From Continuous Affective Space to Continuous Expression Space: Non-verbal Behaviour Recognition and Generation

Junpei Zhong, Lola Cañamero
Email: {j.zhong3, l.canamero}@herts.ac.uk
School of Computer Science, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, United Kingdom

***Abstract*—In this research, a recurrent neural network with parametric bias (RNNPB) was adopted to construct a continuous expression space from emotion caused human behaviours. It made use of the short-term memory ability of the recurrent weights to store spatio-temporal sequences features, while the attached parametric bias units were trained in a self-organizing way and represented as a low-dimensional expression space to capture these non-linear features of the sequences. Three demonstrations were given: training and recognition performances were examined in computer simulations, while the network generated both trained and novel movements were shown in a three-dimensional avatar demonstrations.**

I. Introduction

Although there are various theories concerning the links between emotion and cognition (see e.g. [7], [8], [11] for edited collections presenting a broad sample of models in various disciplines) and even questioning whether the loop between cognition and emotion really exists (e.g. [4]), it is nowadays agreed that emotion mediates various cognitive processes including sensorimotor process, and therefore that emotional states interact with and modulate motor behaviours. As a result of such modulation, emotion is (implicitly or explicitly) expressed in motor behavior. Therefore, some sensorimotor processes, physically expressed as motor behaviours, are partially caused by, or at least co-occur with, emotions.

This link is being increasingly investigated by recent research in psychology and neuroscience. For instance, body languages have been found to constitute a signal of affective information [2], [12], [17], [18]. Conversely, people are able to recognize the emotions that are originated from either static postures [29] or dynamic motions [28], probably from so-called "critical features" [26].

Certain subtle general motor behaviours of the human body can be thought of as expressing emotional states. For instance, the tilt angle of head can be linked inferiority- and superiority-related emotions [23]. Since such behaviour can be accounted for part of the cortical neural responses from neurotransmitter dopamine modulation (such as arousal), these bodily states tend to be relatively slow to arise and slow to dissipate. This facilitates people to differentiate emotions from observing these subtle behaviours.

Based on this hypothesis, different applications for emotion recognition have been developed, mainly focusing on facial expressions, such as [15], [25]. However, compared to the extensive research on analysis of emotion from face expressions, little attention has been placed on affective body movement analysis. This is due to two main challenges that need to be overcome in order to develop an application to recognize an emotion from the affective body expression:

1) there are a great number of possible behaviours (e.g. walking, standing, talking) arising from the same emotion,
2) there are also a great number of dimensions of human body skeleton in one movement.

To tackle the first problem, experimental data was often constrained to a finite set of body movements or hand gestures. For instance, [30] carried out studies only in non-affective gestures such as sign languages. In [10] the authors successfully classified six basic emotion expressions from human postures, but they used only computer-generated static figures.

To address the second problem, one approach is to reduce the data dimensions based on their statistical or geometrical correlations. In [1], [13], [16], PCA (principal component analysis) was used to analyse the gesture features. Computer vision related methods can also be an option to detect the correlation between dimensions within the captured images. For instance, the convexity approach was adopted to reduce the large-dimensional feature vectors into a smaller set of points that represents the hand posture by generating convex hull in the pixel space [3], which could be further used for classification method (e.g. [14], [24]).

In this paper, we propose the following answers to address the two above-mentioned challenges:

1) To design a method to analyse body movement without any facial expressions information, with a view to further develop such a method for the generation of expressive behaviours in humanoid robots in the context of the ALIZE-E project[1]. Consequently, *only* the body movements which are comparable to the basic behaviours of our humanoid robot, e.g. walking and standing, were learnt. After that, we used data that had been collected with a body movement tracker with wearable sensors [22], as this provides precise quantitative data of each part of the body.

[1] http://www.aliz-e.org

2) In order to recognize and generate (both trained and novel) behaviours, we aim to construct an expression space between actual behaviour data and the affective space as [5], [9], [21] did. However, different from previous research, the body joint space contains a larger number of dimensions than the facial expression or the head position space. A learning method should be adopted to reduce features and represent slow-changing profiles of the features in a space with smaller number of dimensions. Furthermore, in agreement with of our previous study [22], this behaviour expression space should not be restricted to basic emotions but should also be *continuous*. To fulfil these requirements in dimensionality reduction, we model the behaviour expression space with a variant version of a recurrent neural network model, called recurrent neural network with parametric bias (RNNPB) [27], because of its low-dimensional PB space which can be trained as a continuous expression space with emotional behaviours data and other features.

The rest of the paper is organized as follows. The RNNPB architecture is presented in section 2. In Section 3 we will demonstrate its capabilities to learn, recognize and generate emotion caused behaviours. We conclude with discussion and conclusions in Section 4 and 5.

II. Recurrent Neural Model

Compared to the generic RNN networks which are difficult to implement multiple attractor dynamics, RNNPB is able to generate and recognize multiple temporal sequence patterns by its self-organizing property within an additional layer, called parametric bias units (PB Units). As shown in Fig. 1, an RNNPB is essentially a recurrent network (Elman or Jordan types) with a set of bias units with adjustable values. The parametric bias (PB) units in this recurrent network are connected to the hidden layer as ordinary bias units, but the internal values of them are also updated through back-propagation. Comparing to the generic RNN, the RNNPB owns the additional PB variables which act as bifurcation parameters (which further multiplied with the adjustable weights) for the nonlinear dynamics. Also, the PB units endow the network to have generalization ability to untrained dynamics. Thus, this network has also been further adopted to mimic the process of language acquisition [31], sensorimotor contingency [19] and other cognitive processes.

In this paper, we employed this model to learn spatio-temporal knowledge from perceived behaviours. This information was represented in the PB units and the trained units could be used as emotion recognition and emotional behaviour regeneration. Moreover, due to the generalization capabilities of the recurrent architectures, the RNNPB was able to generate novel spatio-temporal sequences which were not trained before.

Fig. 1 shows the architecture of the RNNPB model. We use $s^b(t)$ to represent the activation and $PB(t)$ to represent the activation of the PB units at the time-step t. In some of the following equations, the time-index t is omitted if all activations are from the same time-step. The inputs to the hidden units y_j are defined as

$$y_l(t) = \sum_i s_i^b(t) w_{li} + \sum_{l'} s_l(t-1) v_{ll'} + \sum_n PB_n(t) \bar{w}_{ln} \quad (1)$$

where w_{li} represents the weighting matrix between the hidden layer and the input layer, $\bar{w}_{ln}$ represents the weighting matrix between PB units and the hidden layer, and $v_{ll'}$ indicates the recurrent weighting matrix within the hidden layer itself.

The transfer functions in the output layer, hidden layer and the PB units all employ the sigmoid function proposed by [20],

$$s = 1.7159 \cdot tanh(\frac{2}{3} y) \quad (2)$$

$$PB_n = 1.7159 \cdot tanh(\frac{2}{3} \rho_n) \quad (3)$$

where y represents the input vector to the neurons in the hidden and output layers and ρ_n represents the internal values of the PB units.

Three running modes (learning, recognition and generation) can functionally simulate different stages between emotion-driven motor behaviours and emotions.

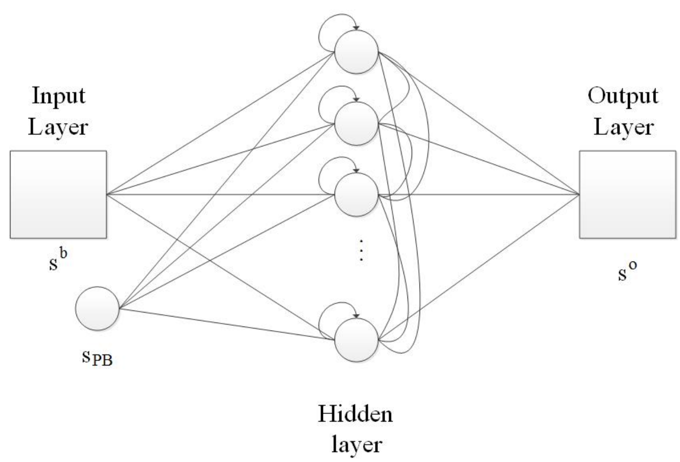


Figure 1: RNNPB with Elman-like connections

Learning mode: The learning is performed off-line, but the internal values in PB units are learned *unsupervised*. When providing the training stimulus for each movement pattern, the weights are updated with BPTT (back-propagation through time). Similarly, the internal values of the PB units are also updated in a self-organizing way from back-propagation. If we refer to one entire learning cycle (all sequences) as an epoch e, in each epoch, the kth PB unit u updates its internal value based on the summation of the back-propagated error from one complete sequence (Eq. 4).

$$\rho_{k,e+1}^{PB} = \rho_{k,e}^{PB} + \eta_l \sum_{t=1}^{T} \delta_{k,t}^{PB} \quad (4)$$

$\delta_{k,t}^{PB}$ represents the back-propagation error of the PB unit u_k at time-step t, $p_{k,e}$ is the output value of parametric units which are transferring into the hidden layer via full connectivity. The learning rate of the PB unit η_l is proportional to the initial learning rates η of the synaptic weights with factor M. T represents the length of the whole training sequence in one epoch e, so the update value of the internal PB value is the error summation from the whole sequence multiplied by the learning rate.

The training progress is basically determined by the cost function:

$$C = \frac{1}{2}\sum_{t}^{T}\sum_{k}^{N}(s_k^b(t+1) - s_k^o(t))^2 \tag{5}$$

where $s_i^b(t+1)$ is the one-step ahead input (as well as the desired output), $s_k^o(t)$ is the current output, T is the total number of available time-step samples in a complete sensorimotor sequence and N is the number of output nodes, which is equal to the number of input nodes. Following gradient descent, each weight update in the network is proportional to the negative gradient of the cost with respect to the specific weight w that will be updated:

$$\Delta w_{ij} = -\eta_{ij}\frac{\partial C}{\partial w_{ij}} \tag{6}$$

Furthermore, all the learning rates η in this network are adaptive, which means that they are adjusted in every epoch. To determine whether the learning rate has to be increased or decreased, we compute the changes of the weighting matrix $w_{i,j}$ between neurons i and j in consecutive epochs:

$$\sigma_{ij} = \frac{\partial C}{\partial w_{ij}}(e-1)\frac{\partial C}{\partial w_{ij}}(e) \tag{7}$$

The update of the learning rate is

$$\eta_{ij}(e) = \begin{cases} min(\eta_{ij}(e-1)\cdot\xi^{+}, \eta_{max}) & \text{if } \sigma_{ij} > 0, \\ max(\eta_{ij}(e-1)\cdot\xi^{-}, \eta_{min}) & \text{if } \sigma_{ij} < 0, \\ \eta_{ij}(e-1) & \text{else.} \end{cases}$$

where $\xi^{+} > 1$ and $\xi^{-} < 1$ represent the increasing/decreasing rate of the adaptive learning rates, with η_{min} and η_{max} as lower and upper bounds, respectively. Thus, the learning rate of a particular weight increases by ξ^{+} to speed up the learning when the changes of that weight from two consecutive epochs have the same sign, and vice versa.

Action recognition mode:

This mode recognizes the types of behaviour sequences by updating the PB units according to the past observation. The information flow in this running mode is mostly the same as in the learning mode, i.e. by back-propagation, except that the synaptic weights are not updated; rather, the error between target and prediction is only back-propagated and updated into the PB units. If a trained sequence is presented to the network, the activation of the PB units will converge to the values that were previously shown in the learning mode, so as to recover the PB values trained before.

The internal values of PB unit are updated by:

$$u_{k,t+1}^{PB} = u_{k,t}^{PB} + \eta_r\sum_{step=t-l}^{t}\delta_{k,step}^{PB} \tag{8}$$

where η_r is the updating rate. The internal values of the PB units are summed from back-propagated error with a specific time length l. The reason of doing a summation is that we expect it can smooth the local fluctuations of the error and decrease its effect to maintain the update of PB unit relatively slow.

Action generation mode: After learning and after the synaptic weights are determined, the RNNPB can act in a closed-loop way, in which the output prediction can be applied as an input for the next time step. In principle, the network can generate a trained sequence by providing initial value of the input and externally setting the PB values (i.e. values in the expression space).

III. DEMONSTRATIONS

In this section, three demonstrations focusing on network learning, movement recognition and movement generation are shown. Computer simulations were done in the first two demonstrations, while in the third demonstration, 3-dimensional avatar movements were generated, which were demonstrated online. These are proof-of-concept demonstrations to use this network as a core model for generating specialised non-verbal robot behaviours in human-robot-interaction. Based on this target,

1) We separated two kinds of training data based on two different behaviours in order to eliminate the subtle differences in different behaviours generation.
2) We kept the network in a small scale since the number of *specialised* movements from one person is limited.

Specifically, the data that used for the network was taken from an inertial motion capture system by Xsens[2], while the actor was performing two sets of basic behaviours (standing and walking) expressing five kinds of emotions (joy, sadness, fear, anger, pride), which was documented in [22]. Three-dimensional skeleton motion capture was used rather than camera-based systems such as Microsoft Kinect, because it would be easier to map the generated motion into humanoid robots for further project developments. Also motions from various body parts could be well-isolated in a quantitative way for network training. Such quantitative data can be reviewed and utilized in external editors and programs (Fig. 2). The sampling frequency of this system was 120Hz.

A. Network Performance

Although the raw data-sets from the system contained various kinds of movement information, such as movement velocity, joint angles, we only utilised the three-dimensional vectors representing the segments' origins in the global coordinate. The sequences were further processed into a 78-dimensional data by calculating the 3-dimensional coordinates of each segment with respect to the centre of mass of the whole body as a velocity-matched reference point (which is also captured and calculate from the system). Then the PCA algorithm was employed as a pre-whitening stage to reduce

[2]www.xsens.com/en/general/mvn

Parameters	Parameter's Descriptions	Value
η	Initial Learning Rate	2.0×10^{-6}
η_{max}	Maximum Value of Learning Rate	1.0×10^{-4}
η_{min}	Minimum Value of Learning Rate	1.0×10^{-8}
η_r	Updating Rate in Recognition Mode	8.0×10^{-3}
M_γ	Proportionality Constant of PB Units Updating Rate	0.001
$n_{b/o}$	Size of Input/Output Layer	4
n_h	Size of Hidden Layer	100
n_{PB}	Size of PB Unit	2
ξ^-	Decreasing Rate of Learning Rate	0.999999
ξ^+	Increasing Rate of Learning Rate	1.000001

Table I: Network parameters

PB	joy1	joy2	pride1	pride2	fear1	fear2	anger1	anger2	sadness1	sadness2
joy1	—	**0.5182**	0.9510	0.9864	2.0909	2.0364	0.7909	0.7591	1.3045	1.4864
joy2	**0.5182**	—	0.7773	0.5364	2.3045	2.3182	1.0773	0.9591	1.4591	1.6773
pride1	0.9510	0.7773	—	**0.1501**	1.7409	1.7636	0.7864	0.6004	0.8273	1.1091
pride2	0.9864	0.5364	**0.1501**	—	1.7682	1.7678	0.7591	0.6136	0.8318	1.1005
fear1	2.0909	2.3045	1.7409	1.7682	—	**0.1227**	1.2818	1.3136	0.6773	0.9227
fear2	2.0364	2.3182	1.7636	1.7678	**0.1227**	—	1.3682	1.2773	0.6682	0.9245
anger1	0.7909	1.0773	0.7864	0.7591	1.2818	1.3682	—	**0.1727**	0.5755	0.7455
anger2	0.7591	0.9591	0.6004	0.6136	1.3136	1.2773	**0.1727**	—	0.6591	0.5545
sadness1	1.3054	1.4591	0.8273	0.8318	0.6773	0.6682	0.5755	0.6591	—	**0.2727**
sadness2	1.4864	1.6773	1.1091	1.1005	0.9227	0.9245	0.7455	0.5545	**0.2727**	—

Table II: PB Values after Training (Standing Behaviour)

PB	joy1	joy2	pride1	pride2	fear1	fear2	anger1	anger2	sadness1	sadness2
joy1	—	**0.2955**	1.2364	1.0682	2.1709	2.3682	1.9136	2.0127	1.4182	1.0545
joy2	**0.2955**	—	0.9773	0.8591	1.9545	2.2002	1.6773	0.7318	1.1227	0.7955
pride1	1.2364	0.9773	—	**0.1636**	0.9909	1.1864	0.8227	0.6818	0.8011	0.2591
pride2	1.0682	0.8591	**0.1636**	—	1.1318	1.3227	0.9955	0.8773	0.8227	0.2318
fear1	2.1709	1.9545	0.9909	1.1318	—	**0.2184**	0.8954	0.7364	1.5502	1.2344
fear2	2.3682	2.2002	1.1864	1.3227	**0.2184**	—	0.8885	0.7509	1.2402	1.4427
anger1	1.9136	1.6773	0.8227	0.9955	0.8954	0.8885	—	**0.1731**	0.8334	0.8995
anger2	2.0127	0.7318	0.6818	0.8773	0.7364	0.7509	**0.1731**	—	0.8975	0.8580
sadness1	1.4182	1.1227	0.2591	**0.2318**	1.5502	1.2402	0.8334	0.8975	—	0.2634
sadness2	1.0545	0.7955	0.8011	0.8227	1.2344	1.4427	0.8995	0.8580	**0.2634**	—

Table III: PB Values after Training (Walking Behaviour)

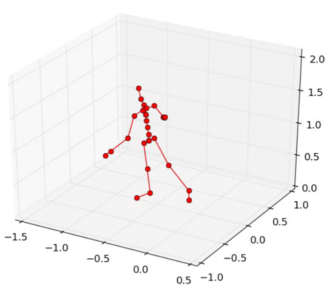


Figure 2: Three Dimensional Skeleton

the dimensions of the dataset to four-dimensional data to obtain a minimal representation by removing the transformed dimensions whose eigenvalue was less than 0.1. The reduced-dimension data was then fed to the RNNPB network, whose parameters were shown in Tab.I. Since we needed to train the PB units as an expression space mapped from emotion space and eliminate the effects from different behaviours, we did two training in two separate RNNPBs (namely, walking RNNPB and standing RNNPB). Each network was trained with the same dataset of one basic behaviour (i.e. walking and standing) but with different emotions.

As we aim to apply different movements in different robot behaviours, we presented the two RNNPBs two sets of training data from two behaviours, walking and standing. To make the training more efficient and practical to map the generated movements into humanoid robots, the sampling rate of the the processed sequence was reduced to 6Hz. Each training set for one network included 10 sequences of the spatio-temporal reduced-dimensional data from 5 basic emotions (i.e. 2 sequences for each emotion) were used for training. Fig. 3a and Fig. 3b showed the PB values of two behaviours. In these figures, the same shape of the markers (except the asterisks) indicated the same emotion. We can observe that PB values from the same emotion were located closer in the PB space after the network was learned, which suggests the PB space became a continuous space for different emotions. Tabs. II and III gave a quantitative measurement of the distance between each PB values in the PB space.

PB	joy	pride	fear	anger	sadness
joy	**0.4909**	0.8858	1.7479	0.8844	0.6358
pride	0.7977	**0.1117**	0.6875	0.7023	0.9044
fear	1.7481	1.3566	**0.3934**	0.8885	0.4349
anger	0.9040	0.8127	1.3534	**0.1935**	0.7175
sadness	1.3579	0.8298	0.9044	0.5324	**0.1612**

Table IV: Distance between Recognized Value and Training Value in the PB Space in Standing Behaviour

PB	joy	pride	fear	anger	sadness
joy	**0.2941**	0.9021	2.0525	1.6026	0.8420
pride	0.8390	**0.2388**	1.3171	0.9807	0.5002
fear	1.9321	0.9808	**0.1916**	0.8895	1.3309
anger	1.6690	0.7975	1.2520	**0.3388**	0.5692
sadness	0.8294	0.3738	1.4204	0.9642	**0.2601**

Table V: Distance between Recognized Value and Training Value in the PB Space in Walking Behaviour

B. Movement Recognition

In the second demonstration, we mainly investigated the recognition mode of the network. Similar to the training, ten new sequences that captured from five emotions and walking/standing behaviours were fed into the walking/standing-trained networks, respectively, to test whether the networks can distinguish the emotion correctly. A stopping criterion is set if the updating of PB values was smaller than a threshold in consecutive 200 times:

$$\|\delta_k^{PB}\| < \epsilon \tag{9}$$

where $\epsilon = 0.001$ is the threshold.

In Figs. 3 we compared the trained PB values and the recognized ones (asterisk markers). Together with the training sets, markers with the same colour implied the same emotions. Tabs. IV and V showed the distance in the PB space between the recognized values and the average training values we obtained in the previous experiment. We can still identify they were locating closer the same emotions in the PB space. In other words, inherited from generic RNN network, the RNNPB network owns generalisation capability to predict/classify untrained sequences.

C. Generation of Behaviours

Given that the PB values representing certain kinds of emotion-related behaviours in the learning and recognition modes, we further adjusted the PB values in the network's generation mode to test the its performance in generating behaviours. These behaviours were represented in a 78-dimensional data-set by reconstructing the network output. These simulations were done by adjusting the PB units into the pre-trained values and the *novel* values as the midpoints between either two pre-trained values in the PB space. In the avatar demonstrations, only the body skeleton were shown by means of the lines connecting to skeleton points. Part of the demonstrations can be viewed online[3]. Apart from the trained emotions, some "interval" emotions from the behaviours can be also perceived when the novel points in the PB spaces were selected.

IV. Discussion and Conclusion

The RNNPB network has recently been applied in various cognitive models since it was proposed a decade ago. It can be accounted for its distinct hierarchical features that can be found in cognitive processes, in which the higher level (PB units) interacts with lower level (input and output layers) in different modes. Particularly, the generation mode is similar as top-down influences in the cognitive processes, in which smaller number of parameters the PB space determine the high DOF (degree of freedom) dynamics in the lower layer. This has also been discovered by motor control in biological system. As Bernstein [6] proposed, generation of certain movements is not so trivial, since the body in most biological system is a highly-redundant manipulator, which is comparable to solve control extreme abundance of DOF systems in a stable manner. In terms of emotional driven behaviours in this paper, certainly those behaviours cannot be solely driven by the emotion states, because behaviours are actually results from numerous factors: such as person decision and situations. Nevertheless, the short-term subtle actions, such as tilt angle of head, are mediated by the emotion states [5], [22].

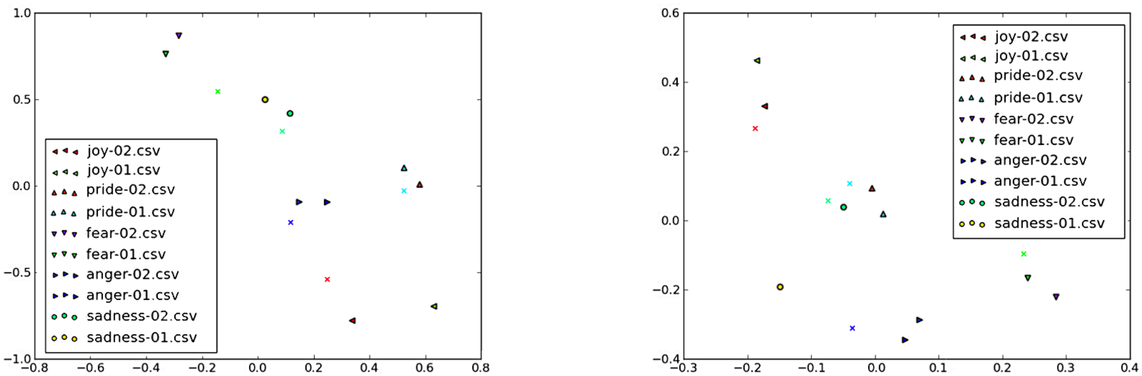
In this paper, we adopted the recurrent neural network with parametric bias (RNNPB) model to train the mapping between an expression space to emotional body movements. The movement data were captured by a movement capture system. A PCA algorithm were first utilised to reduce dimensionality of the raw data as a pre-whitening stage. During training, the values corresponding to the same emotional behaviours are self-organized in the PB space in an unsupervised way. The PB units became an expression space arisen from the same emotion, since the PB space has lower-number of dimensions than the training data. In recognition and generation modes, with various synaptic weight updating schemes, the RNNPB network was also able to recognize and to generate trained behaviour sequences.

Interestingly, as generic recurrent networks, the RNNPB model also showed a generalisation ability during generation mode: using untrained PB values, which located between certain trained behaviours in the expression space, the network could generate certain avatar behaviours that could indicate "interval" emotions of the agent. This also endorsed that the behaviour space should be continuous. In the neural dynamic aspect, the PB units can be regarded as bifurcation parameters with a small number of dimensions to change the topological structure (i.e. different body movements) of the network.

Two future experiments will be conducted:

1) an online quantitative survey will be conducted to investigate the human perception to emotions arisen from the manually generated avatar behaviours;
2) joint angles from motion capture data will be used to train the RNNPB network, in order to allow the humanoid robot NAO to generate non-verbal behaviours

[3] http://youtu.be/9yahOKcEi-A



(a) Coordinates of the PB Vectors of Standing Behaviour

(b) Coordinates of the PB Vectors of Walking Behaviour

Figure 3: PB values of different emotions under two behaviours in learning and recognition modes: the same shape of markers indicate the same emotion; PB values in recognition mode were shown with asterisk markers.

from its expression space.

Acknowledgment

This research was supported by ALIZ-E project (FP7-ICT-248116). The authors are grateful to members of the EECAiA Lab for discussions and feedback.

References

[1] J. Aleotti and S. Caselli. Functional principal component analysis for recogniton of arm gestures and humanoid imitation. *Int. J. of Humanoid Robotics*, 10(04), 2013.

[2] M. Argyle. *Bodily communication*. Routledge, 2013.

[3] P. VA. Barros, N. Junior, J. MM. Bisneto, B. JT. Fernandes, B. LD. Bezerra, and S. MM. Fernandes. Convexity local contour sequences for gesture recognition. In *Proc. of the 28th Ann. ACM Symp. on Appl. Computing*, pages 34–39, 2013.

[4] R. F. Baumeister, K. D. Vohs, C N. DeWall, and L. Zhang. How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. *Personality and Social Psychology Rev.*, 11(2):167–203, 2007.

[5] A. Beck, L. Cañamero, and K. A. Bard. Towards an affect space for robots to display emotional body language. In *Ro-man, IEEE*, pages 464–469, 2010.

[6] N. Bernshtein. Co-ordination and regulation of movements. 1967.

[7] D. Cañamero. Emotional and intelligent: The tangled knot of cognition. In *Proc. of AAAI Fall Symp.*, 1998.

[8] L. D. Cañamero. *Emotional and Intelligent II: the Tangled Knot of Social Cognition*. AAAI Press, 2001.

[9] P. Chopra and J. Meyer. Modeling an infinite emotion space for expressionistic cartoon face animation. In *Comp. Graphics and Imaging*, pages 13–18. Citeseer, 2003.

[10] M. Coulson. Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. *J. of Nonverbal Behavior*, 28(2):117–139, 2004.

[11] T. Dalgleish, M. J. Power, and J. Wiley. *Handbook of cognition and emotion*. Wiley Online Library, 1999.

[12] C. Darwin. *The expression of the emotions in man and animals*. Oxford Uni. Press, 1998.

[13] P. R. De Silva and N. Bianchi-Berthouze. Modeling human affective postures: an information theoretic characterization of posture features. *Comp. Animation and Virtual Worlds*, 15(3-4):269–276, 2004.

[14] M. Dumas. Emotional expression recognition using support vector machines. In *Proc. of Int. Conf. on Multimodal Interfaces*, 2001.

[15] B. Fasel and J. Luettin. Automatic facial expression analysis: a survey. *Pattern Recognition*, 36(1):259–275, 2003.

[16] D. Glowinski, N. Dael, A. Camurri, G. Volpe, M. Mortillaro, and K. Scherer. Toward a minimal representation of affective gestures. *IEEE Tran. on Affective Computing*, 2(2):106–118, 2011.

[17] N. Hadjikhani and B. de Gelder. Seeing fearful body expressions activates the fusiform cortex and amygdala. *Current Biology*, 13(24):2201–2205, 2003.

[18] E. MJ Huis in 't Veld, G. JM Van Boxtel, and B. de Gelder. The body action coding system i: Muscle activations during the perception and expression of emotion. *Social neuroscience*, 9(3):249–264, 2014.

[19] J. Kleesiek, S. Badde, S. Wermter, and A. K. Engel. Action-driven perception for a humanoid. In *Agents and Artificial Intelligence*, pages 83–99. Springer, 2013.

[20] Y. A. LeCun, L. Bottou, G. B. Orr, and K. Müller. Efficient backprop. In *Neural networks: Tricks of the trade*, pages 9–48. Springer, 2012.

[21] H. Lee, J. Park, and M. Chung. A linear affect–expression space model and control points for mascot-type facial robots. *IEEE Trans. on Robotics*, 23(5):863–873, 2007.

[22] M. Lewis and L. Cañamero. Are discrete emotions useful in human-robot interaction? feedback from motion capture analysis. In *2013 ACII*, pages 97–102. IEEE.

[23] A. Mignault and A. Chaudhuri. The many faces of a neutral face: Head tilt and perception of dominance and emotion. *J. of Nonverbal Behavior*, 27(2):111–132, 2003.

[24] J. Nagi, F. Ducatelle, G. A. Di Caro, D. Ciresan, U. Meier, A.o Giusti, F. Nagi, J.n Schmidhuber, and LM. Gambardella. Max-pooling convolutional neural networks for vision-based hand gesture recognition. In *ICSIPA*, pages 342–347. IEEE, 2011.

[25] M. Pantic and L. JM. Rothkrantz. Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, 91(9):1370–1390, 2003.

[26] C. L. Roether, L. Omlor, A. Christensen, and M. A. Giese. Critical features for the perception of emotion from gait. *J. of Vision*, 9(6):15, 2009.

[27] J. Tani, M. Ito, and Y. Sugita. Self-organization of distributedly represented multiple behavior schemata in a mirror system: reviews of robot experiments using RNNPB. *Neural Networks*, 17(8):1273–1289, 2004.

[28] R. D. Walk and C. P. Homan. Emotion and dance in dynamic light displays. *Bulletin of the Psychonomic Society*, 22(5):437–440, 1984.

[29] K. L. Walters and R. D. Walk. Perception of emotion from body posture. 24(5):329–329, 1986.

[30] Y. Wu and T. S. Huang. Human hand modeling, analysis and animation in the context of HCI. In *Proc. ICIP*, volume 3, pages 6–10. IEEE, 1999.

[31] J. Zhong, A. Cangelosi, and S. Wermter. Towards a self-organizing pre-symbolic neural model representing sensorimotor primitives. *Frontiers in Behavioral Neuroscience*, 8:22, 2014.