

Title	Emotional Bodily Expressions for Culturally Competent Robots through Long Term Human-Robot Interaction
Author(s)	Tuyen, Nguyen Tan Viet; Jeong, Sungmoon; Chong, Nak Young
Citation	2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS): 2008-2013
Issue Date	2018-10-01
Type	Conference Paper
Text version	author
URL	<a href="http://hdl.handle.net/10119/16106">http://hdl.handle.net/10119/16106</a>
Rights	This is the author's version of the work. Copyright (C) 2018 IEEE. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018, 2008-2013. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Description	



# Emotional Bodily Expressions for Culturally Competent Robots through Long Term Human-Robot Interaction

Nguyen Tan Viet Tuyen, Sungmoon Jeong, and Nak Young Chong

**Abstract**—Generating emotional bodily expressions for culturally competent robots has been gaining increased attention to enhance the engagement and empathy between robots and humans in a multi-culture society. In this paper, we propose an incremental learning model for selecting the user’s representative or habitual emotional behaviors which place emphasis on individual users’ cultural traits identified through long term interaction. Furthermore, a transformation model is proposed to convert the obtained emotional behaviors into a specific robot’s motion space. To validate the proposed approach, the models were evaluated by two example scenarios of interaction. The experimental results confirmed that the proposed approach endows a social robot with the capability to learn emotional behaviors from individual users, and to generate its emotional bodily expressions. It was also verified that the imitated robot motions are rated emotionally acceptable by the demonstrator and recognizable by the subjects from the same cultural background with the demonstrator.

## I. INTRODUCTION

Human facial and bodily expressions play crucial roles in human-human interaction. Psychological researches have shown that the physical expression of emotion is an integral part of social interactions to better convey the communicator’s emotion which affects social outcomes [1]. Toward understanding this effect, many social robotics studies focused on generating emotional expressions for robots by estimating environmental stimuli and incorporating robot emotional states, which is believed to enhance the social interaction outcomes. In [2], the authors investigated the role of culture in representing the robot’s emotions, where bodily expressions were utilized to convey the robot’s emotional state. This research suggested a way to provide social robots with the capability of learning to behave in alignment with individuals’ cultural traits. Under different cultural environments, robots could generate different emotional and behavioral responses to the same environmental stimuli. Specifically, the Pepper robot’s bodily expressions was motivated by psychological researches about the mapping of human bodily features into affective artifacts [3]. A similar approach was found in [4] for NAO robot, where emotional expressions with bodily movement and eye color was inspired by the work of Meijer [5] and other psychological researchers. In [6], an android head robot imitates human facial expressions with the main goal to improve the emotion recognition capabilities of autistic children. The android robot tracks human expression represented by facial feature points and directly converts

them into corresponding motor movements of the robot. Likewise, the UCLIC Affective Body Posture and Motion Database [7] was utilized to generate emotional expressions for robots in [8]. UCLIC was originally labeled and rated by observers, and the best scoring one was chosen and mapped into the robot model with the provided emotional label.

On the other hand, in order to increase the engagement and empathy between a robot and a human through long term interaction, careful attention should be paid to emotional expressions for robots adapting to the personality and cultural identity of the interacting user. To archive this goal, this research investigates the psychological perspectives about infant social development, where the infant’s interpretation and behavioral responses are highly influenced by their parents through imitative exchanges [9]. Based on social referencing described in our previous paper [10], the robot learns to imitate the user’s emotional and behavioral responses to environmental stimuli. Through long term interaction, the robot generates emotional bodily expressions for specific emotions by taking the following steps: (1) clustering human emotional behavior samples into different groups based on the similarity of body movements, (2) utilizing the human habitual behavior which could be identified by assessing the frequency of similar behaviors [11] as the references for generating emotional bodily expressions and (3) mapping the predicted human habitual behavior into the robot’s motion space.

In this paper, we introduce our approach for generating emotional bodily expression of culturally competent social robots inspired by the infant’s social development process learning their parents’ affective and behavioral responses. In the methodology, we describe our proposed behavior selection model and the model of transformation in order to generate robot emotional behaviors through long term interaction. In the experiment results and discussion section, the research approach is sequentially validated by two different scenarios of interaction. Finally, we summarize the results and our future work in the conclusion and future works section.

## II. GENERATING ROBOT EMOTIONAL EXPRESSIONS THROUGH LONG TERM INTERACTION

### A. Importance of Interacting Partner’s Traits

Social human robot interaction should be treated in a similar way to the interaction with another person [12]. Hence, in order to sustain the user’s engagement, this research sheds light on the interacting partner’s behaviors in order to generate appropriate social behaviors for robots,

The authors are with the School of Information Science, Japan Advanced Institute of Science and Technology, Ishikawa, Japan {ngtvtuyen, jeongsm, nakyounj}@jaist.ac.jp

ensuring that the generated gestures conform to the social norm. There is a strong psychological evidence known as the chameleon effect [13] which is the tendency to mimic the posture, facial expressions, verbal and nonverbal behaviors of others. Likewise, the importance of the interacting partner’s personality traits was emphasized that affected the behavior of social robots in [14], where an interesting experiment was conducted to examine the influence of KMC-EXPR robot’s personalities associated with different facial expressions on the interaction outcomes. The results indicated that interacting with a robot having a similar personality made the user feel more comfortable than interacting with a robot with different personalities.

Summarizing, each individual has their own way to express emotions [7]. The aforementioned studies provide empirical evidences for the need of considering the user’s personality traits when generating behaviors for socially assistive robots and their emotional expressions in particular. While previous researches focus on mimicry of facial expressions [6][12], this research pays special attention to the robot bodily expressions. Understanding and reflecting information of individual users to generate behaviors for social robots, it is believed that their behaviors could be more acceptable in our multi-cultural society.

### B. Habitual Behavior Selection Model

During daily human robot interactions, it is obvious that the number and types of human emotional behaviors vary across cultures and personality traits of individuals that can be identified after long-term relationships. Thus, robots are required to have the capability of learning such behaviors in an unsupervised manner. This idea has been implemented in different contexts. Mohammad [15] used unsupervised learning for association between human gestural commands and robot actions. In [16], the authors made comparisons between different unsupervised learning algorithms such as Self Organizing Maps (SOM), Fuzzy C Means (FCM), and K Means for the recognition of human posture in video sequences. The capability of robot arm trajectory learning from human demonstrations was proposed by [17], where the trajectory clustering and approximation modules take human demonstrative trajectories as the input and then classify the trajectories into groups. For each group, the most consistent trajectory was selected and then a set of generated trajectories can be visualized in a simulated environment, allowing the human user to finally select the desired trajectory. Hence, for everyday social interactions with no *a priori* information about human actions, unsupervised learning is the appropriate approach for classifying various types of actions into different groups based on the similarity of actions.

Human body expressions could be recognized using skeleton features obtained from sensors on-board the robot or motion capture systems. In each emotion space, a set of human emotional behaviors  $A_1, A_2, \dots, A_n$  are gradually received during day-to-day human robot interaction. Action  $A_i = [S_1, S_2, \dots, S_T]$  is the sequence of frames over a period of time  $T$  and  $S_t =$

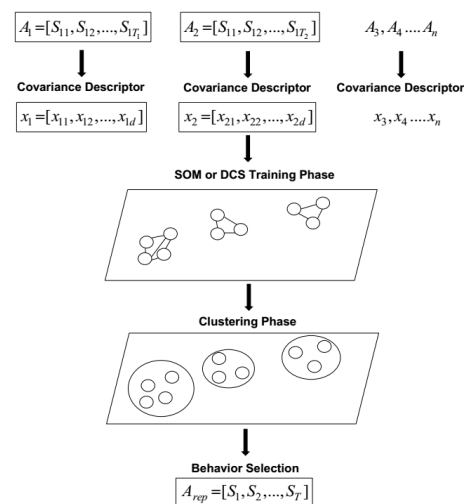


Fig. 1. Emotional behavior selection through long term interaction

$[x_1, x_2, \dots, x_{20}; y_1, y_2, \dots, y_{20}; z_1, z_2, \dots, z_{20}]$  is the human skeleton information including 20 joint positions at time  $t$ . The Covariance Descriptor method [18] is used to encode the sequence of frames  $A_i$  into the fixed length descriptor. Human emotional bodily expression  $A_1, A_2, \dots, A_n$  are classified into clusters through the training and clustering phase. Finally, at the behavior selection phase, by considering the distribution of body movements, the robot can utilize the most frequently observed behavior as the reference for generating its emotional bodily expression. Fig. 1 illustrates the process of selecting an appropriate emotional behavior for social robots interacting with the user in a certain emotional state.

1) *Training and Clustering Phase*: In order to use an unsupervised learning approach without *a priori* knowledge about the number of clusters, a batch version of Self-Organizing Map (SOM) [19] was used for the training phase in our previous paper [10]. It is obvious that topological preservation is the main advantage of SOM for classifying encoded descriptors into different groups based on the similarities. On the other hand, for the scenarios of long term human robot interaction, since the number of human emotional behaviors will be sequentially increased, the robot should be capable of incrementally learning new gestures without corrupting the existing model. However, on the grid of SOM neurons, the number of neurons must be fixed in advance, which makes SOM inappropriate for incremental learning. To satisfy requirement of incremental learning while ensuring the topological preservation of the grid of trained neurons, this research employs a Dynamic Cell Structure (DCS) neural architecture [20] for the training phase. DCS adheres to the Kohonen type learning rule [19] for updating the weight of neural vectors the same as the SOM approach, yet uses the Hebbian learning rule [21] to dynamically update the lateral connection structure (topology of the graph of neurons). At the training phase, new units could be added on the grid of neurons, if the quantization error is higher than the predefined stopping condition. Another

approach of growing neural network by dynamic allocation the feature map in order to evolve its structure are known as Growing Cell Structure (GCS) [22]. DCS works in a similar way to GCS excepts one essential difference: the lateral connections between neuron units are not initially defined, instead, they are dynamically learned during the training phase by Herbian learning rule. DCS has been widely used in many applications for on-line learning purpose. NASA's first generation Intelligent Flight Control System program utilized DCS for on-line learning and estimation of system parameters [23].

After the incremental learning phase with the DCS approach, the grid of trained neurons  $m$  will be classified into different groups at the clustering phase. Here, classifying trained neurons into different groups is conducted with Distance matrix based approach [24]. By clustering the training neurons rather than descriptors directly, significant gains in speed of clustering can be obtained [25]. At the end of the clustering phase, each descriptor  $x$  and its corresponding neuron  $m_i$  was defined by the Best Matching Unit (BMU) function given by

$$\|x - m_i\| = \min\{\|x - m\|\} \quad (1)$$

2) *Behavior Selection Phase*: During the previous phase,  $n$  action data  $\{A_1, A_2, \dots, A_n\}$  was encoded to  $n$  descriptors  $\{x_1, x_2, \dots, x_n\}$  and then classified into different groups  $\{Cluster_1, Cluster_2, \dots, Cluster_k\}$  ( $k \leq N$ ) based on the similarity of actions. At the behavior selection phase, considering the probabilistic distribution of human actions observed by the robot, an appropriate behavior will be selected out of the largest cluster  $Cluster_i$  that contains the highest number of similar actions. This can be considered habitual actions affected by their cultural background. Specifically, we can choose a representative descriptor  $x_{rep}$  located closest to the center of the largest cluster defined as:

$$\|x_{rep} - center\| \leq \|x - center\| \quad \forall x \in Cluster_i, \quad (2)$$

where  $\|x - center\|$  is the Euclidean distance between the center of  $Cluster_i$  to the descriptor  $x$ . Finally, the corresponding action of descriptor  $x_{rep}$  will be detected as  $A_{rep}$ . The robot can select  $A_{rep}$  as a target behavior to generate its emotional bodily expression associated with the corresponding emotion.

### C. Transformation Model

Now the user's target behavior should be mapped into the robot model. It is obvious that the number of Degrees of Freedom (DOFs) and joint configurations are different between the demonstrator and the robot. Therefore, the mapping between two agents should be performed through the transformation model as shown in Fig. 2. This transformation model receives the human pose represented by joint coordinates in Cartesian space as the input and releases a set of corresponding joint angles for the robot subject to its physical constraints.

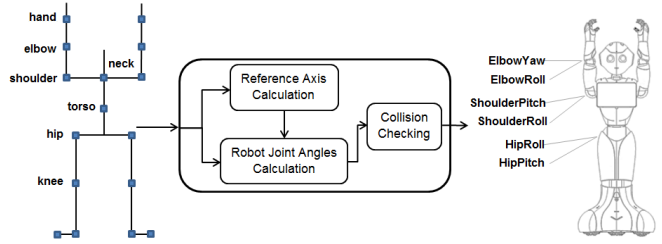


Fig. 2. Mapping of human upper body pose into robot motion space

Depending on specific robot platform, the typical kinematic parameters should be defined [26]. It should be noted that the kinematic model of the human lower body and that of the Pepper robot are completely different from each other. Thus, this paper proposes a transformation model which focuses on the imitation of the human upper body including the movements of Hip, Shoulder, and Elbow on both the Left(L) and Right(R) sides. Consequentially, the transformation model releases a set of joint angle data for the Pepper robot given by  $\theta_{Pepper} = \{(L/R)ElbowRoll, (L/R)ElbowYaw, HipRoll, HipPitch, (L/R)ShoulderRoll, (L/R)ShoulderPitch\}$ .

Specifically, the transformation model starts with calculating the reference axis  $x_{ref}, y_{ref}, z_{ref}$  which describes the orientation of the current human pose. The obtained reference axes are combined with the input human joint positions to calculate the corresponding robot joint angles  $\theta_{Pepper}$ . The self-collision checking is conducted using the off-the-shelf API before releasing the calculated  $\theta_{Pepper}$  to the Pepper robot model.

## III. EXPERIMENT RESULTS AND DISCUSSION

### A. Transferring Human Behaviors into Robot Model

1) *Experiment Scenario*: The first experiment is aimed to qualitatively evaluate, from the viewpoint of ordinary people who are not experienced in robotics, whether initial human actions and key poses are recognizable on Pepper robot model. In order to evaluate that, the subjective evaluation had been setup which made participants have chance to evaluate how appropriately human actions and key poses were displayed on robot. An on-line survey was conducted with a total of 41 participants ranging in the age from 23 to 37 ( $M = 26.1, SD = 2.9$ ). They come from Bangladesh, China, Indonesia, Japan, Thailand, and Vietnam and mostly are not familiar with robots.

Specifically, the demonstrator standing in front of the Pepper robot performs demonstrative actions. The robot then observes the demonstrator's action as time series motion capture data. Through transformation model, skeleton frames represented by joint positions in Cartesian space were sequentially converted into the corresponding robot joint angles. Collision detection was conducted with the calculated joint angles before releasing to the Pepper robot.

2) *Results and Discussion*: The actions demonstrated by the user and imitated by the Pepper robot were firstly evaluated using the recognition rate. Then, we asked the

Action 1	41.00	0.00	0.00	0.00
Action 2	0.00	40.00	1.00	0.00
Action 3	0.00	0.00	38.00	3.00
Action 4	1.00	0.00	9.00	31.00
	Robot Action 1	Robot Action 2	Robot Action 3	Robot Action 4

Fig. 3. Confusion matrix representing the recognition of human action after mapping into robot model



Fig. 4. Human pose and the mapped one on Pepper robot

participants to rate the level of pose similarity between the demonstrator and the Pepper robot ranging in value from 0 to 10. The subjective evaluation was repeated 4 times with 4 different actions and key poses.

Fig. 3 presents the recognition rates of the demonstrated target actions and their imitated actions by the Pepper robot. The result confirmed that human actions can be imitated by the robot and easily recognizable by the participants. Sometimes, they were confused between the mapped action 3 and 4. Hence, the demonstrated actions were well mapped into the robot motion space subject to the robot physical constraints.

In terms of level of pose similarity between the target and imitated actions, the average score was 7.05 out of 10. The most similar pose was scored 8.32 and the lowest one was 6.37. In general, participants agreed that Pepper could imitate human pose with the high similarity. It was noticed that this evaluation gave subjects a chance to carefully evaluate individual parts between the demonstrator's pose and mapped one on robot. Thus, minor comments were received from them about the differences of hand-palm between the user and Pepper in typical poses as shown in Fig. 4. Due to the lack of motion capture data, our current transformation model could not generate  $(L/R)Wrist\ Yaw$  and  $Head\ Pitch/Yaw$  for the Pepper robot. Further investments about utilizing external sensors like Leap Motion to estimate the demonstrator's hand-palm orientation [27] should be conducted in the future.

In general, the evaluation results confirmed the feasibility of transformation model to convert human behaviors into robot motion space while the recognition of demonstrated actions are ensured on robot model. The experiment also

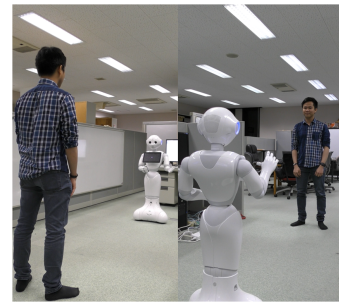


Fig. 5. Scenarios of interaction for learning from the user's emotional behaviors

revealed the promising approach for teaching robot new gestures by demonstration instead of off-line programming as ordinary approach.

### B. Generating Emotional Expression through Scenarios of Interaction

1) *Experiment Scenario*: In this experiment, the target behavior selection model for selecting representative behaviors through long term interaction and the transformation model for behavior mapping were connected to each other. The experiment was set up as shown in Fig. 5, where Pepper interacted with individual users for learning their emotional behaviors. Pepper first detected the user through facial detection<sup>1</sup> and then started the conversation by greeting the user with random questions and non-verbal behaviors using a predefined list of actions. The user then responded to Pepper with their facial and bodily expressions. Pepper obtained the user's motion capture data as a time series similar to the previous experimental setup. At the same time, the robot estimated the user's emotion through their facial expression<sup>2</sup>. The user's bodily expression was associated with the emotion estimated through the user's facial expression. This scenario of interaction is performed repeatedly for 3 consecutive days. In this research, only the user's bodily expressions for *Happy* and *Sad* were stored on the robot memory to simplify the analysis.

In order to evaluate the generated bodily expressions, an on-line survey was conducted with a group of 30 participants aging from 23 to 37 ( $M = 27.4$ ,  $SD = 3.7$ ) who have the same cultural background with the interacting user (Vietnamese). The objective of the survey was to investigate the quality of generated emotional behaviors aligned with the user's cultural background.

2) *Results and Discussion*: After 3 consecutive days of interactions, 52 human emotional actions were labeled "Happy" and 43 actions were labeled "Sad". In each emotion space, the behavior selection model receives the human actions as the input data and releases a representative behavior  $A_{rep}$ . Through the transformation model,  $A_{rep}$  was converted into the robot motion. Figs. 6 and 7 represent screen shots of the generated robot behaviors for the emotion *Happy* and *Sad*, respectively.

<sup>1</sup>[doc.aldebaran.com/2-5/naoqi/peopleperception/alpeopleperception.html](http://doc.aldebaran.com/2-5/naoqi/peopleperception/alpeopleperception.html)

<sup>2</sup>[microsoft.com/cognitive-services/en-us/](http://microsoft.com/cognitive-services/en-us/)

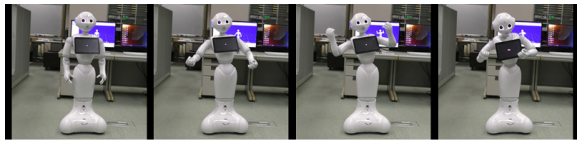


Fig. 6. Screen shots of Pepper’s bodily expression *Happy*

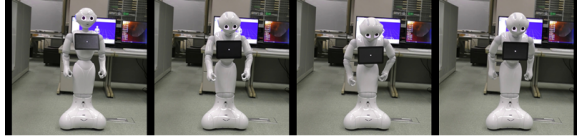


Fig. 7. Screen shots of Pepper’s bodily expression *Sad*

expression Happy	0.00	28.00	2.00
expression Sad	17.00	4.00	9.00
	<i>Sad</i>	<i>Happy</i>	<i>Other</i>

Fig. 8. Recognition rate for Pepper’s expression *Sad* and *Happy*

The generated robot behaviors from this experiment was evaluated by a group of Vietnamese participants. First, they were asked to watch the robot’s bodily expression and then choose the most appropriate emotion label among *Happy*, *Sad*, and other. This approach is similar to the strategy applied in [12], where facial and vocal expressions of Kismet robot were evaluated by best matching emotional labels. Fig. 8 summarizes the recognition rate of *Happy* and *Sad* emotional expressions. At the second phase of survey, subjects were asked to assign appropriate values of arousal and valence [28] using the Self-Assessment Manikin (SAM) five-point scale [29]. This approach allows participants to asses and express their emotional responses to robot behaviors without considering the emotional labels. Participants’ answers were then converted to a group of values in a range of  $[-1, 1]$  as shown in Fig. 9. The distribution of subject’s evaluation along the dimensions of arousal and valence [28] is shown in Fig. 10.

The recognition rate of the robot’s emotional expressions shown in Fig. 8 confirmed that the generated behavior for *Happy* was clearly recognizable by the participants. 28 out of 30 participants believed that Pepper was trying to convey *Happy* cue through its bodily movements. On the other hand, *Sad* was comparatively less distinctive. 17 out of 30 participants thought that Pepper was showing *Sad* cue. The other 9 participants felt that this behavior contains a different meaning such as *Regretful*, *Bored*, or even make them confused. From the viewpoint of the robot’s user, they also agreed that the expression *Happy* was significantly recognizable and the meaning of their behavior was preserved on Pepper robot. However, even the bodily expression *Sad*

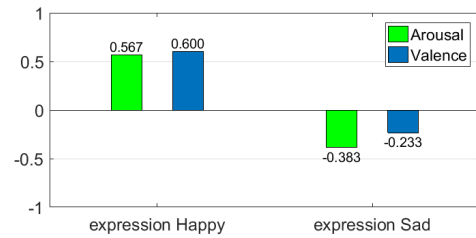


Fig. 9. Mean values of arousal and valence on 2 generated expressions

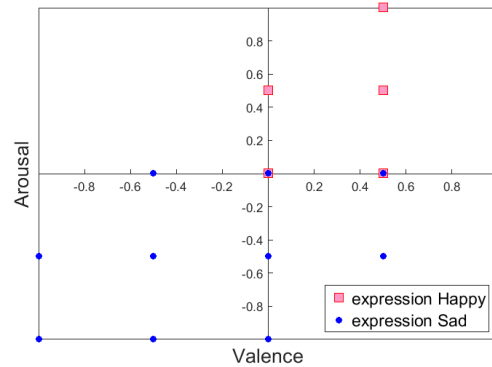


Fig. 10. Distribution of robot expressions on model of affect

was still acceptable, but the user thought that Pepper did not express it as similar as their behavior. It is noted that the lack of the robot’s head movements (*Head Pitch/Yaw*) from the transformation model significantly affected the recognition rate of robot expression *Sad*.

To analyze how different expression cues are within two generated behaviors, the one-way analysis of variance (ANOVA) was conducted in the arousal dimension, followed by the valence dimension. The ANOVA test indicated that there were significant differences ( $F(1, 58) = 126.47, p < 0.001$ ) in the arousal dimension. Similarly, differences were found on the valence dimension ( $F(1, 58) = 79.84, p < 0.001$ ) between *Sad* and *Happy*. This results imply that two generated behaviors were clearly distinctive from each other on the two-dimensional affective space. The distribution of generated robot expressions as shown in Fig. 10 also supported for the differences between 2 generated emotional behaviors. 86 percent of the subjects’ answers indicated that *Sad* lies in the third quadrant of the model of affect, while 100 percent of evaluation for *Happy* belongs to the first quadrant of the model. It is widely understood that the first quadrant is the location of *Happy*, while *Sad* lies in the third quadrant of model [28]. Hence, the generated robot emotional behaviors were correctly located in the two dimensional affective space with the values of arousal and valence obtained from the subjects’ evaluation.

There are strong influences of speed and amplitude of robot motions on the perceived level of arousal and valence [30]. Our experiments also confirmed that participants often assigned higher value of arousal and valence for *Happy* than *Sad* because they though that the robot performed gesture of

*Happy* faster and higher amplitude than the gesture of *Sad*. The reason was that the demonstrator performed different emotional behaviors with different speed of motions, thus, robot motion capture system obtained different number of skeleton frames for different emotional behaviors.

#### IV. CONCLUSION AND FUTURE WORKS

This paper aimed to investigate the importance of the user's cultural background when generating emotional bodily expressions for social robots. In order to meet the requirements for cultural competence, we implemented an incremental learning model to select a representative emotional response through long term human-robot interaction, and the transformation model to convert human behavior into the Pepper robot's motion space. The proposed approach was validated by two example scenarios of human robot interaction. The experiments' results indicated that our research approach provided the robot with the capability of entering into scenario of interaction for imitation learning purpose. Through 3 consecutive days of interaction, the robot utilized the user information to generate its emotional behaviors which were acceptable by the robot's user and recognizable from a group of subjects who share the same cultural background with the robot's user.

In the future work, user emotion estimation from multiple modalities (facial, verbal, heartbeat) as well as segmentation of time series motion capture data will be investigated. Indeed, we will extend this research idea by using the proposed models to generate robot behaviors for other purposes such as robot non-verbal behaviors associated with the verbal content of speech.

#### ACKNOWLEDGMENTS

This work was supported by the EU-Japan coordinated R&D project on Culture Aware Robots and Environmental Sensor Systems for Elderly Support commissioned by the Ministry of Internal Affairs and Communications of Japan and EC Horizon 2020.

#### REFERENCES

- [1] B. N. Vosk, R. Forehand, and R. Figueroa, "Perception of emotions by accepted and rejected children," *Journal of Psychopathology and Behavioral Assessment*, vol. 5, no. 2, pp. 151–160, 1983.
- [2] T. L. Q. Dang, N. T. V. Tuyen, S. Jeong, and N. Y. Chong, "Encoding cultures in robot emotion representation," in *Robot and Human Interactive Communication, IEEE International Symposium on*. IEEE, 2017, pp. 547–552.
- [3] A. Kleinsmith and N. Bianchi-Berthouze, "Affective body expression perception and recognition: A survey," *IEEE Transactions on Affective Computing*, vol. 4, no. 1, pp. 15–33, 2013.
- [4] M. Häring, N. Bee, and E. André, "Creation and evaluation of emotion expression with body movement, sound and eye color for humanoid robots," in *Robot and Human Interactive Communication, IEEE International Symposium on*. IEEE, 2011, pp. 204–209.
- [5] M. De Meijer, "The contribution of general features of body movement to the attribution of emotions," *Journal of Nonverbal Behavior*, vol. 13, no. 4, pp. 247–268, 1989.
- [6] A. Adams and P. Robinson, "An android head for social-emotional intervention for children with autism spectrum conditions," in *Affective Computing and Intelligent Interaction*. Springer, 2011, pp. 183–190.
- [7] A. Kleinsmith, P. R. De Silva, and N. Bianchi-Berthouze, "Recognizing emotion from postures: Cross-cultural differences in user modeling," in *International Conference on User Modeling*. Springer, 2005, pp. 50–59.
- [8] G. Van de Perre, M. Van Damme, D. Lefeber, and B. Vanderborght, "Development of a generic method to generate upper-body emotional expressions for different social robots," *Advanced Robotics*, vol. 29, no. 9, pp. 597–609, 2015.
- [9] S. Feinman and M. Lewis, "Social referencing at ten months: A second-order effect on infants' responses to strangers," *Child Development*, pp. 878–887, 1983.
- [10] N. T. V. Tuyen, S. Jeong, and N. Y. Chong, "Learning human behavior for emotional body expression in socially assistive robotics," in *Ubiquitous Robots and Ambient Intelligence, International Conference on*. IEEE, 2017, pp. 45–50.
- [11] I. Ajzen, "Residual effects of past on later behavior: Habituation and reasoned action perspectives," *Personality and Social Psychology Review*, vol. 6, no. 2, pp. 107–122, 2002.
- [12] C. L. Breazeal, *Designing sociable robots*. MIT press, 2004.
- [13] T. L. Chartrand and J. A. Bargh, "The chameleon effect: The perception-behavior link and social interaction," *Journal of Personality and Social Psychology*, vol. 76, no. 6, p. 893, 1999.
- [14] E. Park, D. Jin, and A. P. del Pobil, "The law of attraction in human-robot interaction," *International Journal of Advanced Robotic Systems*, vol. 9, no. 2, p. 35, 2012.
- [15] Y. Mohammad, T. Nishida, and S. Okada, "Unsupervised simultaneous learning of gestures, actions and their associations for human-robot interaction," in *Intelligent Robots and Systems, IEEE/RSJ International Conference on*. IEEE, 2009, pp. 2537–2544.
- [16] K. K. Htike and O. O. Khalifa, "Comparison of supervised and unsupervised learning classifiers for human posture recognition," in *Computer and Communication Engineering, International Conference on*. IEEE, 2010, pp. 1–6.
- [17] J. Aleotti and S. Caselli, "Robust trajectory learning and approximation for robot programming by demonstration," *Robotics and Autonomous Systems*, vol. 54, no. 5, pp. 409–413, 2006.
- [18] M. E. Hussein, M. Torki, M. A. Gowayyed, and M. El-Saban, "Human action recognition using a temporal hierarchy of covariance descriptors on 3d joint locations," in *Artificial Intelligence, International Joint Conference on*, vol. 13, 2013, pp. 2466–2472.
- [19] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.
- [20] J. Bruske and G. Sommer, "Dynamic cell structure learns perfectly topology preserving map," *Neural Computation*, vol. 7, no. 4, pp. 845–865, 1995.
- [21] T. Martinez, "Competitive hebbian learning rule forms perfectly topology preserving maps," in *Artificial Neural Networks, International Conference on*. Springer, 1993, pp. 427–434.
- [22] B. Fritzke, "Growing cell structures a self-organizing network for unsupervised and supervised learning," *Neural Networks*, vol. 7, no. 9, pp. 1441–1460, 1994.
- [23] M. G. Perhinschi, G. Campa, M. R. Napolitano, M. Lando, L. Massotti, and M. L. Fravolini, "A simulation tool for on-line real time parameter identification," in *AIAA Modeling and Simulation Technologies Conference*, 2002.
- [24] J. Vesanto and M. Sulkava, "Distance matrix based clustering of the self-organizing map," in *International Conference on Artificial Neural Networks*. Springer, 2002, pp. 951–956.
- [25] J. Vesanto and E. Alhoniemi, "Clustering of the self-organizing map," *IEEE Transactions on Neural Networks*, vol. 11, no. 3, pp. 586–600, 2000.
- [26] J.-H. Lee et al., "Full-body imitation of human motions with kinect and heterogeneous kinematic structure of humanoid robot," in *System Integration, IEEE/SICE International Symposium on*. IEEE, 2012, pp. 93–98.
- [27] G. Marin, F. Dominio, and P. Zanuttigh, "Hand gesture recognition with leap motion and kinect devices," in *Image Processing, IEEE International Conference on*. IEEE, 2014, pp. 1565–1569.
- [28] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, p. 1161, 1980.
- [29] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.
- [30] J. Xu, J. Broekens, K. Hindriks, and M. A. Neerincx, "The relative importance and interrelations between behavior parameters for robots' mood expression," in *Affective Computing and Intelligent Interaction, Humaine Association Conference on*. IEEE, 2013, pp. 558–563.