Int. J. Human-Computer Studies 78 (2015) 1–16

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Emotionally expressive dynamic physical behaviors in robots

Mason Bretan [a,*], Guy Hoffman [b], Gil Weinberg [a]

[a] Center for Music Technology, Georgia Institute of Technology, 840 McMillan St., Atlanta, GA 30332, USA
[b] Media Innovation Lab, Interdisciplinary Center Herzliya, 167 Herzliya 46150, Israel

ARTICLE INFO

Article history:
Received 2 April 2013
Received in revised form
22 January 2015
Accepted 23 January 2015
This paper was recommended for publication by E. Motta.
Available online 4 February 2015

Keywords:
Human robotic interaction
Emotion
Affective computing
Expression
Sentiment analysis

ABSTRACT

For social robots to respond to humans in an appropriate manner, they need to use apt affect displays, revealing underlying emotional intelligence. We present an artificial emotional intelligence system for robots, with both a generative and a perceptual aspect. On the generative side, we explore the expressive capabilities of an abstract, faceless, creature-like robot, with very few degrees of freedom, lacking both facial expressions and the complex humanoid design found often in emotionally expressive robots. We validate our system in a series of experiments: in one study, we find an advantage in classification for animated vs static affect expressions and advantages in valence and arousal estimation and personal preference ratings for both animated vs static and physical vs on-screen expressions. In a second experiment, we show that our parametrically generated expression variables correlate with the intended user affect perception. Combining the generative system with a perceptual component of natural language sentiment analysis, we show in a third experiment that our automatically generated affect responses cause participants to show signs of increased engagement and enjoyment compared with arbitrarily chosen comparable motion parameters.

1. Introduction

The ability to express emotion through nonverbal means can be an effective tool for computational and mechanical systems which interact with people. Coulson's component process view of emotions is defined as the "affective significance of a series of evaluations" (Cañamero and Aylett, 2008). This relationship between cognition and evaluation often results in some physical behavior such as a smile or scream. These physical behaviors are usually an unconscious reaction and can be considered representations of one's internal states.

There is a large body of evidence which supports facial expressions and prosodic cues as being indicative of a person's internal emotional state (Ekman, 1993; Fridlund et al., 1987; Schuller et al., 2006, 2011). However, the question of whether pose and body movements are reflections of internal emotional states has been subject to debate for many years. Some studies suggest emotional body language and physical expression are used primarily for social and communicative purposes rather than an unconscious expression of internal emotion (Fridlund, 1991; Kraut and Johnston, 1979). Though in more recent years evidence has been mounting which suggests the opposite is true. There is indeed a direct relationship between external physical behavior and emotional states (Inderbitzin et al., 2011; Walbott, 1998). In fact, Aviezer et al. (2012) declare that body language instead of facial expression better broadcasts what a person is experiencing especially in circumstances of extreme positively or negatively valenced emotions.

Other research also demonstrates that gesture is useful for conveying information other than affect and is a component of the speech planning process (Alibali et al., 2000; Kita et al., 2007). In essence, gesture aids speech generation beyond lexical retrieval by helping speakers to organize and conceptualize spatial information. Movement is also important for interactive scenarios. The timing of visual cues including movement illustrators and gaze plays an important role in the collaborative process of conversation (Bavelas et al., 2002; Bavelas and Chovil, 2006). In fact, Bavelas and Chovil (2000) describe an integrated model of communication which unifies the visible and audible components of face-to-face dialogue.

Though the debate concerning the true function of body pose and body movements remains, and it appears the use of gesture has many functions, it is largely accepted that, at some level, people are able to associate postures and movements with particular emotions (de Gelder, 2006; Nele Dael and Scherer, 2012; Coulson, 2004; Krauss et al., 1991; Kipp and Martin, 2009). There is even evidence of the brain processing emotional body language unconsciously and without reliance on the primary visual cortex (de Gelder and Hadjikhani, 2006). The human ability to recognize emotion through body language is an important trait and quite relevant to the field of affective computing. Additionally, the capacity for processing emotion through

* Corresponding author. Tel.: +1 415 246 6475.
E-mail address: masonbretan@gmail.com (M. Bretan).

http://dx.doi.org/10.1016/j.ijhcs.2015.01.006
1071-5819/

a non-conscious "affective channel" is an important attribute which strengthens the argument for emotionally expressive machines (Picard, 1995). This enables communication without increased cognitive function from the user. This is especially important in machine–human social interactions in which the machine should support the natural tendencies of human communication.

In this paper, we further explore the use of body language and physical motion of a robot as a means for expressing emotion. Expressive movements should not merely be considered a sequence of varied positions, but rather an additional form of communication and a behavior which can carry influence in a socially interactive environment (Frijda, 1987). Physical behaviors have components which are both representative and interactive, often modifying the relationship and emotions among multiple actors.

Robots that demonstrate a sensitivity to people's emotional states by responding with such expressive movements and behaviors enable more natural, engaging, and comfortable human–robotic interactions (Kozima and Yano, 2001; Breazeal and Aryananda, 2002; Castellano et al., 2010). An emotionally responsive and expressive robot can be incredibly effective in social situations such as in learning and teaching environments (Scheutz et al., 2006) by communicating pieces of information including levels of compassion, awareness, accuracy, and competency all through a non-conscious affective channel. For these reasons designing and developing emotionally expressive and perceptive robots has been a focus of many researchers in the field as described in Section 2.

We contribute to this research by examining a robot's ability to (1) express emotion using non-facial physical behaviors and (2) autonomously generate affective behaviors. We investigate several questions. What aspects of human physical emotional expression can be translated to robotics? Can limitations such as range of motion, the number of degrees of freedom (DoFs), velocity restrictions, and non-humanoid design be surmounted so that a robot can still be naturally and intuitively communicative through physical expression? Can emotionally perceptive robots offer increased levels of engagement in human robotic interactions by responding with expressive movements?

We address these questions by first reviewing a number of strategies and efforts to communicate emotion through a nonverbal channel in animation and robotics. Next, we introduce our robot, *Shimi*, and describe our own efforts for enabling Shimi to convey affective behavior through physical gestures and evaluate our efforts with a user study. We then describe a set of variables we believe to be essential to creating emotionally expressive motion and behavior. Using these variables, we introduce a novel computational architecture for algorithmically generating affective physical behaviors in Shimi which we evaluate with additional user studies. Finally, we evaluate the entire system with a final user study based on a human–robotic interaction involving communication.

2. Emotionally expressive systems

2.1. Utility of affective computing

Some debate in the HCI community exists regarding whether machines should attempt to detect and display emotion. Emotional displays can allow observers to interpret a person's beliefs and intentions. However, Muller (2004) argues that there is currently no method for discriminating between parts and wholes. For example, a computer user might demonstrate frustration for any number of reasons such as a faulty mouse or a non-user friendly software interface. An emotionally intelligent machine would detect the frustration, interpret it as resulting from the whole experience, and modify its own behavior accordingly. Should the machine modify its behavior without fully understanding the source of the frustration? Muller postulates that doing so might not be the optimal solution and ultimately could lead to even more distraction and frustration for the user.

Though differentiating between the possible sources and causes of emotions may be difficult, the amount of useful information within the emotional signal is so significant that the common view is that the benefits of emotional intelligence outweigh any potential mishaps. This is especially true in the context of sociable robots in which the machine is a partner and not merely a tool. In social settings, emotions can be used as a means of influence to elicit responses from other contributors and improve group efficiency by minimizing social conflicts (Frank, 1988; Campos et al., 2003; Frijda, 1987; Simon, 1967). Leveraging these benefits requires functionalities for both emotion detection and synthesis and evaluation is based on whether such emotional intelligence benefits an agent's reasoning process, improves performance, and creates human–computer interactions that are effective, productive, and enjoyable.

2.2. Virtual agents

The idea of turning nonliving objects into expressive beings with an abundance of personality has been a hallmark of computer animation for decades. Lasseter (1987) describes the use of pose, motion, and acceleration in animating the iconic Pixar lamp, *Luxo Jr*. Some of the most important techniques Lasseter presents are "staging" and "exaggeration" which together represent the artistry of presenting an idea so that it is unmistakably clear by developing its essence to extreme proportions. "If a character is sad, make him sadder; if he is bright, make him shine; worried, make him fret; wild, make him frantic."

Studies have shown that these ideas are appropriate for robots as well. Gielniak and Thomaz (2012) demonstrate that it is not human-like motions, but rather the exaggerated cartoon-like motions of a robot which are most effective for yielding the benefits of increased partner engagement and entertainment value. The interactive graphics-based agent, *Rea*, is a virtual real-estate agent who uses gestures, gaze, and facial expressions as an additional communicative and expressive layer to accompany her speech (Cassell et al., 2000). Nayak and Turk (2005) describe how emotional expression in virtual agents is achieved through a combination of facial expressions and body language including torso, shoulder, head, and leg movements.

2.3. Robots

Emotionally expressive animations have demonstrated effectiveness by making the agents more relatable and lifelike. This encourages the viewer to empathize with and respond to the virtual agent as if it were human. Using animation techniques is useful in robotics and often animating virtual agents is a first step to robotic motion generation (Salem et al., 2010). Applying the methods of animation to robotics is a challenging task because there is often less mobility, fewer degrees-of-freedom, and slower movement. However, the proximity and presence of a physical robot can have many benefits. In this section, we describe related work in robotic communication through gaze, gesture, and proxemics.

2.3.1. Gaze and facial expression

Gaze and facial expression play an important role in the communicative and interaction abilities of social robotics. Even simple glances and postural shifts can encourage the flow of dialogue (Sidner et al., 2004; Cassell et al., 2001). It has also been shown that people accept robots as proactive communicative agents and respond to a robot's gaze and nods in the same manner as they respond to other humans (Muhl and Nagai, 2007; Sidner et al., 2006). The role of gaze and facial

expression plays an important role in social robotics as a method for communicating emotion as well as eliciting emotion in people. Sometimes emotional expressions are not explicitly designed into the robot, but are rather associated with the robot as a result of its behavior. In Mutlu et al. (2009) a *Robovie R-2* robot adjusted its gaze to focus on specific individuals, which elicited emotions of dislike in people because they believed the robot demonstrated a preference or was ignoring them.

In socially guided learning and shared focus experiences, robots utilize posture and motion to intentionally indicate confusion and attention or to acknowledge understanding. In these scenarios, the robot must explicitly display emotion with the intent of eliciting a response from the user. One example of this is *Kismet* (Breazeal, 2003), which uses facial expression to express emotion to users. Delaunay and Belpaeme (2012) describe a retro-projected "LightHead" robot and argue a projection method allows for more versatility of emotional display because it is not limited to a small number of facial features. Other systems such as *Infanoid* (a humanoid social robot Kozima et al., 2004) and *Leonardo* (a creature-like robot Lockerd and Breazeal, 2005) exhibit emotion through arm and hand gestures, torso and neck movements, in addition to varying facial expressions.

2.3.2. Gesture and faceless robots

Gesture is an important facet of human communication and studies have shown that characteristics such as velocity, acceleration, and location (front versus side-oriented) can influence how people respond to different movement in robots (Riek et al., 2010; Moon et al., 2013). Often gesture accompanies speech to help the speaker to formulate ideas and provide additional information to the observer. Salem et al. (2012) developed an integrated model of speech-gesture production to address this concurrence. In the experiments detailed later in this paper, however, we focus on solely physical gesture without additional auditory cues.

There are several examples of robots which use gesture to explicitly convey posture, motion, and gaze to convey and elicit emotion. Many of these robots are faceless (or facially expressionless) and must rely only on the physical movements of its body and head to achieve this. *RoCo* is a robotic computer designed to move in a subtly expressive manner and influence users' affective states by altering its posture (Breazeal et al., 2007). *AUR* is a robotic lamp which utilizes both human-controlled and autonomous modules to generate expressive gestures and behaviors (Hoffman and Breazeal, 2008).

Keepon, though not faceless (it has eyes and a nose), does not have any motors in its face for it to be facially expressive. Keepon is a creature-like robot designed to facilitate emotional expression through its gaze and movements. It can convey emotions such as pleasure and dislike through bobbing (up and down and back and forth) and vibrating motions (Michalowski et al., 2007).

The *NAO* robot is a humanoid robot which also has a face, but no facial motors. It adjusts its arms, legs, head, and torso to arrange itself in emotionally semantic poses (Monceaux et al., 2009). Similarly, both the *DARwIN* and *Hubo* robots have been programmed to utilize head and arm positions to express mood while dancing to music (Grunberg et al., 2012).

Shimon is a non-humanoid, marimba playing, robotic musician. It has a head with a single eye and elongated neck. Often the expressive nature of musical robots is generated through the use of sound and various modes of musical interaction (Weinberg et al., 2009, 2006). However, physical motion and behavior can additionally be leveraged as a tool for creating an even more expressive robot and engaging interactive experience. Shimon uses posture, motion, and gaze to demonstrate emotions and internal states such as attention and awareness, recognition of the beat and tempo in music, and interest and curiosity (Bretan et al., 2012).

2.3.3. Presence and proximity

Presence and proximity also influence the way people perceive emotion. A physical presence (as opposed to virtual) can lead to more "altruistic and persuasive" perceptions of the robot (Kidd, 2003). Having mechanisms in place which control the proximity of the robot to the user is important because spatial features are used to support certain social behaviors and can be used to trigger levels of comfort and discomfort (Walters et al., 2009; Takayama and Pantofaru, 2009; Mead et al., 2011). However, the robotic platform we use in our experiments is stationary so we can evaluate presence, but cannot test the effectiveness of a dynamically changing distance between collaborators and how this may affect emotion perception and recognition.

2.4. Architectures for algorithmically generating emotions

Before describing how emotion is synthesized it is important to understand the relevant methods of classification. Some methods utilize a set of basic discrete emotions (happiness, sadness, fear, etc.) with more complex emotions being a combination of two or more of these fundamentals (Devillers et al., 2005). Other methods use a continuous dimensional model to classify emotion. These dimensions include valence, arousal, and less commonly, dominance, and stance (Mehrabian, 1996; Russell, 2009).

There is much debate regarding the fundamental issues of emotion in humans. One argument against a theory of basic and discrete emotions is that one would assume people exhibit similar behaviors for individual emotions. Many studies involving facial expressions have demonstrated that this is not the case (Carroll and Russell, 1997; Fernández-Dols and Ruiz-Belda, 1997; Jack et al., 2012). However, recently it has been shown that facial cues are not good indicators for discriminating emotion and, instead, body cues should be used (Aviezer et al., 2012).

Other arguments fueling the debate detail brain activity. There is much evidence showing that discrete brain regions cannot be mapped to discrete emotion categories (Lindquist et al., 2012). Though other recent studies have shown basic emotion views are represented in neural correlates. The mappings are not one-to-one, however, but rather complex distributed networks (Vytal and Hamann, 2010; Hamann, 2012). A newer theory describes dynamical discrete emotions and attempts to address the variability and context-sensitivity of emotions (Colombetti, 2009). This approach has been deemed viable by researchers on both ends of the discrete versus continuous emotion debate (Barrett et al., 2009).

Despite the fact that there is not yet universal agreement on emotion classification and perception, there are specific attributes an emotional intelligence architecture should entail. Picard (1995) describes the constituents of expressive machines as having both instinctual (spontaneous) and communicative (intentional) pathways. Ideally, an affect expression system for a robot should be flexible enough to account for both of these pathways.

A well known system for an emotionally expressive robot is Kismet's implementation described in Breazeal (2003). Breazeal describes a system which enables facial expression transformation on a continuous *valence*, *arousal*, and *stance* scale. Velásquez (1997) describes the system, *Cathexis*, which is based on the six primary emotions (*anger, fear, disgust, sadness, happiness,* and *surprise*). An implementation of expressive dance motions in the NAO robot is described in Xia et al. (2012) which uses automated scheduling of discrete motion primitives driven by beats and emotion in music. These systems connect the agent's motivations to its affective behavior based on the fact that emotions arise out

of one's internal goal (Frijda, 1995; Rolls, 2005). The motivations, or "drives" as they are often referred, may result from a number of factors including the agent's need for energy, survival, communication, or even dancing. The agent's desire to satiate these goals influences its behaviors and thus emotions are born.

A more recent system designed for human–machine companions is presented by Traue et al. (2013). The system integrates the use of both discrete and continuous emotions. External events are subjected to an appraisal process and interpreted as discrete emotional stimuli. These discrete stimuli are mapped to a continuous space and averaged. The resulting value is used to manipulate the agent's behavior in response to the stimuli.

3. Experiments with emotional intelligence

The following sections describe experiments which evaluate our efforts for developing emotional intelligence in our robot, Shimi. The elements of emotional intelligence we investigate are expression and perception. The first two experiments focus primarily on Shimi's ability to express specific emotions using both preprogrammed physical gestures and algorithmically generated gestures. The preprogrammed gestures were designed to test whether it was possible for Shimi to convey discrete emotions when given what we found to be the optimal poses and motions representing each affect. Then, using what we learned from the hand designed gestures, we developed a system for autonomous generation of emotionally expressive movements.

In our second experiment, we measure the correlations of the parameters controlling this algorithm with particular emotions. Finally, we integrate our expression module with a perception module in order to a evaluate the system with a user experiment involving an interaction with Shimi. We attempt to demonstrate the effectiveness of the aforementioned physical behaviors and perception abilities through increased partner engagement and improved user assessment of Shimi.

4. Design and evaluation of emotionally expressive gestures for shimi

Shimi (see Fig. 1) is a facially expressionless, creature-like robot that was designed to be an interactive speaker dock (Hoffman, 2012).[1] It has two speakers on either side of its head, and is powered by a mobile phone running Android OS. Shimi has five DoFs with three in its head, one on the hand that holds the phone, and one on its foot for tapping. The robot itself has no sensors. Instead, the mobile phone's sensors (microphone and camera) are used, and an application running on the Android device sends movement commands to an arduino which controls each DoF. Shimi utilizes the phone's functionalities to become "intelligent" allowing it to dance to music, understand speech, and follow a person so that he or she is always in the stereo field. For the following experiments, we use the Shimi robot as a platform for creating and exploring physical expression of emotion without musical accompaniment.

There are several similarities among Shimi, Keepon, Hubo, and DARwIN (specifically Grunberg's et al. work with humanoid dancing robots), most notably the fact that they are robots which respond to music with physical gestures. Unlike the Hubo and DARwIN robots, the Shimi platform allows us to experiment with a non-humanoid design and focus on physical movement and behavior rather than human posture. The work involving motion and emotive expression with Keepon has provided an excellent starting point into examining how a simple mechanical system can be expressive with movement. Here, we expand on the ideas introduced through Keepon and attempt to achieve emotional expression of more complex and discrete emotions through Shimi. Additionally, unlike Keepon, Shimi's motions involving gaze are more ambiguous because it does not have a face.

[1] Shimi is referred to as "Travis" in this paper.

Fig. 1. Shimi is a faceless, five DoF, smart phone powered, robotic speaker dock which uses musical intelligence to engage users and dance to music. Shimi has five DoFs: three in its head, one on the hand that holds the phone, and one on its foot for tapping.

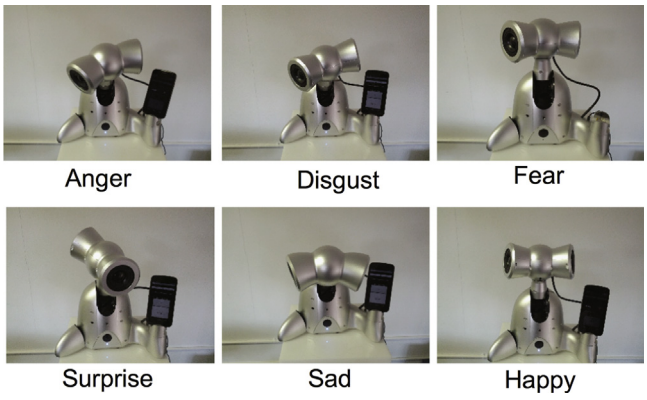

Fig. 2. Static poses representing six basic emotions.

4.1. Expressing emotion

Static postures and dynamic gestures were designed to represent each of the six fundamental emotions: happiness, sadness, anger, fear, surprise, and disgust (see Fig. 2). We chose these emotions, not because we want a discrete emotional system, but rather these emotions can easily be mapped to coordinates on the core affect plane. This also allows us to use more commonly understood emotion terms (opposed to valence, arousal, and stance) when working with participants in a user study. Similarly to the labeling study performed by Breazeal (2003) for robotic facial expressions, we use continuous parameters to algorithmically generate emotional behaviors but use a discrete labeling system. The premise is that if we can make robust representations of specific coordinates on the continuous core affect plane, we can then build a model that allows for smooth and continuous regression from coordinate to coordinate. Ekman's six emotions were chosen because of their straightforward and comprehensible

nature as well as their coordinates being ideally located on the core affect plane for regression.

The postures and gestures were designed to be the quintessential portrayals of each emotion and exaggerated to exemplify the essence of each emotion. Both the postures and gestures were hand-designed using an iterative process with informal user interviews for refinement. We based our dynamic gestures on the findings of Inderbitzin et al. (2011) such that the robot's posture is strongly correlated to valence and the velocities of the movements are correlated with arousal levels. The process entailed designing a set of emotional gestures, performing a small labeling pilot study for evaluation, collecting feedback, and repeating the process until a stable ground truth of gestures had been established. The labeling pilot ensured that there was agreement with regards to the emotions Shimi was expressing with each gesture and pose. This ensured that Shimi was expressing the emotions we thought In the end, each gesture has a duration of 1–4 s.

Though culture can definitely influence how a person interprets emotive facial expressions (Jack et al., 2012), speech accompanying gestures (Kita, 2009), and, to a lesser extent, stand-alone emotive body postures (Kleinsmith et al., 2006), we did not control for this in our design process. However, considering the diversity of those involved in the pilot studies and design process we can safely assume that these emotional gestures cross cultural boundary lines to at least some extent.

4.2. Experiment

We designed an experiment to evaluate the emotionally expressive nature of the postures and dynamic gestures. We make several hypotheses based on our own and others' previous work. Based on the agreement found in our labeling pilot study, we believe that the participants will be able to classify Shimi's expressions with the appropriate emotion.

H1 (Overall Classification)—Participants will perform better than chance for identifying the emotion each posture and gesture represents

Though static facial expressions are often sufficient for conveying emotion, a faceless robotic platform must rely on physical movements and behaviors to effectively convey emotion. Inderbitzin et al. (2011) show that both posture and movement are equally important for representing valence and arousal values. Therefore, we believe there will be perceptual differences between pose and motion that are relevant to emotional expression and make these hypotheses:

H2 (Movement Bias, Classification)—Participants will perform better at classifying the dynamic gestures compared to the postures.
H3 (Movement Bias, Valence)—Participants will more often label the dynamic gesture with the appropriate valence of the emotion portrayed compared to the postures.
H4 (Movement Bias, Arousal)—Participants will more often label the dynamic gesture with the appropriate arousal of the emotion portrayed compared to the postures.
H5 (Movement Bias, Approval)—Participants will rate the dynamic gestures more positively than the postures at portraying the emotions.

As described earlier, proximity and presence can influence a person's perception of a robot. We are able to evaluate the affect of presence as well in these studies. Specifically, we compare the effect of Shimi's different affect expression techniques on human perception of emotion with four conditions: viewing of posture images (PI), viewing of postures in person (PP), viewing of dynamic gesture videos (DV), and viewing of dynamic gestures in person (DP). In addition to investigating the perceptual differences between pose and motion, the experiment was designed to evaluate an additional element relevant to HRI: peoples' responses between viewing a collocated (physically embodied) robot and viewing a video or image of a robot.

Evidence for increased social influence and engagement during the collocated scenario has been demonstrated in children (Kose-Bagci et al., 2009) and adults (Kidd and Breazeal, 2004; Bainbridge et al., 2008), yet another study by Powers et al. (2007) showed that a collocated robot does not always yield better results. These studies involved interactions between the participants and the robot (either virtually or in person) to examine the benefits and social influences of each scenario. In this experiment, participants only view the robot and do not interact with or socially engage it. This allows us to test purely the expressive nature of the robot under the different conditions. We make two more hypotheses based on the findings of these previous studies.

H6 (Co-presence Bias, Posture Classification)—Participants will rate the postures more accurately when they witness them in person rather than as an image.
H7 (Co-presence Bias, Dynamic Gesture Classification)—Participants will rate the dynamic gestures more accurately when they witness them in person rather than as a video

4.2.1. Procedure

The experiment is a 4×1 between subjects design with each participant randomly assigned to one of four testing conditions: (1) viewing of images of the robot's poses (2) viewing of the robot's poses in person (3) viewing of HD videos of the robot's dynamic gestures and (4) viewing of the robot's dynamic gestures in person. Each participant sat at a table either with a computer screen or with Shimi where the proctor introduced the robot and described the purpose of the study. They were told that they would be shown six static postures or animated gestures (in random order) and be asked to classify each as one of the six fundamental emotions. There was no time limit and they were able to view each gestures as many times as they wished. In addition to classification, they were asked to label the postures and animations with a valence on a 7-point discrete visual analog scale (DVAS) where 7 was highly positive, 1 was highly negative, and 4 was neutral. They did this for arousal as well. Finally, they were asked to rate the posture/gesture on a 7-point Likert scale (strongly disagree to strongly agree) describing how well they thought each represented the emotion we had intended for it to portray. The mean of three questions were used for the subjective feelings Likert analysis (Cronbach's alpha=.83). The questions were “This gesture represents [this particular emotion]”, “Shimi shows characteristics of [this particular emotion] in this gesture”, and “It is easy to understand the emotion Shimi is conveying.”

4.2.2. Results

There were 48 undergraduate and graduate Georgia Tech students who participated in the study (30 male and of American, Indian, and Chinese origin). A one-way between subjects analysis of variance (ANOVA) was conducted to compare the effect of different affect expression techniques on affect classification accuracy on the four testing conditions: PI, PP, DV, and DP. There was a significant effect of affective expression techniques on the classification accuracy at the $p < .001$ level for the four conditions [$F(3, 44)=8.36$, $p < .001$]. Post hoc comparisons using the Tukey HSD test indicated that the mean score for the PI (M=.45, SD=.25)

condition was significantly different than the DV (M=.68, SD=.18) and DP (M=.81, SD=.21) conditions. The PP (M=.47, SD=.13) condition was significantly different than the DV and DP conditions. However, the PI condition did not significantly differ from the PP condition and the DV condition did not significantly differ from the DP condition.

These results suggest that our dynamic gestures are better at conveying the six fundamental emotions than the static postures. Specifically, viewing Shimi's dynamic gestures either in person or on video produces better perceptual representations than posture in general does. Contradictory to our hypotheses (H6 and H7), however, the differences between on screen and in person viewings are insignificant. Fig. 3 shows the summarized ANOVA results for the identification accuracies. As shown in Table 1, participants were on average able to accurately identify the emotion of a posture image 46% of the time, 47% of the time for viewings of postures in person, 66% of the time for videos of the gestures, and 80% of the time for in person viewings of the gestures. Table 2 shows the forced choice percentage rates for each emotion. A Pearson's Chi-squared goodness of fit test was performed for each emotion. The results indicate that the frequency distributions for each emotion differ significantly at the $p < .001$ level to a theoretical distribution where all six emotions are considered equally likely to occur. All emotions performed better than the likelihood of chance (16.7%) other than the identification of the *fear* posture which performed at chance. However, the Chi-square results suggest that, though the identification of the fear posture was equal to chance, the overall distribution for the fear posture still significantly differed from a random distribution indicating the posture did influence responses in some manner.

A one-way ANOVA was also conducted for each emotion for the subjective Likert ratings for the effective portrayal for each expression technique. Again, when the results suggested significance at the $p < .05$ level for the four conditions, a Tukey HSD post hoc comparison was done to compare each of the conditions to one another. There were significant findings for the emotions of Disgust, Anger, Fear, and Surprise. We also report the mean DVAS values of the valence and arousal ratings. Table 1 shows the average results for the DVAS values and the Likert ratings and whether there are any statistical differences between two or more of the conditions. Participants also tended to label the animated gestures with valence and arousal values closer to the true values for each emotion. Negative emotions (sadness, anger, fear, and disgust) have a valence less than four, neutral emotions (surprise) have a value close to four, and positive emotions (happiness) have a valence value greater than four. Emotions with positive arousals (fear, surprise, disgust, anger, and happiness) have ratings greater than four and emotions with negative arousals (sadness) less than four.

Fig. 4 shows the ANOVA for participants' ratings of how each posture and dynamic gesture represented the particular emotion. For these effective portrayal results there were significant differences between conditions were found for the emotions of fear, surprise, anger, and disgust. Though the DP condition was rated significantly higher than both the PI and PP conditions for all of these emotions, the DV condition was rated significantly higher

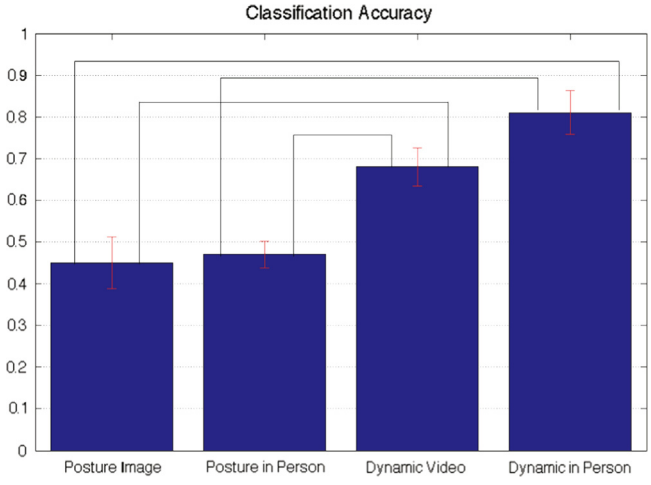

Fig. 3. Mean overall classification accuracies with red standard error bars and black lines indicating significant differences at the $p < .05$ level. The *y*-axis is the accuracy value for each of the different viewing conditions on the *x*-axis including viewing of postures as images (PI), postures in person (PP), dynamic gestures as videos (DV), and dynamic gestures in person (DP). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

Table 1
Average Results. The table shows where significant findings were found from the ANOVA results for the classification accuracy and Likert opinion ratings. We also report the average DVAS values of the arousal and valence ratings for the different emotions.

Task	Emotion	PI	PP	DV	DP	*P*-Value	η^2
Classification		.46	.47	.68	.81	***	.69
Opinion	Happy	4.92	5.42	5.33	5.83		
	Sad	6.59	6.25	6.50	6.67		
	Disgust	4.25	3.42	4.75	5.50	***	.58
	Anger	3.75	3.92	5.25	6.00	***	.66
	Fear	2.83	2.75	3.58	4.42	**	.36
	Surprise	4.59	4.33	5.00	5.58	**	.32
Valence	Happy	5.67	5.25	5.67	6.3	–	–
	Sad	1.67	1.67	2.00	1.75	–	–
	Disgust	3.42	4.00	3.33	3.00	–	–
	Anger	2.92	3.08	2.08	1.83	–	–
	Fear	4.56	4.25	3.25	3.42	–	–
	Surprise	4.92	5.17	4.00	3.75	–	–
Arousal	Happy	5.67	4.75	5.25	5.75	–	–
	Sad	3.33	3.08	1.92	1.58	–	–
	Disgust	3.75	3.83	5.08	4.75	–	–
	Anger	3.67	4.00	6.17	6.33	–	–
	Fear	3.42	4.25	5.41	5.00	–	–
	Surprise	4.67	5.00	5.33	5.67	–	–

Table 2
Identification confusion matrices.

	H	Sa	D	A	F	Su	%Correct
(a) Viewing of posture images							
Happy	41.7				25	33.3	41.7
Sad		91.7			8.3		91.7
Disgust		8.3	41.7	25	25	8.3	41.7
Anger	16.7	33.3		50			50
Fear	41.7	8.3		8.3	16.7	25	16.7
Surprise			16.7		16.7	41.7	41.7
(b) Viewing of postures in person							
Happy	41.7			8.3	41.7	8.3	41.7
Sad		100					100
Disgust			58.3		16.7	25	58.3
Anger	33.3	8.3		33.3	25		33.3
Fear	25		16.7	8.3	16.7	33.3	16.7
Surprise			33.3	33.3		33.3	33.3
(c) Viewing of dynamic gesture videos							
Happy	91.7				8.3		91.7
Sad		100					100
Disgust			58.3	8.3	16.7	16.7	58.3
Anger				75	25		75
Fear	8.3		16.7	8.3	25	41.7	25
Surprise			25	8.3	25	41.7	41.7
(d) Viewing of dynamic gestures in person							
Happy	100						100
Sad		100					100
Disgust			75		8.3	16.7	75
Anger				91.7	8.3		91.7
Fear			16.7		50	33.3	50
Surprise					41.7	58.3	58.3

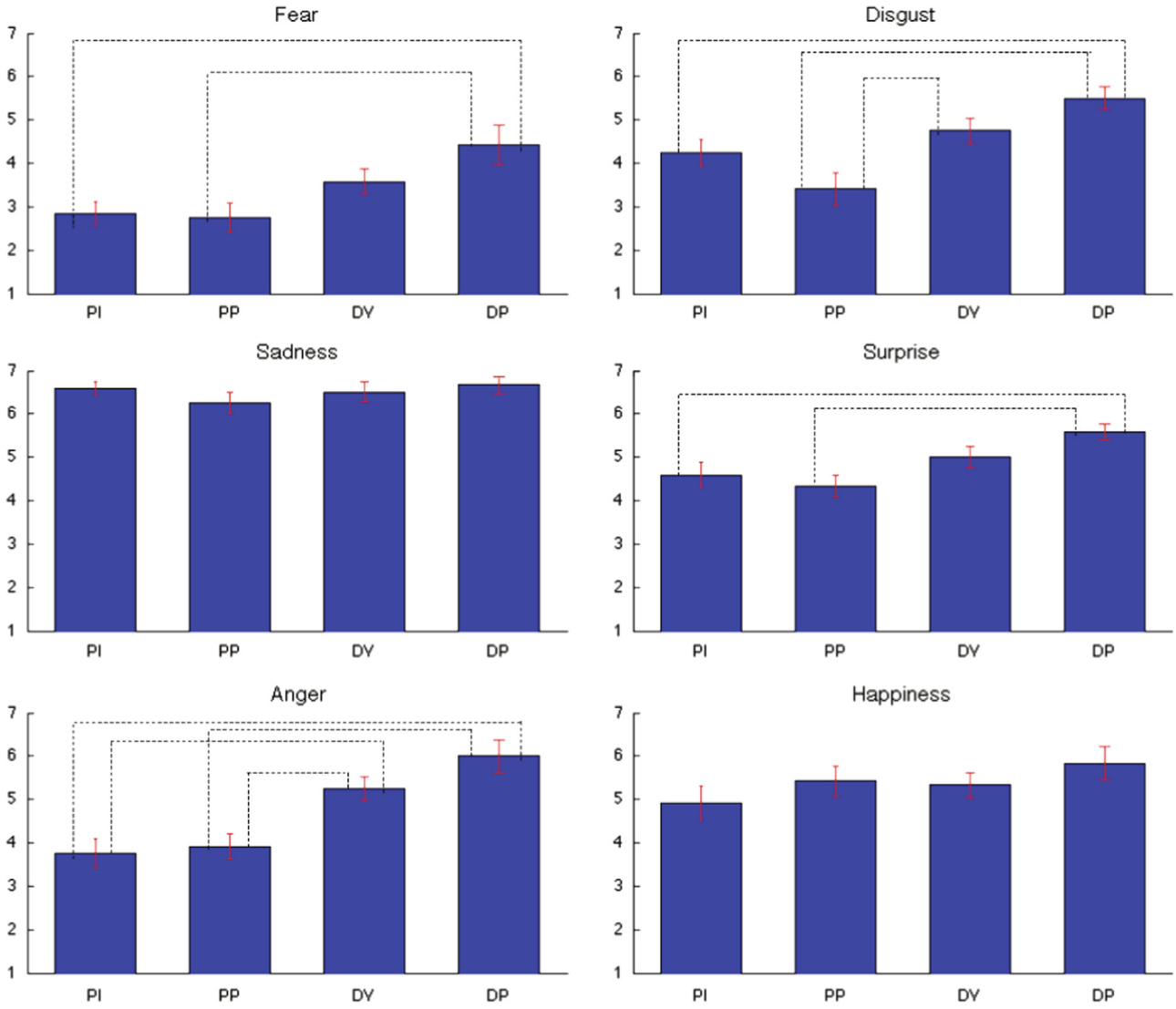

Fig. 4. Mean overall Likert rating results of how well each posture or dynamic gesture represented the particular emotion for each viewing condition with red standard error bars and black lines indicating significant differences at the $p < .05$ level. ANOVA results for participants' ratings of how well each static posture or dynamic gesture represented the particular emotion on a 7-Point Likert scale where 1 is very poor and 7 is very well. The *y*-axis is the 7-Point Likert rating where 1 is very poorly and 7 is very well. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

than the PI and PP conditions for anger and significantly higher than only the PP condition for disgust. This suggests that, though the classification accuracies were not influenced by the differences between on screen and in person viewings, the subjective ratings were influenced by both the differences between static posture and dynamic gesture and between on screen and in person viewings (for at least some of the emotions).

4.2.3. *Discussion*

The findings suggest Shimi was able to convey emotion through both its poses and gestures, though the gestures performed much better. All of our hypotheses were supported by the results other than H6 and H7. There was significant confusion with fear, happiness, and surprise in the posture identification. This is not surprising because the human posture of fear, surprise, and happiness differ most significantly in arm position (De Silva and Bianchi-Berthouze, 2004). Shimi has no arms to imitate these differences and can rely on only the subtle head and neck differences.

Though there were no statistically significant differences between the viewings of Shimi's postures as images and in person, there were differences in Likert ratings between the video and in person viewings of Shimi's gestures. For almost all emotions the in person viewing results had higher averages for identification accuracy, more similar DVAS ratings for valence and arousal to the true values for each emotion, and higher Likert ratings for how well each emotion was represented by the gesture. Additionally, differences did exist between the classification accuracies for the DV and DP conditions. Though the differences were small and not statistically convincing (p-value$=.08$ for overall identification accuracy) the tendencies suggest that, with a larger subject pool, more statistically significant findings may have been found. A post hoc power analysis revealed that on the basis of the mean, between-groups comparison effect size observed between the DP and DV conditions ($d=.67$, $N=24$), an n of approximately 95 would be needed to obtain statistical power of $p < .05$ at the .90 level. We found no previous studies which solely evaluate the ability of a robot to convey semantic information under collocated and virtual conditions. However, according to studies which evaluate perceived quality of speech and video in multimedia conferencing applications there is no perceptible loss of information (even when the signals are degraded by current video or audio conferencing technology) (Watson and Sasse, 1998; Anderson et al., 2000; Claypool and Tanner, 1999).

However, these studies do not evaluate the perception of individual's affective states when communicating through multimedia conferencing technology. Additionally, Coulson (2004) reports that viewpoint can influence the perception of emotion in static pose. This may be a possible reason for the differences between perception of video and in person viewings. However, there was no statistical difference between perception of emotion between in person viewings of the static poses and the images. We believe this is an avenue of research worth more investigation.

Fear had the lowest identification accuracies for both the postures and animations. This is indeed partly due to the fact that humans can use shoulder and arm position cues to determine whether somebody is afraid or not. However, fear also comes in different forms, and we did not specify what type of fear Shimi was attempting to convey. For example, we believe our animation much better represents a startled type of fear which is triggered by some sudden surprise or disbelief. It does not represent the type of psychological fear experienced when anxiously anticipating some impending event. Both of these factors may have contributed to the lower classification accuracies of fear.

Overall, the identification accuracies suggest that there are certain characteristics of motion and pose which people associate with particular emotions. This is in contrast to the *quantity but not quality* view which conforms to the idea that body language is merely a reflection of the intensity of the emotion being experienced and not the actual emotion. The findings support our hypothesis that motion can be used to successfully convey emotion when facial expression, sound, numerous DoFs, and humanoid design are not available.

5. A system for generating emotional behaviors

The previous section demonstrated that it is possible for Shimi to accurately convey both a discrete set of emotions and levels of valence and arousal. In this section, we attempt to establish a finite set of parameters and mathematical functions defining pose and motion, which when fitted with the appropriate numerical values, will create movements that correlate with specific emotions. Hardcoding short gestures as was done in the first experiment can be very useful in determining how best to apply the traits of pose and motion to represent each emotion. However, this is time consuming and limits the robot to portraying only a finite number of emotions. Additionally, people can experience and express different levels of any particular emotion. For example, the list of words to describe varying levels of happiness (joyous, content, exuberant, satisfied, ecstatic, pleased, jubilant, overjoyed, etc.) is immense yet each word is unique. If a robot is to be truly expressive and exhibit a rich set of emotions it must have a system which allows for it. Breazeal (2003) describes transitions between facial expressions on a continuous scale which allows Kismet to express a wide range of emotion with varying degrees of intensities. We were influenced by this system and have expanded on its methods to create a system for expressive emotional behaviors (using posture and motion) on a continuous scale, while still flexible enough to generate the discrete emotive gestures designed in the previous section.

5.1. Control variables

As described earlier, non-verbal behavior such as gaze and proxemics can be used to convey information or influence the behavior of collaborators (Hüttenrauch et al., 2006; Kendon, 1990). Architectures have been developed for synthetically generating these communicative movements (Salem et al., 2012). In this section, we describe our method for algorithmically generating emotive behaviors.

In order to computationally define what it means for something to behave in a manner representative of a particular emotion, we must first understand the variables which constitute an emotional behavior. Head orientation, position, and body postures have been observed as having an effect on facial expressions presented during feelings of specific emotions (Hess et al., 2007; Aviezer et al., 2008; Krumhuber et al., 2007). We must design a system which can sufficiently convey varied levels of emotion using only the DoFs available to Shimi. Research in facial expression and limb position is not as relevant for us. Instead, we must consider how something moves such smooth versus jerky motions.

Fig. 5 presents a summary provided by Walbott (1998) of observations from Darwin (1916). The relationships between posture and the movements inherent to specific emotions are described. Though these descriptions may no longer be considered the quintessential physical human behaviors exhibited for each emotion (as evidence shows behaviors differ across cultures and even genders), we can at least examine the descriptions to determine which characteristics (such as velocity, head position, etc.) can be used to control motion. A model can then be built to generate different types of behavior based on mappings of these particular characteristics to different emotions or coordinates on the core affect plane. We generalize Darwin's observations as a set of relevant features for differentiating the behaviors.

1. *Posture Height*—representative of the relative distance between the height of a person's chest and height of the waist. In essence, how erect or crouched a person's torso is.
2. *Shoulder Height*—representative of the relative distance between the waist and the shoulders.
3. *Arm Position*—the location of the arms in respect to the rest of the body (“close to sides” vs “over the head”).
4. *Gaze (Head Position)*—the direction and angle at which the head is positioned is indicative of where somebody's attention is and how welcoming or rejecting somebody is.
5. *Body Activation*—the type of motion the body and arms exhibit (slow vs fast) including
   (a) *Up and Down Activation*
   (b) *Left and Right Activation*
   (c) *Rotational Activation*—twisting of the body and arms
6. *Head Activation*—the type of motion the head exhibits is classified into 2 parts
   (a) *Positive Head Activation*—head nodding up and down

Table 1. Body movements and postures accompanying specific emotions (citations from Darwin, 1872/1965)

Joy	Various purposeless movements, jumping, dancing for joy, clapping of hands, stamping, while laughing head nods to and fro, during excessive laughter whole body is thrown backwards and shakes or almost convulsed, body held erect and head upright (pp. 76, 196, 197, 200, 206, 210, 214)
Sadness	Motionless, passive, head hangs on contracted chest (p. 176)
Pride	Head and body held erect (p. 263)
Shame	Turning away the whole body, more especially the face, avert, bend down, awkward, nervous movements (pp. 320, 328, 329)
Fear/terror/ horror	Head sinks between shoulders, motionless or crouches down (pp. 280, 290) convulsive movements, hand alternately clenched and opened with twitching movement, arms thrown wildly over the head, whole body often turned away or shrinks, arms violently protruded as if to push away, raising both shoulders with the bent arms pressed closely against sides or chest (pp. 291, 305)
Anger/rage	Whole body trembles, intend to push or strike violently away, inanimate objects struck or dashed to the ground, gestures become purposeless or frantic, pacing up and down, shaking fist, head erect, chest well expanded, feet planted firmly on the ground, one or both elbows squared or arms rigidly suspended by the sides, fists are clenched, shoulders squared (pp. 74, 239, 243, 245, 271, 361)
Disgust	Gestures as if to push away or to guard oneself, spitting, arms pressed close to the sides, shoulders raised as when horror is experienced (pp. 257, 260)
Contempt	Turning away of the whole body, snapping one's fingers (pp. 254, 255, 256)

Fig. 5. A table presented by Walbott (1998) describing observations from Darwin (1916).

(b) *Negative Head Activation*—head shaking left and right
7. *Volatility or Periodicity*—a function of variation in the movements in terms of the positions a movement oscillates between and the rate at which this is done. A highly periodic movement would indicate a motion with a low volatility. (This idea is supported by Linda A. Camras and Michel (1993) who show anger to be accompanied by more spasmodic movements compared to sadness)
8. *Exaggeration*—the range of motion exhibited by each DoF (smaller vibrations vs massive convulsions)

Though these parameters were derived from Darwin's observations of humans, we believe that all, or a subset of these parameters, can be useful for designing expressive movement in non-humanoid robots as well. On a robot the *posture, shoulder, arm,* and *gaze* parameters can be considered positional values for the DoFs which correspond to the relevant body parts. The *activation* parameters describe how the DoF positions change in time in relation to their "center values" described by these positional values. *Volatility* and *exaggeration* are modifying parameters which can be used to manipulate both the positional and activation values in time. In the following sections, we describe the system and control parameters in more detail. For Shimi, we use only the parameters relevant to its design and DoFs (see Fig. 8) which include posture height, gaze, head activation, volatility, and exaggeration.

5.2. Design

Our system for generating emotional behaviors consists of three main sections: Emotion Tagging, Interpretation, and Expression.

1. *Emotion Tagging* The system takes in some sensor data (such as sound or language) and represents it as a particular emotion. This may be valence and arousal values or an emotional word such as "anger" or "happy".
2. *Interpretation* During the interpretation phase, the system maps the emotional tags to each of the control parameters.
3. *Expression* Emotive behaviors based on the control parameters are generated using motion and body language.

An outline of this architecture is shown in Fig. 6. We begin by describing the expression section of the system and in subsequent sections discuss the tagging and interpretation phases.

The expression portion of the system is made up of two essential parts: a growth and decay function and motion primitives. When the expression system receives a set of control parameters it attempts to

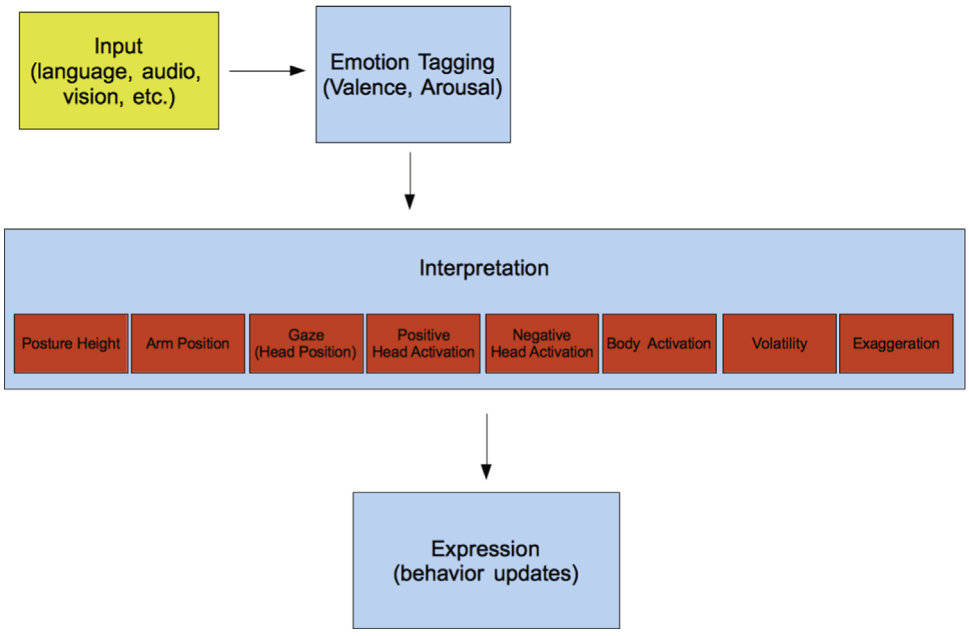

Fig. 6. . The system for generating emotional behaviors consists of three main phases: emotion tagging or identification, interpretation of how best to characterize the emotion using the predefined parameters, and synthesis for effectively expressing the emotion.

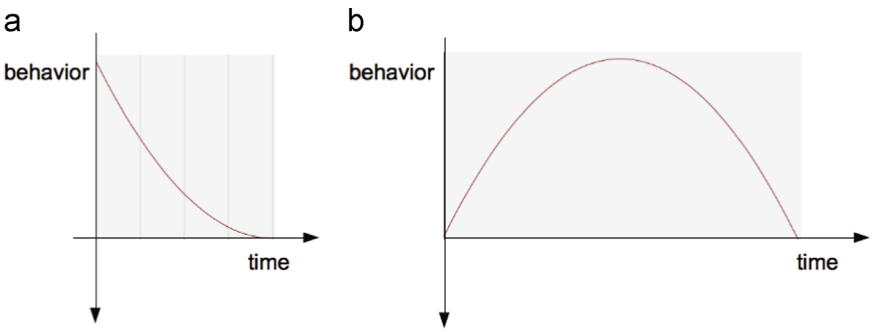

Fig. 7. Example decay rates where $y=0$ is the homeostatic state and $y=1$ is the new emotional state. (a) shows an immediate transition and exponential decay (b) shows a growth and decay.

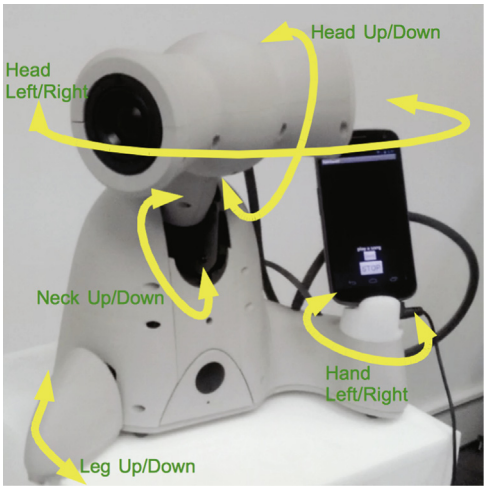

Fig. 8. Shimi's five degrees of freedoms. For our experiment we use the two head DoFs and the neck DoF.

move based on those parameters. The growth and decay function allows robots to fluidly transition between emotional states. Breazeal (2003) describes a homeostatic regime in which Kismet "wants" to maintain a certain intensity level. Here, a homeostatic state represented by particular control parameter values is implemented. This state describes the emotional behavior that the robot "wants" to be at. If a robot is to be inherently happy it's homeostatic values would be representative of a happy behavior. If it is to be inherently sad then these values would be representative of a sad behavior. In our implementation we gave Shimi a homeostatic nature of calm and content.

5.2.1. Growth and decay function

The growth and decay function describes the rate at which the robot "grows" to a new behavior (defined by the control parameters) and "decays" back to its homeostatic state. This is crucial based on the assumption that it is not natural for a person in a joyous and excited state to immediately change to a calm and still state unless triggered by some external force. The decay function allows for the robot to naturally make transitions. Multiple growth and decay functions can be used depending on the situation. For example, Fig. 7a represents an immediate transition from the robot's current state to the new emotional state with an exponential decline back to the homeostatic state. This function is useful when abrupt physical action is important such as for expressing surprise or startled fear. Fig. 7b might be used to express an emotion such as disappointment where it takes a moment for the emotion to "sink in" and peak disappointment to be reached. Then it slowly decays back to the homeostatic state. The growth and decay function can be given by

$$f(P_i) = \text{decay}(H, S, T, t_i - t_0) \tag{1}$$

where P is the set of n control parameters at current time i such that $P = [p_{i0}, p_{i1}, \ldots, p_{in}]$, t_i is the current time, t_0 is the initial time of activation, T is the total time necessary to decay back to the homeostatic state, H is the set of control parameters for the homeostatic state such that $H = [h_0, h_1, \ldots, h_n]$, and S is the set of received control parameters from the interpretation phase such that $S = [s_0, s_1, \ldots, s_n]$.

The total time, T, is a value which must be determined by the situation. For example, in terms of motion the human expression of laughter can be thought of as a series of contiguous chuckles. The physical traits between a discrete segment of laughter and a chuckle are similar, but where laughter is repetitive and lengthy a chuckle is fleeting. The same notion applies to the temporal length of growth and decay. In our current implementation we have a fixed T for different emotions. For future work an algorithmically determined T can be determined by a function of the relevant social display norms.

5.2.2. Beat synchronized update

The expression system is driven by a timer which updates at a given interval, α where $\alpha = t_{i+1} - t_i$. The timer triggers an update of Eq. (1) every α seconds. This interval can be adjusted depending on how frequently the robot should update. Obviously, for smaller α values the transitions between emotional states will more closely follow the contours of the decay function. A larger α will ease computation expense, but transitions may not be as smooth.

Shimi, is first and foremost a musical robot (see Hoffman, 2012) and the use of an α is a consequence of this fact. In music and dance the concept of a "beat" is essential and signifies the perceptible pulse exhibited by the rhythm. When Shimi is dancing to music it is important for it to align its motions with this pulse. If Shimi were to use this emotion expression system for dance its α would be set to the temporal interval which signifies the length of a beat.

5.2.3. Motion primitives

In the system the robot's movements can be defined by a number of motion primitives. The motion primitives we have defined for Shimi are head nod (up and down) and head shake (left and right). Fig. 8 shows the range of motion for each of Shimi's DoFs. The motion primitive for the head nod involves the head up/down and neck up/down DoFs. The motion primitive for the head shake involves the head left/right DoF. How the motion primitive is performed is defined by a subset of the control parameters such that:

$$P_{nod} = [p_{posture}, p_{gaze}, p_{posHeadActivation}, p_{volatility}, p_{exaggeration}] \tag{2}$$

$$P_{shake} = [p_{gaze}, p_{negHeadActivation}, p_{volatility}, p_{exaggeration}] \tag{3}$$

These parameters control the rate at which the primitive is performed, the maximum range the motion exhibits, and the center location of the motion.

The notion of volatility is to introduce variation into a repetitive series. We define volatility as a function of variation which manipulates a value based on a Gaussian distribution about p_η where p_η is some control parameter. The Gaussian probability density function is the statistical distribution:

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{[]2\pi}} e^{-(x-\mu)^2/2\sigma^2} \tag{4}$$

where $\mu = mean$, and $\sigma^2 = variance$. An increase in the value $p_{volatility}$ results in an increase of variance. The volatility manipulation function is thus defined as:

$$\text{volatility}(p_\eta) = f(x; p_\eta, p_{volatility}) \tag{5}$$

where x is a random number. The rate, range, and location of the motion primitives are defined as functions of the remaining original control parameters and the volatility manipulated control parameters.

$$rate = g(\text{volatility}(p_{headActivation})) \tag{6}$$

$$range = g(\text{volatility}(p_{exaggeration})) \tag{7}$$

$$location_{nod} = g(p_{gaze}, p_{posture}) \tag{8}$$

$$location_{shake} = g(p_{gaze}) \tag{9}$$

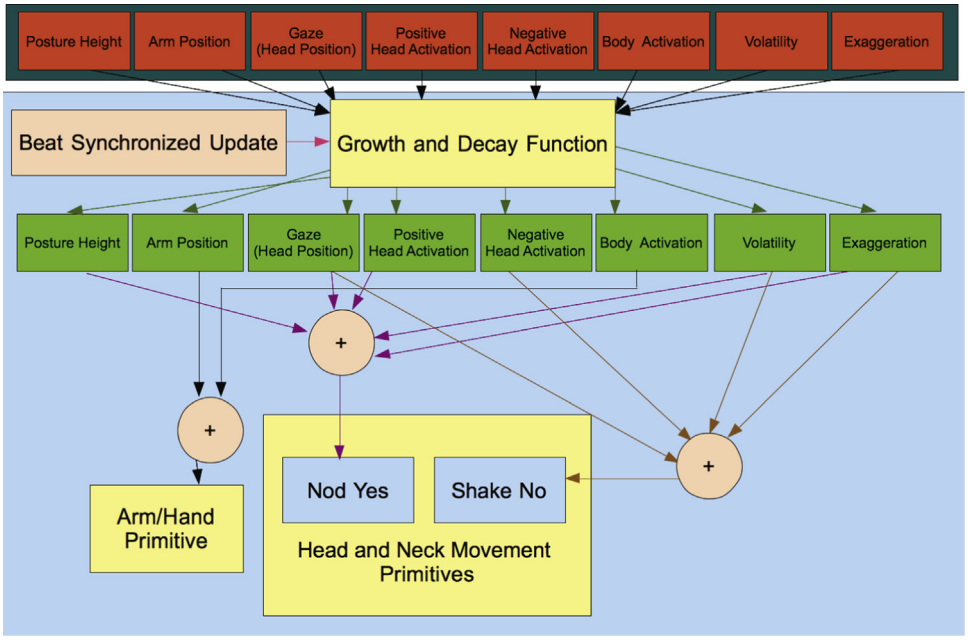

Fig. 9. The initial control parameters are received by the growth and decay function. This function allows for different types of transitions between emotional states (such as slow and smooth vs abrupt) and outputs new parameters as a function of time and the homeostatic state. The homeostatic state represents a set of constant parameters which the robot will always come back to as determined by the decay function. The function is run at intervals determined by an arbitrary timer in the beat synchronized update. This allows a dancing robot to align its movements with the perceived beat of music. Finally, motion primitives are performed in a manner determined by the outputs of the decay function.

where *rate* is always a whole number multiple of α in order to keep the motions synchronized to the beats. Fig. 9 gives an overview of the architecture for the expression portion of the system.

5.3. Evaluating control parameters

An experiment was conducted to measure the significance of our control parameters as emotional contributors by determining the correlation between each parameter and participants' perceptions of emotions. For the experiment the system for generating emotional expression was implemented on Shimi.

5.3.1. Procedure

Ten participants (7 male) were asked to watch six 45 second performances of Shimi generating motions using the system. For each performance Shimi had five emotional state updates with each having random values for the control parameters (values were randomly generated using Java's Random.nextFloat() function). The real time values (the decay function outputs) for each parameter were recorded. The α update rate for the decay function was 200 ms, thus, parameter values were recorded every 200 ms).

During each performance participants were tasked with rating the level of a particular emotion in each performance by moving a slider up and down on a MIDI control board. Slider values were also recorded every 200 ms. Participants were told that a slider in its lowest position indicated the absence of the emotion and anything higher indicated the emotion was present with the level describing how well the emotion was being represented. Before each performance they were told which particular emotion they would be evaluating so that they coded for only one emotion for each performance. This meant each participant moved the slider a total of six times rating each of the six fundamental emotions. Because parameter values were generated randomly there were no identical performances. A message was sent over a network in order to synchronize the recording of Shimi's parameter values and the participant slider values. There were a total of ten participants (different from the first experiment).

5.3.2. Hypotheses

We made several hypotheses based on Darwin's observations and our empirical findings regarding dynamic affective expression design from the previous experiment:

- H8 (Posture vs Valence)—Posture height will be positively correlated with positive valence emotions and negatively correlated with negative valence emotions
- H9 (Head Nodding vs Arousal and Valence)—Positive head activation (increased nodding rate) will be positively correlated with emotions of positive arousal and positive valence (happiness and surprise) and negatively correlated for the other emotions
- H10 (Head Shaking vs Arousal and Valence)—Negative head activation (increased shaking rate) will be positively correlated with emotions of positive arousal and negative valence (anger, fear, and disgust) and negatively correlated for the other emotions
- H11 (Gaze vs Arousal and Valence)—Gaze will be positively correlated with emotions of positive valence and positive arousal (happiness and surprise) and negatively correlated for other emotions
- H12 (Volatility vs Emotions with “Purposeless Movements”)—Volatility will be positively correlated with emotions which Darwin describes as being expressed with “purposeless movements” (happiness and anger)
- H13 (Exaggeration vs Arousal)—Exaggeration will be positively correlated with emotions which have a positive arousal level (anger, happiness, disgust, surprise, fear) and negatively correlated with emotions which have a negative arousal level (sadness)

5.3.3. Results and discussion

We evaluated the results by calculating the correlation coefficients between the slider values and control parameters for each of the emotions. The results are summarized in Fig. 10. The *p*-values for the majority of the coefficients were below .05 suggesting the null hypothesis

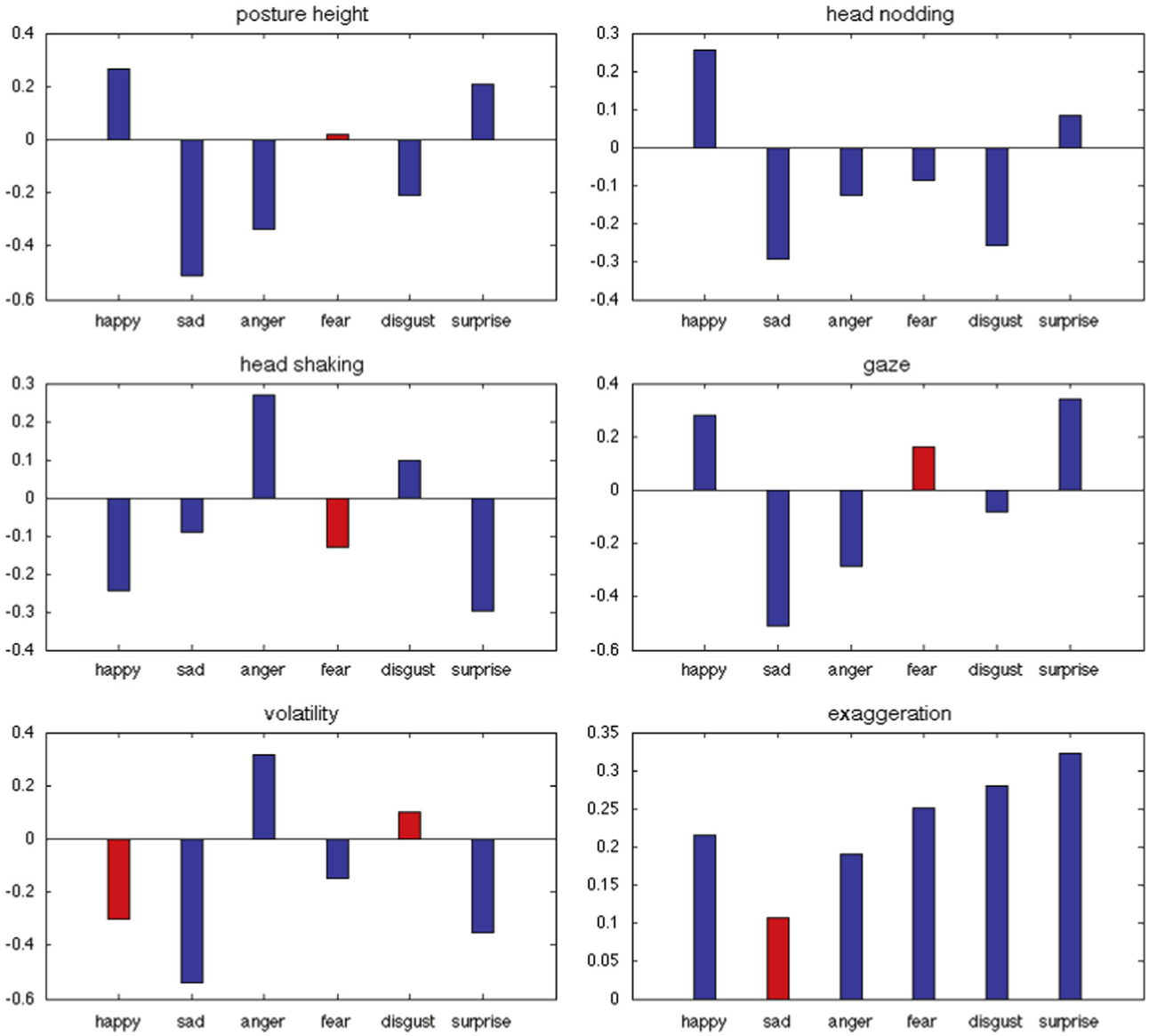

Fig. 10. Correlation coefficients for the parameter control variables and each of the six fundamental emotions. Blue bars indicate results supporting our hypotheses and red bars indicate results contrary to our hypotheses. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

can be rejected. The complete correlation values are shown in Table 3. The coefficients with $p > .05$ are shown in red.

The results largely supported our hypotheses and provide additional evidence that specific characteristics of body language are indicative of specific emotions. However, with only 10 participants (and no controls for culture) we are only cautiously optimistic that our parameters demonstrated relevance to expressing emotion. Specifically, posture height and gaze demonstrated direct correlations with an emotion's valence. Fear, however, was the one exception and showed weak positive correlations for both of these parameters despite its negative valence. Head activation (nodding and shaking) demonstrated direct correlations with an emotion's valence and arousal. Though, again, fear demonstrated a negative correlation with head shaking which was contradictory to our hypothesis. Volatility, or the amount of variation exhibited in a behavior, showed positive correlations with anger and disgust and negative correlations with the remaining four emotions. Finally, exaggeration, or the range of motion exhibited by a movement, showed positive correlations with all of the emotions. These results can only offer broad ideas concerning the relationship between the control parameters and the emotions. There may be stronger or weaker correlations for each individual parameter depending on the values of the other parameters. Additional studies and analyses can be helpful in describing the covariance between multiple control parameters and the emotion ratings.

6. Evaluating the expressive generative system using an interactive experiment

Both human–human (Nagai and Rohlfing, 2009) and human–robot (Mutlu et al., 2009) studies have shown social interactions to be more engaging when actors demonstrate a certain degree of physical responsiveness to the actions of one another. Mutlu et al. demonstrate that even a response as simple as gaze can influence how fondly a robot is perceived. Such feedback responses are helpful in reading others' goals, intentions, and levels of understanding. In this section we evaluate the efficacy of our expressive

Table 3
Correlation values for each parameter where bold values indicate no statistical significance or $p > .05$.

Parameter	Happy	Sad	Anger	Fear	Disgust	Surprise
Posture height	.26	−.51	−.34	**.02**	−.21	.21
Head nodding	25	−.30	−.13	**−.09**	−.26	**.08**
Head shaking	−.25	**−.09**	.27	−.13	.1	−.30
Gaze	.28	−.51	−.29	.16	**−.08**	.34
Volatility	−.3	−.54	.32	−.15	.1	−.35
Exaggeration	.21	.11	.19	.25	.28	.32

generative system's ability to provide coherent and appropriate physical responses during a human–robot interaction.

6.1. Experiment

The experiment evaluates an implementation of the entire emotion intelligence system on Shimi through a user study. In the implementation the user speaks a phrase that is recorded by the mobile phone and then sent to Google's servers for speech recognition analysis. The resulting text representation of the speech is then evaluated for its emotional content using six parameters (representing levels for happiness, sadness, anger, disgust, fear, and surprise). In order to find the weights of each parameter an averaged perceptron classifier was trained using emotionally tagged datasets of news headlines (Strapparava and Mihalcea, 2008), children's stories, Alm (2005), and an original corpus acquired from social media sources including Facebook, Tumblr, and Twitter. It is assumed that Google returns an accurate representation of what was spoken (though this is not always the case). When the text is classified Shimi alters its behavior from a calm, breathing state to something more representative of the emotion it has detected based on the control parameter values representing happiness, sadness, anger, disgust, fear, or surprise. In summary, the participant speaks to the robot and the robot responds with emotive expressive physical behaviors.

Contingency plays an important role in engagement and interactive scenarios (Fischer et al., 2013). The level of engagement and enjoyment can be quantified by measuring the time a person spends on a task or interaction (Powers et al., 2007). Therefore we make the following hypotheses:

H14 (Time Spent Interacting)—Participants whose emotions are acknowledged using the emotional intelligence system will enjoy the interaction experience more so than those whose emotions were randomly acknowledged. This will result in a longer time spent interacting with Shimi.

H15 (Number of Phrases Spoken)—Participants whose emotions are acknowledged using the emotional intelligence system will enjoy the interaction experience more so than those whose emotions were randomly acknowledged. This will result in an increased number phrases spoken during the interaction.

6.1.1. Experimental design

The experiment was a between groups design. There were two groups of participants with 22 total participants (15 male) made up of undergraduate and graduate students (different from the previous experiments). Twelve students received course credit for participation. Each group had 11 members. In the first group, participants interacted with a version of Shimi that utilized the emotional intelligence system. In the other group, Shimi did not use any recognition or perception abilities and instead randomly chose an emotive behavior to respond with. Each participant was told that Shimi could recognize phrases as being one of the six basic emotions. They were then asked to interact with Shimi using speech twice. The first time, each participant read from a script following a narrative in which they read a phrase, Shimi responded, read another phrase, Shimi responded, and so on until the script was finished. In the second interaction, participants were asked to follow the same turn taking narrative, but instead of reading from a script they were asked to freely speak for as long as they like given only the fact that Shimi will respond to emotional content in language.

During the free speak interaction we kept track of the number of phrases spoken and the length of time of the interaction as objective measures. After completing surveys for both parts of the study each participate was interviewed by the proctor and asked to discuss their experience with Shimi.

6.1.2. Evaluation and results

The data was analyzed using a one-way ANOVA where significance was found at the $p < .05$ level. Fig. 11 shows that on average participants interacted with Shimi for a longer period of time [$F(1, 16) = 10.94$, $p < .01$] and said more phrases [$F(1, 16) = 6.39$, $p < .05$] when Shimi used the emotional intelligence system. Overall, the results suggest that participants perceived Shimi as being more attentive and responsive when using the emotional intelligence system. Participants also preferred to say more and interact for longer periods of time when Shimi utilized emotional intelligence as opposed to randomly responding.

6.2. Discussion

The content provided by the participants during the free speak interaction varied much more than anticipated. Some really thought about what they said and attempted to elicit certain emotions in Shimi. Others said things which seemingly did not have any emotional content at all such as “I am 21 years old.” Though the script reading was incorporated to control for this variation, it proved difficult to encourage people to interact with

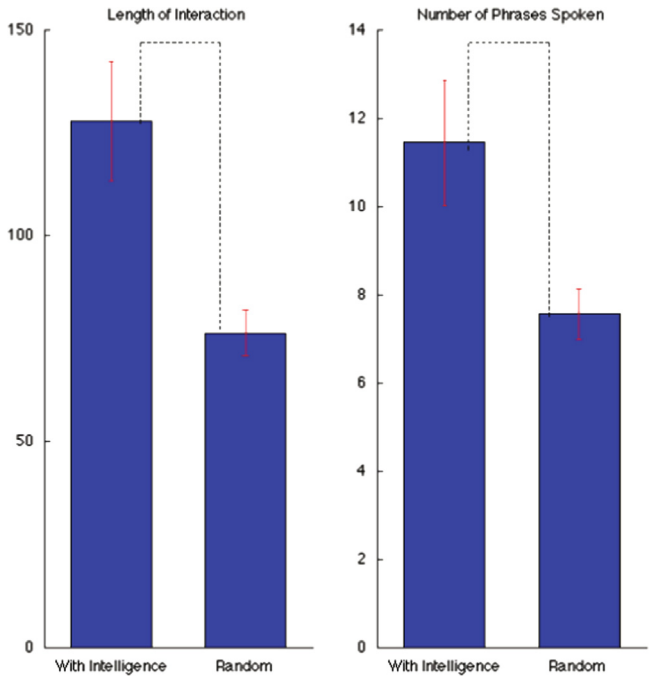

Fig. 11. Mean results for length of time (in seconds) spent interacting with Shimi and number of phrases spoken during the free speak interactions for Group 1 (with intelligence) and Group 2 (random). For both objective measures the mean differences between groups were statistically significant.

Shimi without biasing the results by explicitly telling them what to say. Constraining the free speak interaction to a number of suggested topics may be useful in the future. The ANOVA results of the objective measurements do, however, indicate there was a preference to interact with Shimi for longer periods of time when it used the emotional intelligence system.

After each trial the we interviewed the participant to gauge their experience. The interviews further validated Shimi's ability to convey emotion through its movements in a manner which is understandable to people. For example, during one participant's interaction, Shimi, by chance, responded with angry gestures many times. The participant did not think that this was a consequence of Shimi misunderstanding the emotional content of what was being said, but rather "Shimi was just an angry robot." Another participant described Shimi as seeming "really surprised" to what was being said. Another said "I love how happy Shimi is." On one final note, we did not formally measure whether Shimi could induce specific emotions in people using its gestures. However, in this study we noticed that participants seemed to exhibit an expression of satisfaction and joy when Shimi appropriately responded to an emotional phrase despite the emotion being conveyed. This suggests Shimi can successfully communicate emotion with its gestures, though inducing emotion (as humans do) seems less promising with the current system and interaction.

7. Future work

Future work involves both the expression and perception aspects of robots and emotion. The perception system described here assumes everything that is said has emotional content and does not attempt to distinguish emotionally neutral statements. A neutral classification can help the interaction to be more engaging and may provoke more emotional content from the participants in an attempt to get Shimi to respond.

Another goal is to expand Shimi's recognition beyond the six basic emotions to include a wider range of variation. This can be done by getting additional content for the corpus which represents different emotions such as love, interest, exhaustion, or boredom. Twitter data is naturally suited for discrete emotion classification and it is somewhat of a challenge to map onto a core affect dimensional space. One method is to use the probabilities the perceptron calculates for each emotion and classify a phrase as a mixture of emotions (i.e. 80% anger and 20% disgust) and perform a weighted average to estimate the proper coordinates. Using such a method would allow us to better take advantage of the more continuous nature of the expression system. Additionally, some emotional gestures are more easily identifiable than others (happy, sad, and anger gestures received over 90% classification accuracy) and we can dynamically change these gestures to convey different levels of these emotions. The literature shows that not every version of happy or sad is equivalent so perhaps levels of discrete emotions can be useful in achieving optimal emotive displays.

Deciding how Shimi should respond to the emotions it recognizes in language is another necessary aspect of the system that needs to be developed further. Mirroring can play a role in displaying empathy (Pfeifer et al., 2008), but some emotions, such as anger, are believed to be tools for manipulating others in a social context. Mirroring would not be appropriate in this context. Discriminating between possible sources of a person's emotions seems an essential next step in affective computing and is one of the challenges (Muller, 2004) describes. Emotional perception also varies across genders and cultures so it is reasonable to think that Shimi's emotional intelligence system should be personalized for a particular user. Though exactly how much is unclear. The results of these study indicate a certain level of ubiquity in perception and recognition so perhaps the current parameters could be used as starting points from which the system evolves and learns.

Emotional intelligence can also be used to automate some of Shimi's current functions. For example, currently one can explicitly ask Shimi to play a happy song. Using affective recognition Shimi can autonomously choose the music to play based on a person's mood or emotional state. We have also only been examining motion and posture as a method for expressing emotion. Shimi is a speaker dock and using bimodal expression system which incorporates sound and motion can be very useful for enhancing and reinforcing emotional expression.

8. Conclusion

Emotion is something inherent to all people and the perceptual and communicative attributes of emotion can be salvaged to establish more engaging and comfortable human–machine interactions. Therefore, it is important for social robots to be responsive to people and in a manner perceived as appropriate and natural. This can be challenging depending on the DoFs and physical capabilities available to the robotic platform.

In this article, we presented several experiments evaluating our robot's ability to express emotion through physical behavior. Our robot, Shimi, allowed us to explore affective expression in a robot characterized by a small number of DoFs, non-humanoid design, and no face. Through several experiments we found that it is possible for a robot to still be emotionally expressive despite such constraints.

In the first experiment we found:

1. The identifications of both static poses and dynamic physical behaviors were significantly better than chance. This suggests that certain characteristics of motion and pose are associated with particular emotions. This is unlike the *quantify but not quality* theory that depicts the use of body language as merely reflecting the intensity of the emotional experience, rather than the actual emotion.
2. The dynamic physical behaviors demonstrated better classification accuracy than static poses. This supports previous studies comparing emotion recognition of pose and movements of humans (Gunes and Piccardi, 2007). However, unlike the previous work, which focuses on identifying the emotions of people and claims that specific body parts (torso, arms, legs, etc.) are vital for recognition (De Meijer, 1989), we examined the phenomena using a non-humanoid robot with only a faceless head. Despite these physical constraints of the robot, the classification rates of the dynamic physical behaviors were quite high, thus, suggesting the nature in which a DoF moves (exhibited by its velocity, acceleration, changes in direction, etc.) plays a more significant role in emotion expression than the specific DoF itself and its location on the body.
3. Though the differences between the video and in person classification rates were not statistically significant, there was an increase in recognition accuracy for each emotion for the in person viewings (except for sadness, which achieved 100% accuracy in both cases). There was also greater internal consistency (i.e. less confusion with other emotions) for the in person viewings compared to the video viewings. Previous studies in human–robot interaction have shown collocated experiences to have both positive and zero influence on the interaction compared to a virtual experience (see Section 4.2). The trends from our experiment suggest that physical presence may influence the ability to recognize emotion expressed through physical behaviors and should be researched in more depth in the future.

4. Seeing the robot in person versus a video resulted in an increase of perceived magnitude of the expressed emotion. For example, happy was happier, sad was sadder, and angry was angrier. This was shown through more extreme user ratings of arousal and valence for each emotion. It seems presence has a stronger affect on the perceived intensity of the expressed emotion, rather than the identification of the emotion.

In Section 5, we described a system for algorithmically generating emotionally expressive movement using variables inspired by human expressive physical tendencies. These variables included specific DoF positions, activations, volatility, and exaggeration.

In this section:

1. We described a generative architecture with methods for representing each of the control variables as a mathematical function.
2. Based on a user study evaluating the generative system's control variables, we found a correlation exists between the variables manipulating physical motion and the perceived emotion state of Shimi. These findings support the notion that body language and certain physical behaviors can be indicative of specific emotions.

In Section 6, a final user study was conducted to evaluate the utility of expressive physical behaviors on a social interaction between a person and robot. The results of this study demonstrate increased user engagement when the robot utilizes our system to express emotion. This is shown through an increase in the duration of the interaction and the number spoken phrases.

In summary, we have shown that it is possible for a physically constrained robot to successfully express emotion through the use of dynamic physical behaviors. Though attributes such as facial expression and humanoid form can be undoubtedly useful, the absence of such features does not prevent a robot from coherently communicating emotion in an expressive manner. In lieu of a face, limbs, and torso it is possible to express emotion through dynamic physical behaviors by manipulating parameters of DoF activation, volatility, gaze, exaggeration, and posture. Though such behaviors can be useful for most robots interacting with people, they are especially useful when using robotic platforms with very few DoFs that may result from various constraints such as size or cost. Additionally, designing robots with the ability to create such meaningful affective behaviors is beneficial to human–robotic interaction tasks as it provides increased user engagement. We hope our architecture for autonomous generation of expressive behaviors can be useful for others who design affective robots.

References

Alibali, M.W., Kita, S., Young, A.J., 2000. Gesture and the process of speech production: we think, therefore we gesture. Lang. Cognit. Process. 15, 593–613.

Alm, C.O., 2005. Emotions from text: Machine learning for text-based emotion prediction. In: Proceedings of HLT/EMNLP, pp. 347–354.

Anderson, A.H., Smallwood, L., MacDonald, R., Mullin, J., Fleming, A., O'MALLEY, C., 2000. Video data and video links in mediated communication: what do users value?. Int. J. Hum. Comput. Stud. 52, 165–187.

Aviezer, H., Hassin, R.R., Ryan, J., Grady, C., Susskind, J., Anderson, A., Moscovitch, M., Bentin, S., 2008. Angry, disgusted, or afraid? Studies on the malleability of emotion perception. Psychol. Sci. 19, 724–732.

Aviezer, H., Trope, Y., Todorov, A., 2012. Body cues, not facial expressions, discriminate between intense positive and negative emotions. Science 338, 1225–1229.

Bainbridge, W.A., Hart, J., Kim, E.S., Scassellati, B., 2008. The effect of presence on human-robot interaction. In: The 17th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2008, IEEE, pp. 701–706.

Barrett, L.F., Gendron, M., Huang, Y.M., 2009. Do discrete emotions exist? Philos. Psychol. 22, 427–437.

Bavelas, J.B., Chovil, N., 2000. Visible acts of meaning an integrated message model of language in face-to-face dialogue. J. Lang. Soc. Psychol. 19, 163–194.

Bavelas, J.B., Chovil, N., 2006. Nonverbal and verbal communication: hand gestures and facial displays as part of language use in face-to-face dialogue.

Bavelas, J.B., Coates, L., Johnson, T., 2002. Listener responses as a collaborative process: the role of gaze. J. Commun. 52, 566–580.

Breazeal, C., 2003. Emotion and sociable humanoid robots. Int. J. Hum. Comput. Stud. 15, 119–155.

Breazeal, C., Aryananda, L., 2002. Recognition of affective communicative intent in robot-directed speech. Auton. Rob. 12, 83–104.

Breazeal, C., Wang, A., Picard, R., 2007. Experiments with a robotic computer: body, affect and cognition interactions. In: 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE, pp. 153–160.

Bretan, M., Cicconet, M., Nikolaidis, R., Weinberg, G., 2012. Developing and composing for a robotic musician. In: Proceedings of International Computer Music Conference on (ICMC'12), Ljubljana, Slovenia.

Campos, J.J., Thein, S., Owen, D., 2003. A darwinian legacy to understanding human infancy. Ann. NY. Acad. Sci. 1000, 110–134.

Cañamero, L., Aylett, R., 2008. Animating Expressive Characters for Social Interaction, vol. 74. John Benjamins Publishing Company, Philadelphia, PA.

Carroll, J.M., Russell, J.A., 1997. Facial expressions in hollywood's protrayal of emotion. J. Pers. Soc. Psychol. 72, 164.

Cassell, J., Bickmore, T., Campbell, L., Vilhjdlm&&on, H., 2000. Designing embodied conversational agents. Embodied Convers. Agents 29.

Cassell, J., Nakano, Y.I., Bickmore, T.W., Sidner, C.L., Rich, C., 2001. Non-verbal cues for discourse structure. In: Proceedings of the 39th Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, pp. 114–123.

Castellano, G., Leite, I., Pereira, A., Martinho, C., Paiva, A., McOwan, P.W., 2010. Affect recognition for interactive companions: challenges and design in real world scenarios. J. Multim. User Interfaces 3, 89–98.

Claypool, M., Tanner, J., 1999. The effects of jitter on the peceptual quality of video. In: Proceedings of the Seventh ACM International Conference on Multimedia (Part 2), ACM, pp. 115–118.

Colombetti, G., 2009. From affect programs to dynamical discrete emotions. Philos. Psychol. 22, 407–425.

Coulson, M., 2004. Attributing emotion to static body postures: recognition accuracy, confusions, and viewpoint dependence. J. Nonverb. Behav. 28, 117–139.

Darwin, C., 1916. The Expression of the Emotions in Man and Animals. D. Appleton and Co., New York, URL ⟨http://www.biodiversitylibrary.org/item/24064⟩. ⟨http://www.biodiversitylibrary.org/bibliography/4820⟩.

De Meijer, M., 1989. The contribution of general features of body movement to the attribution of emotions. J. Nonverbal Behav. 13, 247–268.

De Silva, P.R., Bianchi-Berthouze, N., 2004. Modeling human affective postures: an information theoretic characterization of posture features. Comput. Anim. Virtual Worlds 15, 269–276.

Delaunay, F., Belpaeme, T., 2012. Refined human-robot interaction through retro-projected robotic heads. In: 2012 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO), IEEE, pp. 106–107.

Devillers, L., Vidrascu, L., Lamel, L., 2005. 2005 special issue: challenges in real-life emotion annotation and machine learning based detection. Neural Netw. 18, 407–422.

Ekman, P., 1993. Facial expression and emotion. Am. Psychol. 48, 384.

Fernández-Dols, J.M., Ruiz-Belda, M.A., 1997. 11. spontaneous facial behavior during intense emotional episodes: artistic truth and optical truth. Psychol. Facial Expr. 255.

Fischer, K., Lohan, K., Saunders, J., Nehaniv, C., Wrede, B., Rohlfing, K., 2013. The impact of the contingency of robot feedback on hri. In: International Conference on Collaboration Technologies and Systems (CTS), IEEE, pp. 210–217.

Frank, R.H., 1988. Passions within Reason: The Strategic Role of the Emotions. WW Norton & Co., New York, NW.

Fridlund, A.J., 1991. The sociality of solitary smiles: effects of an implicit audience. J. Pers. Soc. Psychol. 60, 229–240.

Fridlund, A.J., Ekman, P., Oster, H., 1987. Facial expressions of emotion.

Frijda, N., 1987. The Emotions. Cambridge University Press, Cambridge, England.

Frijda, N., 1995. Emotions in robots. In: Roitblat, H.L., Meyer, J.-A. (Eds.), Comparative Approaches to Cognitive Science, pp. 501–516.

de Gelder, B., 2006. Towards the neurobiology of emotional body language. Nat. Rev. Neurosci. 7, 242–249.

de Gelder, B., Hadjikhani, N., 2006. Non-conscious recognition of emotional body language. Neuroreport 17, 583–586.

Gielniak, M.J., Thomaz, A.L., 2012. Enhancing interaction through exaggerated motion synthesis. In: Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction, ACM, New York, NY, USA, pp. 375–382. doi: http://dx.doi.org/10.1145/2157689.2157813.

Grunberg, D.K., Batula, A.M., Schmidt, E.M., Kim, Y.E., 2012. Synthetic emotions for humanoids: perceptual effects of size and number of robot platforms. Int. J. Synth. Emotions (IJSE) 3, 68–83.

Gunes, H., Piccardi, M., 2007. Bi-modal emotion recognition from expressive face and body gestures. J. Netw. Comput. Appl. 30, 1334–1345.

Hamann, S., 2012. Mapping discrete and dimensional emotions onto the brain: controversies and consensus. Trends Cognit. Sci. 16 (9), 458–466.

Hess, U., Adams Jr., R.B., Kleck, R.E., 2007. Looking at you or looking elsewhere: the influence of head orientation on the signal value of emotional facial expressions. Motiv. Emot. 31, 137–144.

Hoffman, G., 2012. Dumb robots, smart phones: a case study of music listening companionship. In: RO-MAN, 2012 IEEE, IEEE, pp. 358–363.

Hoffman, G., Breazeal, C., 2008. Anticipatory perceptual simulation for human-robot joint practice: theory and application study. In: Proceedings of the 23rd

National Conference on Artificial Intelligence—vol. 3, AAAI Press, pp. 1357–1362. URL ⟨http://dl.acm.org/citation.cfm?id=1620270.1620285⟩.
Hüttenrauch, H., Severinson Eklundh, K., Green, A., Topp, E.A., 2006. Investigating spatial relationships in human-robot interaction. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, 2006, IEEE, pp. 5052–5059.
Inderbitzin, M., VŠljamŠe, A., Calvo, J.M.B., Verschure, P.F.M.J., Bernardet, U., 2011. Expression of emotional states during locomotion based on canonical parameters. In: Ninth IEEE International Conference on Automatic Face and Gesture Recognition (FG 2011), Santa Barbara, CA, USA, 21–25 March 2011, IEEE, pp. 809–814. doi: http://dx.doi.org/10.1109/FG.2011.5771353.
Jack, R.E., Garrod, O.G., Yu, H., Caldara, R., Schyns, P.G., 2012. Facial expressions of emotion are not culturally universal. Proc. Natl. Acad. Sci. 109, 7241–7244.
Kendon, A., 1990. Conducting Interaction: Patterns of Behavior in Focused Encounters, vol. 7. CUP Archive.
Kidd, C.D., 2003. Sociable Robots: the Role of Presence and Task in Human-Robot Interaction (Ph.D. thesis), Massachusetts Institute of Technology.
Kidd, C.D., Breazeal, C., 2004. Effect of a robot on user perceptions. In: Proceedings. 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2004, (IROS 2004), IEEE, pp. 3559–3564.
Kipp, M., Martin, J.C., 2009. Gesture and emotion: can basic gestural form features discriminate emotions?. In: 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009, IEEE, pp. 1–8.
Kita, S., 2009. Cross-cultural variation of speech-accompanying gesture: a review. Lang. Cognit. Processes 24, 145–167.
Kita, S., Özyürek, A., Allen, S., Brown, A., Furman, R., Ishizuka, T., 2007. Relations between syntactic encoding and co-speech gestures: implications for a model of speech and gesture production. Lang. Cognit. Processes 22, 1212–1236.
Kleinsmith, A., De Silva, P.R., Bianchi-Berthouze, N., 2006. Cross-cultural differences in recognizing affect from body posture. Interact. Comput. 18, 1371–1389.
Kose-Bagci, H., Ferrari, E., Dautenhahn, K., Syrdal, D.S., Nehaniv, C.L., 2009. Effects of embodiment and gestures on social interaction in drumming games with a humanoid robot. Adv. Robot. 23, 1951–1996.
Kozima, H., Nakagawa, C., Kawai, N., Kosugi, D., Yano, Y., 2004. A humanoid in company with children. In: 4th IEEE/RAS International Conference on Humanoid Robots, IEEE, pp. 470–477.
Kozima, H., Yano, H., 2001. In search of otogenetic prerequisites for embodied social intelligence. In: International Conference on Cognitive Science Proceedings of the Workshop on Emergence and Development on Embodied Cognition, pp. 30–34.
Krauss, R.M., Morrel-Samuels, P., Colasante, C., 1991. Do conversational hand gestures communicate? J. Pers. Soc. Psychol. 61, 743.
Kraut, R.E., Johnston, R.E., 1979. Social and emotional messages of smiling: an ethological approach. J. Pers. Soc. Psychol. 37, 1539–1553.
Krumhuber, E., Manstead, A.S., Kappas, A., 2007. Temporal aspects of facial displays in person and expression perception: the effects of smile dynamics, head-tilt, and gender. J. Nonverbal Behav. 31, 39–56.
Lasseter, J., 1987. Principles of traditional animation applied to 3d computer animation. SIGGRAPH Comput. Graph. 21, 35–44. http://dx.doi.org/10.1145/37402.37407.
Linda A. Camras, J.S., Michel, G., 1993. Do infants express discrete emotions? adult judgments of facial, vocal, and body actions. J. Nonverbal Behav. 17, 171–186.
Lindquist, K.A., Wager, T.D., Kober, H., Bliss-Moreau, E., Barrett, L.F., 2012. The brain basis of emotion: a meta-analytic review. Behav. Brain Sci. 35, 121–143.
Lockerd, A., Breazeal, C., 2005. Tutelage and socially guided robot learning. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE/RSJ.
Mead, R., Atrash, A., Mataric, M.J., 2011. Recognition of spatial dynamics for predicting social interaction. In: Proceedings of the 6th International Conference on Human-Robot Interaction, ACM, pp. 201–202.
Mehrabian, A., 1996. Pleasure-arousal-dominance: a general framework for describing and measuring individual differences in temperament. Curr. Psychol. 14, 261–292.
Michalowski, M.P., Sabanovic, S., Kozima, H., 2007. A dancing robot for rhythmic social interaction. In: 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE. pp. 89–96.
Monceaux, J., Becker, J., Boudier, C., Mazel, A., 2009. Demonstration: first steps in emotional expression of the humanoid robot nao. In: Proceedings of the 2009 International Conference on Multimodal Interfaces, ACM, pp. 235–236.
Moon, A., Parker, C.A., Croft, E.A., Van der Loos, H., 2013. Design and impact of hesitation gestures during human-robot resource conflicts. J. Hum. Rob. Interact. 2, 18–40.
Muhl, C., Nagai, Y., 2007. Does disturbance discourage people from communicating with a robot?. In: The 16th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2007, IEEE, pp. 1137–1142.
Muller, M., 2004. Multiple paradigms in affective computing. Interact. Comput. 16, 759–768.
Mutlu, B., Yamaoka, F., Kanda, T., Ishiguro, H., Hagita, N., 2009. Nonverbal leakage in robots: communication of intentions through seemingly unintentional behavior. In: Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction, ACM, New York, NY, USA, pp. 69–76. doi: http://dx.doi.org/10.1145/1514095.1514110.
Nagai, Y., Rohlfing, K., 2009. Computational analysis of motionese toward scaffolding robot action learning. IEEE Trans. Auton. Mental Dev. 44–54.
Nayak, V., Turk, M., 2005. Emotional expression in virtual agents through body language. Adv. Vis. Comput., 313–320.
Nele Dael, M.M., Scherer, K.R., 2012. The body action and posture coding system (bap): development and reliability. J. Nonverbal Behav. 36, 97–121.
Pfeifer, J.H., Iacoboni, M., Mazziotta, J.C., Dapretto, M., 2008. Mirroring others' emotions relates to empathy and interpersonal competence in children. Neuroimage 39, 2076–2085.
Picard, R.W., 1995. Affective computing.
Powers, A., Kiesler, S., Fussell, S., Torrey, C., 2007. Comparing a computer agent with a humanoid robot. In: 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE, pp. 145–152.
Riek, L.D., Rabinowitch, T.C., Bremner, P., Pipe, A.G., Fraser, M., Robinson, P., 2010. Cooperative gestures: effective signaling for humanoid robots. In: 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2010, IEEE, pp. 61–68.
Rolls, E., 2005. Emotion Explained. Oxford University Press, USA.
Russell, J.A., 2009. Emotion, core affect, and psychological construction. Cognit. Emot. 23, 1259–1283.
Salem, M., Kopp, S., Wachsmuth, I., Joublin, F., 2010. Generating robot gesture using a virtual agent framework. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, pp. 3592–3597.
Salem, M., Kopp, S., Wachsmuth, I., Rohlfing, K., Joublin, F., 2012. Generation and evaluation of communicative robot gesture. Int. J. Soc. Robot. 4, 201–217.
Scheutz, M., Schermerhorn, P., Kramer, J., 2006. The utility of affect expression in natural language interactions in joint human-robot tasks. In: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction, ACM, pp. 226–233.
Schuller, B., Batliner, A., Steidl, S., Seppi, D., 2011. Recognising realistic emotions and affect in speech: state of the art and lessons learnt from the first challenge. Speech Commun. 53, 1062–1087.
Schuller, B., Stadermann, J., Rigoll, G., 2006. Affect-robust speech recognition by dynamic emotional adaptation. In: Proceedings of the Speech Prosody.
Sidner, C.L., Kidd, C.D., Lee, C., Lesh, N., 2004. Where to look: a study of human-robot engagement. In: Proceedings of the 9th International Conference on Intelligent User Interfaces, ACM, pp. 78–84.
Sidner, C.L., Lee, C., Morency, L.P., Forlines, C., 2006. The effect of head-nod recognition in human-robot conversation. In: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, ACM, pp. 290–296.
Simon, H.A., 1967. Motivational and emotional controls of cognition. Psychol. Rev. 74, 29.
Strapparava, C., Mihalcea, R., 2008. Learning to identify emotions in text. In: Proceedings of the 2008 ACM symposium on Applied Computing, ACM, New York, NY, USA, pp. 1556–1560. doi: http://dx.doi.org/10.1145/1363686.1364052.
Takayama, L., Pantofaru, C., 2009. Influences on proxemic behaviors in human-robot interaction. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2009, IEEE, pp. 5495–5502.
Traue, H.C., Ohl, F., Brechmann, A., Schwenker, F., Kessler, H., Limbrecht, K., Hoffmann, H., Scherer, S., Kotzyba, M., Scheck, A., et al., 2013. A framework for emotions and dispositions in man-companion interaction. Coverbal Synchr. Hum. Mach. Interact. 99.
Velásquez, J.D., 1997. Modeling emotions and other motivations in synthetic agents. In: Proceedings of the National Conference on Artificial Intelligence, Citeseer, pp. 10–15.
Vytal, K., Hamann, S., 2010. Neuroimaging support for discrete neural correlates of basic emotions: a voxel-based meta-analysis. J. Cognit. Neurosci. 22, 2864–2885.
Walbott, H.G., 1998. Bodily expression of emotion. Eur. J. Soc. Psychol. 28, 879–896.
Walters, M.L., Dautenhahn, K., Te Boekhorst, R., Koay, K.L., Syrdal, D.S., Nehaniv, C.L., 2009. An empirical framework for human-robot proxemics. In: Proceedings of New Frontiers in Human-Robot Interaction.
Watson, A., Sasse, M.A., 1998. Measuring perceived quality of speech and video in multimedia conferencing applications. In: Proceedings of the Sixth ACM International Conference on Multimedia, ACM, pp. 55–60.
Weinberg, G., Blosser, B., Mallikarjuna, T., Raman, A., 2009. The creation of a multi-human, multi-robot interactive jam session. In: Proceedings of NIME, pp. 70–73.
Weinberg, G., Driscoll, S., Thatcher, T., 2006. Jamaa: a percussion ensemble for human and robotic players. In: ACM International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2006), ACM Boston, MA.
Xia, G., Dannenberg, R., Tay, J., Veloso, M., 2012. Autonomous robot dancing driven by beats and emotions of music. In: Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems – vol. 1, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC. pp. 205–212. URL ⟨http://dl.acm.org/citation.cfm?id=2343576.2343605⟩.