Mina Marmpena*, Angelica Lim, and Torbjørn S. Dahl

How does the robot feel? Perception of valence and arousal in emotional body language

https://doi.org/10.1515/pjbr-2018-0012
Received November 30, 2017; accepted May 29, 2018

Abstract: Human-robot interaction in social robotics applications could be greatly enhanced by robotic behaviors that incorporate emotional body language. Using as our starting point a set of pre-designed, emotion conveying animations that have been created by professional animators for the Pepper robot, we seek to explore how humans perceive their affect content, and to increase their usability by annotating them with reliable labels of valence and arousal, in a continuous interval space. We conducted an experiment with 20 participants who were presented with the animations and rated them in the two-dimensional affect space. An inter-rater reliability analysis was applied to support the aggregation of the ratings for deriving the final labels. The set of emotional body language animations with the labels of valence and arousal is available and can potentially be useful to other researchers as a ground truth for behavioral experiments on robotic expression of emotion, or for the automatic selection of robotic emotional behaviors with respect to valence and arousal. To further utilize the data we collected, we analyzed it with an exploratory approach and we present some interesting trends with regard to the human perception of Pepper’s emotional body language, that might be worth further investigation.

Keywords: social robots, human robot interaction, dimensional affect, robot emotion expression

1 Introduction

Humanoid robots are increasingly populating social environments like homes, schools and commercial facilities. As a consequence, their ability to interact with humans in a social way is becoming increasingly important [1]. The ability to mimic emotional expressions is an essential feature for a socially interactive robot [2] and appropriate displays of affect can drastically augment engagement on the side of the user during interaction [3]. Moreover, a robot can employ displays of virtual emotion as a means to provide feedback about its internal state, goals or intentions [2], e.g., an expression of fear could be triggered by the robot as part of a self-protection mechanism [4]. Robots with the ability to mimic emotional expressions could be employed in various human-robot interaction (HRI) settings and applications, be it human-robot collaboration tasks, robot-assisted therapy with kids or elderly, education, or entertainment.

Body language and in general non-verbal cues like gestures and non-linguistic speech are indispensable in the communication of emotion. The integration of such social traits in robot behavior during HRI can enhance the user experience and improve the collaboration between humans and robots.

From a holistic point of view, robotic emotion expression can be seen as a component of an affective loop [5], that integrates successive stages of human emotion detection, emotion synthesis (parameterized with external input signals and perhaps other internal parameters), and finally emotion expression that is based on the emotion synthesis output. All these functionalities can be implemented as independent modules, but the exchange of information between them could be potentially benefited by a continuous representation of high sensitivity that captures many subtle variations of input and output.

With that in mind, we conducted an initial experiment to collect data and knowledge that can help to structure an emotion expression module. We used a set of high quality animations (context-free, non-verbal, dynamic expressions of emotion) that were designed by expert animators for the Pepper humanoid robot to convey emotions. We
had them assessed as part of a user-experience setup that also provided us with insights on human perception of robotic emotion expression, and finally, after performing a reliability study, we derived the affect labels of valence and arousal, in a continuous two-dimensional space, that can potentially allow for higher sensitivity of the labels with respect to subtle variations of the emotions conveyed.

To the best of our knowledge, this is the first effort to assign such labels to emotional body language animations for a humanoid robot, at least in terms of the reliability study support and high discretization of the affect space. Such a dataset could potentially be re-used by robotic application developers to select outgoing emotional expressions based on continuous input signals, or even as a ground truth from researchers that would like to study different setups related to human perception of robotic emotion expression.

The paper is organized as follows. Section 2 describes different approaches to representing emotions as well as our motivation for selecting a dimensional and continuous space representation. This section also reviews related work on emotion annotation and human perception of robot expressions of emotion. Section 3 describes the experiment we conducted. Section 4 presents the analysis of the resulting data and the derivation of the labels. Section 5 discusses the results, limitations and future work.

## 2 Representing emotions

Emotion modeling research has been dominated by three major approaches in emotion representation: the categorical, the dimensional, and the componential (appraisal-based). According to the categorical approach, there are at least six basic emotion modules, that are discrete and separate states, different not only in expression, but also in other fundamental aspects (e.g., appraisal and physiological reaction) [6]. The dimensional approach postulates that affect states can be represented as points in an underlying affect space of two or more dimensions and thus emotions are not independent, but instead they are related forming a synergistic phenomenon. According to the componential approach, emotions as the results of evaluations (appraisals) of external stimuli (events) filtered by the internal state (goals and beliefs) of the subject [8, 9].

Significant efforts have been made to compare their benefits and limitations [10–13], but the question of how to select an appropriate model for a given problem remains open. In our present study, our main objective was to derive affect labels of valence and arousal, in a continuous and highly discretized space that would be intuitive and convenient to use from an engineering point of view. Such labels are potentially more effective in capturing different subtle variations of the emotional state [14], than categorical tags that aggregate many expressions in a set number of categories. Additionally, the dimensional approach is not affected by the variability of the individual semantic interpretation of the categorical labels.

Therefore, in our present work, as a basis for an emotion representation, we adapt the dimensional theories of emotion and more specifically the Circumplex Model and the core affect concept [7]. Russel and Feldman describe core affect as the conscious affective feeling, that is always present along with its neurophysiological correlates, but is nevertheless non-reflective and not directed to something, as opposed to the prototypical emotional episode, the full-blown emotion, which encapsulates core-affect but also involves more complicated cognitive structures that emerge from the property of directionality. According to the Circumplex Model, emotional states can be represented on a two-dimensional surface defined by a *valence* (pleasure/displeasure) and an *arousal* (activation/deactivation) axis. The two-dimensional approach has been criticized on the grounds that subtle variations between certain emotions that share common core affect, e.g., fear and anger, might not be captured in less than four dimensions [15]. Nevertheless, it can provide a fair representation of non-directional core affect which is considered as a valid low-dimensional component of the full-blown emotion [16].

### 2.1 Representing emotional body language in robots

Affective body expressions are very important for conveying and perceiving emotion [17], especially for humanoid robots with constrained facial expression. The problem of mimicking expressions of emotions in humanoid robots has attracted much attention in the recent years and several efforts have been made to design effective expressions and test their impact on humans. However, it is difficult to juxtapose different efforts due to differences in the design principles adopted, the robotic platform level of embodiment, the modalities involved in the expressions, the static or dynamic nature of the expressions, and the evaluation methodology.

From the design point of view, a prominent approach is to model expressions of emotion on how humans use body language to express themselves emotionally. Sources of such information can be found in studies coding be-
behavior patterns of movement or posture from recordings of actors performing emotional expressions [18–20], computer generated mannequin figures [21], or human motion capture data [22]. These studies provide lists of validated discriminative movement indicators, such as body orientation, symmetry of limbs, joint positions, force, velocity, etc., that can be employed for the design of robotic body language expressions of emotions.

Another influential design approach derives from applying the artistic principles and practices used in cartoon animation to generate robot expressions of emotion [23]. Monceaux and colleagues [24] describe their methodology of creating a big library of dynamic multi-modal emotion expressions as an iterative process involving imitation from direct observation, and abstraction of key positions to adjust them to the Nao robot morphology.

Previous research shows that humans can recognize the emotions expressed through a humanoid robot’s body language. An evaluation conducted using static body postures and head positions representing the six basic emotions on a Nao robot by Beck et al. [25], showed that such cues can be interpreted at better than chance levels. In the same experiment, valence, arousal and stance were also rated by the participants in a 10-point Likert scale and it was found that the head position had an effect on these ratings for all basic emotions except fear.

Tsiouri et al. [26], tested three basic emotions (happy, sad and surprised) with animated body expressions on a Pepper robot. The evaluation followed a categorical approach. Happiness was recognized more accurately, while sadness was poorly recognized. According to the authors, this effect could be explained by the robot’s facial configuration, and mainly the mouth that is designed to resemble a smile. Furthermore, the same authors tested separately the expressiveness of sound, but in this case too, the recognition of sadness remained lower than the recognition of happiness.

Destephe et al. [27, 28] used motion capture features to transform a humanoid robot’s (WABIAN-2R) walking pattern so that it conveys four basic emotions (happiness, sadness, anger, fear) and neutral, as well as three different intensities: intermediate, high, exaggerated. The authors achieved a high recognition rate for the emotions (72.32%). The intensity recognition rate though was not as high (41.07% for intermediate intensity was the highest), but they found that it modulates the emotion recognition rate (higher intensity results to higher emotion recognition rate).

Animated expressions of emotions have also been evaluated on a Nao robot by Härning et al. [29], again with a modality perspective. Two different versions of four basic emotions; anger, fear, sadness and joy, were used to create expressions based on body motion, sound, and four different eye LED colors. All three modalities were tested separately. The evaluation was dimensional, including pleasure, arousal and dominance (the PAD model). The results did not show any significant effect for the eye LED color modality, half of the sounds used were categorized appropriately, and all the body movements were rated in the right octant of the PAD model except sadness which was rated with positive arousal.

Embgen et al. [30] tested the recognition rate for three basic emotion (happiness, sadness, fear) and three secondary ones (curiosity, embarrassment, disappointment) and they found that participants were able to identify them when displayed by the humanoid robot Daryl using motion, color and lights.

In an approach that explores how professionals from performing arts can influence the design and perception of emotional body language, Li et al. [31] tested two groups of basic emotions expressions, one designed by professional puppeteers and one from engineers. The expressions were designed for RobotPHONE, a robot with the size and form of a teddy bear. For some of the emotions (fear, disgust), the puppeteers group succeeded to get better recognition rates, but the overall recognition was low.

Another interesting approach for analyzing features of body language and designing robotic expressions draws inspiration from Laban Motion Analysis methodologies, a framework initially conceived for the performing arts [32]. More often a subset of features is used, the Laban Effort System, that can represent how movement is performed with respect to the inner intentions with four parameters: space, weight, time, and flow. These parameters have been used to handcraft features and generate locomotion trajectories that express affect for non-humanoid robots with low degrees of freedom [33–36], with promising results in terms of readability by humans. A similar approach but for a humanoid robot is described in [37], where Masuda et al. use Laban Effort System features to modify three basic movements so that they convey emotion. Their evaluation, shows successful rates of recognition for some of the modulated movements, and with regard to the emotions, sadness appeared to have the highest modulating effect.

Finally, there is an important body of work that is concerned with the generation of emotional expressions based on computational architectures. Since the animations we used for this study were not created this way, we will not discuss this body of work in detail, but we will refer to one approach to give a sense of the diversity in the field. The SIRE model [38] was proposed for generating robot expressions of happiness, sadness, anger and
fear based on speed, intensity, regularity and extent. These features were extracted from human expressions, in particular vocal expressions. The model was evaluated on a Nao robot using the dimensional PAD model, and results showed successful recognition of happiness and sadness. The SIRE model was also adapted in another study to produce expressions of emotions on a Nao robot that would offer movie recommendations [39]. This Nao was found preferable and more persuasive than a Nao without emotional expression.

The studies and results we presented above are very informative for advancing the field of robotic emotional body language, especially from the aspect of proposing and evaluating features and design principles that have an effect in the readability of emotions. However, most of the studies focus on the categorical approach using the basic emotions, both for the design phase and the evaluation. This is, of course, a legitimate choice since the debate between different representations of emotion is still open. Nevertheless, we believe that further study of the dimensional approach can also contribute to the body of knowledge.

On the other hand, some of the related work we presented used dimensional models for the evaluation phase, but again, the main objective in these approaches was to propose design principles, and evaluate them in coarse classes defined by the combination of two or three levels from each dimension (e.g., +/- Pleasure, +/- Arousal, +/- Dominance in [29]). Our objectives lie more on the side of creating a reliably labeled set of animations that can be reused as the ground truth for behavioral experiments or for defining automatic outgoing robot behaviors in applications. For that reason, we want to exploit the properties of the dimensional models which due to their continuous space representation can allow for fine-grained differentiation, even between expressions that only slightly vary. Therefore, we use highly discretized scales to collect our evaluation results, and we conduct a reliability analysis to aggregate the evaluations we collected into labels for the rather diverse set of animations we use.

The work we present here does not propose a new approach for designing robot expressions of emotion, since we are using pre-designed dynamic emotion expressions selected from the Softbank Robotics animation library for the Pepper robot. These animations have been created by professional animators in a category-oriented way to convey emotions, but to the best of our knowledge, they have not been formally evaluated, and they are only labeled with subjective categorical tags. With this choice, we want to take advantage of the professional animators expertise to create a baseline set of reliably labeled animations with advanced expressivity that can be reused.

3 Methods and materials

3.1 Participants

The study included 20 volunteers, 9 women and 11 men with a mean age of 23.6 years and a standard deviation of 4.09. The youngest participant was 19 and the oldest 35. The participants were mainly students at the University of Plymouth, UK. Nine participants were British and the rest were European nationals studying in the UK. The participants had minimal or no experience with the Pepper robot. The experimental protocol was granted ethical approval by the University of Plymouth, Faculty of Science and Engineering Research Ethics Committee, and written informed consent was obtained from all participants prior to the study. The participants were reimbursed £5 for their time.

3.2 Robot platform and hardware

For this study we used the Pepper robot, a humanoid robot created by SoftBank Robotics. The Pepper robot is 120 cm tall, weighs 28 kg and it has 20 degrees of freedom, including a head, two arms and a wheeled base. The operating system running on the robot was NAOqi 2.5.5.5 and the experiment used the NAOqi Python SDK to control the robot’s actions. The participant interface was implemented as a web application with Django and presented to the participants on a touch screen.

Table 1: Nine classes reflecting possible combinations of valence-arousal levels.

<table>
<thead>
<tr>
<th>Valence</th>
<th>Arousal</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Excited</td>
<td>Neutral</td>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>Neutral/Excited</td>
<td>Neutral/Excited</td>
<td>Positive/Excited</td>
<td></td>
</tr>
<tr>
<td>Calm</td>
<td>Negative/Calm</td>
<td>Neutral/Calm</td>
<td>Positive/Calm</td>
<td></td>
</tr>
<tr>
<td>Tired</td>
<td>Negative/Tired</td>
<td>Neutral/Tired</td>
<td>Positive/Tired</td>
<td></td>
</tr>
</tbody>
</table>
3.3 Animations

The set of emotional expressions used for this study consisted of 36 animations designed by Aldebaran/SoftBank Robotics animators. Each animation has a different duration (from 1.96 to 10.36 s, mean = 4.39 s), and consists mainly of body motion without locomotion. Some of the animations involved additional interactive modalities, in particular eye LED color patterns, sound or both. Essentially, each animation is a set of key body poses (frames) with their timestamps, and then intermediary frames are generated with interpolation. The interpolated frames are executed with a speed of 25 frames per second, but depending on the number of key frames, each animation appears to have different speed. For a more detailed description of the animations, please refer to Appendix A, Table 4, which presents the categorical tag of each animation assigned by the animators during the designing phase, a short description of the movements that comprise each animation, the duration in seconds, the speed calculated as total frames divided by key frames, and the different modalities (motion, sound, LEDs). A video depicting them is available online.

The selection was done from the broader animation libraries for Pepper, with the aim to have a good spread of the animations across the affect space defined by valence and arousal. We predefined nine different classes of valence/arousal combinations presented in Table 1, and we selected 4 animations for each class. The selection was conducted based on the display of the animations by the authors and a professional animator involved in the creative process of designing emotion animations for Pepper, i.e., researchers highly accustomed to robot’s motion and an expert in expression of emotions with animations for the particular platform. Although we aimed to have a good representation for each class, the selection has been subjective and the pre-assignment in the nine classes is not considered as a ground truth. Maximizing the coverage of the affect space with our pre-assignment has been a desirable property, and we do look into how the preassigned classes match the final ratings, but the final ratings provided by the participants were validated independently of the categorical tags.

3.4 Ratings interface

After carefully considering the available options for collecting valence and arousal ratings, we decided to use the Affective Slider [40], a modern digital tool that has been validated as a reliable mechanism for the quick measurement of the emotion dimensions of valence and arousal. The Affective Slider enables the collection of ratings using continuous scales with customizable discretization. This study used scales from 0 to 1, with a step of 0.01 resulting in a resolution of 100 points. Furthermore, the Affective Slider is an intuitive interface that does not require significant training.

For our experiment we replicated the tool as part of a Django web application based on the design guidelines recommended by its authors. We integrated it into the experimental interface along with a play button for the participants to control the robot, and a submit button for them to indicate that the sliders had been set. The interface is presented in Fig. 1.

---

1 Animation set video: https://www.labtube.tv/video/dimensional-affect-ratings-for-emotion-animations-performed-by-a-humanoid-robot
3.5 Questionnaires

We administered an affect measuring questionnaire [41] ahead of the subject’s interaction with the robot in order to indicate the subject’s positive and negative mood at the time of the experiment. Furthermore, two five-point Likert scales (from 1 = "Not confident at all" to 5 = "Very confident") were presented for the participants to indicate their level of confidence in their arousal and valence assessment after each trial.

3.6 Experimental procedure

3.6.1 Instructions and training

Initially the participants were provided with an information sheet and asked to give their written consent if they have no objections. After registering their demographic data (age, sex and nationality), they were seated in front of the robot at a distance of two meters with the touchscreen right in front of them. This setup is shown in Fig. 2.

The participants’ first task was to complete the affect measuring questionnaire presented on the touchscreen. Consequently, the experimenter gave a brief explanation of the main task. Participants were told that they had to rate the robot’s emotional expression in terms of valence, referred as "pleasure" during the experimental procedure, and arousal. Valence was defined as "how positive, neutral or negative the robot’s expression was" and arousal as "how excited, calm or tired the robot appeared". The participants’ responses were solicited by the question "How does the robot feel?". After this introduction, the participants went through a training session of three trials during which the experimenter was interacting with them to make sure they fully understood the task and the concepts of valence and arousal.

3.6.2 Main session

During the main session of the experiment, the participants were presented with the Affective Sliders interface and a play button for each trial. First, they would click on the play button for the robot to perform the animation, and then submit their ratings. The participants were told that they could replay the animation if they needed to. When the participants clicked the submit button, the sliders would be removed and two Likert scale questions would appear requesting them to indicate how confident they had been in their valence and arousal assessments.

When these questions had been answered, the next trial would begin. The main session consisted of 39 trials; 36 original animations plus three repetitions. The order of the original animations was randomized for each participant to avoid order effects, and the first three animations were repeated at the end of each participant’s session in order to evaluate intra-rater consistency. Moreover, the two Affective Sliders were presented in different order between trials. Between trials the robot resumed its initial, neutral, motionless standing posture, with eye LEDs switched off, so that participants could clearly perceive the onset and offset of the animation.

4 Results

4.1 Affect questionnaire

The momentary affect scores from the PANAS questionnaire administered to the participants before the experiment were within the expected range [41], so our ratings analysis did not exclude any of the participants. Positive affect mean was 31.75 (SD = 6.27) and Negative affect mean was 12.75 (SD = 3.16).

4.2 Descriptive statistics

The descriptive statistics by gender are presented in Table 2. An independent sample t-test did not reveal any statistically significant gender differences (valence: t(18) = −0.62, p = .53, arousal: t(18) = −0.68, p = .49).

Table 2: Descriptive statistics by gender. Ratings range from 0 to 1 with 100 points resolution.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Gender</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arousal</td>
<td>Women</td>
<td>0.60</td>
<td>0.29</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Men</td>
<td>0.61</td>
<td>0.27</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Group</td>
<td>0.61</td>
<td>0.28</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Valence</td>
<td>Women</td>
<td>0.49</td>
<td>0.28</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Men</td>
<td>0.49</td>
<td>0.24</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Group</td>
<td>0.49</td>
<td>0.26</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
4.3 Exploratory analysis

In terms of the initial pre-assignment of the animations in nine different classes of valence/arousal combination levels, we plotted the means and the standard deviations of the collected ratings per class in Fig. 3, and the detailed descriptive statistics are presented in Table 5 of Appendix B.

Since the animations comprising each class were subjectively chosen as described in Subsection 3.3, there is no ground truth for the class centers (mean ratings per class). Interestingly though, the figure shows that the class centers appear in the order expected according to the pre-assignment (see Table 1 for the order). However, almost half of the class centers are not spread out in the affect space as we had expected with the pre-assignment. For example, all the "Tired" class centers are pushed to the "Calm" subspace with respect to arousal, and to the "Neutral" subspace with respect to valence. The rest of the class centers appear in the expected levels of the affect space. It appears that the animations preassigned as of low arousal were in average perceived as of medium arousal and their valence as neutral. For a visualization of the descriptive statistics of the animations within each class, see the box-plots in Appendix B, Fig. 7.

Fig. 4 shows all the ratings plotted in the valence-arousal space. The sparsity of data points in the low arousal subspace is more pronounced in the high valence area. Moreover, an accumulation of ratings can be observed in the area of neutral valence (valence = 0.5). More specifically, 98 ratings out of 720 in total are accumulated at 0.5 valence. The same is not true for arousal (only 33 ratings out of 720 were found at 0.5). While for both variables equal to 0.5 (origin of the space), only 10 out of 720 ratings were found. To understand this effect better, we plotted the kernel density estimates (KDE) for the ratings with respect to the five different levels of raters' confidence in Fig. 5.

The KDEs in the valence graph in Fig. 5 show a peak at 0.5 (neutral valence) for the ratings that were submitted with low confidence, i.e., from 1 to 3 ("not confident at all" to "somewhat confident"). This trend appears clearer for male raters in confidence levels 1 and 3, and for female
raters in confidence level 2. No such accumulation is detected for arousal.

With respect to the arousal graph in Fig. 5, the KDEs are negatively skewed for all confidence levels for both male and female raters. The negative skew increases with confidence, suggesting that the effect of sparse ratings in the low arousal subspace is not a result of difficulties in interpreting the animations.

### 4.4 Arousal is easier to rate than valence

Furthermore, we examined for statistical differences in raters’ confidence levels between valence and arousal. Our hypothesis, was that raters’ confidence would be higher for arousal than valence, considering valence as harder to interpret on a robot with constrained facial expression. A one-sided Wilcoxon Signed Rank Test supported this hypothesis by showing that the confidence in the arousal ratings \( z = -1.90, p = .03 \) with a medium effect size \( r=0.3 \).

### 4.5 Intra-rater reliability

To test intra-rater consistency, we repeated the presentation of the three first animations at the end of the main session and we measured the intra-rater mean squared error (MSE) and the intraclass correlation.

#### 4.5.1 Intra-rater reliability results

The average MSE across 20 participants was 5.26% (SD = 5.96, 50th percentile = 2.87, 95th percentile = 14.13) for arousal, and 3.97% (SD = 4.42, 50th percentile = 2.91, 95th percentile = 10.24) for valence. Hence, MSE was less than 25% in all cases and less than 15% in 95% of the cases.

The intraclass correlation coefficient (ICC) for intra-rater reliability was estimated as a 2-way mixed-effects model with the definition of absolute agreement since we are testing a paired group. The formula for the estimation is in Eq. 1:

\[
ICC(2, k) = \frac{MS_R - MS_E}{MS_R + \frac{MS_C - MS_R}{n}}
\]  

where \( k \) denotes the number of repeated samples, \( MS_R \) is the mean square for rows, \( MS_E \) is the mean square for error, \( MS_C \) is the mean square for columns, and \( n \) is the number of items tested [48].

The result was \( ICC = 0.72 \) with 95% confidence interval 0.52-0.83, and \( F(59, 60) = 3.5, p < .001 \), which is interpreted as moderate to good reliability. We discuss more details about the interpretation of the ICC coefficients in the next section.

### 4.6 Inter-rater reliability

To assess the inter-rater reliability among participants we used intraclass correlation, one of the most widely-used statistics for assessing ratings reliability [42, 46], applicable to ratings of interval type. For fully crossed design, i.e., each item (animation) was rated by all the participants, ICC can provide a good estimation of the reliability with adequate control of systematic bias between our coders [47].

The selection of the appropriate ICC form depends on three different parameters; the model, the type and the definition of ICC [48]. The model we have chosen is the two-way random effects model, since we randomly selected our raters from a larger population and we want to be able to generalize the results to a larger population with similar characteristics. The type of ICC is the mean of the ratings, since we are going to use the aggregated ratings. Finally, the definition of the ICC form is that of consistency (instead of agreement) since we are not interested in absolute agreement between the raters, but on how well do the ratings correlate in an additive manner. The formula is given in Eq. 2 and is defined for consistency instead of agreement:

\[
ICC(3, k) = \frac{MS_R - MS_E}{MS_R + MS_E}
\]  

where \( k \) denotes the number of the raters, \( MS_R \) is the mean square for error. The ICC estimates we present in the results section were calculated using the function icc (parameters: “twoWay”, “consistency”, “average”) from ‘irr’ package of R (v3.4.1, CRAN, http://CRAN.R-project.org/). Finally, for the interpretation of the results we are following Koo et al. [48]: poor for values < 0.5, moderate for values between 0.5 and 0.75, good for values between 0.75 and 0.9, and excellent for values > 0.9. The evaluation thresholds are applied on the 95% confidence intervals and not the ICC value itself.

#### 4.6.1 Inter-rater reliability results

The resulting ICC (Table 3) for the whole group of the participants is in a high range for both valence and arousal.
Table 3: Intra-class correlation coefficients indicate the inter-rater reliability for: a) the whole group, b) female and male participants, and c) low and high confidence in the judgment ($p < .001$).

<table>
<thead>
<tr>
<th></th>
<th>Intra-class Correlation</th>
<th>95% Confidence Interval</th>
<th>F Test With True Value 0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lower Bound</td>
<td>Upper Bound</td>
</tr>
<tr>
<td>Group</td>
<td>0.95</td>
<td>0.93</td>
<td>0.97</td>
</tr>
<tr>
<td>Female</td>
<td>0.91</td>
<td>0.86</td>
<td>0.95</td>
</tr>
<tr>
<td>Male</td>
<td>0.92</td>
<td>0.87</td>
<td>0.95</td>
</tr>
<tr>
<td>Low Conf.</td>
<td>0.81</td>
<td>0.58</td>
<td>0.94</td>
</tr>
<tr>
<td>High Conf.</td>
<td>0.98</td>
<td>0.95</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Figure 6: The final set of valence and arousal labels derived from the means across raters. They are plotted with the original categorical tags of the animations (given by the animators in the creation phase) and color-coded according to our pre-assigned classes. (lower confidence interval is > 0.9) indicating high consistency and that raters perceived the core-affect expressed by the animations similarly. We also estimated the ICC for males and females, as well as with respect to the confidence in the ratings, by taking the ratings of the 10 highest in confidence and 10 lowest in confidence animations. In the case of the lower confidence ratings, ICC is lower, indicating lower consistency among raters compared to that of higher confidence, but still in acceptable range.

The Feldt test [43, 44] for statistical inference between the ICC coefficients for female and male participants did not reveal any statistical differences (valence: $F(1, 18) = 0.03$, $p = .88$, arousal: $F(1, 18) = 0.08$, $p = .75$). Applying the same test for paired samples revealed significant differences between ICC coefficients for low and high confidence in the ratings (valence: $t(18) = 17.26$, $p < .001$, arousal: $t(18) = 9.38$, $p = .001$). The results for the Feldt test were obtained with R (v3.4.1, CRAN, http://CRAN.R-project.org/), and the cocron package [45].

Supported from the inter-rater reliability results, we derive the final set of the aggregated annotations by taking the mean across raters for each animation. The final anno-
tations are plotted in the valence-arousal space in Fig 6, along with the original tags of the animations (assigned by the animators during the creation phase) and color-coded according to the pre-assigned class. Detailed statistics and raters' confidence levels for the final annotation set can be found in Appendix C, Table 6.

5 Discussion

The work presented in this paper used a set of pre-designed, non-verbal animations created in a category-oriented way for the Pepper robot to convey emotions. Our objective was to annotate the animations with continuous interval, dimensional affect labels of valence and arousal in order to facilitate the use of the animations to convey emotions in robots. To validate the collected ratings we tested the inter-rater reliability with intra-class correlation which was found in an excellent range (ICC > 0.90) for both valence and arousal. This result indicates very high consistency among raters and shows that the different animations are perceived similarly.

The intra-rater MSE was below 23.8% for all participants and below 14.13% for 95% of the participants. Intrarater ICC was found moderate to good, indicating an acceptable within-rater consistency. The final set of aggregated labels for each animation was derived from the mean across all participants. The means are plotted in Fig. 6 and the complete descriptive statistics are given in Table 6.

We also aimed to explore trends in the ratings and found that with respect to valence when participant confidence levels were low they tended to evaluate the expression as neutral. We then tested the hypothesis that valence is harder to rate than arousal which was confirmed. This observation appears in agreement with other evidence indicating that arousal is more easily perceived from bodily expressions than valence [17], while facial expression appears as a more stable visual modality for valence recognition [49]. Taking also into consideration findings supporting that context has significant influence on perceivers' judgments of valence [50], the fact that the animations were displayed on a robot with constrained facial expression and in a context-free setup could indicate two factors that might be worth further investigation as possible explanations of our observation.

For the arousal ratings, we observed that the lower arousal subspace was more sparsely covered compared to the medium- and high-arousal subspaces, especially in the higher-valence subspace ("Positive/Tired"). This trend could be related to the subjective pre-assignment of the animations in the nine classes; it might be the case that the pre-assignment did not succeed to fully cover the affect space. Another possibility could be that the perception of arousal differs between the group who conducted the pre-assignment (two researchers and an animator highly accustomed to the robot), and the participants who had minimal or no experience with Pepper. In the second case, a question that arises is whether the novelty effect contributes to inflated ratings of arousal. Finally, a different explanation could be that this trend might be indicative of the challenges related to designing animations that map to this subspace of low arousal and high valence.

5.1 Limitations

With regards to the limitations of this work, the choice of the two-dimensional model of emotion might be considered as too limited since it has been shown that subtle variations between certain emotions that share common core affect, e.g., fear and anger, might not be captured in less than four dimensions [15]. Nevertheless, core affect is still considered valid as a low dimensional structure that represents aspects of the full-blown emotional space [7]. In our future work we would like to explore how additional dimensions, like dominance and unpredictability, contribute in terms of enabling user discrimination of emotional expressions.

Another limitation is the fact that we did not systematize and balance the animation set in terms of the interactive modalities; motion, eye LEDs and sound. Nevertheless, with this experiment it was not our intention to examine how different combinations of modalities impact the perception of emotion. This is a very interesting question, but our present work treated each animation in a holistic manner.

Finally, the present study only involved one robotic platform, the Pepper robot, and consequently the observations might not transfer to other robots with different design, level of embodiment or expressive capabilities. Emotion expression with body language is inherently dependent on the embodiment level and the design, especially the more complex the expressions are, and the particular characteristics of Pepper might have evoked impressions that would be different in another platform. This is a native challenge in studying the perception of emotion expression in HRI, and comparison between different approaches appears challenging as we discussed in Section 2. Moreover, the animations we used are designed for the particular platform by professional animators and their structure cannot be translated and adapted for other robots auto-
matically. Although, our work can not directly generalize to other platforms, hopefully it could potentially serve as a case-study from which other researchers could shape hypotheses to test with different robots.

5.2 Future work

In our future work we plan to study how humans perceive emotional animations as responses to a given emotion-inducing context. The context will be presented in the form of images labeled with valence and arousal values. The annotated set we derived in the present study will allow us to test different aspects of this problem, such as the impact of a robot responding with appropriate or inappropriate emotional expressions given the context. Such an experiment can reveal new information about humans perception of, and expectations from, humanoid robots with the ability to mimic expressions of emotion.

Finally, aiming to address the last limitation we discussed in the previous subsection, another step we would like to take in our future work is the exploration of ways to extract high-level features from the labeled animations. Such features could be potentially transferred to different robotic platforms and be utilized with a dimensional affect perspective.

Acknowledgement: The authors would like to thank Olivier Landais for helping with the selection of the animation subset, and Samuel Lascombe for offering his advice in terms of the experimental design.

This project has received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 674868 (APRIL).

References

[1] C. Breazeal, Role of expressive behaviour for robots that learn from people, Philosophical Transactions of the Royal Society B: Biological Sciences, 2009, 364(1535), 3527–3538


[29] M. Häring, N. Bee, E. André, Creation and evaluation of emotion expression with body movement, sound and eye color for humanoid robots, Proceedings of International Symposium in Robot and Human Interactive Communication (2011, Atlanta, USA), 204–209


### Table 4: Informal descriptions and properties of the animations: categorical tag for the identification of the animation (assigned by the animators during the creation phase), informal descriptions of the movements displayed, duration of the animation in seconds, speed calculated as total frames divided by key frames, and modalities coded as M = Motion, L = LEDs, S = Sound.

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
<th>Duration</th>
<th>Speed</th>
<th>Modality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy_4</td>
<td>Slightly raises both hands, elbows bent, head nods up and down</td>
<td>2.4</td>
<td>6.67</td>
<td>ML</td>
</tr>
<tr>
<td>Interested_2</td>
<td>Bends right elbow and brings hand near waist, while left hand raises near mouth. Head raises looking upward and on the left</td>
<td>2.56</td>
<td>2.34</td>
<td>MLS</td>
</tr>
<tr>
<td>Joy_1</td>
<td>Both hands in front of the torso, elbows bent, then open wide, head looks upward</td>
<td>2.8</td>
<td>3.93</td>
<td>M</td>
</tr>
<tr>
<td>Loving_01</td>
<td>Both hands brought behind the body, torso bends to right and then to the left</td>
<td>4.36</td>
<td>6.65</td>
<td>ML</td>
</tr>
<tr>
<td>Confident_1</td>
<td>Bends both elbows slightly while nodding head up and down</td>
<td>4.96</td>
<td>3.83</td>
<td>ML</td>
</tr>
<tr>
<td>Heat_1</td>
<td>Looks downward, hands open slightly, bends torso slightly forward</td>
<td>4.68</td>
<td>1.92</td>
<td>M</td>
</tr>
<tr>
<td>Optimistic_1</td>
<td>Slightly nodding and then looks upward to the left</td>
<td>6.6</td>
<td>2.58</td>
<td>ML</td>
</tr>
<tr>
<td>Peaceful_1</td>
<td>Bends both elbows and raises hands slightly up and then down. Head moves downward and then up respectively.</td>
<td>3.44</td>
<td>2.62</td>
<td>ML</td>
</tr>
<tr>
<td>Content_01</td>
<td>Head looks up, torso bends right and left, hands open and close repeatedly</td>
<td>2.44</td>
<td>14.34</td>
<td>ML</td>
</tr>
<tr>
<td>Excited_01</td>
<td>Elbows bent, hands raise in front midway and move slightly up and down repeatedly, while head slightly nodding</td>
<td>2.24</td>
<td>16.07</td>
<td>M</td>
</tr>
<tr>
<td>Happy_01</td>
<td>Swings left and right, with hands on the sides and head nodding up and down</td>
<td>3.48</td>
<td>6.61</td>
<td>M</td>
</tr>
<tr>
<td>Joyful_01</td>
<td>Swings left and right, while moving arms up and down in front</td>
<td>3.56</td>
<td>8.99</td>
<td>M</td>
</tr>
<tr>
<td>Angry_3</td>
<td>Turns head to the right and brings left hand further away from body to the left</td>
<td>2.32</td>
<td>4.31</td>
<td>MS</td>
</tr>
<tr>
<td>Frustrated_1</td>
<td>Turns head to the right, nods right and left, hands brought in front, slightly bent</td>
<td>3.08</td>
<td>4.22</td>
<td>MLS</td>
</tr>
<tr>
<td>Hot_1</td>
<td>Head nods right and left</td>
<td>1.96</td>
<td>2.04</td>
<td>M</td>
</tr>
<tr>
<td>SadReaction_01</td>
<td>Moves torso back, then forward to the right, head nods left and right</td>
<td>2.88</td>
<td>9.38</td>
<td>M</td>
</tr>
<tr>
<td>Angry_4</td>
<td>Moves backward, raises left hand up and shakes it while head is looking up</td>
<td>2.68</td>
<td>7.84</td>
<td>MS</td>
</tr>
<tr>
<td>Fear_1</td>
<td>Arms raise slightly in front, head looks up, then right and left</td>
<td>4.4</td>
<td>5.23</td>
<td>MLS</td>
</tr>
<tr>
<td>Fearful_1</td>
<td>Both hands raise in front of the head, left hand higher</td>
<td>6.52</td>
<td>3.68</td>
<td>ML</td>
</tr>
<tr>
<td>Sad_01</td>
<td>Hands raise slightly in front with elbows bent, shakes head right and left and downward on the right direction</td>
<td>3.88</td>
<td>7.47</td>
<td>ML</td>
</tr>
<tr>
<td>Bored_01</td>
<td>Right hand is brought in front of the mouth and slightly moves back and forth</td>
<td>3.68</td>
<td>4.08</td>
<td>ML</td>
</tr>
<tr>
<td>Disappointed_1</td>
<td>Torso moves forward and slightly downward, while head nods right and left</td>
<td>2.88</td>
<td>3.82</td>
<td>ML</td>
</tr>
<tr>
<td>Lonely_1</td>
<td>A slight movement of hands and head up and then downward. Hands on the sides</td>
<td>7.04</td>
<td>2.98</td>
<td>ML</td>
</tr>
<tr>
<td>Shocked_1</td>
<td>Hands and head move slightly up and then downward. Hands in front</td>
<td>4.2</td>
<td>3.33</td>
<td>ML</td>
</tr>
<tr>
<td>AskForAttention_3</td>
<td>Brings right hand in front of the mouth, and then down while looking right</td>
<td>4.24</td>
<td>2.59</td>
<td>MLS</td>
</tr>
<tr>
<td>Chill_01</td>
<td>Slightly nodding and swinging left and right</td>
<td>6.72</td>
<td>3.57</td>
<td>M</td>
</tr>
<tr>
<td>Puzzled_1</td>
<td>Brings right hand in front of the mouth and left hand on waist. Head to the left</td>
<td>4</td>
<td>2.75</td>
<td>ML</td>
</tr>
<tr>
<td>Relaxation_2</td>
<td>Raises both hands midways in front with elbows bent and looks up</td>
<td>6.56</td>
<td>2.44</td>
<td>ML</td>
</tr>
<tr>
<td>Curious_01</td>
<td>Bends torso forward left and then right, hands open on the sides</td>
<td>2.68</td>
<td>2.24</td>
<td>M</td>
</tr>
<tr>
<td>SurprisedBig_01</td>
<td>Hands open on the sides, looks up</td>
<td>2.52</td>
<td>9.13</td>
<td>ML</td>
</tr>
<tr>
<td>Surprised_01</td>
<td>Slightly raises hands and looks up</td>
<td>3.08</td>
<td>3.57</td>
<td>M</td>
</tr>
<tr>
<td>Surprised_1</td>
<td>Raises both hands midst in front with elbows bent, looks around</td>
<td>4.88</td>
<td>4.92</td>
<td>M</td>
</tr>
<tr>
<td>Alienated_1</td>
<td>Torso slightly bent forward, head up, arms hanging</td>
<td>10.36</td>
<td>1.93</td>
<td>M</td>
</tr>
<tr>
<td>Hesitation_1</td>
<td>Repeatedly looks up to the left and down to the right, while right hand raises and falls again</td>
<td>9.04</td>
<td>2.32</td>
<td>MLS</td>
</tr>
<tr>
<td>Innocent_1</td>
<td>Hands meet in front of the torso and head looks left and right upward</td>
<td>7.68</td>
<td>2.47</td>
<td>M</td>
</tr>
<tr>
<td>Stretch_2</td>
<td>Hands open on the sides, torso bends backward and head looks up</td>
<td>7.28</td>
<td>2.61</td>
<td>M</td>
</tr>
</tbody>
</table>
Appendix B

Figure 7: Box plots of the ratings collected from the whole group of participants (N=20) per class of animations (9 classes with 4 animations each). The classes are labeled as combinations of valence and arousal levels. Individual animations are referenced with the original categorical tag assigned by the animators during the creation phase.

Table 5: Descriptive statistics by animation class of valence/arousal levels combinations (also visualized in Figure 4).

<table>
<thead>
<tr>
<th>Animation Class (V/A)</th>
<th>Arousal</th>
<th>Valence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Negative/Excited</td>
<td>0.83</td>
<td>0.19</td>
</tr>
<tr>
<td>Negative/Calm</td>
<td>0.48</td>
<td>0.28</td>
</tr>
<tr>
<td>Negative/Tired</td>
<td>0.40</td>
<td>0.26</td>
</tr>
<tr>
<td>Neutral/Excited</td>
<td>0.80</td>
<td>0.17</td>
</tr>
<tr>
<td>Neutral/Calm</td>
<td>0.54</td>
<td>0.21</td>
</tr>
<tr>
<td>Neutral/Tired</td>
<td>0.45</td>
<td>0.27</td>
</tr>
<tr>
<td>Positive/Excited</td>
<td>0.82</td>
<td>0.13</td>
</tr>
<tr>
<td>Positive/Calm</td>
<td>0.66</td>
<td>0.21</td>
</tr>
<tr>
<td>Positive/Tired</td>
<td>0.47</td>
<td>0.26</td>
</tr>
</tbody>
</table>
### Appendix C

Table 6: The final affect labels (means across 20 raters for valence and arousal) with the averaged confidence in judgment, as well as the original categorical tags of the animations and the pre-assigned classes of valence/arousal levels combinations. The ratings range from 0 to 1, with 100 points resolution.

<table>
<thead>
<tr>
<th>ID</th>
<th>Animation</th>
<th>Class</th>
<th>Arousal</th>
<th></th>
<th></th>
<th>Valence</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Conf.Mean</td>
<td>SD</td>
<td>Conf.Mean</td>
<td>Conf.SD</td>
</tr>
<tr>
<td>1</td>
<td>Happy_4</td>
<td>Positive/Calm</td>
<td>0.69</td>
<td>0.15</td>
<td>3.95</td>
<td>0.76</td>
<td>0.73</td>
<td>0.19</td>
</tr>
<tr>
<td>2</td>
<td>Interested_2</td>
<td>Positive/Calm</td>
<td>0.79</td>
<td>0.10</td>
<td>4.15</td>
<td>0.49</td>
<td>0.78</td>
<td>0.18</td>
</tr>
<tr>
<td>3</td>
<td>Joy_1</td>
<td>Positive/Calm</td>
<td>0.55</td>
<td>0.20</td>
<td>3.50</td>
<td>1.15</td>
<td>0.65</td>
<td>0.15</td>
</tr>
<tr>
<td>4</td>
<td>Loving_01</td>
<td>Positive/Calm</td>
<td>0.63</td>
<td>0.26</td>
<td>3.45</td>
<td>1.23</td>
<td>0.57</td>
<td>0.22</td>
</tr>
<tr>
<td>5</td>
<td>Confident_1</td>
<td>Positive/Tired</td>
<td>0.67</td>
<td>0.18</td>
<td>3.55</td>
<td>0.76</td>
<td>0.70</td>
<td>0.15</td>
</tr>
<tr>
<td>6</td>
<td>Heat_1</td>
<td>Positive/Tired</td>
<td>0.33</td>
<td>0.22</td>
<td>3.60</td>
<td>0.60</td>
<td>0.41</td>
<td>0.14</td>
</tr>
<tr>
<td>7</td>
<td>Optimistic_1</td>
<td>Positive/Tired</td>
<td>0.49</td>
<td>0.24</td>
<td>3.70</td>
<td>0.66</td>
<td>0.62</td>
<td>0.16</td>
</tr>
<tr>
<td>8</td>
<td>Peaceful_1</td>
<td>Positive/Tired</td>
<td>0.40</td>
<td>0.25</td>
<td>3.45</td>
<td>1.32</td>
<td>0.51</td>
<td>0.12</td>
</tr>
<tr>
<td>9</td>
<td>Content_01</td>
<td>Positive/Excited</td>
<td>0.85</td>
<td>0.14</td>
<td>4.05</td>
<td>0.94</td>
<td>0.71</td>
<td>0.23</td>
</tr>
<tr>
<td>10</td>
<td>Excited_01</td>
<td>Positive/Excited</td>
<td>0.81</td>
<td>0.16</td>
<td>4.15</td>
<td>0.67</td>
<td>0.71</td>
<td>0.27</td>
</tr>
<tr>
<td>11</td>
<td>Happy_01</td>
<td>Positive/Excited</td>
<td>0.82</td>
<td>0.12</td>
<td>4.30</td>
<td>0.73</td>
<td>0.86</td>
<td>0.12</td>
</tr>
<tr>
<td>12</td>
<td>Joyful_01</td>
<td>Positive/Excited</td>
<td>0.80</td>
<td>0.13</td>
<td>3.85</td>
<td>0.99</td>
<td>0.64</td>
<td>0.25</td>
</tr>
<tr>
<td>13</td>
<td>Angry_3</td>
<td>Negative/Calm</td>
<td>0.63</td>
<td>0.23</td>
<td>3.90</td>
<td>0.55</td>
<td>0.40</td>
<td>0.20</td>
</tr>
<tr>
<td>14</td>
<td>Frustrated_1</td>
<td>Negative/Calm</td>
<td>0.61</td>
<td>0.23</td>
<td>3.25</td>
<td>0.97</td>
<td>0.20</td>
<td>0.17</td>
</tr>
<tr>
<td>15</td>
<td>Hot_1</td>
<td>Negative/Calm</td>
<td>0.35</td>
<td>0.20</td>
<td>3.45</td>
<td>1.00</td>
<td>0.32</td>
<td>0.18</td>
</tr>
<tr>
<td>16</td>
<td>SadReaction_01</td>
<td>Negative/Calm</td>
<td>0.32</td>
<td>0.29</td>
<td>3.90</td>
<td>0.97</td>
<td>0.19</td>
<td>0.15</td>
</tr>
<tr>
<td>17</td>
<td>Angry_4</td>
<td>Negative/Excited</td>
<td>0.78</td>
<td>0.25</td>
<td>4.05</td>
<td>1.00</td>
<td>0.19</td>
<td>0.22</td>
</tr>
<tr>
<td>18</td>
<td>Fear_1</td>
<td>Negative/Excited</td>
<td>0.89</td>
<td>0.17</td>
<td>4.15</td>
<td>0.75</td>
<td>0.43</td>
<td>0.32</td>
</tr>
<tr>
<td>19</td>
<td>Fearful_1</td>
<td>Negative/Excited</td>
<td>0.90</td>
<td>0.11</td>
<td>4.40</td>
<td>0.60</td>
<td>0.14</td>
<td>0.16</td>
</tr>
<tr>
<td>20</td>
<td>Sad_01</td>
<td>Negative/Excited</td>
<td>0.75</td>
<td>0.17</td>
<td>3.80</td>
<td>0.83</td>
<td>0.33</td>
<td>0.22</td>
</tr>
<tr>
<td>21</td>
<td>Bored_01</td>
<td>Negative/Tired</td>
<td>0.48</td>
<td>0.31</td>
<td>3.65</td>
<td>1.09</td>
<td>0.49</td>
<td>0.13</td>
</tr>
<tr>
<td>22</td>
<td>Disappointed_1</td>
<td>Negative/Tired</td>
<td>0.37</td>
<td>0.19</td>
<td>3.40</td>
<td>0.94</td>
<td>0.21</td>
<td>0.14</td>
</tr>
<tr>
<td>23</td>
<td>Lonely_1</td>
<td>Negative/Tired</td>
<td>0.28</td>
<td>0.23</td>
<td>3.65</td>
<td>0.81</td>
<td>0.28</td>
<td>0.19</td>
</tr>
<tr>
<td>24</td>
<td>Shocked_1</td>
<td>Negative/Tired</td>
<td>0.48</td>
<td>0.28</td>
<td>3.70</td>
<td>0.80</td>
<td>0.56</td>
<td>0.17</td>
</tr>
<tr>
<td>25</td>
<td>AskForAttention_3</td>
<td>Neutral/Calm</td>
<td>0.49</td>
<td>0.25</td>
<td>3.45</td>
<td>0.89</td>
<td>0.38</td>
<td>0.22</td>
</tr>
<tr>
<td>26</td>
<td>Chill_01</td>
<td>Neutral/Calm</td>
<td>0.56</td>
<td>0.18</td>
<td>3.65</td>
<td>0.88</td>
<td>0.64</td>
<td>0.18</td>
</tr>
<tr>
<td>27</td>
<td>Puzzled_1</td>
<td>Neutral/Calm</td>
<td>0.60</td>
<td>0.19</td>
<td>3.80</td>
<td>0.62</td>
<td>0.49</td>
<td>0.12</td>
</tr>
<tr>
<td>28</td>
<td>Relaxation_2</td>
<td>Neutral/Calm</td>
<td>0.50</td>
<td>0.22</td>
<td>3.15</td>
<td>1.14</td>
<td>0.56</td>
<td>0.15</td>
</tr>
<tr>
<td>29</td>
<td>Curious_01</td>
<td>Neutral/Excited</td>
<td>0.89</td>
<td>0.13</td>
<td>4.25</td>
<td>0.79</td>
<td>0.62</td>
<td>0.17</td>
</tr>
<tr>
<td>30</td>
<td>SurprisedBig_01</td>
<td>Neutral/Excited</td>
<td>0.82</td>
<td>0.16</td>
<td>4.00</td>
<td>0.73</td>
<td>0.40</td>
<td>0.19</td>
</tr>
<tr>
<td>31</td>
<td>Surprised_01</td>
<td>Neutral/Excited</td>
<td>0.78</td>
<td>0.19</td>
<td>4.00</td>
<td>0.79</td>
<td>0.46</td>
<td>0.22</td>
</tr>
<tr>
<td>32</td>
<td>Surprised_1</td>
<td>Neutral/Excited</td>
<td>0.71</td>
<td>0.15</td>
<td>3.45</td>
<td>0.94</td>
<td>0.60</td>
<td>0.15</td>
</tr>
<tr>
<td>33</td>
<td>Alienated_1</td>
<td>Neutral/Tired</td>
<td>0.20</td>
<td>0.24</td>
<td>4.05</td>
<td>0.60</td>
<td>0.41</td>
<td>0.16</td>
</tr>
<tr>
<td>34</td>
<td>Hesitation_1</td>
<td>Neutral/Tired</td>
<td>0.70</td>
<td>0.16</td>
<td>3.55</td>
<td>1.00</td>
<td>0.32</td>
<td>0.22</td>
</tr>
<tr>
<td>35</td>
<td>Innocent_1</td>
<td>Neutral/Tired</td>
<td>0.52</td>
<td>0.21</td>
<td>4.05</td>
<td>0.39</td>
<td>0.69</td>
<td>0.15</td>
</tr>
<tr>
<td>36</td>
<td>Stretch_2</td>
<td>Neutral/Tired</td>
<td>0.40</td>
<td>0.21</td>
<td>3.40</td>
<td>1.05</td>
<td>0.60</td>
<td>0.18</td>
</tr>
</tbody>
</table>