HEFES: an Hybrid Engine for Facial Expressions Synthesis to control human-like androids and avatars

Daniele Mazzei, Nicole Lazzeri, David Hanson and Danilo De Rossi

***Abstract*— Nowadays advances in robotics and computer science have made possible the development of sociable and attractive robots. A challenging objective of the field of humanoid robotics is to make robots able to interact with people in a believable way. Recent studies have demonstrated that human-like robots with high similarity to human beings do not generate the sense of unease that is typically associated to human-like robots. For this reason designing of aesthetically appealing and socially attractive robots becomes necessary for realistic human-robot interactions.**

In this paper HEFES (Hybrid Engine for Facial Expressions Synthesis), an engine for generating and controlling facial expressions both on physical androids and 3D avatars is described. HEFES is part of a software library that controls a human robot called FACE (Facial Automaton for Conveying Emotions). HEFES was designed to allow users to create facial expressions without requiring artistic or animatronics skills and it is able to animate both FACE and its 3D replica.

The system was tested in human-robot interaction studies aimed to help children with autism to interpret their interlocutors' mood through facial expressions understanding.

I. INTRODUCTION

In the last years, more and more social robots have been developed due to rapid advances in hardware performance, computer graphics, robotics technology and Artificial Intelligence (AI).

There are various examples of social robots but it is possible to roughly classify them according to their aspect in two main categories: human-like and not human-like. Human-like social robots are usually associated to the pernicious myth that robots should not look or act like human beings in order to avoid the so-called 'Uncanny Valley' [1]. MacDorman and Ishiguro [2] explored observers' reactions to gradual morphing of robots and humans pictures and found a peak in judgments of the eeriness in the transition between robot and human-like robot pictures according to the Uncanny Valley hypothesis. Hanson [3] repeated this experiment morphing more attractive pictures and found that the peak of eeriness was much smoother, approaching to a flat line, in the transition between human-like robot and human beings pictures. This indicates that typical reactions due to the Uncanny Valley were present only in the transition between classic robots and cosmetically atypical human-like robots. Although more studies demonstrate the presence of the Uncanny Valley effect, it is possible to design and create human-like robots that are not uncanny using innovative technologies that integrate movies and cinema animation with make-up techniques [4].

Daniele Mazzei, Nicole Lazzeri and Danilo De Rossi are with Interdepartmental Research Center 'E. Piaggio', Faculty of Engineering - University of Pisa, Via Diotisalvi 2, 56126 Pisa, Italy. (mazzei@di.unipi.it)

David Hanson is with Hanson Robotics, Plano Tx, USA. (david@hansonrobotics.com)

The enhancement of the believability of human-like robots is not a pure aesthetic challenge. In order to create machines that look and act as humans, it is necessary to improve the robot's social and expressive capabilities in addition to the appearance. Therefore, facial expressiveness is one of the most important aspect to be analyzed in designing human-like robots since it is the major emotional communication channel used in interpersonal relationships together with facial and head micro movements [5].

Since the early 70's, facial synthesis and animation have raised a great interest among computer graphics researchers and numerous methods for modeling and animating human faces have been developed to reach more and more realistic results.

One of the first models for the synthesis of faces was developed by Parke [6], [7]. The Parke parametric model is based on two groups of parameters: conformation parameters which are related to the physical facial features, such as the shape of the mouth, nose, eyes, etc., and expression parameters which are related to facial actions such as wrinkling the forehand for anger or open the eyes wide for surprise.

Differently, physically-based models manipulate directly the geometry of the face to approximate real deformations caused by the muscles including skin layers and bones. Waters [8], using vectors and radial functions, developed a parameterized model based on facial muscles dynamic and skin elasticity.

Another approach used for creating facial expressions is based on interpolation methods. Interpolation-based engines use a mathematical function to specify smooth transitions between two or more basic facial positions in a defined time interval [9]. One, two or three-dimensional interpolations can be performed to create an optimized and realistic facial morphing. Although interpolations are fast methods, they are limited in the number of realistic facial configurations they can generate.

All geometrically-based methods described above can generate difficulties in achieving realistic facial animations since they require artistic skills. On the other hand, animation skills are required only for creating a set of basic facial configurations since an interpolation space can be use to generate a wide set of new facial configurations starting from the basic ones.

In this work a facial animation engine called HEFES was implemented as fusion of a muscle-based facial animator and an intuitive interpolation system. The facial animation system is based on the Facial Action Coding System (FACS) in order to make it compatible with both physical robots and 3D avatars and usable in different facial animation scenarios. The FACS is the most popular standard for describing facial behaviors in terms of muscular movements. The FACS is based on a detailed study of the facial muscles carried out by Ekman and Friesen in 1976 [10] and is aimed at classifying the facial muscular activity according to Action Units (AUs). AUs are defined as visually discernible component of facial movements which are generated through one or more underlying muscles. AUs can be used to describe all the possible movements that a human face can express. Therefore an expression is a combination of several AUs, each of them with their own intensity measured in 5 discrete levels (A:Trace, B:Slight, C:Marked pronounced, D:Severe, E:Extreme maximum).

II. Materials and Methods

A. FACE

FACE is a robotic face used as emotions conveying system (Fig. 1). The artificial skull is covered by a porous elastomer material called Frubber™ that requires less force to be stretched by servo motors than other solid materials [11]. FACE has 32 servo motors actuated degrees of freedom which are mapped on the major facial muscles to allow FACE to simulate facial expressions.

Fig. 1. FACE and the motor actuation system

FACE servo motors are positioned following the AUs disposition according to the FACS (Fig. 2) and its facial expressions consist of a combination of many AUs positions. Thanks to the fast response of the servo motors and the mechanical properties of the skin, FACE can generate realistic human expressions involving people in social interactions.

B. SYSTEM ARCHITECTURE

HEFES is a subsystem of the FACE control library deputed to the synthesis and animation of facial expressions and includes a set of tools for controlling FACE and its 3D avatar. HEFES includes four modules: synthesis, morphing, animation and display. The synthesis module is designed to allow

Fig. 2. Mapping between servo motors positions and Action Units of FACS

users to manually create basic facial expressions that are normalized and converted according to the FACS standard. The morphing module takes the normalized FACS-based expressions as input and generates an emotional interpolation space where expressions can be selected. The animation module merges concurrent requests from various control subsystems and creates a unique motion request resolving possible conflicts. Finally, the display module receives the facial motion request and converts it in movements according to the selected output display.

1) The synthesis module allows users to generate new facial expressions through the control of the selected emotional display, i.e. FACE robot or 3D avatar. Both modules provide a graphical user interface (GUI) with as many slider controls as the number of servo motors (FACE robot) or anchor points (3D avatar) which are present in the corresponding emotional display.

In the Robot editor, each slider defines a normalized range between 0 and 1 for moving the corresponding servo motor which is associated to an AU of the FACS. Using the Robot editor, the six basic expressions, i.e. happiness, sadness, anger, surprise, fear and disgust, defined as 'universally accepted' by Paul Ekman [12], [13], were manually created. According to the "Circumplex Model of Affect" theory [14], [15], each generated expression was saved as an XML file including the set of the AUs values, the expression name and the corresponding coordinates in terms of Pleasure and arousal. In the Circumplex Model of Affect expressions are associated with Pleasure that indicates the pleasant/unpleasant feelings and with Arousal which is related to a physiological activation.

The 3D virtual editor is a similar tool used to deform a facial mesh. The 3D editor is based on a user interface on which a set of slider controls is used to actuate various facial muscles. Expressions are directly rendered on the 3D avatar display and saved as XML files as in the Robot Editor.

2) The morphing module generates, on the base of the Posner's theory, an emotional interpolation space, called Emotional Cartesian Space (ECS) [16]. In the ECS the x coordinate represents the valence and the y coordinate represents the arousal. Each expression e(v, a) is consequently

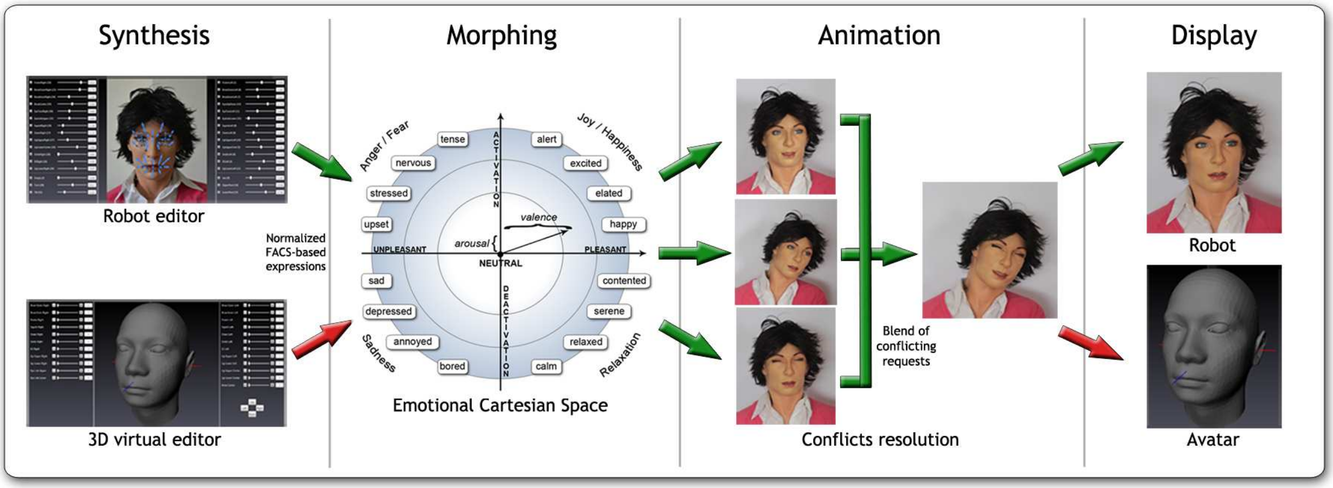

Fig. 3. The architecture of the facial animation system based on four main modules: synthesis, morphing, animation and display.

associated with a point in the valence-arousal plane where the neutral expression e(0, 0) is placed in the origin (Fig. 3, Morphing module). The morphing module takes the set of basic expressions as input and generates the ECS applying a shape-preserving piecewise cubic interpolation algorithm implemented in Matlab™. The output of the algorithm is a three-dimensional matrix composed of 32 planes corresponding to the 32 AUs. As shown in Fig. 4, each plane represents the space of the possible positions of a single AU where each point is identified by two the coordinates, valence and arousal. The coordinates of each plane range between -1 and 1 with a step of 0.1 therefore the generated ECS produces 21x21 new normalized FACS-based expressions that can be performed by the robot or the 3D avatar independently. Since the ECS is not a static space, each new expression manually created through the synthesis module can be used to refine the ECS including it in the set of expressions used by the interpolation algorithm. The possibility of updating the ECS with additional expressions allows users to continuously adjust the ECS covering zones in which the interpolation algorithm could require a more detailed description of the AUs (II-B.1).

3) The animation module is designed to combine and merge multiple requests coming from various modules which can run in parallel in the robot/avatar control library. The facial behavior of the robot or avatar is inherently concurrent since parallel requests could interest the same AU generating conflicts. Therefore the animation module is responsible for mixing movements, such as eye blinking or head turning, with requests of expressions. For example, eye blinking conflicts with the expression of amazement since normally amazed people react opening the eyes wide.

The animation module receives as input a motion request, which is defined by a single AU or a combination of multiple AUs, with an associated priority. The animation engine is implemented as a Heap, a specialized tree-based data structure used to define a shared timer that is responsible for orchestrating the animation. The elements of the Heap,

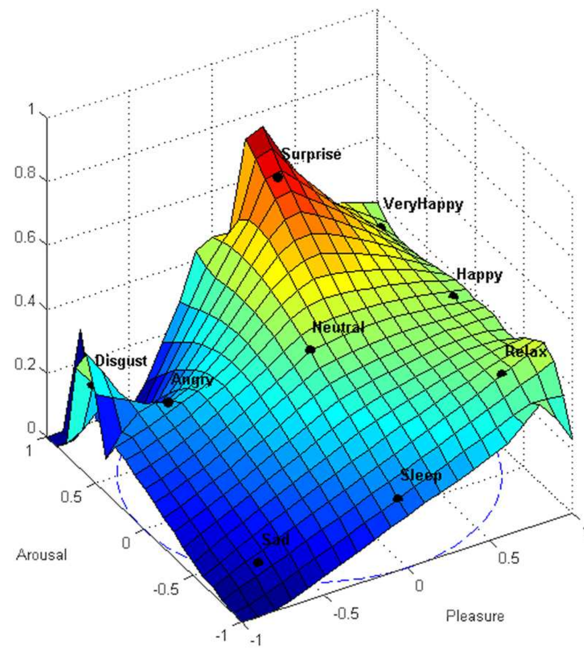

Fig. 4. The emotional Cartesian plane for the right eyebrow (motor #24 corresponding to the AU 1 in Fig. 2).

called Tasks, are ordered by their due time therefore the root of the Heap contains the first task to be executed. In the Heap there can be two types of tasks, Motion Task and Interpolator Task, that are handled in a different way. Both types of tasks are defined by the expiring time, the duration of the motion and the number of steps in which the task will be divided. A Motion Task also includes 32 AUs, each of them with their associated values and a priority. When a movement request is generated, a Motion Task is sent to the Animation Engine and inserted into the Heap which will be reordered according to the due time. The animation engine is always running to check whether some tasks into the Heap are expired. For each expired task, the animation engine removes it from the Heap and executes it. If the task is a Motion Task, the animation engine calculates the amount of movement to be performed at the current step, stores the result in correspondence to the relative AU and reschedules the task into the Heap if the task is not completed. If the task is an Interpolation Task, the animation engine calculates the new animation state by blending all the steps, previously calculated, for each AUs according to their priority. At the end, the Interpolator Task is automatically rescheduled into the Heap with an expiring time of 40ms.

The output of the animation module is a motion task composed of 32 normalized AUs values that is sent to the emotional display module.

4) The display module represents the output of the system. We implemented two dedicated emotional displays: the FACE android and the 3D avatar. According to a calibration table, the FACE android display converts normalized AUs values into servo motor positions that are expressed as duty cycles in the range 500-2500. Each motor has a different range of movements due to its position inside the FACE. For this reason, the display module includes a control layer to avoid the exceeding the servo motor limits according to minimum and maximum values stored in the calibration tables.

The 3D avatar display is a facial animation system based on a physical model described in [17] that approximates the anatomy of the skin and the muscles. The model is based on a non-linear spring system which can simulate the dynamics of human face movements while the muscles are modeled as mesh of force deformed springs. Each skin point of the mesh is connected with its neighbors by non-linear springs. Human face includes a wide range of muscles types, e.g. rectangular, triangular, sheet, linear, sphincter. Since servo motors act as linear forces, the type of muscle satisfying this condition is the linear muscle that is specified by two points: the attachment point which is normally fixed and the insertion point which defines the area where the facial muscle performs its action. Facial muscle contractions pull the skin surface from the area of the muscle insertion point to the area of the muscle attachment point. When a facial muscle contracts, the facial skin points in the influence area of the muscle change their position according to the distance from the muscle attachment point and the elastic properties of the mass-spring system. Facial skin points not directly influenced by the muscle contraction are in a sort of unbalanced state that is stabilized through propagation of other unbalanced elastic forces.

The elastic model of the skin and the mathematical implementation of the muscles have been already developed while the manual mapping of the 3D mesh anchor points to AUs is still under development.

C. *ANIMATION TOOL*

Generally facial animation softwares are tools that require a certain level of knowledge in design, animation and anatomy. Often users only need to easily animate facial expressions without having these specific skills. Therefore the system was designed to be used both by experts in facial design and animation which can directly create or modify expressions and users that are interested in quickly designing HRI experimental protocols selecting a set of pre-configured expressions.

The ECS Timeline is a tool of the HEFES system that is intended to meet the needs of different users. The timeline is a Graphical User Interface (GUI) with two use modalities: ”Auto Mode” and ”Advanced Mode”. In Auto Mode, users can create sequences of expressions selecting the corresponding points in the ECS and dragging them into the timeline. Sequences can be saved, played and edited using the timeline control. When a sequence is reproduced, motion requests are sent to the animation module that resolves conflicts and forwards them to the robot or the avatar display. The ECS Timeline GUI includes a chart that visualizes the motors positions during an animation for a deeper understanding of the facial expression animation process (Fig. 5). In Advanced Mode, a sequence of expressions can be displayed as editable configurations of all AUs values in a multitrack graph where each AU is expressed as a motion track and can be manually edited. In the Advanced Mode is possible to use ECS expressions as starting point for creating more sophisticated animations in which single AUs can be adjusted in real-time.

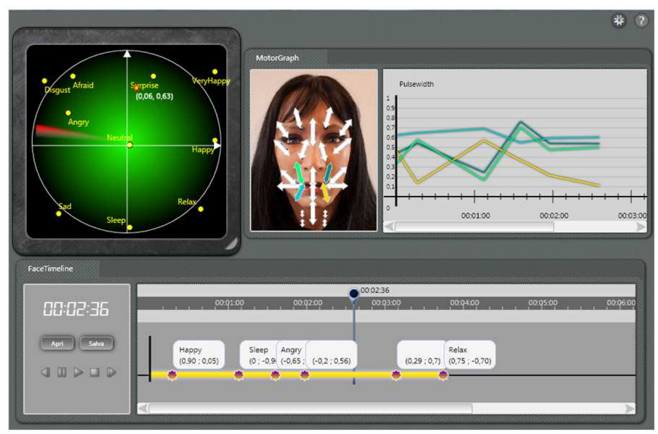

Fig. 5. The ECS Animation in the Auto Mode configuation.

III. RESULTS AND DISCUSSION

HEFES was used as emotions conveying system within the IDIA (Inquiry into Disruption of Intersubjective equipment in Autism spectrum disorders in childhood) project in collaboration with the IRCCS Stella Maris (Calambrone, Italy) [16], [18].

In particular, the ECS Animation tool was used by the psychologist in Auto Mode to easily design the therapeutic protocol creating facial animation paths without require FACE android direct motor configuration and calibration. The tool does not required skills in facial animation and human anatomy and allowed therapist to intuitively create therapeutic scenarios adding expressions to the timeline dragging them from the ECS. Moreover the Manual Mode

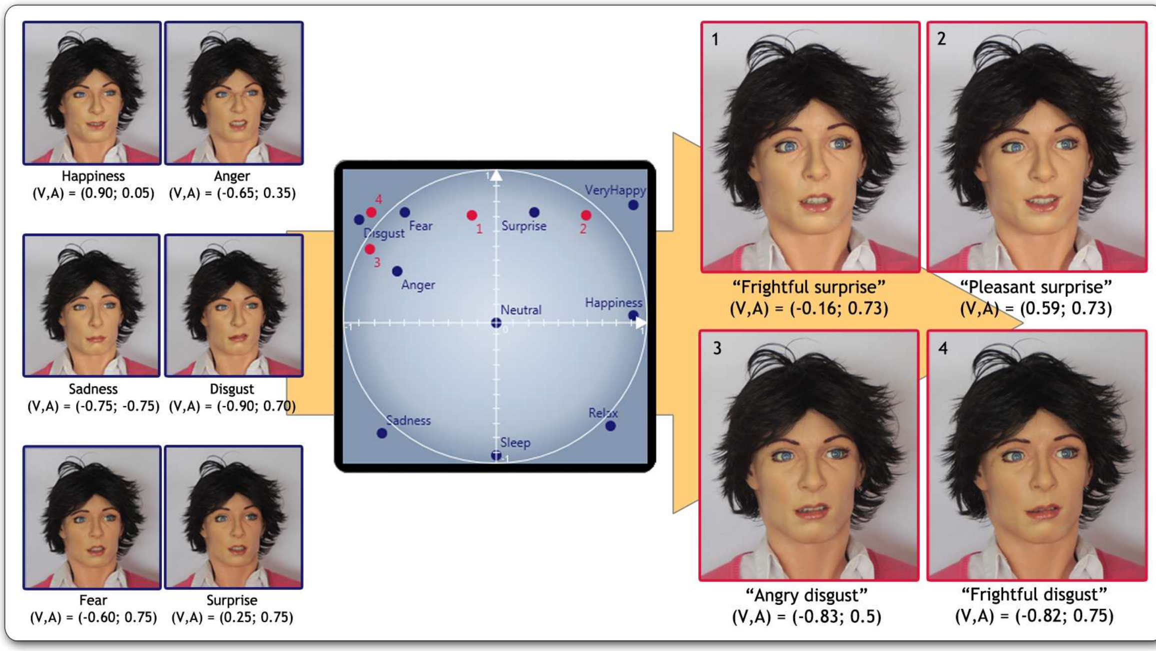

Fig. 6. The morphing module used for creating new 'mixed' expressions (right side) selecting (V,A) points (red dots) from the ECS. The module takes in input a set of basic expressions (left side) with their (V,A) values (blue dots).

configuration was used to create specific patterns of movements such as the turning of the head. Head movements was oriented to watch a little robot used by the therapist to test children's shared attention capabilities.

Recent study demonstrated that people with Autism Spectrum Disorders (ASDs) do not perceive robots as machine but as "artificial partners" [19]. On the base of this theory the IDIA project aimed to the study of alternative ASD treatment protocol involving robots, avatars and other advanced technologies. One of the purposes of the protocol was to verify the capability of the FACE android to convey emotions to children with ASD. Figure 6 shows examples of expressions generated by the morphing module. It takes the six basic expressions as input (expressions on the left side of the figure corresponding to the blue dots in the ECS) and generates 'half-way' expressions (right side of the figure corresponding to the red dots in the ECS) by clicking on the ECS. All these generated expressions are identified by their corresponding pleasure and arousal coordinates.

FACE base protocol was tested on a panel of normally developing children and children with Autism Spectrum Disorders (ASDs) (aged 6-12 years).

The test was conducted on a panel of 5 children with ADSs and 15 normally developing interacting with the robot individually under therapist supervision. The protocol was divided in phases and one of these concerned evaluating the accuracy of emotional recognition and imitation skills. In this phase children were asked to recognize, label and then imitate a set of facial expressions performed by the robot and subsequently by the psychologist. The sequence of expressions included happiness, anger, sadness, disgust, fear and surprise. Moreover, the protocol included a phase called "free play" where the ECS tool was directly used by the psychologist to control the FACE android in real-time.

The subjects' answers in labeling an expression were scored as correct or wrong by a therapist and used for calculating the percentage of correct expressions recognition. As shown in Fig. 7 both children with ASDs and normally developing children were able to label Happiness, Anger and Sadness performed by FACE and by the psychologist without errors. Otherwise Fear, Disgust and Surprise performed by FACE and by the psychologist have not been labeled correctly, especially by subjects with ASDs. Fear, Disgust and Surprise are emotions which convey empathy not only through stereotypical facial expressions but also with body movements and vocalizations. The affective content of this emotions is consequently dramatically reduced if expressed only through facial expressions.

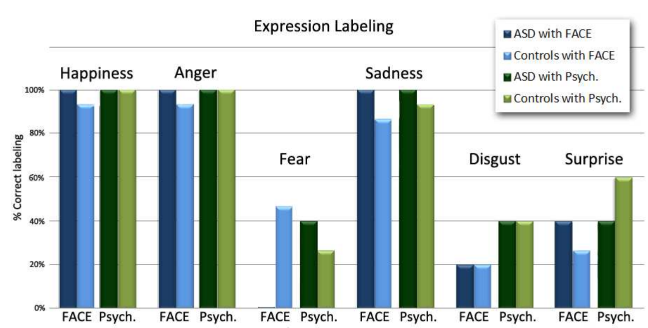

Fig. 7. Results of the labeling phase for ASD and control subjects observing FACE and psychologist expressions.

In conclusion HEFES allows operators and psychologists to easily model and generate expressions following the current standards of facial animations. The morphing module provides a continuous emotional space where it is possible to select a wide range of expressions, most of them difficult to be manually generated. The possibility to continuously add new expressions to the ECS interpolator allows users to refine the expressions generation system for reaching a high expressiveness level without requiring animation or artistic skills.

Through HEFES is possible to control robot or avatar creating affective based human-robot interaction scenarios on which different emotions can be conveyed. Facial expressions performed by FACE and by the psychologist have been labeled by children with ASDs and normally developed children with the same score. This analysis demonstrates that the system is able to correctly generate human-like facial expressions.

IV. FUTURE WORKS

HEFES was designed to be used both with a physical robot and with a 3D avatar. The actual state of the 3D editor includes the algorithm to animate the facial mesh according to the model described in Sec. II and the definition of some anchor points. In future all the AUs will be mapped on the 3D avatar mesh for a complete control of the avatar. HEFES will be used to study how human beings perceive facial expressions and emotion expressed by a physical robot in comparison with its 3D avatar for understanding if the physical appearance has an emphatic component in conveying emotions.

Moreover the synthesis module will include the control of facial micro movements and head dynamics that are associated with human moods. For example, blinking frequency and head speed are considered to be indicators of discomfort. These micro movements will be designed and controlled using an approach similar to the one designed for facial expressions. A set of basic head and facial micro movements will be generated and associated with corresponding behaviors according to their pleasure and arousal coordinates. The set of basic behaviors will be used as input of the morphing module which will generate a Behavioral Cartesian Space (BCS). Future experiment on emotion labeling and recognition will be conducted including the facial micro movement generator and a face tracking algorithm in order to investigate the contribute of this affective related activities on emotions conveying FACE capabilities.

REFERENCES

[1] M. Mori, "Bukimi no tani (the uncanny valley)," *Energy*, 1970.

[2] K. F. MacDorman and H. Ishiguro, "The uncanny advantage of using androids in cognitive and social science research," *Interaction Studies*, vol. 7, no. 3, pp. 297–337, 2006.

[3] D. Hanson, "Exploring the aesthetic range for humanoid robots," in *Proceedings of the ICCS CogSci 2006 Symposium Toward Social Mechanisms of Android Science*. Citeseer, 2006, p. 1620.

[4] H. Ishiguro, "Android science - toward a new cross-interdisciplinary framework," *Development*, vol. 28, pp. 118–127, 2007.

[5] P. Ekman, "Facial expression and emotion," *American Psychologist*, pp. 384–392, 1993.

[6] F. I. Parke, "Computer generated animation of faces," in *ACM '72: Proceedings of the ACM annual conference*. New York, NY, USA: ACM, 1972, pp. 451–457.

[7] F. I. Parke, "A parametric model for human faces," Ph.D. dissertation, The University of Utah, 1974.

[8] K. Waters, "A muscle model for animation three-dimensional facial expression," *SIGGRAPH Computer Graphics*, vol. 21, pp. 17–24, August 1987.

[9] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, and D. H. Salesin, "Synthesizing realistic facial expressions from photographs," in *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, ser. SIGGRAPH '98. New York, NY, USA: ACM, 1998, pp. 75–84.

[10] P. Ekman and W. V. Friesen, "Measuring facial movement," *Journal of Nonverbal Behavior*, vol. 1, no. 1, pp. 56–75, Sep. 1976.

[11] D. Hanson, "Expanding the design domain of humanoid robots," in *Proceedings of ICCS CogSci Conference, special session on Android Science*, 2006.

[12] P. Ekman, "Are there basic emotions?" *Psychological Review*, vol. 99, no. 3, pp. 550–553, Jul 1992.

[13] P. Ekman, *Handbook of Cognition and Emotion: 3 Basic emotions*. New York: John Wiley & Sons Ltd, 1999, ch. 3, pp. 45–60.

[14] J. A. Russell, "The circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, pp. 1161–1178, 1980.

[15] J. Posner, J. A. Russell, and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Development and Psychopathology*, vol. 17, no. 3, pp. 715–734, 2005.

[16] D. Mazzei, L. Billeci, A. Armato, N. Lazzeri, A. Cisternino, G. Pioggia, R. Igliozzi, F. Muratori, A. Ahluwalia, and D. De Rossi, "The face of autism," in *RO-MAN 2009. The 18th IEEE International Symposium on Robot and Human Interactive Communication, 2009*, 2010, pp. 791–796.

[17] Y. Zhang, E. C. Prakash, and E. Sung, "Real-time physically-based facial expression animation using mass-spring system," in *Computer Graphics International 2001*, ser. CGI '01. Washington, DC, USA: IEEE Computer Society, 2001, pp. 347–350.

[18] D. Mazzei, N. Lazzeri, L. Billeci, R. Igliozzi, A. Mancini, A. Ahluwalia, F. Muratori, and D. De Rossi, "Development and evaluation of a social robot platform for therapy in autism," in *EMBC 2011. The 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 4515–4518.

[19] J. Scholtz, "Theory and evaluation of human robot interactions," in *Proc. 36th Annual Hawaii Int System Sciences Conf*, 2003.