting commercial success. Even Aibo, the previously somewhat successful entertainment robot, has been discontinued by Sony in 2006.

A problem that all these artificial entities have to deal with is, that while their expression processing has reached an almost sufficient maturity, their intelligence has not. This is especially problematic, since the mere presence of an animated face raises the expectation levels of its user. An entity that is able to express emotions is also expected to recognize and understand them. The same holds true for speech. If an artificial entity talks then we also expect it to listen and understand. As we all know, no artificial entity has yet passed the Turing test or claimed the Loebner Prize. All of the examples given in Table 1 presuppose the existence of a strong AI as described by John Searle (1980).

The reasons why strong AI has not yet been achieved are manifold and the topic of lengthy discussion. Briefly then, there are, from the outset, conceptual problems. John Searle (1980) pointed out that digital computers alone can never truly understand reality because it only manipulates syntactical symbols that do not contain semantics. The famous 'Chinese room' example points out some conceptual constraints in the development of strong AIs. According to his line of arguments, IBM's chess playing computer "Deep Blue" does not actually understand chess. It may have beaten Kasparov, but it does so only by manipulating meaningless symbols. The creator of Deep Blue, Drew McDermott (1997), replied to this criticism: "Saving Deep Blue doesn't really think about chess is like saying an airplane doesn't really fly because it doesn't flap its wings." This debate reflects different philosophical viewpoints on what it means to think and understand. For centuries philosophers have thought about such questions and perhaps the most important conclusion is that there is no conclusion at this point in time. Similarly, the possibility of developing a strong AI remains an open question. All the same, it must be admitted that some kind of progress has been made. In the past, a chess-playing machine would have been regarded as intelligent. But now it is regarded as the feat of a calculating machine - our criteria for what constitutes an intelligent machine has shifted.

In any case, suffice it to say that no sufficiently intelligent machine has yet emerged that would provide a foundation for our example applications given in Table 1. The point we hope to have made with the digression into AI is that the application dreams of researchers sometimes conceal rather unrealistic assumptions about what is possible to achieve with current technology.

# TOWARDS AN "ART OF THE SOLUBLE"

The outcome of the workshop we organized was unexpected in a number of ways. Most striking was the vast mismatch between the concrete and fairly realistic description of the available FP technology and its limitations arrived at by one

		Intention		
		Persuade	Being a companion	Educate
Application domain	Entertainment	Advertisement: REA (Cas- sell, Sullivan, Prevost, & Churchill, 2000), Greta (Pelachaud, 2005)	Aibo (Sony, 1999), Tama- gotchi (Bandai, 2000)	My Real Baby (Lund & Nielsen, 2002)
	Communication	Persuasive Technology (Fogg, 2003), Cat (Zanbaka, Goolkasian, & Hodges, 2006)	Avatar (Biocca, 1997)	Language tutor (Schwien- horst, 2002)
	Health	Health advisor Fitness tutor (Mahmood & Ferneley, 2004)	Aibo for elderly (Tamura et al., 2004), Attention Capture for Dementia Pa- tients (Wiratanaya, Lyons, & Abe, 2006)	Autistic children (Robins, Dautenhahn, Boekhorst, & Billard, 2005)

Table 1. Examples of face processing applications in HCI and HRI

of the focus groups, and the blue-sky applications discussed by the second group. Another sharp contrast was evident at the workshop. The actual presentations given by participants were pragmatic and showed effective solutions to real problems in HCI not relying on AI.

Perhaps the most salient aspect of our observation on the problem of FP is that HCI technology can often get by with partial solutions. A system that can discriminate between a smile and frown, but not an angry versus disgusted face, can still be a valuable tool for HCI researchers, even if it is not regarded as a particularly successful algorithm from the pattern recognition standpoint. Putting this more generally, components of algorithms developed in the pattern recognition community, may already have sufficient power to be useful in HCI, even if they do not yet constitute general facial expression analysis systems.

This led us to the reflection that scientific progress often relies on what the Nobel prize winning biologist Peter Medawar called "The Art of the Soluble" (Medawar, 1967). That is, skill in doing science requires the ability to select a research problem which is soluble, but which has not yet been solved. Very difficult problems such as strong AI may not yield to solution over the course of decades, so for most scientific problems it is preferable to work on problems of intermediate difficulty, which can yield results over a more reasonable time span, while still being of sufficient interest to constitute progress. Some researchers of course are lucky or insightful enough to re-frame a difficult problem in such a way as to reduce its difficulty, or to recognize a new problem which is not difficult, but nevertheless of wide interest. In the following sections we make several proposals for the application of this strategy to future research into robotic facial expression synthesis and facial expression analysis.

## Facial Expression Analysis: Continuously Update Benchmarks

If any real progress is going to be made towards the hard-AI problem of building machine which can read minds and understand emotions by looking at facial expressions, researchers need to acknowledge the vital importance of updating the methods used to test the performance of their systems. Failure to continuously update the benchmarks used to measure the performance of facial expression systems leads to algorithms which may be highly optimized for a particular set of data and testing conditions, but fail miserably when asked to generalize to more realistic conditions. As discussed elsewhere in this article

the problem of obtaining adequate data for training and testing facial expression analysis systems has long been one of the major bottle necks to progress in the field.

For purposes of concreteness we given here a specific and concrete example of a testing paradigm which has not been adequately explored: instead of using nominal labels in terms of basic categories, facial expression images, or image sequences could be more richly described to reflect empirical data on human perception. A approach used by Lyons et al. (1998) is to use semantic ratings on a set of emotion labels rather than a single emotion category. Training and testing an automatic system with semantic ratings data is more complex than if nominal categorical labels are used. Moreover, collecting the ratings data can also require much time and effort. However a continuous description based on real data has still not been fully explored. So it is not known whether the widespread use of nominal category labels for expression data may be hampering progress towards the development of systems which can be useful in the real world, as opposed to the world of facial expressions artifically posed, selected, or elicited under contrived laboratory conditions.

#### Facial Analysis for Direct Gesture-Based Interaction

A further illustration of the "Art of the Soluble" strategy comes from the analysis of facial expression and movements for direct gesture-based interaction. While there is a large body of work on automatic facial expression recognition and lip reading within the computer vision and pattern recognition research communities, relatively few studies have examined the possible use of the face in direct, intentional interaction with computers. However, the complex musculature of the face and extensive cortical circuitry devoted to facial control suggest that motor actions of the face could play a complementary or supplementary role to that played by the hands in HCI (Lyons, 2004).

One of us (MJL) has explored this idea through a series of related projects which make use of vision-based methods to capture movement of the head and facial features and apply these to intentional, direct interaction with computers. For example we have designed and implemented systems which make use of head and mouth motions were for the purposes of hands-free text entry (De Silva, Lyons, Kawato, & Tetsutani, 2003) and single-stroke text character entry on small keyboards such as those found on mobile phones (Lyons, Chan, & Tetsutani, 2004). In other projects we have used action of the mouth and face for digital sketching (Chan, Lyons, & Tetsutani, 2003) and musical expression (Lyons & Tetsutani, 2001).

One of the systems we developed tracked the head and position of the nose and mapped the projected position of the nose tip in the image plane to the coordinates of the cursor. Another algorithm segmented the area of the mouth and measured the visible area of the cavity of the user's mouth in the image plane. The state of opening/closing of the mouth could be determined robustly and used in place of mouse-button clicks. This simple interface allowed for text entry using the cursor to select streaming text. Text entry was started and paused by opening and closing the mouth, while selection of letters was accomplished by small movements of the head. The system was tested extensively and found to permit comfortable text entry at a reasonable speed. Details are reported in (De Silva, Lyons, Kawato, & Tetsutani, 2003).

Another project used the shape of the mouth to disambiguate the multiple letters mapped to the keys of a cell phone key pad (Lyons, Chan, & Tetsutani, 2004). Such an approach works very well for Japanese, which has a nearly strict CV (consonant-vowel) phoneme structure, and only five vowels. The advantage of this system was that it took advantage of existing user expertise in shaping the mouth to select vowels. With some practice, users found they could enter text faster than with the standard multi-tap approach.

The unusual idea of using facial actions for direct input may find least resistance in the realm

Facial Expression Analysis, Modeling and Synthesis

of artistic expression. Indeed, our first explorations of the concept were with musical controllers using mouth shape to control timbre and other auditory features (Lyons & Tetsutani, 2001). Of course, since many musical instruments rely on action of the face and mouth, this work has precedence, and was greeted with enthusiasm by some musicians. Similarly, we used a mouth action-sensitive device to control line properties while drawing and sketching with a digital tablet (Chan, Lyons, & Tetsutani, 2003). Here again our exploration elicited a positive response from artists who tried the system.

The direct action facial gesture interface serves to illustrate the concept that feasible FP technology is ready to be used as the basis for working HCI applications. The techniques used in all the examples discussed are not awaiting the solution of some grand problem in pattern recognition: they work robustly in real-time under a variety of lighting conditions.

### Artificial Expressions and Other Computational Scaffolds for Emotion

A radical reformulation of facial expression research results from the observation that our most meaningful interactions with computers, when scrutinized carefully, usually turn out to be human-computer-human interactions, or, in other words, machine-mediated human-interactions. Bearing this in mind allows designers to sidestep strong-AI issues, leaving the task of interpreting and understanding emotions to humans. This design strategy assigns machines to the tasks they can perform well - automatically reproducing, processing, and displaying information. An example of the application of such a design philosophy is the "Artificial Expressions" system (Lyons, Kluender, & Tetsutani, 2005). In this system, information from biosensors such as galvanic skin response, respiration, and pulse, are shared amongst users participating in a networked cooperative learning task. The physiological system is visualized in real-time using visual displays (see Figure 3 and Figure 4) that were designed to be so simple and intuitive that they require almost no explanation. For example, the pulse is represented by a pulsating red circle. Users learn to attribute meaning to these physiological displays during the course of meaning interaction on a shared task; we studied a situation in which one user tutored another on written Chinese. While the interpretation of the artificial expressions is not fixed, but must be learned constructively through engagement in a meaningful situation, neither is it completely arbitrary - the expressions are linked directly to the physiological status of the users. One of the outcomes of this study was the observation users quickly learned to make use of the galvanic skin response as a measure of the level of stress of their partner. Specifically, the tutor found the galvanic skin response signal helpful in adjusting the pace of instruction avoiding a situation where

Figure 3. Biosensors for artificial expressions (Lyons, Kluender, & Tetsutani, 2005)



Figure 4. Artificial expression displays (Lyons, Kluender, & Tetsutani, 2005)



the learner was struggling to keep up.

It is reasonable to consider an analogous approach in the design of systems making use of facial expressions. For example, face detection and tracking methods are now sufficiently advanced that they can be robustly used under standard office lighting conditions. Hence, an automatic approach can be used to normalize and scale a video display of an expressive face. Removing variations in location and scale could help a user to focus their attention on the actual movements of a face. This approach has been used in the design of a facial expression data navigation system (Lyons, Funk, & Kuwabara, 2005). In this work, video data of long-term observation of the face could be browsed efficiently and quickly by normalizing the position and scale of the face extracted from a video sequence. In addition, optical flow was calculated at several locations on the face and converted to an aggregate measure of non-rigid facial movement. This measure could be used to highlight possible hotspots of facial expression activity, to further ease the task of navigating long term behavioral data. The system, as described, was developed as part of a project to assist in the long term care of dementia patients, to provide a tool allowing physicians and caregivers to more easily understand long term trends in the well being of a patient. Again, the difficult AI problem of understanding the patients emotions is left to a human, but their task is made much more efficient by leveraging a soluble problem of automatic face processing - face detection and tracking.

#### **Expressive Robots**

One of the most engaging robots that use the "Art of the Soluable" approach is KeepOn. This robot has a very limited repertoir of movements, which consists of bouncing, tilting and rotating (Michalowski, Sabanovic, & Kozima, 2007). With this set of movements, KeepOn is able to express a considerable variety of internal states, including emotions. The interaction with KeepOn does also not require a strong AI. So far, KeepOn has been used for rhythm imitation games. The user beats a drum and KeepOn dances to the rhythm. KeepOn demonstrates that a simple robot can stimulate highly engaging interaction between itself and a user. Another example of a simple robot that effectively interacted with a user is eMuu (Bartneck, 2002). The robot has only four degrees of freedom, but its emotional expressions improved the interaction. This robot demonstrate not only that a limited repertoire of iconic facial expression are sufficient to communicate emotional states, but it also demonstrated that a radically simplified OCC model has been sufficient to create believable expressions.

#### CONCLUSION

In this paper we have argued in favour of an "Art of the Soluble" approach in HCI. Progress can often be made by sidestepping long-standing difficult issues in artificial intelligence and pattern recognition. This is partly intrinsic to HCI:

Figure 5. eMuu



the presence of a human user for the system being developed implies leverage for existing computational algorithms. Our experience and the discussions that led to this article have also convinced us that HCI researchers tend towards an inherently pragmatic approach even if they are not always self-conscious of the fact. In summary, we would like to suggest that skill in identifying soluble problems is already a relative strength of HCI and this is something that would be worth further developing.

### REFERENCES

Bandai. (2000). Tamagotchi. Retrieved January 2000, from http://www.bandai.com/

Bartneck, C. (2001). How convincing is Mr. Data's smile: Affective expressions of machines. *User Modeling and User-Adapted Interaction, 11*, 279-295. | DOI: 10.1023/A:1011811315582

Bartneck, C. (2002). *eMuu - an embodied emotional character for the ambient intelligent home.* Ph.D. thesis, Eindhoven University of Technology, Eindhoven.

Bartneck, C. (2003). Interacting with an embodied emotional character. In *Proceedings of the Design for Pleasurable Products Conference (DPPI2004), Pittsburgh* (pp. 55-60). | DOI: 10.1145/782896.782911

Bartneck, C., Kanda, T., Ishiguro, H., & Hagita, N. (2007). Is the uncanny valley an uncanny cliff? In *Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2007, Jeju, Korea* (pp. 368-373). | DOI: 10.1109/ROMAN.2007.4415111

Bartneck, C., & Lyons, M. J. (2007). HCI and the Face: Towards an Art of the Soluble. In J. Jacko (Ed.), *Human-computer interaction, part 1, HCII2007, LNCS 4550* (pp. 20-29). Berlin: Springer. | DOI: 10.1007/978-3-540-73105-4\_3 Bartneck, C., & Okada, M. (2001). Robotic user interfaces. In *Proceedings of the Human and Computer Conference (HC2001), Aizu* (pp. 130-140).

Bartneck, C., & Reichenbach, J. (2005). Subtle emotional expressions of synthetic characters. *The International Journal of Human-Computer Studies, 62*(2), 179-192. | DOI: 10.1016/j.ijhcs.2004.11.006

Bartneck, C., Reichenbach, J., & Breemen, A. (2004). In your face, robot! The influence of a character's embodiment on how users perceive its emotional expressions. In *Proceedings of the Design and Emotion 2004, Ankara*.

Bartneck, C., & Suzuki, N. (2005). Subtle Expressivity for Characters and Robots. *International Journal of Human Computer Studies*, *62*(2), 159-160. | DOI: 10.1016/j.ijhcs.2004.11.004

Biocca, F. (1997). The cyborg's dilemma: embodiment in virtual environments. In *Proceedings of the Second International Conference on Cognitive Technology - "Humanizing the Information Age", Aizu* (pp. 12-26). | DOI: 10.1109/CT.1997.617676

Bondarev, A. (2002). *Design of an emotion management system for a home robot*. Master, Eindhoven University of Technology, Eindhoven.

Bratman, M. E., Israel, D. J., & Pollack, M. E. (1988). Plans and resource-bounded practical reasoning. *Computational Intelligence*, *4*(4), 349-355. DOI: 10.1111/j.1467-8640.1988.tb00284.x

Breemen, A., Yan, X., & Meerbeek, B. (2005). iCat: an animated user-interface robot with personality. In *Proceedings of the Fourth International Conference on Autonomous Agents & Multi Agent Systems, Utrecht.* | DOI: 10.1145/1082473.1082823

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (2000). *Embodied conversational agents*. Cambridge: MIT Press.

Cerami, F. (2006). Miss Digital World. Retrieved August 4th, from http://www.missdigitalworld. com/

Chan, C., Lyons, M. J., & Tetsutani, N. (2003). Mouthbrush: drawing and painting by hand and mouth. In *Proceedings of the Fifth International Conference on Multimodal Interfaces, Vancouver* (pp. 277-280). | DOI: 10.1145/958432.958482

Cycorp. (2007). Cyc. Retrieved February 2007, from http://www.cyc.com/

Dailey, M. N., Cottrell, G. W., Padgett, C., & Adolphs, R. (2002). EMPATH: A neural network that categorizes facial expressions. *Journal of Cognitive Neuroscience*, *14*(8), 1158-1173. | DOI: 10.1162/089892902760807177

Darwin, C. (1872). *The expression of the emotions in man and animals*. London,: J. Murray.

De Silva, C. G., Lyons, M. J., Kawato, S., & Tetsutani, N. (2003). Human factors evaluation of a vision-based facial gesture interface. In *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop* (pp. 52). | DOI: 10.1109/CVPRW.2003.10055

Ekman, P. (1985). *Telling lies: Clues to deceit in the marketplace, politics, and marriage*. New York: W.W. Norton.

Ekman, P. (1999). Basic emotions. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 45–60). Chichester, England; New York: Wiley.

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face*. Englewood Cliffs: Prentice Hall.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the human face: guidelines for research and an integration of findings*. New York: Pergamon Press.

Elliott, C. D. (1992). *The affective reasoner: A process model of emotions in a multi-agent system.* Ph.D. thesis, The Institute for the Learning

Sciences, Northwestern University, Evanston, Illinois.

Ellis, P. M., & Bryson, J. J. (2005). The significance of textures for affective interfaces. In J. G. Carbonell & J. Siekmann (Eds.), *Intelligent virtual agents* (Vol. 3661/2005, pp. 394-404). Berlin: Springer. | DOI: 10.1007/11550617 33

Fasel, B., & Luettin, J. (2003). Automatic facial expression analysis: a survey. *Pattern Recognition*, *36*(1), 259-275. | DOI: 10.1016/S0031-3203(02)00052-3

Fogg, B. J. (2003). *Persuasive technology: using computers to change what we think and do.* Amsterdam; Boston: Morgan Kaufmann Publishers.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems, 42*, 143-166. | DOI: 10.1016/S0921-8890(02)00372-X

Hara, F., & Kobayashi, H. (1996). A face robot able to recognize and produce facial expression. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems '96, IROS 96, Osaka* (pp. 1600-1607). | DOI: 10.1109/ IROS.1996.569026

Honda. (2002). Asimo. Retrieved from, http://www.honda.co.jp/ASIMO/

Ishiguro, H. (2005). Android science - Towards a new cross-interdisciplinary framework. In *Proceedings of the CogSci Workshop Towards Social Mechanisms of Android Science, Stresa* (pp. 1-6).

Koda, T. (1996). Agents with faces: A study on the effect of personification of software agents. Master Thesis, MIT Media Lab, Cambridge.

Lund, H. H., & Nielsen, J. (2002). An edutainment robotics survey. In *Proceedings of the Third International Symposium on Human and Artificial Intelligence Systems: The Dynamic Systems Approach for Embodiment and Sociality, Fukui.* 

Lyons, M. J. (2004). Facial gesture interfaces for expression and communication. In *Proceedings* of the IEEE International Conference on Systems, Man and Cybernetics, The Hague (pp. 598-603).

| DOI: 10.1109/ICSMC.2004.1398365

Lyons, M. J., Akamatsu, S., Kamachi, M., & Gyoba, J. (1998). *Coding Facial Expressions with Gabor Wavelets*. Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara pp. 200-205. | DOI: 10.1109/AFGR.1998.670949

Lyons, M. J., Budynek, J., & Akamatsu, S. (1999). Automatic Classification of Single Facial Images. *IEEE Transactions Pattern Analysis and Machine Intelligence, 21*(12), 1357-1362. | DOI: 10.1109/34.817413

Lyons, M. J., Chan, C., & Tetsutani, N. (2004). *MouthType: Text Entry by Hand and Mouth.* Proceedings of the Conference on Human Factors in Computing Systems CHI2004, Austria pp. 1383-1386. | DOI: 10.1145/985921.986070

Lyons, M. J., Funk, M., & Kuwabara, K. (2005). Segment and browse: A strategy for supporting human monitoring of facial expression behaviour. In *Lecture notes in computer science - humancomputer interaction - INTERACT 2005* (Vol. 3585/2005, pp. 1120-1123). Rome: Springer. |DOI: 10.1007/11555261 119

Lyons, M. J., Kluender, D., & Tetsutani, N. (2005). Supporting empathy in online learning with artificial expressions. *Journal of Educational Technology & Society*, 8(4), 22-30.

Lyons, M. J., & Tetsutani, N. (2001). Facing the music: a facial action controlled musical interface. In *Proceedings of the Conference on Human Factors in Computing Systems CHI2001, Seattle* (pp. 309-310). | DOI: 10.1145/634067.634250

MacDorman, K. F. (2006). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. In *Proceedings of the ICCS/CogSci-2006* Long Symposium: Toward Social Mechanisms of Android Science, Vancouver.

Mahmood, A. K., & Ferneley, E. (2004). Can avatars replace the trainer? A case study evaluation. In *Proceedings of the the International Conference on Enterprise Information Systems (ICEIS), Porto* (pp. 208-213).

McDermott, D. (1997, May 14th). Yes, Computers Can Think. *New York Times*.

Medawar, P. B. (1967). *The art of the soluble*. London: Methuen.

Michalowski, M. P., Sabanovic, S., & Kozima, H. (2007). A dancing robot for rhythmic social interaction. *In Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction, Arlington, Virginia, USA* (pp. 89-96). | DOI: 10.1145/1228716.1228729

Mori, M. (1970). The uncanny valley. *Energy*, *7*, 33-35.

NEC. (2001). PaPeRo. Retrieved from, http:// www.incx.nec.co.jp/robot/

O'Reilly, W. S. N. (1996). *Believable social and emotional agents*. Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, PA.

Ortony, A. (2003). On making believable emotional agents believable. In R. P. Trapple, P. (Ed.), *Emotions in humans and artefacts*. Cambridge: MIT Press.

Ortony, A., Clore, G., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge: Cambridge University Press.

Pantic, M., & Rothkrantz, L. J. M. (2000). Automatic analysis of facial expressions: the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(12), 1424 - 1445 [DOI: 10.1109/34.895976

Pelachaud, C. (2005). Multimodal expressive embodied conversational agents. In *Proceedings of the 13th Annual ACM International Conference on Multimedia, Hilton, Singapore* (pp. 683 - 689). | DOI: 10.1145/1101149.1101301

Picard, R. W. (1997). *Affective computing*. Cambridge: MIT Press.

Robins, B., Dautenhahn, K., Boekhorst, R., & Billard, A. (2005). Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills? *Universal Access in the Information Society*, *4*(2), 105-120. | DOI: 10.1007/s10209-005-0116-3

Roseman, I. J., Antoniou, A. A., & Jose, P. E. (1996). Appraisal determinants of emotions: constructing a more accurate and comprehensive theory. *Cognition and Emotion*, *10*(3), 241-278. | DOI: 10.1080/026999396380240

Schiano, D. J., Ehrlich, S. M., Rahardja, K., & Sheridan, K. (2000). Face to interface: facial affect in (hu)man and machine. In *Proceedings of the CHI 2000, Den Hague* (pp. 193-200).

Schiano, D. J., Ehrlich, S. M., & Sheridan, K. (2004). Categorical imperative not: facial affect is perceived continously. In *Proceedings* of the CHI2004, Vienna (pp. 49-56). | DOI: 10.1145/985692.985699

Schlossberg, H. (1952). The description of facial expressions in terms of two dimensions. *Journal of Experimental Psychology*, 44(2).

Schlossberg, H. (1954). Three dimensions of emotion. *Psychological Review*, *61*, 81-88.

Schulz, F. v. T. (1981). *Miteinander Reden - Stoerungen und Klaerungen*. Reinbeck bei Hamburg: Rowolth Taschenbuch Verlag GmbH.

Schwienhorst, K. (2002). The state of VR: A meta-analysis of virtual reality tools in second language acquisition. *Computer Assisted Lan*-

*guage Learning, 15*(3), 221 - 239. | DOI: 10.1076/ call.15.3.221.8186

Searle, J. R. (1980). Minds, brains and programs. *Behavioral and Brain Sciences*, *3*(3), 417-457.

Shugrina, M., Betke, M., & Collomosse, J. (2006). Empathic painting: interactive stylization through observed emotional state. In *Proceedings of the 4th International Symposium on Non-Photorealistic Animation and Rendering, Annecy, France* (pp. 87 - 96). | DOI: 10.1145/1124728.1124744

Sloman, A. (1999). Architectural requirements for human-like agents both natural and artificial. In K. Dautenhahn (Ed.), *Human cognition and social agent technology, advances in consciousness research*. Amsterdam: John Benjamins Publishing Company.

Sony. (1999). Aibo. Retrieved January 1999, from http://www.aibo.com

Studdard, P. (1995). *Representing human emotions in intelligent agents*. Master Thesis, The American University, Washington DC.

Tamura, T., Yonemitsu, S., Itoh, A., Oikawa, D., Kawakami, A., Higashi, Y., et al. (2004). Is an entertainment robot useful in the care of elderly people with severe dementia? *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 59:M83-M85* 

Tobii Technology. (2007). Tobii Technology. Retrieved February 2007, from http://www.tobii. com/

Wehrle, T. (1998). Motivations behind modeling emotional agents: Whose emotion does your robot have? In C. Numaoka, L. D. Cañamero & P. Petta (Eds.), *Grounding emotions in adaptive systems*. Zurich: 5th International Conference of the Society for Adaptive Behavior Workshop Notes (SAB'98).

Wiratanaya, A., Lyons, M. J., & Abe, S. (2006). An interactive character animation system

for dementia care. In *Proceedings of the ACM SIGGRAPH 2006 Research posters, Boston, Massachusetts* (pp. Article No. 82). | DOI: 10.1145/1179622.1179717

Zanbaka, C., Goolkasian, P., & Hodges, L. (2006). Can a virtual cat persuade you? The role of gender and realism in speaker persuasiveness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Montreal, Quebec, Canada.* | DOI: 10.1145/1124772.1124945

Zhai, S., Morimoto, C., & Ihde, S. (1999). Manual and gaze input cascaded (MAGIC) pointing. In Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit, Pittsburgh (pp. 246-253). | DOI: 10.1145/302979.303053

#### **KEY TERMS**

Art of the Soluble (AOTS): Scientific research strategy advocated by Nobel laureate biologist Peter Medawar. Specifically, the AOTS strategy emphasizes skill in in the recognition of scientific problems which have not yet been solved but are reasonably amenable to solution with reasonable time and resources. Here we have suggested that, in some cases, the introduction of facial expression technology into HCI may be hindered by excessive concentration on research problems which fall into the domain of strong A.I. and that it is time to consider AOTS approaches.

Artificial Expressions: This term relates to a somewhat radical proposal to reframe the goals of affective computing towards the construction of new machine-mediated channels for the communication of affect between humans, or artificial expressions as we call them. The affected intended by these artificial expressions is not to be defined a prior, but to be learned and evolved through ongoing situational interaction in humanmachine-human communication.

**Expressive Robots:** are robots that use facial expressions, gestures, posture and speech to communicate with the human user. This communication might not only include factual information, but also emotional states.

**Facial Expression Classification:** In machine vision, the automatic labelling of facial images or sequences of images with a semantic label or labels describing affect portrayed by the face. Our paper suggests that research has come to focus on a narrowly defined version of this problem: namely the hard classification of facial images (or sequences) into the stereotypical Ekman universal facial expressions, and that researchers in pattern recognition and human-computer interaction could profit by more broadly framing the research domain.

**Human-Robot Interaction (HRI):** is the study of interactions between people (users) and robots. HRI is multidisciplinary with contributions from the fields of human-computer interaction, artificial intelligence, robotics, natural language understanding, and several social sciences.

Weak A.I. (contrast with hard A.I.): This term has connotations in the context of practical work in artificial intelligence, as well as for theoretical studies of A.I. and the philosophy of mind. In the current article we are primarily concerned with the former usage of the term, namely with that domain of approaches to machine intelligence which do not take, as a primary goal, an attempt to match or exceed human intelligence, this latter goal being the hallmark of "strong A.I." research.