Research Article

INTERNATIONAL JOURNAL OF
ADVANCED ROBOTIC SYSTEMS

International Journal of Advanced Robotic Systems
July-August 2018: 1–16
© The Author(s) 2018
DOI: 10.1177/1729881418783158
journals.sagepub.com/home/arx
SAGE

The influence of dynamics and speech on understanding humanoid facial expressions

Nicole Lazzeri[1], Daniele Mazzei[2], Maher Ben Moussa[3], Nadia Magnenat-Thalmann[4] and Danilo De Rossi[1]

Abstract

Human communication relies mostly on nonverbal signals expressed through body language. Facial expressions, in particular, convey emotional information that allows people involved in social interactions to mutually judge the emotional states and to adjust its behavior appropriately. First studies aimed at investigating the recognition of facial expressions were based on static stimuli. However, facial expressions are rarely static, especially in everyday social interactions. Therefore, it has been hypothesized that the dynamics inherent in a facial expression could be fundamental in understanding its meaning. In addition, it has been demonstrated that nonlinguistic and linguistic information can contribute to reinforce the meaning of a facial expression making it easier to be recognized. Nevertheless, few studies have been performed on realistic humanoid robots. This experimental work aimed at demonstrating the human-like expressive capability of a humanoid robot by examining whether the effect of motion and vocal content influenced the perception of its facial expressions. The first part of the experiment aimed at studying the recognition capability of two kinds of stimuli related to the six basic expressions (i.e. anger, disgust, fear, happiness, sadness, and surprise): static stimuli, that is, photographs, and dynamic stimuli, that is, video recordings. The second and third parts were focused on comparing the same six basic expressions performed by a virtual avatar and by a physical robot under three different conditions: (1) muted facial expressions, (2) facial expressions with nonlinguistic vocalizations, and (3) facial expressions with an emotionally neutral verbal sentence. The results show that static stimuli performed by a human being and by the robot were more ambiguous than the corresponding dynamic stimuli on which motion and vocalization were associated. This hypothesis has been also investigated with a 3-dimensional replica of the physical robot demonstrating that even in case of a virtual avatar, dynamic and vocalization improve the emotional conveying capability.

Keywords

Human–robot interaction, humanoid robot, facial expression recognition, social interaction

Date received: 20 July 2016; accepted: 25 May 2018

Topic: Human Robot/Machine Interaction
Topic Editor: Chrystopher L Nehaniv

Introduction

Our ability to process facial information is so quick and apparently effortless that most people take it for granted. Faces are incredibly important to us and thanks to their innate expressiveness they are used for a wide variety of purpose. A brief glance at a human face can give us important information about age, gender, or social status used to identify the person, and about emotional state, intentions,

[1] Research Center E. Piaggio, Faculty of Engineering, University of Pisa, Pisa, Italy
[2] Computer Science Department, University of Pisa, Pisa, Italy
[3] Computer Science Centre, University of Geneva, Geneva, Switzerland
[4] MIRALab, CUI, University of Geneva, Geneva, Switzerland

Corresponding author:
Nicole Lazzeri, University of Pisa, Largo Lucio Lazzarino 1, Pisa 56122, Italy.
Email: n.lazzeri@centropiaggio.unipi.it

or attentions used to understand the person, especially during social situations. This has led to a huge interest in the scientific study of faces since the 1800s with the first books concerning systematic descriptions of the movements of the facial muscles.[1–3]

One of the most remarkable books about the study of facial expressions in humans is *The Expression of the Emotions in Man and Animals* written by Charles Darwin in 1872.[2] The aim of the book was "to ascertain, independently of common opinion, how far particular movements of the features and gestures are really expressive of certain states of the mind."[2] Darwin was mainly interested in investigating the universality of emotional expressiveness and hypothesized that the so-called six basic expressions, that is, happiness, sadness, anger, fear, surprise, and disgust, contain emotion-specific patterns of facial elements which make them biologically determined and universally recognizable by all people in spite of race and culture.

The universality of facial expressions led to the further work and study of Paul Ekman and Wallace Friesen who nearly 100 years later proposed to encode and decode the facial expressions. Based on the evidence that some facial expressions of emotion were universal,[4–6] in 1976, Ekman and Friesen developed a procedure for measuring visibly different facial movements based on an anatomical analysis of facial actions.[7] This method called Facial Action Coding System (FACS) aimed at describing any facial expression a human being can make in terms of anatomically based Action Units (AUs), that is, the unit of measurement defining an observable independent movement of the face. There is not always a 1:1 correspondence between AUs and muscle movements: An AU can include more than one muscle and vice versa, one muscle can be described by more than one action. By now, the FACS has become a standard widely used by scientists in emotion research field.

Recently, the commonly held hypothesis about the six basic emotions as universally recognized and easily interpreted by all has been questioned from different points of view. Jack and his team[8] suggested that there are only four basic emotions. Their research demonstrates that dynamic FACS-based facial expressions transmit an evolving hierarchy of signals over time, from simpler and biologically innate face signals in the early stage of the dynamics supporting the discrimination of four categories, that is, happy, sad, fear/surprise, and disgust/anger, to more complex specific signals that finely discriminate the six facial expressions of emotion. This result is argued by observing that the confusions between surprise and fear and between disgust and anger are due to common transmission of the same AUs. In particular, both surprise and fear involve the activation of the Upper Lid Raiser (AU5) and the Jaw Drop (AU26) AUs followed up by the Upper Lid Raiser (AU5). The discrimination of surprise from fear is achieved due to the activation of the Lip Stretcher (AU20) AU. On the other hand, disgust and anger are initially confused due to the similarity of the Nose Wrinkler (AU9) AU which appears in disgust and the Brow Lowerer (AU4) AU which appears in anger. Disgust is discriminated from anger mainly by the activation of the Lip Corner Depressor (AU15) and the Lower Lip Depressor (AU16) AUs. In a recent paper, Crivelli and colleagues[9] highlighted that the assumption that the facial expression interpretations are pan-cultural derives largely from Western societies. They studied two different cultures, Spaniards and the Trobrianders (a tribe of Papua New Guinea), and found that a wide-eyed gasping face was interpreted as fear by Spaniards, as commonly happened in western culture, and as anger by Trobrianders. As counter-check, by asking to select the face that was threatening, Spaniards chose an angry scowling face, whereas Trobrianders chose the fear gasping face. These studies led the question about the interpretation of facial expressions open to new theories about how humans interpret and categorize facial expressions that could defy the universality of facial expressions. On the other hand, the whole idea of existence of basic emotions is also being questioned in the field of affective science. The field of affective science is characterized by the existence of various schools of thoughts and each one of these school of thought has a different look at emotions. Constructivist theorists such as Barrett[10] and Russell[11] contest the existence of basic emotions and consider the neurophysiological states of valence and arousal as the building blocks for the human emotional system. Appraisal theorists such as Scherer[12] and Frijda[13] consider appraisal components as the building blocks for the human emotional system. A more comprehensive survey about the various emotion theories can be found at Moors.[14]

Nevertheless, it is not the focus of this article to engage in the ongoing debate regarding the nature of emotions. The study covered in this article is based on Ekman's basic emotion theory as due to the popularity and maturity of Ekman's FACS theory a sufficient quantity of material, guidelines, and measurement tools is available which makes the basic emotion theory more convenient for designing facial expression for robot and virtual human as well as for evaluating the facial expressions.

The emotion recognition ability has been widely studied both by using static stimuli,[15–19] that is, photographs of the apex or peak of an expression, and by using dynamic stimuli,[20–24] that is, neutral faces gradually unfolding into emotional expressions. Indeed, real-life social situations are characterized by dynamic facial behaviors; therefore, it had been hypothesized that the motion inherent to facial expressions over the time plays a crucial role in discriminating and understanding them correctly.[7,25,26] This hypothesis has been also confirmed by neurological studies which found that patients were able to normally recognize dynamic facial expressions, although they were not able to recognize emotions shown as static pictures by suggesting that our brain elaborates static and dynamic expressions separately.[27–31]

As dynamics is intrinsic in facial movements, auditory information is instinctive in conveying facial expressions. Through the vocal channel, the meaning of a facial expression can be reinforced with nonlinguistic vocalizations, for example, crying, hums, grunts, laughter, or shrieks, and verbal expressions that carry explicit linguistic content. Our ability to recognize emotions from facial expressions integrated with nonlinguistic vocalizations[32–36] or with verbal information[15,37–39] has been widely studied demonstrating that the average performance in recognizing the basic emotions is generally higher than the case of facial expressions without auditory information. Indeed, acoustic features of tone, pitch, intensity, and duration contribute to convey the meaning of the emotions such as anger, fear, happiness, and sadness.[38,40,41] Most research studies have been conducted using human stimuli, that is, emotions conveyed by human beings.[16],[42–45] However, especially in the last decade, thanks to the advances in computer graphics and robotics, anthropomorphic virtual and robotic characters have been used as research tools for investigating aspects of human social communication. Virtual avatars with advanced expressive capabilities are typically used as tutors, storytellers, or caregivers.[46–50] Relative few studies have been focused on investigating these aspects on anthropomorphic robots.[19,51–53] Instead, robots endowed with highly anthropomorphic body equipped with sensors raise more challenges due to technical limitations, for example, there are some important differences in the way the human muscles and the robot servomotors actuate the face, and mechanical constraints, for example, the reduced space inside an artificial skull or a body where servomotors are placed. Moreover, creating such artificial empathic machines means making them able to communicate with humans in a believable way.

The concept of believability introduces a long-standing debate started by the roboticist Masahiro Mori in the late 1970s when he proposed the "Uncanny Valley" hypothesis: The acceptance of a humanoid robot in terms of perceived familiarity increases hand in hand with its human-likeness until a certain point where the excessive realism causes an eerie sensation evoking a negative effect.[54,55] More recently, Tinwell and collegues[56] working on human-like virtual characters, proposed the Uncanny Wall rather than the Uncanny Valley. Starting from the question "We will ever overcome the Uncanny Valley?," they changed the point of view and pointed out that rather than scrambling out of the valley, it would be more right to think about it as an unsurpassable wall as she states "We may continue to scale the uncanny wall as new human-like characters are introduced in games and animation, but there will never be the opportunity to peak the extending wall."[57]

The question of believability is increasingly important since robots have already become part of our daily life. The type of interaction is definitely based on the nature of the robot itself in order to maintain the illusion of dealing with a real human being. Thus, the aesthetic aspect is the first significant element that impacts a communication. Then the behavioral aspect is a crucial factor in evaluating the ongoing interaction. Indeed, we have more expectations when interacting with anthropomorphic robots and we tend to define them believable if they respect human social conventions. Therefore, human–robot interaction (HRI) researchers are focused both on increasingly anthropomorphizing the embodiment of the robots and on giving the robots a realistic expressive behaviour.[58,59]

The research hypotheses

A previous experiment organized at the University of Pisa, Italy was focused on evaluating the contribution of the physical embodiment of a humanoid robot in expressing emotions.[60] The robot used in that experiment was mechanically similar to the one used in this current study with human-like expressive capability and aesthetically resembles a real woman. This study compared 2-D pictures and 3-D models of human and robot expressions with the expressions performed by the robot itself in real time. The first result showed a similar participants' performance in understanding the human and robot static expressions, that is, 2-D pictures and 3-D models. A second result highlighted that the expressions performed by the physical robot in real time obtained a higher recognition rate compared to the same expressions showed as 2-D pictures and 3-D models. These results support the hypothesis that the physical embodiment and the motion inherent to the robot's expressions could improve the participants' performance in understanding emotions conveyed by a humanoid robot.

On the basis of the previous experiment, the hypotheses that guided this new experiment are as follows.

> **H1**: Evaluating whether the expressions performed by a humanoid robot are positively influenced by the dynamic aspect as it happens in case of human facial expressions by comparing static and dynamic expressions of a human female and a humanoid robot.
>
> **H2**: Examining whether the auditory information that plays an important role in understanding the emotional meaning of facial expressions of humans during their social interactions has the same positive effect in discriminating the expressions performed by a humanoid robot compared to a virtual avatar which represents its 3-D replica endowed with human-like expressive skills.

Dynamics and vocalizations are innate in human communication and are commonly considered obvious factors that contribute to discriminate and interpret the meaning of facial expressions. Conversely, these factors are not obvious in designing a robot that resembles a human being and mimics human gestures. This field of research has to face a wide variety of challenges that depends both on the nature of the robot itself, for example, linking the motor control for performing facial expressions with a vocal engine for producing vocalizations aligned with the meaning of the face is still a challenging aspect, and on the domain-specific context where the robot has to perceive and react to the human activity in real time with a natural and lifelike behavior. Therefore, the novelty of this research is focused on understanding the contribution on dynamics and vocalization in facial expressions performed by a robot and demonstrating that a robot physically similar to a human being in shape and aesthetics but with more mechanical limits due its nature can become an emphatic machine, that is, a machine able to express emotions in a believable way as humans do and associable with affective meanings.

Related works

A growing number of studies are dedicated to understanding how people perceive and elaborate emotional information and which features are relevant elements in recognizing facial expressions. Most of these studies are focused on recognizing or scoring static stimuli shown as photographs, for example, facial expressions with different levels of intensity, both of human beings and of various kinds of expressive robots.[19,47,48,61] Verbal and nonverbal communications, however, are clearly dynamic except in rare situations. Therefore, it is reasonable to hypothesize that dynamic stimuli can be detected and decoded more easily or naturally than static stimuli.[20,21,62,63] However, despite evidences that motion improves the recognition of a facial expression, other studies found no differences between dynamic and static conditions[64] which could be explained by the fact that some facial expressions can be characterized by few distinctive features, for example, a smiling mouth for a happy expression, which allows to discriminate them. Indeed, when the details that are typical of the target expression are evident, it is possible to recognize the expression even before it reaches its apex.

Due to the nature of the stimuli and the different methodologies used in the experiments, it is difficult to state a well-defined result. However, the experimental studies focused on investigating the difference between static and dynamic stimuli have raised the question whether dynamic facial expressions produce different results from static expressions.

In addition to dynamics, human communication normally includes verbal messages with a linguistic meaning and a variety of paralinguistic speech features that do not carry linguistic content such as speech rate, loudness, and pitch. These communicative aspects can reinforce the meaning of a bodily expression if they are congruent, that is, they communicate the same emotional message. For instance, the interpretation of a verbal message of agreement “Sure!” may be interpreted differently depending on the tone of the voice and the facial expression.

Virtual characters

The rapid growth in virtual reality and computer graphics has made possible to develop highly realistic virtual faces with convincing facial expressions. Indeed, the main goal of these social agents is to be user-friendly and to be able to engage people by following human social behaviors and rules.

Faita and colleagues[65] investigated the correlation between dynamism and realism of virtual faces performing facial expressions. In their study, two groups with different expertise in virtual reality were asked to associate a score 1–5 to each of the emotion rendered on the virtual character. They measured the level of intensity in the correspondence between facial expressions of virtual avatars and emotional stimuli perceived by an observer and found a high level of intensity in this correspondence in both groups through the evaluation of two variables: time response and the score assigned to each emotion.

Dyck et al.’s research team[66] investigated whether basic emotions expressed by virtual avatars are recognized as well as the emotions expressed by natural human faces and found consistent results with this hypothesis. However, they also found that the disgust was difficult to convey on the virtual avatar, whereas sadness and fear were better understood compared to the corresponding natural faces. The advances in virtual reality made possible to reach high level of realism even if improvements such as better modeling of the nasolabial area may lead to even better results as compared to trained actors.

In the Mower et al.’s study,[67] participants were asked to recognize facial expressions of an animated display with congruent vocal expression, that is, happy face and happy voice, or conflicting vocal expression, that is, happy face and angry voice. They found that the congruent combination of facial and vocal expressions was more accurately recognized than both the video-only and audio-only stimuli. Differently from the findings by De Gelder and Vroomen,[68] results also showed that the emotions presented in the audio-only evaluation data were more differentiable than in the video-only evaluations probably due to the limited expression in the animated face used in this analysis.

Social robots

In the field of social robots, relative few studies have been conducted to evaluate the ability of a humanoid robot to express emotional states through different communication channels.

Trovato and his colleagues[53] developed a facial expression generator aimed at producing thousands of combinations of facial and neck movements for a humanoid robot called KOBIAN-R. They evaluated their system through a web survey where participants were asked to label a certain number of facial expressions. Results showed that people are able to interpret the meaning of the most basic facial expressions performed by the robot KOBIAN-R. The second survey was focused on evaluating the KOBIAN-R's nonverbal abilities. Participants were asked to evaluate basic facial expressions with and without congruent and incongruous sentences. Their results proved that KOBIAN-R's nonverbal communication influenced the overall meaning in a similar way to human nonverbal cues.

Berns and Hirth[51] presented the development of a behavior-based control to realize a humanoid robotic head capable of performing realistic human facial expressions. They conducted an experimental study to evaluate the capability of the robotic head to convey the emotional meaning of its facial expressions. Participants were asked to classify the presented expression shown as photo or video as one of the six basic facial expressions with a level between 1 (weak correlation) and 5 (strong correlation). Their results showed a correct recognition for anger, happiness, and sadness while fear and disgust were not identified. Moreover, they did not find differences in recognizing facial expressions shown as pictures and through videos.

Becker-Asano and Ishiguro[19] investigated the capability to express facial emotions of a humanoid robot called Geminoid F through an online survey. Users were asked to choose among angry, fearful, happy, neutral, sad, surprise, or "none of these labels" to label a set of photos of Geminoid F's facial expressions. Results showed that participants were more confused in recognizing facial expressions of Geminoid F than the corresponding human expressions.

Materials and methods

The aim of this study was evaluating the expressiveness of highly anthropomorphic robots endowed with human-like expressive facial skills. Facial Automaton for Conveying Emotions (FACE)[69] and Eva [70] are two aesthetically equivalent human-like female robotic heads based on the same mold developed by Hanson Robotics[71] through a life-casting technique. Both robots consist of an artificial skull covered by a porous silicone elastomer called Frubber™ which is an extremely soft, supple, and strong silicone that makes it closely correlated with a living facial tissue.[72] Flexible rubber cloth anchors are designed and cast directly into the Frubber™ material and are strategically placed to simulate the stress distribution of the facial muscles on the human skin surface to reproduce realistic human-like wrinkles, folds, and bunches. The anchors are connected by yarns to the 32 servomotors that are integrated into the skull and the upper torso similarly to the major facial muscles and represent the actuation system.

FACE is a believable facial display system endowed with a passive articulated body (Figure 1(a)–left). The animation system that controls FACE has been developed by the University of Pisa, Italy[69] and it is based on FACS.[7] Figure 1 shows the robot FACE and the mapping between the major human facial muscles and the position of the servomotors inside the skull of the robot.

In a previous experiment, FACE was used to investigate the influence of its physical embodiment in conveying emotions.[60] In particular, this study aimed at evaluating the capability of FACE to perform facial expressions in terms of recognition rate and response time in comparison with static 2-D photos and 3-D models of a human female and of the robot itself. The results showed that the recognition rates of expressions performed by the physical robot were generally higher than the ones obtained by showing static 2-D photos and 3-D models both of the human female and the robot. As preliminary study, its results are encouraging and support the hypothesis that the embodiment of physical social humanoids can positively influence the discrimination of emotions in comparison with static 2-D and 3-D stimuli.[21,23,73]

The present work can be considered an extension of this previous experiment[60] in which Eva, the twin of FACE, has been used for investigating the influence of facial dynamics and utterance in conveying emotions through facial expressions. Technically and aesthetically similar to FACE, Eva is a female robotic head endowed with a human-like expression capability (Figure 2(a)) through an animation system developed by MIRALab at University of Geneva, Switzerland.[70]

Differently from FACE, this animation system is based on Moving Pictures Experts Group (MPEG)-4 Facial Animation (FA) standard,[74] a standard set of facial parameters widely used in the domain of computer graphics to animate faces of virtual characters. In the late 1990s, the MPEG introduced the facial animation parameters (FAPs), a standard to represent virtual human-like characters endowed with speech intelligibility and gesture capabilities through a very low bit-rate compression and transmission of animation parameters.[75] As a result, Eva has also a corresponding virtual avatar with its same expressive capabilities (Figure 2(b)).

The MPEG-4 FA standard defines a set of facial definition parameters (FDPs) as feature points on the 3-D facial skin mesh (Figure 3). Many of these feature points correspond to FAPs, which are used to modify the FDPs and consequently to animate the 3-D face. Each FAP value represents the displacement of a particular feature point from its neutral position, and in computer graphics, it causes the geometric deformation of the related face area. The FAP value is normalized by the facial animation parameter units, which correspond to fractions of distances between key facial features, for example, the distance between the eyes. MPEG-4 defines 84 FDPs that can be used to feature the face and 68 FAPs that can be used to animate it.

To animate the robotic head, FAP parameters and FAP ranges should be mapped to corresponding servomotors and servo ranges, respectively.[70] First of all, FAPs have

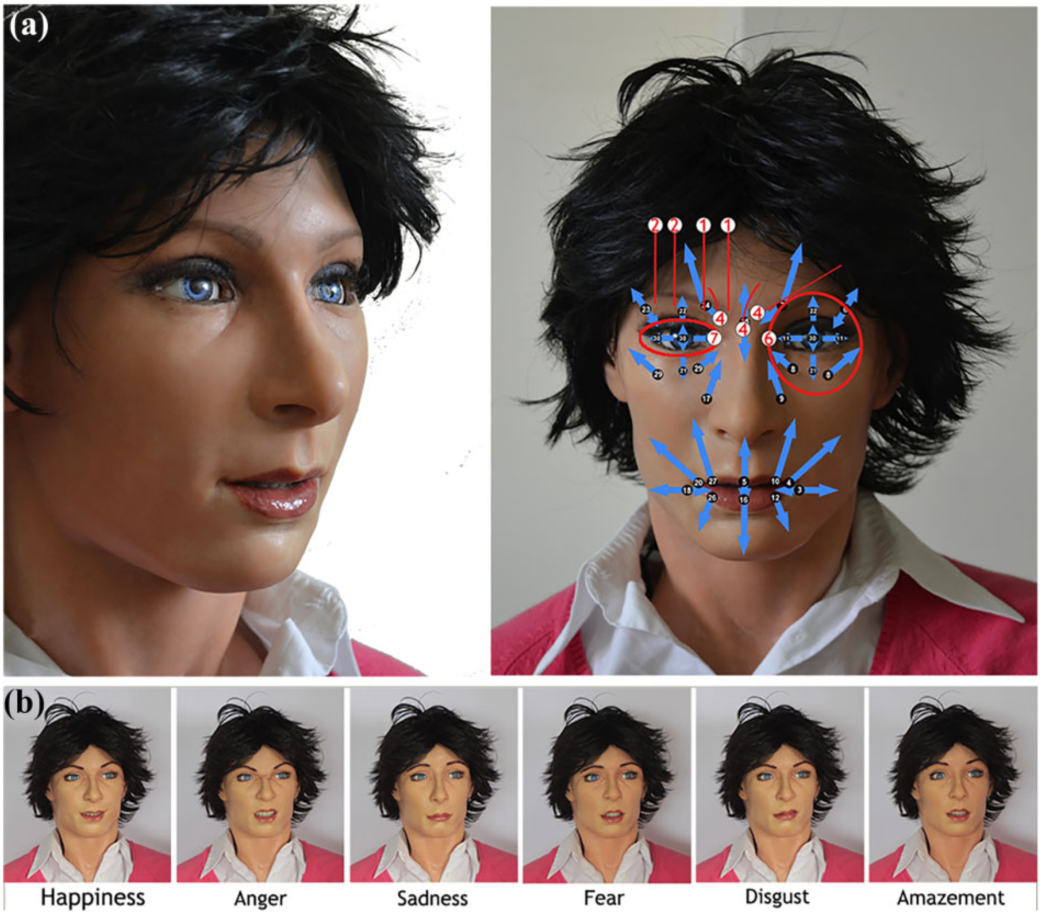

Figure 1. (a) The robot FACE (left), the major facial muscles taken into consideration (middle), and the servomotors positions inside the skull with an example of mapping between servomotor positions and action units of FACS (right); and (b) basic facial expressions performed by FACE. FACE: Facial Automaton for Conveying Emotions; FACS: Facial Action Coding System.

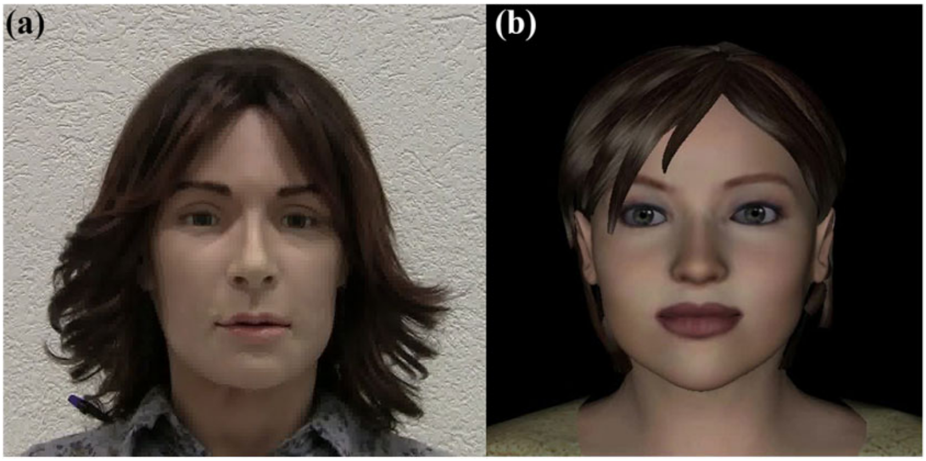

Figure 2. (a) The humanoid robot Eva and (b) the Eva's avatar.

been mapped to the corresponding servomotors by observing the effects of the movements of each servo on the robot face. The result is a table of FAP-servo conversion rules. Successively, undesirable situations such as two FAP parameters controlling the same servo at the same time were avoided by applying conditional statements. At the

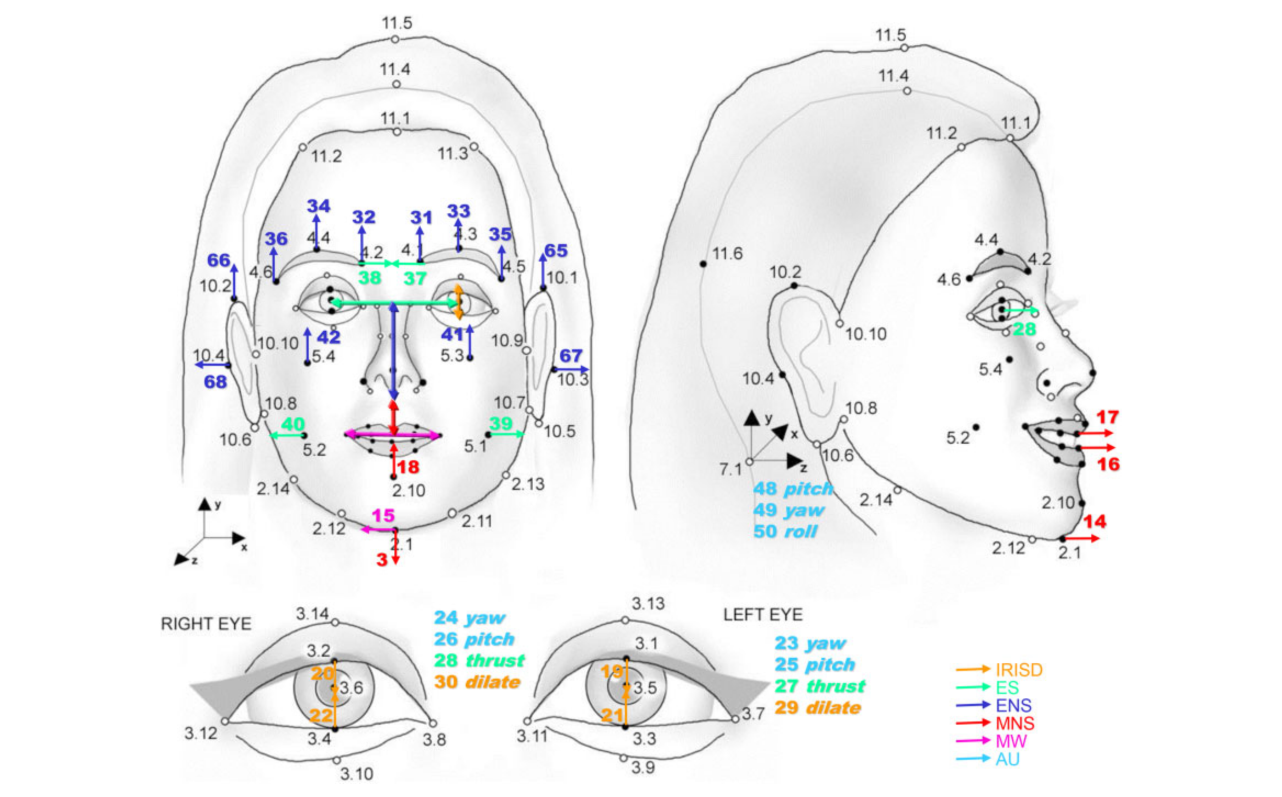

Figure 3. MPEG-4 FDPs and FAPs definition. MPEG: Moving Pictures Experts Group; FDP: facial definition parameter; FAP: facial animation parameter.

end, due to mechanical limitations, the minimum and maximum values of the servos and the FAP parameters were calculated by moving the servos to its extremities and creating the corresponding FAP animations on a virtual human to match the effect of the servo on the robot face. The neutral face corresponding to 0 for FAP parameters was reproduced on the robot face and the resulting values of the servos were considered as the neutral face values for the servomotors. The final mapping is a matrix used to generate the servomotor values at each animation frame through an interpolation between manually set ranges of both robot and virtual human (Figure 4).

The Eva animation system consists of several components that make it possible to animate the robotic head or the virtual character based on prerecorded animations acquired from motion capture systems or from dynamically generated animations.[70] At its core, there is an advanced animation component able to play FAP animation sequences, to interpolate between FAP key frames and to blend different animations together based on several techniques,[76,77] making it possible for Eva to speak and to express emotions at the same time as well as to produce realistic and natural transitions between facial expressions. In more interactive setting, Eva is equipped with components responsible for dynamic generation of lip movements, eye movements, and emotional expressions based on input from the dialogue manager as well as from commercial text-to-speech technologies.[78] This line of research has given Eva the capability to interact with the users in a natural way.

Although FACE and Eva are controlled by two different animation systems, both robots are similar in terms of mechanical and technical features. The motor controller of the robots has a maximum frame rate of 25 fps, and therefore, the maximum update frequency of the servo positions is 25 fps. The maximum speed of the servomotor used for the pushing and pulling of the face tendons of the robots is 0.23 s/60° with a maximum applicable torque of 10.3 kg-cm. Moreover, their expressions are based on the same anatomical basis. Similarly to the philosophy behind the AUs in FACS, servos in the robot are positioned according to the major facial muscles and their skin displacement behavior is based on what the muscles allow the face. The computer graphics standard for animation MPEG-4 also follows the same principle. The positions and directions of MPEG-4 FAP points also correspond to the anatomical basis of the facial muscles.

The virtual avatar is controlled by an MPEG-4 computer animation engine that runs at 30 fps with a resolution of 1024 × 768 pixel. The virtual face is animated by geometrical deformation of the facial mesh based on

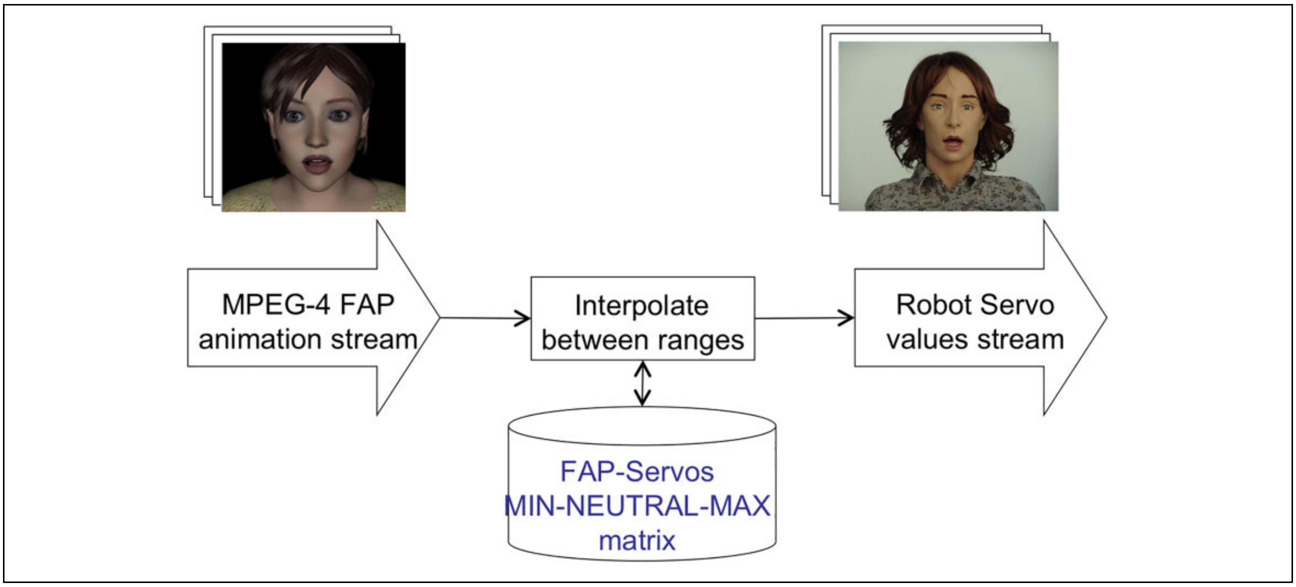
MPEG-4 FAP animation stream
Interpolate between ranges
Robot Servo values stream
FAP-Servos MIN-NEUTRAL-MAX matrix

Figure 4. Conversion from MPEG-4 FAPs to servomotor positions. MPEG: Moving Pictures Experts Group; FAP: facial animation parameter.

the displacement of the MPEG-4 FAP points. As will be explained later in this study, MPEG-4 animations are created using a motion capture system with the goal of generating more human-like facial expression. The selected animations are first applied and tested on the virtual avatar and then converted to animate the robot using a specific algorithm developed by the Eva's research team.[70]

Both systems have their advantages and disadvantages. The virtual human does not exist in the physical world and its skin deformation technique may not produce results that look as realistic as the deformation of the physical skin in the robotic system. The appearance and disappearance of facial wrinkles in the robotic skin in certain facial expression can be considered as an advantage comparing to the virtual human system. On the other hand, the advantage of the virtual human system is that it has smoother movements during the animation comparing to the robot. Due to the mechanical nature of the servomotors, the resulting skin animation of the robot can be considered a bit shaky in comparison with the virtual human. Another disadvantage of the use of servomotors is the mechanical/electrical noise that they produce during the movements which can be annoying the human counterpart.

Static and dynamic stimuli

The set of stimuli of the first phase included six photographs (static stimuli) and six video recordings (dynamic stimuli) of the basic expressions, that is, anger, disgust, fear, happiness, sadness, and surprise, performed by a female amateur actress and by Eva, the robot chosen for this study. Dynamic stimuli were shown as video recordings of the transition from the neutral to the selected expression (10 s) followed by 20 s of black screen. Static stimuli were shown as photographs taken from the peak of the expression of the video recordings (at about 10 s).

The second and third phases compared the recognition of stimuli performed by the Eva's avatar counterpart and performed by the physical robot Eva in three different conditions: (1) muted facial expressions; (2) facial expressions with nonlinguistic vocalizations, that is, “Hey!” (happiness), “Mmhh” (sadness), “Aahh” (fear), “Oohh” (surprise), “Bleah!” (disgust), and “Grrr!” (anger); and (3) facial expressions with an emotionally neutral verbal sentence, that is, “What is happened?”. Indeed, vocal expressions of emotions can occur overlaid on speech in the form of affective prosody together with a range of nonverbal vocalizations often referred to as “interjection” which do not carry linguistic content, for example, screams, laughs, yawns, and other such vocal outbursts, or verbal sentences that explicitly communicate linguistic information.

In order to create the set of the stimuli, an optical tracking system (VICON 8) with six cameras was used to capture the facial movements of an amateur actress.[70] According to the MPEG-4 FA standard, 27 markers were positioned on the actress' face following the FDPs. The output of the motion capture system was the 3-D trajectories of the marker points for each facial expression which were converted to FAPs-based files through a specific algorithm. In order to animate the robot with the same stimuli, the FAPs-based files were converted to be compatible with the robot animation system, as previously mentioned.

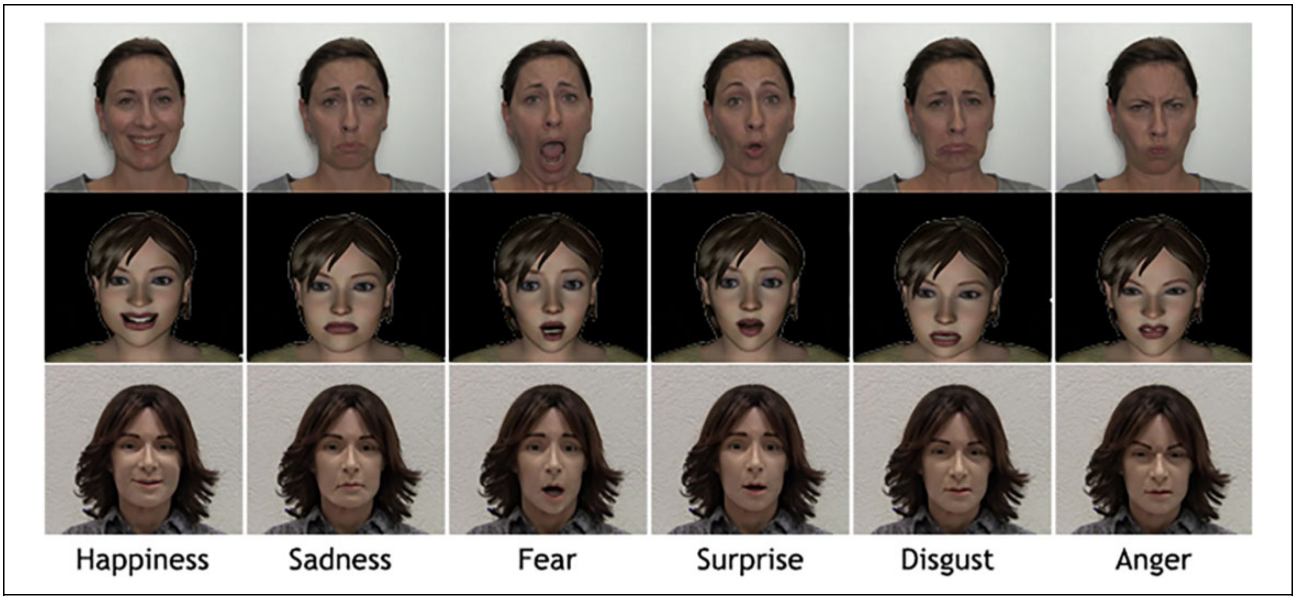

Figure 5. Stimuli used in the experiment: human (first row), avatar (second row), and robot (third row).

Separately, the audio tracks were extracted by the video recorded during the motion capture process. Each audio track was manually synchronized with the corresponding expression of both the avatar and the robot. In the first case, each audio track was synchronized with a video recording of the avatar performing the corresponding expression using a software video editing tool. In the latter case, the audio track was manually aligned with the robot animation through its software control system.

The resulting stimuli for the second and third phases were six animations of Eva's avatar and six real-time animations of the robot Eva performing the basic expressions in the three different conditions in which the corresponding audio was synchronized with the mouth movements.

Figure 5 shows an example of stimuli presented during the entire experiment for the human female, the avatar, and the physical robot (a video summary of the stimuli used in the experiment is available at https://www.dropbox.com/sh/7jqbmax3marv2g4/AADd9QpWeIedaO4rz2sSNXCHa?dl=0).

Participants

A total of 25 voluntary students and researchers (14 males, 11 females) aged 19–37 years (mean age 28.3 ± 5.8) and working in the scientific area were recruited for the experiment. All participants gave a written informed consent for participating in the experiment.

Limitations

This study considers a small number of participants, mainly for the following reason. While long-term studies are very desirable to evaluate various aspects of HRI, such as user's perception and reaction to robots, only a relatively small amount of long-term studies with numerous groups of participants have been published in this field. The main reason is the cost in terms of research time, hardware and software development, data acquisition and analysis, experiment organization, funding for the equipment, and sometimes the time for person necessary to perform the entire experiment. Therefore, normally the first step in this field is organizing a pilot study with a small group of participants to evaluate the feasibility and the effectiveness of the experiment and in a second moment, on the basis of the results of the pilot study, designing a long-term experiment that involves a large group of people.

Procedure

Participants were seated comfortably at a desk about 0.5 m far from a monitor (during all the phases) and the robot (only during the third phase). The avatar has been shown as a full screen application on a 32 inches 16:9 PC monitor with a resolution of 1920 × 1080 pixels. The monitor was plugged via high-definition multimedia interface connection and the application resolution upscale was managed by the PC graphic card driver installed on a Windows 8 OS. Before the start of the experiment, participants were asked to fill in a form with their demographic information. During the experiment, at each phase, the subjects were asked to label the facial expressions shown on the screen or performed by the robot, by selecting on the screen one of the seven labels, that is, anger, disgust, fear, happiness, sadness, surprise, and I do not know.

The experimental protocol included three phases.

Phase 1.

- *Stimuli*: Six photographs of the robot and of the human female face performing an expression in random order (10 s of stimulus + 20 s of black screen).
- *Stimuli*: Six recorded videos of the robot and of the human female face performing an expression in random order (10 s of stimulus + 20 s of black screen).

Phase 2.

- *Stimuli*: Six animations of the avatar performing an expression without sound in random order (10 s of stimulus + 20 s of neutral expression).
- *Stimuli*: Six animations of the avatar performing an expression with a congruent nonlinguistic vocalization in random order (10 s of stimulus + 20 s of neutral expression).
- *Stimuli*: Six animations of the avatar performing an expression together with a verbal sentence in random order (10 s of stimulus + 20 s of neutral expression).

Phase 3.

- *Stimuli*: Six real-time animations of the physical robot performing an expression without sound in random order (10 s of stimulus + 20 s of neutral expression).
- *Stimuli*: Six real-time animations of the physical robot performing an expression with a congruent nonlinguistic vocalization in random order (10 s of stimulus + 20 s of neutral expression).
- *Stimuli*: Six real-time animations of the physical robot performing an expression together with a verbal sentence in random order (10 s of stimulus + 20 s of neutral expression).

The experiment was conducted in a controlled laboratory environment and the setup included one laptop for controlling the video animation and one desktop for controlling the robot. The robot was placed in a different area of the same room and covered with a blanket during the first two phases in order to do not influence the participants' evaluation.

Data analysis and results

The Cohen's κ coefficient[79] was used as a measure for inter-rater reliability. This method is commonly used in this kind of study to evaluate the agreement among all subjects on the assignment of labels to a categorical variable.[19,67,80] According to Landis and Koch,[81] with a significance level of 0.05, κ can be classified in the following ranges: $\kappa \leq 0.00$ less than chance agreement; $0.01 < \kappa < 0.20$ slight agreement; $0.21 < \kappa < 0.40$ fair agreement; $0.41 < \kappa < 0.60$ moderate agreement; $0.61 < \kappa < 0.80$ substantial agreement; and $0.81 < \kappa \leq 1$ almost perfect agreement. Results are presented in the form of confusion matrix, that is, a specific table which contains information about the presented models (on the columns) against the selected labels (on the rows). The statistical inference was carried out using the OriginLab software [version 2015].[82]

Table 1. Confusion matrix ($N = 25$) of the recognition rates (in %) of the six human facial expressions with the presented models (columns) against the selected labels (rows).[a]

Phase 1: Recognition rates (in %) of human stimuli												
	Static						Dynamic					
	A	D	F	Sa	H	Su	A	D	F	Sa	H	Su
A	*76*	0	0	0	0	0	*96*	0	0	0	0	0
D	12	12	0	0	0	4	4	16	0	0	0	0
F	0	0	*92*	0	0	4	0	0	*96*	0	0	0
Sa	4	*76*	0	*92*	0	0	0	*76*	0	*88*	0	0
H	0	0	0	0	*92*	0	0	0	0	0	*100*	0
Su	0	0	8	0	0	*88*	0	0	0	0	0	*96*
No	8	12	0	8	8	4	0	8	4	12	0	4

A: anger; D: disgust; F: fear; H: happiness; Sa: sadness; Su: surprise; No: I do not know.

[a] Highest values are set in italics.

Static versus dynamic stimuli

Table 1 shows the confusion matrix of the subjects' answers for the human static and dynamic stimuli. For both categories, there was a substantial agreement in judging the facial expressions: $K_{\text{HumStatic}} = 0.768$ ($p < 0.001$, 95% CI (0.690–0.847)) for static stimuli and $K_{\text{HumDynamic}} = 0.832$ ($p < 0.001$, 95% CI (0.764–0.900)) for dynamic stimuli. The human disgust was the only not well-recognized expression since it was labeled as "sadness" (76%) both in static and dynamic stimuli. For the other human expressions, the recognition rate was higher than 76% for both types of stimuli.

Table 2 shows the confusion matrix of the subjects' answers for the robot static and dynamic stimuli. In both categories, there was a moderate agreement in judging the facial expressions: $K_{\text{RobStatic}} = 0.573$ ($p < 0.001$, 95% CI (0.470–0.675)) for static stimuli and $K_{\text{RobDynamic}} = 0.602$ ($p < 0.001$, 95% CI (0.503–0.701)) for dynamic stimuli. The best recognition rate was achieved for robot anger and happiness both for static stimuli (92% and 88%, respectively) and dynamic stimuli (92% and 96%, respectively). The worst recognized expression was the robot disgust in both categories with a low agreement among all the subjects. The expression intended to convey fear was confused with "surprise" (60% in static stimuli, 92% in dynamic stimuli). Finally, the robot sadness in the dynamic

Table 2. Confusion matrix ($N = 25$) of the recognition rates (in %) of the six robot facial expressions with the presented models (columns) against the selected labels (rows).[a]

Phase 1: Recognition rates (in %) of robot stimuli												
	Static						Dynamic					
	A	D	F	Sa	H	Su	A	D	F	Sa	H	Su
A	*92*	12	0	16	0	0	*92*	*36*	0	0	0	0
D	8	24	0	4	0	4	0	28	0	12	0	0
F	0	8	24	0	0	24	0	4	8	0	0	8
Sa	0	12	4	*40*	0	12	0	4	0	36	0	4
H	0	0	0	0	*88*	0	0	4	0	4	*96*	0
Su	0	4	*60*	0	4	*48*	0	0	*92*	0	0	*84*
No	0	*40*	12	40	8	12	8	24	0	*48*	4	4

A: anger; D: disgust; F: fear; H: happiness; Sa: sadness; Su: surprise; No: I do not know.

[a] Highest values are set in italics.

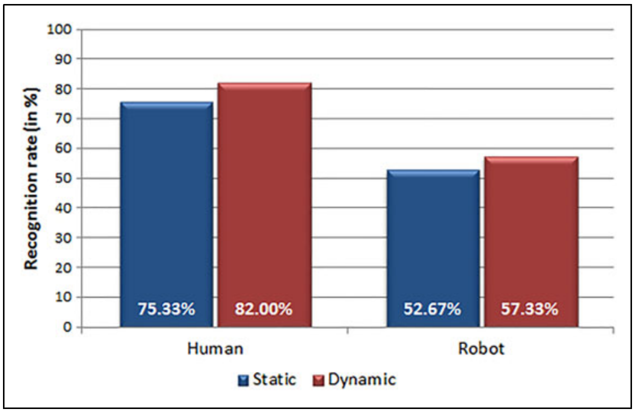

Figure 6. Recognition rates (in percentage) of 25 subjects for human and robot expressions in the two different conditions: static and dynamic stimuli.

condition was not well understood by choosing "I do not know" (48%).

Figure 6 shows the recognition rate grouped by the type of stimulus: in both cases, that is, human stimuli and robot stimuli, there was a tendency to slightly better recognize dynamic expressions than static expressions.

Muted stimuli versus stimuli with auditory information

Table 3 shows the confusion matrix of the subjects' answers for the Eva's avatar stimuli. The Cohen's κ statistics show a moderate agreement $K_{\text{AvatMute}} = 0.590$ ($p <$ 0.001, 95% CI (0.491–0.688)) in recognizing muted stimuli, a very good agreement $K_{\text{AvatSound}} = 0.842$ ($p <$ 0.001, 95% CI (0.774–0.910)), and a good agreement $K_{\text{AvatSent}} = 0.739$ ($p < 0.001$, 95% CI (0.655–0.823)), for stimuli performed together with nonlinguistic vocalizations and verbal sentence, respectively. A first comparison among the three categories shows a higher degree of confusion in recognizing muted stimuli than stimuli with nonlinguistic vocalization and verbal sentence. More specifically, muted anger was confused with "disgust" most often (28%) or was not recognized at all (32%). Similarly, muted fear was confused with "surprise," and in half of the cases, muted sadness was not recognized (48%). Moreover, the expressions intended to convey fearful without sounds were labeled as "surprised" in the most cases (92%). Recognition rates for stimuli with nonlinguistic vocalization and verbal sentence show a trend which looks better than for the rates with the muted stimuli. Only disgust expressed together with a verbal sentence was confused with "anger" most often (64%).

Figure 7 shows the recognition rates of the avatar stimuli in three different conditions, that is, muted, with nonlinguistic vocalization and with verbal sentence. This result highlights a tendency to better recognize expressions combined with auditory information (nonlinguistic vocalization or verbal sentence) than muted expressions.

Table 4 shows the confusion matrix of the subjects' answers for the robot Eva's stimuli. Results of the Cohen's κ statistics for stimuli performed by Eva are similar to the previous case. There is a moderate agreement in recognizing muted stimuli with $K_{\text{RobotMute}} = 0.671$ ($p < 0.001$, 95% CI (0.582–0.760)), a very good agreement with $K_{\text{RobotSound}} = 0.807$ ($p < 0.001$, 95% CI (0.735–0.879)), and a good agreement with $K_{\text{RobotSent}} = 0.797$ ($p < 0.001$, 95% CI (0.722–0.871)) for stimuli performed with nonlinguistic vocalizations and verbal sentence, respectively. The trend of the recognition rates in the three categories is similar to the case of the Eva's avatar stimuli. In case of muted stimuli, disgust was confused with "anger" (52%) while fear was labeled as "surprise" in half of the cases (52%) and was correctly recognized in the other half (48%). The robot stimuli performed with nonlinguistic vocalization and verbal sentence were generally better recognized than the only visual stimuli. Even in this case, disgust with a verbal sentence was confused with "anger" most often (64%).

Similar to the avatar stimuli, there was a tendency to better recognize expressions combined with audio information, that is, nonlinguistic vocalization or verbal sentence, than muted expressions as shown in Figure 8.

Conclusions

The experiment aimed at investigating whether (hypothesis 1) the dynamics underlying human facial expressions entails advantages even in the case of an expressive humanoid robot and whether (hypothesis 2) nonlinguistic vocalizations and verbal information influence the recognition of facial expressions performed by a humanoid robot in comparison with the same visual stimuli without auditory information.

Regarding the first hypothesis (hypothesis 1), that is, whether the dynamics underlying human facial expressions entails advantages even in the case of an expressive

Table 3. Confusion matrix ($N = 25$) of the recognition rates (in percentage) of the six facial expressions performed by the Eva's avatar in three different conditions (muted, with nonlinguistic vocalization, and with verbal sentence) with the presented models (columns) against the selected labels (rows).[a]

Phase 2: Recognition rate (in %) for avatar stimuli																		
	Muted						Nonlinguistic vocalization						Verbal sentence					
	A	D	F	Sa	H	Su	A	D	F	Sa	H	Su	A	D	F	Sa	H	Su
A	28	44	0	0	0	0	*80*	0	0	0	0	0	*84*	*64*	0	0	0	0
D	*36*	*56*	0	4	8	0	4	*88*	0	0	0	0	4	20	0	0	0	0
F	0	0	20	0	0	0	0	0	*52*	0	0	16	0	0	*76*	0	0	4
Sa	0	0	0	*48*	0	0	0	0	0	*76*	0	0	0	0	20	*92*	0	0
H	0	0	0	0	*92*	0	0	0	0	0	*96*	0	0	0	0	0	*56*	0
Su	4	0	*80*	0	0	*100*	4	0	48	0	0	*84*	0	0	0	4	20	*92*
No	32	0	0	*48*	0	0	12	12	0	24	4	0	12	16	4	4	24	4

A: anger; D: disgust; F: fear; H: happiness; Sa: sadness; Su: surprise; No: I do not know.

[a] Highest values are set in italics.

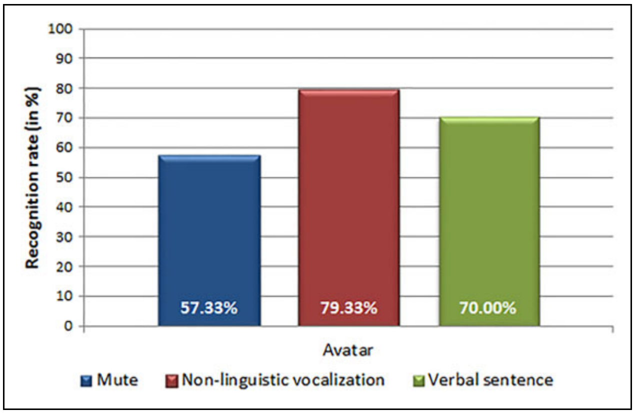

Figure 7. Recognition rates (in percentage) of 25 subjects for the expressions performed by the avatar in the three different conditions: stimuli without auditory information, with nonlinguistic vocalization, and with verbal sentence, respectively.

humanoid robot, preliminary results related to the recognition scores showed that the static stimuli were more ambiguous than the dynamic stimuli both for human and robot facial expressions even if with a different degree, that is, the agreement in the case of human stimuli was higher than the one of robot stimuli. More specifically, recognition rates of facial expressions performed by the robot revealed a difficulty in recognizing "disgust," "fear," and "sadness" which was present only for the "disgust" in case of human expressions. The fact that these negative expressions were more difficult to recognize in comparison with the positive expressions corresponds to findings in literature which state that positive emotions may be visually simpler to be recognized than negative facial expressions.[83,84] However, this does not apply to the expression "anger" which was well recognized negative emotional expression by the subjects in the current study. This suggests that we should not over-simplify emotional expression in just positive and negative emotions, but also consider other factors and affect dimension.[85]

Regarding the second hypothesis (hypothesis 2), that is, whether nonlinguistic vocalizations and verbal information influence the recognition of facial expressions performed by a humanoid robot in comparison with the same visual stimuli without auditory information, there was a general ambiguity in recognizing muted facial expressions in comparison with facial expressions associated with nonlinguistic vocalization and verbal information both for stimuli generated by an virtual avatar and for stimuli performed by a physical robot. As in the previous case, muted stimuli of negative expressions, that is, "anger," "fear," and "sadness" shown by the avatar and "disgust" and "fear" performed by the robot, were more confused than the other expressions as demonstrated in literature.[79,80] Stimuli with nonlinguistic vocalization and verbal information were generally well recognized with a recognition score higher than 52% and 56% for avatar stimuli and 64% and 84% for robot with nonlinguistic vocalization and verbal information, respectively. The only exception was the "disgust" with the verbal sentence that was confused with "anger" both for avatar and robot stimuli.

Further, the results showed a strong confusion between the emotions disgust and anger in both static and dynamic expressions of the robot and virtual avatar, also found in previous studies.[51,66] If we consider Ekman's and Friesen's FACS,[7] we think the reason for the confusion is mainly the fact that the so-called AU 4 (Brow Lowerer) which is present in anger and AU 9 (Nose Wrinkler) which is present in disgust are easily confused with each other as both produce wrinkles around the nose region. Furthermore, the results also showed that the emotion fear had often been recognized as the emotion surprise, also in both static and dynamic expressions of the robot and virtual avatar. This is also understandable as both emotions share the AUs: 1 (Inner Brow Raiser), 2 (Outer Brow Raiser), 5 (Upper Lid Raiser), and 26 (Jaw Drop). The fact that there is a strong confusion between disgust and anger and between fear and surprise does not come as a big surprise, as a recent study

Table 4. Confusion matrix ($N = 25$) of the recognition rates (in percentage) of the six facial expressions performed by the robot Eva in three different conditions (muted, with nonlinguistic vocalization, and with verbal sentence) with the presented models (columns) against the selected labels (rows).[a]

Phase 3: Recognition rate (in %) for robot stimuli																		
	Muted						Nonlinguistic vocalization						Verbal sentence					
	A	D	F	Sa	H	Su	A	D	F	Sa	H	Su	A	D	F	Sa	H	Su
A	*92*	*52*	0	0	0	0	*76*	0	0	0	0	0	*96*	*64*	0	0	0	0
D	0	36	0	12	0	0	12	*88*	0	4	0	0	4	24	0	0	0	0
F	0	0	48	0	0	36	0	0	*64*	0	0	32	0	0	*84*	0	0	8
Sa	0	0	0	*76*	0	0	0	4	0	*84*	0	0	0	0	4	*96*	0	8
H	0	0	0	0	*88*	0	0	0	0	0	*100*	0	0	0	0	0	*88*	0
Su	0	0	*52*	0	0	*64*	4	0	36	0	0	*68*	0	0	8	0	0	*84*
No	8	12	0	12	12	0	8	8	0	12	0	0	0	12	4	4	12	0

A: anger; D: disgust; F: fear; H: happiness; Sa: sadness; Su: surprise; No: I do not know.
[a]Highest values are set in italics.

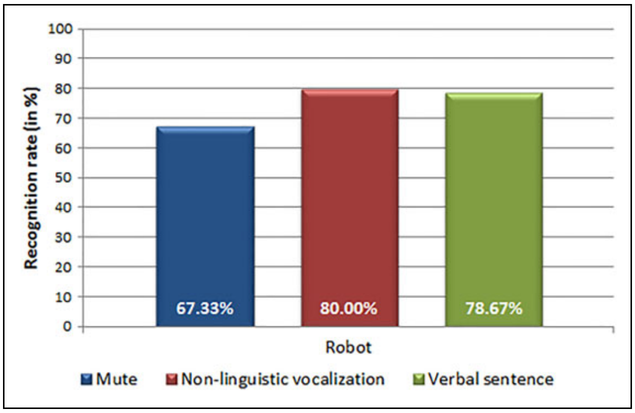

Figure 8. Recognition rates (in percentage) of 25 subjects for the expressions performed by the robot in the three different conditions: stimuli without audio information, with nonlinguistic vocalization, and with verbal sentence, respectively.

from affective science has shown that there are strong similarities between these emotions.[8] Jack and colleagues argue that only four basic emotions (happiness, sadness, fear/surprise, and anger/disgust) exist and that the distinctions between fear and surprise and between anger and disgust appeared later for social reasons. An interesting observation from our results is that the subjects mainly distinguished disgust from anger when a nonlinguistic voice accompanies the dynamic animation, differently from the study by Mower et al.[67] where emotions were better recognized in audio-only mode. We think this can be mainly explained with the fact that disgust has very prototypical nonlinguistic vocal expressions such as “Bleah!” that distinguish it from other emotions. Surprise and fear, on the other hand, are mainly distinguishable with verbal sentences, where one clearly can hear the different between the strongly negative valenced fear and slightly positive valenced surprise.

In conclusion, these results demonstrate that the presence of motion improves and makes it easier to recognize facial expressions even in case of a humanoid robot. Moreover, auditory information (nonlinguistic vocalizations and verbal sentence) helps to discriminate facial expressions both in case of a virtual 3-D avatar and a humanoid robot and seems to be very important for distinguishing fear from surprise and anger from disgust. Negative expressions which resulted in more ambiguous than the positive ones will be improved for future studies and a wider set of expressions will be created in order to enrich the robot’s and avatar’s expressiveness.

This work is a preliminary but encouraging step demonstrating that advanced high-tech tools, like humanoid robots and virtual characters, can potentially engage and entertain social interactions. Adding motion and vocalization made the expressiveness of the robot more reliable. Whether they are physical or virtual, these social agents can be used in various fields ranging from entertainment and education to human assistance and health care[86,87] to engage people to interact and communicate with others by following our own social behaviors and rules.

Future work

The results of this preliminary experiment have demonstrated the promising social capabilities of the robot EVA to perform human-like expressions which is an essential starting point for the development of real social agents. Future work will focus on studies with a large group of participants which was one of the limitations of this experiment.

This experiment highlighted that generating effective and sympathetic emotional facial expressions requires a high-fidelity reproduction and animatronic- and artistic-related expertise. Therefore, future works should consider to redefine and improve the generation of facial expressions performed by the robot taking care of these two important factors, that is, the motion to make the movement more natural and human-like and the nonlinguistic vocalization and verbal information to improve the effectiveness and empathy of the expressions. Consequently, other factors, such as speed and frequency, will be considered in the animation of the expressions, to reflect the real dynamism of the human expressivity.

Acknowledgements

Authors would like to thank Marlene Arevalo for her availability as an actress in recording the set of stimuli for the experiment and Niels Nijdam for his technical support during the motion capture acquisition process.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research has been funded by the European research projects EASEL (Expressive Agents for Symbiotic Education and Learning) (FP7-ICT-611971), CaMe-Li (AAL-2012-5), and Miraculous-Life (FP7-ICT-611421).

References

1. Bell C. *Essays on the anatomy and philosophy of expression*. London: John Murray, 1824.
2. Darwin C. *The expression of the emotions in man and animals*. London: John Murray, 1872.
3. Crichton-Browne J. On emotional expression. *Trans J Proc Dum Gall Natural History Antiq Soc* 1895; 11: 72–77.
4. Ekman P and Friesen WV. The repertoire of non verbal behaviour. *Semiotica* 1969; 1: 49–98.
5. Ekman P. Universal and cultural differences in facial expressions of emotion. In: Cole J (ed) *Nebraska symposium on motivation, 1971, pp. 207–283*. Lincoln: University of Nebraska Press.
6. Ekman P and Friesen WV. Constants across cultures in the face and emotion. *J Pers Soc Psychol* 1971; 17(2): 124–129.
7. Ekman P and Friesen WV. *Pictures of facial affect*. Palo Alto: Consulting Psychologists Press, 1976.
8. Jack RE, Garrod OGB, and Schyns PG. Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Curr Biol* 2014; 24(2): 187–192.
9. Crivelli C, Russell JA, Jarillo S, et al. The fear gasping face as a threat display in a Melanesian society. *P Natl A Sci* 2016; 113(44): 12403–12407.
10. Barrett LF. Are emotions natural kinds? *Perspect Psychol Sci* 2006; 1(1): 28–58.
11. Russell JA. Emotion, core affect, and psychological construction. *Cogn Emot* 2009; 23(7): 1259–1283.
12. Scherer KR. Emotions are emergent processes: they require a dynamic computational architecture. *Philos Trans R Soc Lond B Biol Sci* 2009; 364(1535): 3459–3474.
13. Frijda NH. Emotion, cognitive structure, and action tendency. *Cogn Emot* 1987; 1(2): 115–143.
14. Moors A. Theories of emotion causation: a review. *Cogn Emot* 2009; 23(4): 625–662.
15. Ekman P. *Emotion in the human face*. New York: Pergamon Press, 1982.
16. Matsumoto D and Ekman P. American-Japanese cultural differences in intensity ratings of facial expressions of emotion. *Motiv Emot* 1989; 13(2): 143–157.
17. Ekman P. Are there basic emotions? *Psychol Rev* 1992; 99(3): 550–553.
18. Matsumoto D, LeRoux J, Wilson-Cohn C, et al. A new test to measure emotion recognition ability: Matsumoto and Ekman's Japanese and Caucasian brief affect recognition test (JACBART). *J Nonverbal Behav* 2000; 24(3): 179–209.
19. Becker-Asano C and Ishiguro H. Evaluating facial displays of emotion for the android robot Geminoid F. In: *IEEE workshop on affective computational intelligence (WACI)*, Paris, France, 11–15 April 2011, pp. 1–8. IEEE.
20. Bassili JN. Facial motion in the perception of faces and of emotional expression. *J Exp Psychol Human* 1978; 4(3): 373–379.
21. Wehrle T, Kaiser S, Schmidt S, et al. Studying the dynamics of emotional expression using synthesized facial muscle movements. *J Pers Soc Psychol* 2000; 78(1): 105–119.
22. Kamachi M, Bruce V, Mukaida S, et al. Dynamic properties influence the perception of facial expressions. *Percep* 2001; 30(7): 875–887.
23. Ambadar Z, Schooler J, and Cohn JF. Deciphering the enigmatic face: the importance of facial dynamics in interpreting subtle facial expressions. *Psychol Sci* 2005; 16(5): 403–410.
24. Cunningham DW and Wallraven C. Dynamic information for the recognition of conversational expressions. *J Vis* 2009; 9(13): 1–17.
25. Bruce V and Young A. Understanding face recognition. *Brit J Psychol* 1986; 77(3): 305–327.
26. Bruce V and Valentine T. When a nod's as good as a wink: the role of dynamic information in facial recognition. In: Gruneberg MM, Morriss PE, and Syke RN (eds) *Practical aspects of memory: Current research and issues*, Wiley: Chichester, 1988, pp. 169–217.
27. Humphreys GW, Donnelly N, and Riddoch MJ. Expression is computed separately from facial identity, and it is computed separately for moving and static faces: neuropsychological evidence. *Neuropsychologia* 1993; 31(2): 173–181.
28. Adolphs R, Tranel D, and Damasio AR. Dissociable neural systems for recognizing emotions. *Brain Cognition* 2003; 52(1): 61–69.
29. Kevin SL, Crupain MJ, Voyvodic JT, et al. Dynamic perception of facial affect and identity in the human brain. *Cereb Cortex* 2003; 13(10): 1023–1033.
30. Schwaninger A, Wallraven C, Cunningham DW, et al. Processing of facial identity and expression: a psychophysical, physiological, and computational perspective. In: Anders S, Ende G, Junghofer M, Kissler J, and Wildgruber D (ed), *Understanding emotions, volume 156 of progress in brain research*. New York: Elsevier, 2006, pp. 321–343.

31. Schultz J and Pilz KS. Natural facial motion enhances cortical responses to faces. *Exp Br Res* 2009; 194(3): 465–475.
32. Schröder M. Experimental study of affect bursts. *Speech Commun* 2003; 40(1–2): 99–116.
33. Simon-Thomas ER, Keltner DJ, Sauter D, et al. The voice conveys specific emotions: evidence from vocal burst displays. *Emotion* 2009; 9(6): 838–846.
34. Hawk ST, van Kleef GA, Fischer AH, et al. "Worth a thousand words": absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion* 2009; 9(3): 293–305.
35. Sauter DA, Eisner F, Calder AJ, et al. Perceptual cues in nonverbal vocal expressions of emotion. *Q J Exp Psychol* 2010; 63(11): 2251–2272.
36. Lima CF, Castro SL, and Scott SK. When voices get emotional: a corpus of nonverbal vocalizations for research on emotion processing. *Behav Res Methods* 2013; 45(4): 1234–1245.
37. Scherer KR, Banse R, Wallbott HG, et al. Vocal cues in emotion encoding and decoding. *Motiv Emot* 1991; 15(2): 123–148.
38. Juslin PN and Laukka P. Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion* 2001; 1(4): 381–412.
39. Ekman P and Rosenberg EL. *What the face reveals: basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. 2nd ed. New York: Oxford University Press, 2005.
40. Pakosz M. Prosodic features and emotive meaning. *Lingua* 1982; 58(3–4): 309–326.
41. Scherer KR. Vocal communication of emotion: a review of research paradigms. *Speech Commun* 2003; 40(1–2): 227–256.
42. Gilles K and Doré FY. Judgment of facial expressions of emotion as a function of exposure time. *Percept Motor Skills* 1984; 59(1): 147–150. PMID: 6493929.
43. Andrew WY, Rowland D, Calder AJ, et al. Facial expression megamix: tests of dimensional and category accounts of emotion recognition. *Cognition* 1997; 63(3): 271–313.
44. Elaine F, Lester V, Russo R, et al. Facial expressions of emotion: are angry faces detected more efficiently? *Cogn Emot* 2000; 14(1): 61–92.
45. Manuel GC and Lundqvist D. Facial expressions of emotion (KDEF): identification under different display-duration conditions. *Behav Res Methods* 2008; 40(1): 109–115.
46. De Rosis F, Pèlachaud C, Poggi I, et al. From Greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent. *Int J Hum Comput St* 2003; 59: 81–118.
47. Gockley R, Forlizzi J, and Simmons R. Interactions with a moody robot. In: *Proceedings of the 1st ACM SIGCHI/SIGART conference on human-robot interaction, HRI '06*, Salt Lake City, UT, USA, 2–3 March 2006, pp. 186–193. New York, USA: ACM.
48. Ochs M, Pèlachaud C, and Sadek D. An empathic virtual dialog agent to improve human-machine interaction. In: *Proceedings of the 7th international joint conference on autonomous agents and multiagent systems, volume 1 of AAMAS '08*, Estoril: International Foundation for Autonomous Agents and Multiagent Systems, Estoril, Portugal, 12–16 May 2008, pp. 89–96. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.
49. McQuiggan SW, Rowe JP, and Lester JC. The effects of empathetic virtual characters on presence in narrative-centered learning environments. In: *Proceedings of the SIGCHI conference on human factors in computing systems, CHI '08*, Florence, Italy, 5–10 April 2008, pp. 1511–1520. Florence, New York: ACM.
50. Arellano D, Varona J, and Perales FJ. Generation and visualization of emotional states in virtual characters. *Comput Animat Virt W* 2008; 19(3–4): 259–270. Chichester, UK: John Wiley and Sons Ltd.
51. Berns K and Hirth J. Control of facial expressions of the humanoid robot head ROMAN. In: *International conference on intelligent robots and systems, 2006 IEEE/RSJ*, Beijing, China, 9–15 October 2006, pp. 3119–3124. IEEE.
52. Hegel F, Spexard T, Wrede B, et al. Playing a different imitation game: interaction with an empathic android robot. In: *6th IEEE-RAS international conference on humanoid robots*, December 2006, pp. 56–61.
53. Trovato G, Zecca M, Kishi T, et al. Generation of humanoid robot's facial expressions for context-aware communication. *Int J Hum Robot* 2013; 10(1): 1350013.
54. Mori M. Bukimi no tani [the uncanny valley]. *Energy* 1970; 7: 33–35.
55. Mori M, MacDorman KF, and Kageki N. The uncanny valley [from the field]. *Robot Autom Magaz IEEE* 2012; 19(2): 98–100.
56. Tinwell A, Grimshaw M, and Williams A. The uncanny wall. *Int J Arts Technol* 2011; 4(3): 326–341.
57. Tinwell A. *The uncanny valley in games and animation*. Boca Raton: AK Peters/CRC Press, 2014.
58. Mike B, Dautenhahn K, Appleby A, et al. The art of designing robot faces: dimensions for human-robot interaction. In: *Proceedings of the 1st ACM SIGCHI/SIGART conference on human-robot interaction, HRI '06*, Salt Lake City, UT, USA, 2–3 March 2006, pp. 331–332. New York, NY: ACM.
59. Breazeal C. Emotion and sociable humanoid robots. *Int J Hum Comput St* 2002; 59: 119–155.
60. Lazzeri N, Mazzei D, Greco A, et al. Can a humanoid face be expressive? A psychophysiological investigation. *Frontiers Bioeng Biotech* 2015; 3: 64.
61. Saldien J, Goris K, Vanderborght B, et al. Expressing emotions with the social robot probo. *Int J Soc Robot* 2010; 2(4): 377–319.
62. Bassili JN. Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face. *J Pers Soc Psychol* 1979; 37(11): 2049–2058.
63. Wallraven C, Breidt M, Cunningham DW, et al. Evaluating the perceptual realism of animated facial expressions. *ACM Trans Appl Perce* 2008; 4(4): 1–20.
64. Fiorentini C and Viviani P. Is there a dynamic advantage for facial expressions? *J Vis* 2011; 11(3): 1–15.

65. Faita C, Vanni F, Lorenzini C, et al. Perception of basic emotions from facial expressions of dynamic virtual avatars. In: De Paolis L and Mongelli A (eds) *Augmented and virtual reality. AVR 2015., volume 9254 of lecture notes in computer science*, Cham: Springer, 2015, pp. 409–419.

66. Dyck M, Winbeck M, Leiberg S, et al. Recognition profile of emotions in natural and virtual faces. *PLoS One* 2008; 3(11): e3628.

67. Mower E, Lee S, Matarić MJ, et al. Human perception of synthetic character emotions in the presence of conflicting and congruent vocal and facial expressions. In: *IEEE international conference on acoustics, speech, and signal processing*, Las Vegas, NV, USA, 31 March–4 April 2008, pp. 2201–2204. IEEE.

68. De Gelder B and Vroomen J. The perception of emotions by ear and by eye. *Cogn Emot* 2000; 14(3): 289–311.

69. Mazzei D, Lazzeri N, Hanson D, et al. Hefes: a hybrid engine for facial expressions synthesis to control human-like androids and avatars. In: *Proceedings of 4th IEEE RAS/EMBS international conference on biomedical robotics and biomechatronics*, BIOROB, 2012, pp. 195–200.

70. Ben Moussa M, Kasap Z, Magnenat-Thalmann N, et al. MPEG-4 FAP Animation Applied to Humanoid Robot Head. In: *Proceedings of the summer school engage*, Zermatt, Switzerland, 2010.

71. Hanson D. Exploring the aesthetic range for humanoid robots. In: *Proceedings of the ICCS CogSci symposium toward social mechanisms of android science*, Vancouver, Canada, 26 July 2006, pp. 16–20. Cognitive Science Society, Inc.

72. Hanson D and White V. Converging the capabilities of EAP artificial muscles and the requirements of bio-inspired robotics. In: *Proceedings of SPIE 5385, smart structures and materials: electroactive polymer actuators and devices (EAPAD)*, San Diego, CA, United States, 27 July 2004, vol. *5385*, 2004, pp. 29–40.

73. Bould E and Morris N. Role of motion signals in recognizing subtle facial expressions of emotion. *Brit J Psychol* 2008; 99: 167–189.

74. Igor SP and Forchheimer R. *MPEG-4 facial animation: the standard, implementation and applications*. New York: John Wiley & Sons, Inc, 2003.

75. Hai T, Chen HH, Wu W, et al. Compression of MPEG-4 facial animation parameters for transmission of talking heads. *IEEE Trans Circ Syst Video Technol* 1999; 9: 264–276.

76. Kshirsagar S, Garchery S, and Magnenat Thalmann N. Feature point based mesh deformation applied to mpeg-4 facial animation. In: *Proceedings of the IFIP TC5/WG5.10 DEFORM'2000 workshop and AVATARS'2000 workshop on deformable avatars* (eds Magnenat-Thalmann N and Thalmann D), *DEFORM '00/AVATARS '00*, 2001, pp. 24–34. Boston, MA: Springer.

77. Garchery S, Egges A, and Magnenat Thalmann N. Fast facial animation design for emotional virtual humans. In: *Proc. measuring behaviour* (eds Noldus LPJJ, Grieco F, Loijens LWS and Zimmerman PH), Wageningen, The Netherlands, 30 August–2 September 2005. Wageningen: Noldus Information Technology.

78. Ben Moussa M and Magnenat-Thalmann N. Toward socially responsible agents: integrating attachment and learning in emotional decision-making. *Computer animation and virtual worlds journal* 2013; 24(3–4): 327–334.

79. Cohen J. Quantitative methods in psychology: a power primer. *Psychol Bull* 1992; 112: 155–159.

80. Shayganfar M, Rich C, and Sidner CL. A design methodology for expressing emotion on robot faces. In: *IROS*, Vilamoura, Portugal, 7–12 October 2012, pp. 4577–4583. IEEE.

81. Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–174.

82. OriginLab Corporation. Originlab. http://www.originlab.com/ (2012, accessed 15 June 2018).

83. Leppänen JM and Hietanen JK. Positive facial expressions are recognized faster than negative facial expressions, but why? *Psychol Res* 2004; 69: 22–29.

84. Adolphs R. Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behav Cogn Neurosci Rev* 2002; 1(1): 21–62.

85. Scherer KR. What are emotions? And how can they be measured? *Soc Sci Inform* 2005; 44(4): 695–729.

86. Christensen HI. *A roadmap for U.S. robotics from internet to robotics*. Computing Community Consortium. http://www.roboticscaucus.org/Schedule/2013/20March2013/2013%20Robotics%20Roadmap-rs.pdf (2013, accessed 15 June 2018).

87. Terrence F, Nourbakhsh I, and Dautenhahn K. A survey of socially interactive robots. *Robot Auton Syst* 2003; 42(3–4): 143–166.