Effects of Bodily Mood Expression of a Robotic Teacher on Students

Junchao Xu, Joost Broekens, Koen Hindriks*, and Mark A. Neerincx, *Member, IEEE

Abstract—This paper reports our investigation into the effects of bodily mood expression of a humanoid robot in a scenario close to real life. To this end, we used the NAO robot to perform as a lecturer in a university class. To display either a positive or a negative mood, we modulated 41 co-verbal gestures by adjusting behavior parameters that control spatial extent and motion dynamics, without modifying gesture function. Unique in this study is that (a) the robot gave an actual lecture to real students, (b) the interaction is one-to-many and relatively long (30 min), and (c) mood modulation was applied to a large set of behaviors. The robot presented the same lecture either in a positive or a negative mood to two audiences (between subjects). Although statistical analysis does not show that participants consciously recognized the robot mood, the results do show that participants in the positive mood condition rated their own arousal significantly higher than in the negative condition. Further, video annotation showed increased valence and arousal of the audience in the positive condition. Finally, participants’ ratings of the lecturing quality and the gesture quality of the robot are higher in the positive condition, demonstrating the importance of robot mood expression in a one-to-many interaction setting.

I. INTRODUCTION

Bodily expression is important for social robots to interact with humans naturally and intuitively [1]. In human-robot interaction (HRI), expressive body language of a robot facilitates a human’s understanding of the robot’s affective state, beliefs, and motives [2], and it improves the perception of a robot as trustworthy, reliable, and life-like [3]. Expressive language also increases the efficiency of human-robot task performance and robustness [4]. Bodily affective expression is in particular important for humanoid robots that lack facial features such as the NAO, ASIMO, and QRIO. In this study, we applied our bodily mood expression model [5] to a robotic tutor (RoboTutor) application. A NAO robot performs as a lecturer who gives a university introductory course on robotics. The NAO uses itself as an example and interacts with the audience using quizzes. The robot showed either a positive or a negative mood using mood-modulated body language during the lecture. We investigated whether students recognize the robot’s mood, and whether robot mood influences students’ mood, task performance, and perception of lecturing quality. This study differs from previous human-robot interaction studies because it combines (a) a “real-life” setting, (b) a broad range of modulated robot gestures that express mood over a longer period of time (30 min), and (c) interaction with a large group of individuals.

First, we used a “real-life” setting of teaching a lecture. We simulated a real classroom as closely as possible, where the main change was that we replaced a human teacher by a robot. The lecture room was familiar to the students, and the lecture content was part of an actual course that they were enrolled in. It is important to bring robots out of laboratory environments and evaluate them in more realistic settings, in which conditions and contexts are closer to where a robot is eventually applied [6], [7]. Several studies have investigated robot applications in public space, in which multiple people interacted with robots at the same time. A study, in which a robotic receptionist that can display mood facially interacted with people in a public region for a long term [8], showed that the average interaction time correlated positively with the number of visitors, as long as the robot displayed facial expressions no matter positive or negative. Social acceptance was studied by putting an ACE robot in a street and having it ask for directions [9]. Results indicated that limitations of the robot were less tolerated by people in a public area right from the start of the interaction. This suggests that the first impression is important for extending human-robot interaction over time. In another study [10], Robovie was placed in a classroom and interacted with students during a two-month period. It was found that students showed increased affinity with the robot and treated the robot more like a class member. A Geminoid robot replaced a lecturer in a university course and presented a lecture to a large audience [11]. This study suggested that the perception of the robot varies with the distance to the robot and with gender. In sum, placing robots in real-life like scenarios might elicit responses from people that are more realistic and varied. Moreover, people’s expectations and requirements for social acceptance are higher in a real life scenario than in a laboratory setup [9]. For example, in our setting, an audience might expect a robot to demonstrate teaching skills similar to those of a human or else would not accept the robot as a teacher.

Second, in this study we aimed to enable a robot to express mood consistently, using body language of the robot over a longer period of time. To this end, we generalized our mood expression model [5] to a broad range of robot gestures. We needed such a variety to ensure that the robot’s body language would be perceived as more or less natural over a longer period of time (30 min) during the lecture. 41 co-verbal gestures were used, each of which was modulated to display mood using our bodily mood expression model [5]. As such, this is the first study into the effects of robot mood expression over an extended period by means of a large variety of mood-modulated bodily gestures.

Third, in our RoboTutor application, the robot interacts with an audience of about 20 students. This allows us to study the effects of robot mood expression on multiple individuals that are part of a group. We previously investigated robot mood expression in dyadic interaction settings [12]. It is
known, however, that different interaction patterns have been reported for affective robot interaction with few and with many people [8], and it is known that there exist many factors that influence emotional contagion between individuals within a group (see [13] for a review) such as group membership [14], affective context of the group [15], and social power relations. This means we need to investigate to what extent our previous findings can be replicated in a group setting.

II. RoboTutor and Mood Expression Model

One of the ideas behind the RoboTutor application is to design a robot lecturer that takes itself as an example to introduce various aspects (e.g., sensors, effectors, and system) related to robotics. We chose to use the humanoid robot NAO and provided the robot with various capabilities such as changing slides in a presentation and asking multiple choice quiz questions. MS PowerPoint 2010 and TurningPoint version 5 (a plugin for MS PowerPoint) were used for presenting slides and quizzes. Students used TurningPoint clickers, which communicate wirelessly with the computer, to provide their quiz answers. They were asked to provide their answers within short intervals of time of about 10-15 seconds, and the time remaining to answer was indicated on the quiz slide. The robot responds to answers, and the percentages of students that selected a particular answer is shown on the quiz slide. If most students selected an incorrect answer, the robot provided additional details to explain the correct answer.

A script engine was designed to enable course instructors to orchestrate the robot gestures, speech, and slides by editing a text file (script). The robot speech was generated by a Text To Speech engine shipped with the robot. We designed co-verbal gestures based on our parameterized model. The script engine synchronizes the starting points of a sentence and its co-verbal gestures automatically, while we need to adjust the sentence length to guarantee the speech and gestures to finish roughly at the same time. The robot selected leg movements from a predefined corpus in real time and performed them between hand gestures, to avoid a long time of no movement.

Using our mood expression model, robot movements can be modulated to display the robot mood by changing behavior parameters with respect to both spatial extent and motion dynamics [5], [16]. The modulation of behavior parameters does not change the function associated with the behavior. Some parameters, such as speed and amplitude of a movement, are generic, and can be used to modulate arbitrary behaviors. Other parameters are associated with a particular body part of the robot such as head, hand palm, and finger. To express mood while performing a specific behavior, one only needs to specify which parameters should be varied. Thus, a range of affective states can be expressed, and affect can be displayed throughout a series of behaviors. We evaluated mood expression with the NAO in a laboratory setup without context [17] and in a game setting [12]. Results show that participants are able to identify correctly valence and arousal levels of a robot mood in both settings. To enable the RoboTutor to express mood while presenting a lecture, we applied our mood expression model to a range of co-verbal gestures. 41 gestures were parameterized using 12 parameters. The design principles for mood expression are obtained and refined from our previous study [5]. Decay-speed was used in [5] to control the speed of movements when robot actuators return to their initial poses. In this study, we used motion-speed as decay-speed because decay-speed was found to correlate with motion-speed in [16]. It was found in [17] that the parameters for motion-speed, hold-time (fluency), repetition, and head-vertical correlate with arousal. The modulated gestures thus do not only display the valence of the robot mood but also arousal. Videos of the modulated gestures that show stepwise changed mood and a video recording the robot during the experiment are available from our website, along with the source code, slides, and scripts.

III. Research Foci and Hypotheses

The main questions addressed in this study are (a) whether audiences can recognize positive and negative mood from robot body language integrated with co-verbal gestures during a lecture, and (b) whether expressed robot mood influences listener mood and perception of lecturing quality.

As previous studies showed that people are able to recognize robot mood, we hypothesize that in this study individuals that are part of the audience will also be able to rate the robot mood as positive/negative when the robot gestures are modulated as positive/negative.

H1. Participants rate a robot’s mood more positively when the robot’s gestures are modulated to display positive mood (“positive condition”) than when they are modulated to display negative mood (“negative condition”).

In previous work [12], we showed that mood can be transferred from a robot to a human in a dyadic interaction in a game setting. We are interested in establishing whether mood contagion also occurs during a RoboTutor lecture. In particular, we are interested in establishing whether contagion can be reproduced in a lecture and in a one-to-many interaction.

H2. Participants’ affective states are influenced by the robot mood: participants’ affective self-reports are more positive in the positive condition than in the negative condition; video annotation of the audiences should confirm this.

Mood induction has been reported to positively influence task performance [18], [19], [20], [14]. We are interested in whether mood induction caused by a robot would also have this effect: the students in a more positive mood (in the positive condition) are expected to answer the quizzes better than the students in a more negative mood (in the negative condition).

H3. Participants’ task performance (correctness of quiz answers) is significantly different in the positive condition from the negative condition.

Fortunato and Mincy [21] showed that induced positive mood increased students’ ratings of teachers. We therefore expect participants to give higher ratings for lecturing quality and gestures of the robot tutor in the positive condition than in the negative condition, resulting in our fourth hypothesis:

1 http://www.turningtechnologies.com/polling-solutions/turningpoint
2 http://i.tudelft.nl/SocioCognitiveRobotics/index.php/RoboTutorMood
H4. Participants’ ratings of the lecturing quality and gestures of the robot are higher in the positive condition than in the negative condition.

Other factors may also influence results. First, the distance and angle between the robot and participants may influence attention paid to the robot and the perception of body language [11], and social distance may influence people’s interpretation of, attitude towards, and preferred type of body language [22]. In our case, participants sitting close to the robot may recognize the robot mood better (an effect related to H1) and may be affected more by the robot mood (an effect related to H2), than those sitting further away. As a result, distance may also influence the participants’ task performances (H3) and the ratings of the lecturing (H4). Second, the attention a participant pays to the robot may mediate the mood contagion process and, as a result, influence task performance [14].

Third, there is evidence showing that social power (“dominance”) influences the contagion process [13], [23]. Affect is more likely to transfer from superiors to subordinates. In our case, the contagion effect may be stronger for participants that rate the robot as dominant.

IV. EXPERIMENTAL METHOD

A. Experimental Design

a. Independent Variables and Factors

The robot mood was manipulated as an independent variable at two valence levels: positive and negative. Note that the arousal is also different for the two mood levels (discussed Section II). The same lecture was given twice, once for each condition. That is, the script (see Section II), the lecture content (presentation and spoken text) and the types of gestures were the same for both conditions. Only the appearance of the gestures was modulated to vary the mood conditions. Slides did not contain too many details to prevent students from paying too much attention to the slides instead of to the robot. Participants were divided into two groups, and were assigned to a lecture in either the positive or the negative condition, making the experiment a between-subject design with one independent variable, the robot mood.

The distance and angle between robot and each participant was estimated geometrically (Fig. 1), according to the seat position (row and column) reported by students in the pre-experiment questionnaire. Experiments aligned the seats in a grid pattern (Fig. 1) and recorded the dimensions beforehand. The center of the rear edge of the desk (the solid circle in Fig. 1) was taken as the position of each participant, despite of their postures (e.g., leaning back on the chair, leaning forward on the desk).

b. Dependent Variables and Measures

Participants assessed the robot mood (valence, arousal, and dominance) using Self-Assessment-Manikins (SAM) [24] on 9-point Likert scales (H1), during a break in the middle of the lecture (T1) and at the end of the lecture (T2). This is because the scenario (lecture) lasts for a long time (about 30 minutes) and we anticipated effects caused by the robot mood expression to vary over time.

The affective states (valence and arousal) of participants were measured at the beginning of the lecture (T0), at T1, and at T2, using SAM on 9-point Likert scales. The difference between T1 and T0 and between T2 and T0 were taken as measures of the induced affective states (H2). By measuring participants’ affect and the perceived robot mood both at T1 and at T2, we obtain a measure for mood change over time. In addition, the lecture of each mood condition was video-recorded. Videos were manually annotated by two annotators (details in Section V) to assess the overall valence, arousal (H2), and attention distribution of the audiences.

The task performance was assessed using the answers to the quiz questions (H3). These questions, which are multiple-choice questions, relate to the lecture content taught by the robot just before the questions were asked. Each student was requested to provide answers independently.

The post-experiment questionnaires at T2 asked for participants’ ratings (H4) about the robot lecturer on 5-point Likert scales, using six items about the lecturing quality including 1) maintenance of participants’ interests, 2) maintenance of participants’ engagement, 3) enthusiasm, 4) friendliness, 5) maintenance of participants’ attention, and 6) overall satisfaction. The items were designed based on [25].

Moreover, we also asked participants to rate the robot gestures (H4) on 5-point Likert scales, using four items including 1) whether the robot communicated information by gestures efficiently, 2) whether the gestures helped the participants to follow the speech, 3) whether the robot organized speech, gestures, and slide switching well, and 4) whether the gestures were natural.

In addition, participants were asked to report their self-assessment of their own attention paid to the robot, slides, or other in percentages, to rate the consistency of the robot mood on a 5-point Likert scale, and to answer an open question about how they rated the valence, arousal, and dominance of the robot mood separately.
B. Assignments to Participants

First, participants were requested to listen carefully to the robot. They were informed that the lecture materials used by the robot were prepared by the course instructors, and that the content would be part of the exam. Second, they were requested to answer quiz questions that were posed by the robot and presented on the slides. We encouraged participants to obtain a high quiz score, by telling them beforehand that the first place of each group would receive a prize.

C. Participants

Participants were recruited from the students who enrolled in the Artificial Intelligence course at the Delft University of Technology. We asked the students to register, and asked permission from the students for this experiment in advance. They were told that each would receive a bonus course credit. 36 students registered, and all except one are master students. They were randomly assigned to each group, but we ensured that the background of participants (department and master program) was roughly equal within each group. 34 students (28 males and 6 females) whose ages range from 21 to 36 (Mean=23.8, SD=2.7) participated in this experiment: 18 of them for the positive condition; 16 for the negative condition. They came from 11 different countries, but 18 of them are Dutch. The pre-experiment questionnaire showed that some participants (12 in each group) had taken courses related to robotics such as “Humanoid Robots”, “Artificial Intelligence”, and “Computer Vision”, or attended projects related to robotics during their bachelor. Participants reported that they were open to technology-assisted education (Mean=3.941, SD=0.736, on a 1 to 5 scale).

D. Materials and Setup

A small lecture room that contains about 26 seats (with desks) was selected for the study. This setup has the advantage that participants sit relatively close to the robot so that they were able to notice the details of robot movements more easily and thus more likely to be influenced by the robot mood, even though we did not assign seats to participants (the experiment setup was identical to a usual classroom setup). The shutters of the window were closed. The screen for showing slides was located on the upper part of the wall behind the robot. The regular human lecturer (experimenter) sat in front of the classroom to protect the robot (e.g., from falling down) and to organize the experiment. Other experimenters were seated at the back of the classroom.

A grey NAO robot (NaoQi version 1.14, head version 4.0, body version 3.3) was used with LED lights switched off. The robot was connected with a laptop using cables, via a router and a gigabit switch, to guarantee sufficient speed of data transmission. The robot, which is 58 cm tall, was positioned on a desk (Fig. 1) while giving the lecture, which ensured that participants could see the robot by looking straight ahead.

Three video cameras were used for video data acquisition. Two cameras were placed on desks at the front of the classroom, on each side of the robot. Each camera recorded half of the classroom. The heights and angles of the two cameras were adjusted to guarantee that participants do not hide each other. The third camera placed at the back of the classroom (Fig. 1) was used to record the robot.

The course material (see footnote 2 for a link to the slides) was part of the course curriculum and designed by course instructors. It contains five parts: 1) an introduction to robotics, 2) the sensors, 3) the effectors, and 4) the programming of the NAO robot, and 5) an introduction to the RoboTutor system. The lecture included seven quiz questions. The first one was a trivial question for a warm-up exercise at the beginning of the lecture. The other six were used for assessing students’ performance. The second question about what a robot is was asked at the end of the part 1; three questions about sensors at the end of part 2, just before a break; the other two at the end of part 3 and 4 each.

E. Procedure

Before the lecture, experimenters aligned the seats and desks, measured and recorded the dimensions needed (Fig. 1), and set up the robot, laptop, projector, and cameras. Students (participants) were allowed to select seats, but not allowed to rearrange desks. Experimenters gave each student an experiment description and a TurningPoint clicker. Students were requested to fill in a consent form, a demographic form, a general questionnaire about previous experiences with robotics, and a pre-experiment SAM questionnaire to report their own affect (valence, arousal) before the start of the lecture (T0). An explanation sheet for valence, arousal, and dominance was provided to them. The human lecturer briefly described the experiment and answered questions. Students were told that they could not ask the robot questions during the lecture. We did not emphasize the robot mood or gestures to eliminate a demand effect.

The experiment started immediately after experimenters collected the pre-experiment forms from all students. Three experimenters started the camera recording manually, and another experimenter started the lecture program. The program started the PowerPoint presentation automatically, and sent the lecture script (Section II) to the robot. The robot then started to talk and perform gestures.

In the middle of the lecture (after part 2 of the slides), the robot asked participants to take a 5 minutes break and fill in the mid-term questionnaires (T1). Experimenters handed out the mid-term questionnaires (Section IV.A.b), but did not collect the questionnaires during the break to save time. The robot resumed the lecture after 5 minutes.

The whole lecture including the break took about 30 minutes. After the robot finished the lecture (T2), experimenters stopped the cameras. Experimenters first collected the mid-term questionnaires, and then handed out the post-experiment questionnaires (Section IV.A.b). After 10 minutes, experimenters collected all forms and questionnaires. The regular human lecturer explained the experiment in details to the participants, and requested them not to tell anything related to the experiment to the second group of participants. Experimenters checked whether all fields had been filled in when they collected questionnaires.

V. ANALYSIS AND RESULTS

A. Check of Participants’ Initial Affective States

As the data are not normally distributed, Mann-Whitney U tests were used to analyze the initial affective states of participants (at T0). The results showed no significant
difference in the self-reports of participants’ own valence and arousal between mood conditions. Thus, we assumed that participants in both groups had similar affective states at T0.

B. Perceived Robot Mood

Participants rated the robot valence, arousal, and dominance at T1 and at T2. As the data violated the normal distribution assumption, we illustrate the median values and interquartile range in Fig. 2. We compare the ratings of valence, arousal, and dominance between the positive and negative conditions using Mann-Whitney U test. The results did not show significant differences between the positive and negative conditions either at T1 or at T2.

C. Induced Affective States of Participants

The self-reports of participants’ own valence and arousal are not normally distributed. Friedman test and Wilcoxon Signed Ranks test were used to analyze for each mood condition whether participants’ affective states (H2) were different over time and how the affective states changed from T0 to T1 and from T1 to T2. Results showed the valence changed over time (\( \chi^2(2)=7.774 \), \( p=0.020 \)), and showed a significant increase of arousal at T1 (Med\(_{T0}=1\), Med\(_{T1}=2\), \( Z=-2.698, p=0.006, \) two-tailed) and a marginally significant drop at T2 (Med\(_{T2}=1\), \( Z=-1.931, p=0.066, \) two-tailed) in the positive condition only (Fig. 4 left) and only for arousal (Fig. 3 left and Fig. 4 left). This shows that there is mood induction in the positive condition.

To find out if this induction is significantly different between mood conditions, i.e., if positive mood induces significantly more arousal than negative mood, we calculated two induction measures by subtracting the T0 from T1 and T0 from T2 (Fig. 3 right and Fig. 4 right). Mann-Whitney U test was used to analyze the induction effect. Results showed that the induced arousal at T1 in the positive condition was significantly larger than in the negative condition (Med\(_{T1}\)neg=0, Med\(_{T1}\)pos=1, \( U=73.5, Z=-2.486, p=0.012, \) two-tailed), and the induced arousal at T2 in the positive condition was larger than in the negative condition at a marginally significance level (Med\(_{T2}\)neg=\(-1\), Med\(_{T2}\)pos=1, \( U=91.5, Z=-1.857, p=0.064, \) two-tailed).

The videos of each condition (29 minutes) were annotated by two experienced annotators to assess participants’ valence and arousal on a 9-point Likert scale, and attention distribution (rank order of robot, slides, and other). We employed the interval coding method: the annotators assessed the overall valence, arousal, and attention of the audiences as a whole from every one minute of the video. Spearman’s rho was used to test the inter-coder reliability. Results showed that the correlation is significant and has a large effect size: rho=0.675, \( p<0.001 \) for valence, and rho=0.511, \( p<0.001 \) for arousal. It indicates a strong consistency between the two coders. Thus, we took the average of the ratings of valence and arousal per minute from both coders as final ratings for valence, \( U=107.0, Z=-4.962, p<0.001, \) two-tailed for valence, \( U=163.5, Z=-4.052, p<0.001, \) two-tailed for arousal. The results indicate that the robot mood expression had an effect on both the valence and arousal of the audiences. In addition, both coders reported that there was more laughter in the positive condition, and only in the positive condition the participants applauded at the end of the lecture.

D. Task Performance

As the data of the answers to quiz questions (2-7) is not normally distributed, Mann-Whitney U test was used to analyze the difference between mood conditions. Results showed no significant difference between the positive (Med=4) and negative (Med=4) conditions. That is, we did not observe an effect of robot mood on participants’ task performance in terms of question answering (H3).

E. Perception of Lecturing and Gesture Quality

The average ratings of the items about lecturing quality of the robot and the items about the robot gesture quality are illustrated in Fig. 5 and Fig. 6. There is a trend that the ratings of all items about the lecturing quality are higher in the positive condition than in the negative condition (Fig. 5).
Independent t test showed that the sum of the ratings about the lecturing quality in the positive condition were significantly higher than the negative condition: \( t(32)=2.210, p=0.034 \). That is, the participants thought that lecturing quality of the robot was higher in the positive condition.

For the ratings of the robot gesture quality, we excluded an outlier (lower than the lower inner fence of the boxplot) from analysis. There is a trend that the ratings of all items about the gesture quality are higher in the positive condition than in the negative condition (Fig. 6). Independent t test showed that the sum of these items in the positive condition were higher than in the negative condition at a marginal significance level: \( t(31)=1.920, p=0.064 \). This suggests that participants thought that the robot gesture quality was higher in the positive condition.

**F. Attention, Dominance, Distance and Angle**

We checked for influences of attention, distance, angle, and perceived robot dominance. We did not observe a significant difference of the attention between conditions from \( t \) tests based on self-reports. According to these self-reports, participants in the positive mood condition paid attention to the robot for 51.94% (SD=15.82%) of the time and 38.61% (SD=16.34%) to the slides on average, and participants in the negative mood condition paid attention to the robot for 48.75% (SD=18.66%) of the time and 40.00% (SD=16.23%) to the slides on average. This is consistent with video annotation of attention distribution. Mann-Whitney U test did not show significant differences between mood conditions for attention focus annotated from videos (inter-coder reliability was strong for attention to robot, \( \rho=0.635, p<0.001 \), and for attention to slides, \( \rho=0.605, p<0.001 \)). In addition, correlation analyses did not show significant relations between distance, angle, or perceived robot dominance on the one hand, and recognition of the robot mood or induced participant mood on the other.

**VI. DISCUSSION**

Participants’ self-reported SAM ratings about the robot mood (H1) did not show a clear difference of the perceived robot mood between conditions. Thus, H1 is rejected. It was discussed in [26] that participants would not spontaneously pay attention to another person’s expression if they were not required to. In our case, we did not prime the participants to pay attention to the robot mood expression in advance, and the analysis of participants’ attention distribution showed that participants paid almost 40% of their attention to the lecture slides in both conditions. This may explain the absence of the significant differentiation of positive and negative mood. Another possible explanation for the absence of a significant difference in reported robot mood is that participants attributed the gesture modulation to teaching quality directly instead of to robot mood. Last, open participant feedback showed that variation in robot mood was attributed to various factors that were not manipulated such as tone and volume of the robot voice, speaking speed, lecture pace, and gaze/eye-contact. This may also explain the absence of a significant difference in reported robot mood: these factors disturbed the conscious recognition of the robot mood through the gesture modulation. Additional work is needed to check how these factors would influence our results.

Interestingly, participants’ self-reports show that the robot expression influenced participants’ arousal, and objective assessment in the form of video annotation revealed that both valence and arousal of participants’ mood were influenced. This means that the mood of the audiences (students) was influenced. It is known that mood contagion can happen automatically or subconsciously [26], [14]. This study showed that automatic mood contagion could also occur from a robot to audiences. We thus did find support for H2.

There is no significant difference in the performance of quiz answering between the positive and negative conditions. That is, H3 is rejected. A possible reason could be that the answering to in-class quizzes is primarily course content oriented. The performance of the quiz answering is less influenced by lecturing quality, i.e., in our case, presenting the same course content with different "moody" gestures. Even so various studies (e.g., [27], [28]) support that a moderate arousal may increase learning performance.

The ratings of the robot in terms of lecturing quality and the gesture quality were influenced by the robot mood. Although each individual item did not show significance, which may be due to the small sample size, trends towards perceiving both qualities higher in the positive condition are clear and the sum of those ratings showed significance or marginal significance. Thus, there is support for H4.

**VII. CONCLUSION AND FUTURE WORK**

To the best of our knowledge, this study is the first to investigate the body language of a humanoid robot in an interaction scenario of one robot versus multiple people. Our study is unique in that a) mood expression was realized by
modulating parameterized functional behaviors, b) robot mood expression was evaluated and studied outside a lab, c) the participants were not primed to pay attention to any form of affective expression, and d) the experiment is a between-subject design, which means that the goal is to test whether participants can recognize the robot mood, rather than differentiate between mood levels, as, e.g., [12].

This study shows that robot mood expression integrated with functional behaviors can influence the affective states (valence and arousal) of the persons who interact with the robot in a more or less real life application. Although it was not clear whether participants (consciously) recognized the robot mood, results show effects of the robot mood expression on the participants: valence and arousal, and lecturing quality ratings of the robot were higher in the positive condition than in the negative condition. These results indicate that robot mood expression can be used to shape human-robot interaction affectively in a real life context. These findings signify the value of robot bodily expression in robot-enhanced education, on the basis of the arousal-learning studies, e.g., [27], [28]. Future work is needed to provide support for the suggestion that the mood of a robot teacher may affect learning. Finally, we cannot completely rule out that the robot body language had no effects on the perception of naturalness, friendliness, or sociability that may have influenced our results. Further study is needed to address these aspects.

Through this study, we identified several aspects that we would like to address in future work. First, as mood is a long-term and stable affective state, expressing a constant mood via multiple behaviors is important. Although participants’ ratings of the consistency of the robot mood is acceptable (Mean=1.1, SD=0.9 on a -2 to 2 Likert scale), modeling a constant mood remains a challenge. Second, a mechanism that coordinates the modulation of temporal parameters and the synchronization between gestures and speech is needed. In this study, temporal parameters were manually adjusted to align gestures with speech. A model is needed to align the timing of speech and gestures automatically. Third, we deliberately limited the levels of positive and negative mood expression to ensure body language would remain acceptable within context of a lecture. It is interesting to explore other settings that allow for a broader range of mood expression. Finally, an ongoing effort is put into making the Robotutor more responsive, to provide students with opportunities of active interactions, such as asking questions by raising hands. Different findings may be obtained when students are more actively involved in the class. Moreover, their interaction pattern may be different in different robot mood conditions, and thus can be used as a behavior measure to evidence the effects of the robot mood.

REFERENCES


