Affect-Expressive Movement Generation with Factored Conditional Restricted Boltzmann Machines

Omid Alemi, William Li, Philippe Pasquier
School of Interactive Arts and Technology
Simon Fraser University
Vancouver, Canada
{oalemi,dla135,pasquier}@sfu.ca

***Abstract*—The expressivity of virtual, animated agents plays an important role in their believability. While the planning and goal-oriented aspects of agent movements have been addressed in the literature extensively, expressing the emotional state of the agents in their movements is an open research problem. We present our interactive animated agent model with controllable affective movements. We have recorded a corpus of affect-expressive motion capture data of two actors, performing various movements, and annotated based on their arousal and valence levels. We train a Factored, Conditional Restricted Boltzmann Machine (FCRBM) with this corpus in order to capture and control the valence and arousal qualities of movement patterns. The agents are then able to control the emotional qualities of their movements through the FCRBM for any given combination of the valence and arousal. Our results show that the model is capable of controlling the arousal level of the synthesized movements, and to some extent their valence, through manually defining the level of valence and arousal of the agent, as well as making transitions from one state to the other. We validate the expressive abilities of the model through conducting an experiment where participants were asked to rate their perceived affective state for both the generated and recorded movements.**

Keywords—artificial agents; affective movement; full-body movement generation; machine learning

I. Introduction

Human movement is a form of non-verbal communication, which can be characterized along three dimensions: *function*, *execution*, and *expression*. The function dimension of a movement, at a cognitive level, defines the task that the movement is achieving, such as walking to a destination or picking up a cup from a table. This is taken into account by the means-end reasoning in virtual agent literature [1]. Note that for some movements such as dancing, the functional aspect may not be the most relevant characterization. The execution dimension of a movement reflects the pattern of the individual limb motions that constitute a movement. For example, walking is executed through locomotion or picking up a cup can be performed with either the *right* or the *left* arm's motion. The expressive dimension of movement represents the affective qualities that the movement is conveying, reflecting the emotional states felt or communicated by an agent or animated character. In computer animation, the expression of emotions is necessary for increasing the believability of virtual agents [2].

Agent animation can be created manually by animators, or computational models can be used to generate new animation automatically. The shift from linear (e.g., films and comics books) to non-linear (e.g., video games and interactive systems) media has increased the desire to build models for movement animation generation. Non-linear media requires a larger number of assets due to its dynamic and interactive nature. For movement animation, in particular, one needs to create variations of the same movement in order to respond to the need for a diverse set of movements performed in different forms and with different internal emotional states. However, creating a large number of assets manually is costly and time-consuming. Therefore, automatic generation can increase the efficiency in the production of such media. The motivation for automatic movement generation is two-fold: to serve as a computer-assisted creativity tool, and as the motor control for virtual agents.

Computer-Assisted Creativity: traditionally, animators use segments of recorded movements of real actors from a database of motion capture (mocap) data in order to create natural-looking movements. However, this method limits the movements to those that exist in the database, and recording all the possible variations of the same movement is not feasible. By using a generative model, animators can specify the characteristics of the movement segments they desire and the model would generate such movements, not limited to the existing set of movements.

Virtual Agent Movement: movements of an agent reflect its inner beliefs, goals, plans, as well as its affective state. While the literature has extensively addressed the relationship of the first three components with movement [3], [4], modelling the mapping between the affective state and the movement is still an open problem [5], [6].

We present an affect-expressive movement controller based on the generative capabilities of the Factored, Conditional Restricted Boltzman Machines (FCRBMs) [7] trained on a corpus of motion capture data that is tailored specifically for this project. FCRBM has recently been applied to model the style of human movement such as the gait or the speed of the movement [7], [8]. In this paper, we extend the previous models by adding explicit control over the expressive qualities of the movement.

We have recorded the movements of two professional actors performing standing, walking, sitting on a chair, and expressive arm gestures. Following the model of valence and arousal [9] for representing affect, each movement type is performed with 9 different expressive combinations of the valence and arousal (Figure 2), which covers more emotional states than similar existing data sets (e.g., 4 emotions in the University of Glasgow's database [10]). Our choice of movements reflects our intention on building a generative movement model which synthesizes novel movement sequences, according to the goal-directed and the affective behaviour of an agent represented by a set of desired movement characterizations. Our corpus of movements adds emotional variations to those movements used by Motion-Graph-like structures [3] that create streams of movements, making transitions from one type to the other based on a given set of requirements (e.g., following an arbitrary path). In this paper, we present the first stage of this model that addresses the expressive dimension of the agent movements, while adding the support for the function and execution dimensions is among the future direction of our work.

We annotate the mocap data based on their valence and arousal levels and train the FCRBM with its context variable set to the annotations during the learning procedure. Our intention is to use these 9 combinations of valence and arousal in order to build a generalized space of emotions, which allows us to induce any emotional state even if it is not within the original 9 combinations that were captured.

Experiments show that the model is capable of controlling the affective qualities of the generated movements through manually defining the level of valence and arousal of the agent. Furthermore, the model can interpolate and extrapolate between and beyond any two points in the affect space and generalizes well to unseen combinations. This feature can be used to create smooth transitions between two affective states or exaggerate certain states. We validate the ability of the model to convey the intended affective states through an experiment where human observers rated their perceived valence and arousal levels from both the recorded and generated movements. Note that the levels of the valence and arousal on the labels of the training data reflect the emotional states that were instructed to the actors. As the instructed or felt emotions might differ from the emotions felt and perceived by independent observers, we also collect and study the perceived emotions, which can be used as the ground truth for future experiments.

The rest of the paper is organized as follows: Section II reviews the related work in statistical movement generation. Section III explains the background of the machine learning model we use. Section IV outlines our approach to model affect and our design decisions in the choice of movements for the training data set. Section V presents the results of our model. Finally, Section VI summarizes the paper and outlines the future directions.

II. Related Work

Approaches to generate movement animation can be divided into physics-based and data-driven categories. While physics-based methods can successfully generate physically-valid and robust movements, it is challenging to capture the expressive qualities of movement using physical simulation. Data-driven methods, on the other hand, use pre-recorded movement data of real human actors and thus can better capture the expressive qualities of the movement that are visible in the data.

Data-driven movement generation has been approached by interpolating two movement sequences [11], by concatenating short movement clips to make longer, functional movements [12], [3], and by using statistical and machine learning models. While the first two techniques are mostly limited to the movements that are available in the recorded data or their combinations, machine learning models are capable of generalizing movement qualities and generating novel movements.

Hidden Markov Models (HMMs) are used to generate human movement: the style machine extracts the stylistic variations of movements in an unsupervised manner and controls the generation using a set of stylistic degrees of freedom variables [13]. Wang et al. [14] use HMM with mixtures of Stylized Decomposable Triangulated Graph (SDTG) as the probability distribution of its visible states in order to model movement using a supervised style variable. Hidden Semi-Markov Models (HSMM) are used to parameterize the movement pace as well as its style [15]. Another study models expressive gaits during walking by training an HMM on a reduced-dimension space derived using Principal Component Analysis [16].

Gaussian process models are also used in order to separate the stylistic characteristics of movements from their content. Multifactor Gaussian Process Models [17] are able to generate movements with stylistic variations (walking and running) that do not exist in the training data by learning those variations from other types of movement.

Extensions to Restricted Boltzman Machines (RBMs) have been recently applied to model human movement. Conditional RBM [18] is used to generate human movement. The Factored Conditional RBM (FCRBM)[7] extends the CRBM and includes a context unit which modulates the interactions between the hidden and visible units as well as the units from past time steps, which allows for controlling the variations of movements while sampling new sequences. In another work [8], the authors propose a two-layer model called Hierarchical Factored Conditional Restricted Boltzmann Machine (HFCRBM) for learning and interpolating movement style using the middle hidden layer.

The works mentioned above only model movements based on their functional aspect or based on arbitrary expressive variations (e.g., chicken walk vs dinosaur walk). Furthermore, simply choosing between discrete *categories* of gait style does not provide a flexible method of controlling the expressivity of the characters or to producing the transitions from one

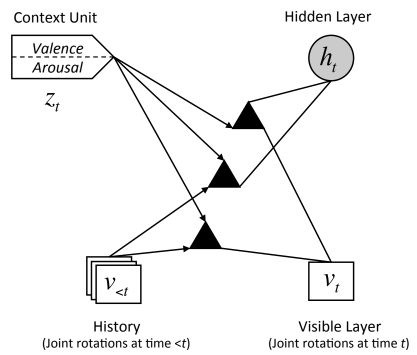

Fig. 1. FCRBM's architecture with valence and arousal labels modulating the interactions between the past visible, current visible, and current hidden units.

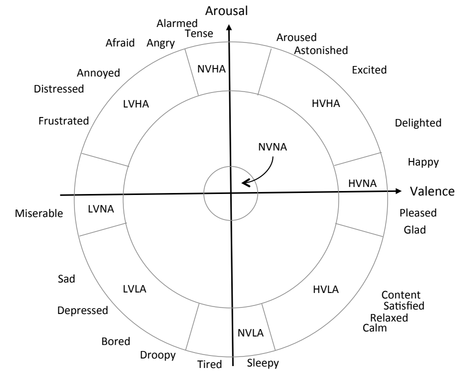

Fig. 2. The affect model described by valence and arousal dimensions with the 9 zones recorded in the training data. The mapping to the categorical emotion labels are based on [20].

emotional state to the other. To the best of our knowledge, no previous study has addressed building a generative machine learning model that allows controlling the expressive qualities of movement using a set of semantically valid variables such as valence and arousal of the affective state of a character.

III. Machine Learning Background

Factored, Conditional Restricted Boltzman Machine [7] is an energy-based machine learning model for capturing the contextual information of time-series data. FCRBM, as shown in Figure 1, consists of a set of visible units, which represent the output at the current time-step, a set of past visible units which represent the history of the output, and a set of hidden units that represent a non-linear interpretation of the output and its past in order to learn the temporal patterns of the training data. It also uses a context unit which controls the interactions between each pair of units. By setting the context unit to the annotation values, the energy landscape of the model changes which allows the model to learn the relationship between the contextual information provided by the annotations and the weights of the connections between the units in an efficient manner.

We use the FCRBM in modelling the expressive qualities of movement as it has a number of advantages over other approaches [19]. First, the hidden states of an FCRBM provide more representational power over HMMs. Second, using a feature variable, FCRBM provides the ability to control the characteristics of generated samples. Third, unlike the Gaussian process models, FCRBM does not need the training data set while sampling new sequences (except a few frames for initializing the model).

The multiplicative, three-way interactions make the parameters of the model cubic. However, by factoring the weight tensors into a product of pair-wise interactions (factors), we can approximate the weight tensor and reduce the order of the parameters to $O(N^2)$. This results in an energy function of the following form:

$$E(v_t, h_t|v_{<t}, z_t) = \frac{1}{2}\sum_i (v_{i,t} - \hat{a}_{i,t})^2 - \sum_f \sum_{ijl} W^v_{if} W^h_{jf} W^z_{lf} v_{i,t} h_{j,t} z_{l,t} - \sum \hat{b}_{j,t} h_{j,t} \tag{1}$$

where v_t and h_t are the visible and hidden units at the current time-step, $v_{<t}$ represents the past visible units, z_t represents the context information (labels) at the current time-step, f represents the index of the factors, $\hat{a}_{i,t}$ and $\hat{b}_{j,t}$ represent the dynamic biases of the visible and hidden units respectively, and W denotes the weight matrix between each unit and a factor. The model can be trained using the Contrastive Divergence algorithm and new samples can be generated by performing alternating Gibbs sampling. For more detailed explanation of the algorithms, refer to the work of Taylor and Hinton [7].

IV. Affect-Expressive Movement

A. Affect Representation

Affective states are typically represented using categorical or dimensional models [21]. Categorical representations define emotions using a set of labels that come from the everyday language uses [22]. Examples are anger, happiness, sadness, surprise, disgust, and fear. Dimensional models break down the affective states into two or more factors, which are represented as a point within a space defined by those factors [20]. The most common example is the PAD model of affect, which defines emotional states based on arousal, valence, and in case of social situations, dominance [9].

We use the arousal and valence dimensions (shown in Figure 2) in order to describe the affective state of movements, as they define the affective state with two degrees of freedom, each across a continuum. This makes it more suitable for learning a generalized model of affect than the categorical models and allows us to create smooth transitions that are essential for interactive applications. In future, we plan to explore other affect dimensions [23]. For example, using dominance to model the interactive and multi-agent scenarios.

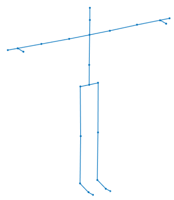

Fig. 3. The skeleton used for the training data

B. Data Gathering

We have captured the movements of two professional actors (one female, one male). The actors were asked to perform standing, walking in different directions, and sitting on a chair as well as expressive arm gestures. These specific movements are chosen in order to be used to build a generative movement model that is capable of synthesizing movements, both on-line and off-line, given any arbitrary set of movement characterizations along the three dimensions of function, execution, and expression. In this paper, we only address the expression dimension. All of the recorded motion capture data and the reference videos are available at *http://moda.movingstories.ca/projects/22-affective-motion-graph*.

Each movement sequence is performed in 9 different expressive modulations, as indicated in Figure 2. Low, neutral, and high levels of valence and arousal are considered. Each modulation of the emotions is expressed by full body movements through mainly the body posture (its shape), the body parts' effort changes, and occasional arm gestures. Each modulation was also repeated 4 times in order to increase the variability of the training data.

Before training, we annotate each sequence using a two-dimensional variable representing their valence and arousal levels. The variable uses a continuous interval to represent the low, neutral, and high levels with the values of 1, 2, and 3 respectively. We decided to use this specific range of numbers after performing experiments with multiple ranges. Note that the values cannot be zero as such condition can cancel the weights in the model. Although the chosen modulations in the training data are discrete points, we rely on the generalization ability of the model and operations such as interpolation and extrapolation in order to generate each intended affective state anywhere in the two-dimensional space and thus the nature of the annotations is continuous.

The movements were recorded with a Vicon motion capture (mocap) system and 53 reflective markers. The final mocap data is mapped to a skeleton with 26 joints as shown in Figure 3. After consideration with movement experts from the Laban Institute of Movement Studies in New York, we decided to use more markers on the spine as it plays an important role in capturing the body shape changes and postures relevant to the expressive dimensions.

C. Data Processing

In order to use the data for the machine learning purposes, we change the representation of the recorded data. The raw mocap data contains a sequence of joints' rotations as well as the position and orientation of the root of the skeleton. The root defines the global position and orientation of the body within a reference coordinate system. The rest of the joints at each time frame are represented by their rotations relative to their parent joint in the skeleton hierarchy. In total, each frame is represented with 72 degrees-of-freedom (DOF).

Initially, the joint rotations were encoded using the Euler angles parameterization which defines the rotations about each axis in a local coordinate system. While widely used, Euler angles parameterization does not always guarantee correct interpolations and can result in the loss of degrees of freedom where different combinations of each of its three components can lead to the same 3D rotation (also known as gimbal lock). Therefore, we convert the root orientation and the joints with 3 DOF to exponential maps [24] in order to avoid gimbal lock and the discontinuities that occur with Euler parameterization of rotations. The final representation of the training data, after removing the dimensions that are constant, contains 52 dimensions.

D. Controlling the Affect in Movement

Our model is defined by a function of the form:

$$M_t = f(M_{<t}, Fun, Exe, Exp) \tag{2}$$

where M_t represents the movement data at time t, $M_{<t}$ represents a set of past movement data, and the Fun, Exe, and Exp represent the function, execution, and expression dimensions of the movement, respectively. This corresponds to our goal to build a parametric motion graph which is capable of generating streams of movement that can be controlled across the three dimensions mentioned above. For example, one can specify that the agent start *walking* from a standing posture and follow a given path, while expressing a *highly aroused emotion*, and then transitioning to *sitting* on a chair with a *neutral arousal level*. In this paper, we only address the expression dimension, which is defined by:

$$Exp = (v, a) \tag{3}$$

where v denotes the level of valence, and a denotes the level of arousal.

In the case of computer-assisted creativity, the animator specifies his or her desired v and a values and provides some initial frames. The initial frames indicate the few poses that the movement starts from and provides a smooth continuation, where the generated movements make a transition from the emotional qualities of the initial frames to the given emotions. In this case, the initial frames can be the last frames of the previous segment of the animation or some frames from the recorded data. By calling the function continuously for the required amount of frames, it generates movement segments that are expressive according to the v and a values.

In the case of controlling a virtual agent's movements, this function can be used as the motor control for the agent, receiving its valence and arousal values directly from the

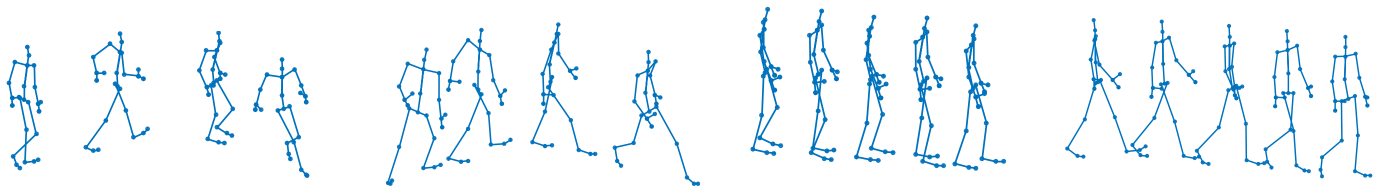

Fig. 4. The generated walking movements. From left to right: trained on the male actor, HVHA; trained on the female actor, LVHA; trained on the male actor, LVLA; trained on the male actor, HVLA.

agent's affective state. At each time step, the function produces the subsequent frame of movement based on the current affective state of the agent while smoothly continuing the movement from previous time steps.

V. Experiments

A. Controlling the Expressivity of Movements

In order to test the model's ability to generate movements based on any given affective state of an agent, we trained an FCRBM with the movements of the actor walking a figure-8-shaped path, down-sampled to 30 frames-per-second, resulting in around 18000 frames that contain all the combinations of the valence and arousal levels of one actor. The data included all the combinations of high, neutral, and low levels of valence and arousal. We used an FCRBM with 400 hidden units, 300 factors for each three-way connection, and at each time-frame, the model was conditioned on six past frames of the data. After 600 epochs, we were able to generate good-quality new samples except for the low valence and low arousal (LVLA) movements from the female actor. We believe that this is due to the very low speed and low energy movements of the female actor for this specific combination in the training data which cannot be captured as well as other combinations by conditioning only on the past six frames. Another shortcoming of the results is the occasional foot sliding, which is due to the lack of constraints on the foot movements in the data-driven approaches.

For the generation, the context unit of the FCRBM is set to the agent's affective state while the state was fixed for each generated sequence. The model was initialized with six frames of the movement from the same affective state from the training data. The results demonstrate that the model was successful in generating new samples as shown in Figure 4. The videos of both the training and the generated movements can be found at http://goo.gl/hL5kJa.

B. Generating Transitions

The affective state of the agent can change gradually over time, and thus we are especially interested in expressing such transitions in the movements of the agent smoothly. For this experiment, we use the same model as above except that we train the model for 2000 epochs. In order to generate transitions, we gradually change the affective state of the agent, and consequently, the feature unit of the FCRBM from one point (e.g., high arousal) to another (e.g., low arousal).

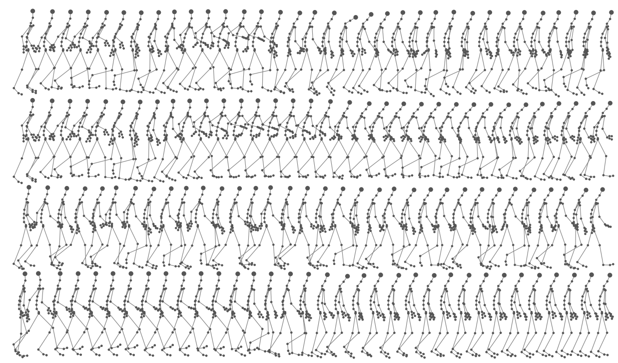

Fig. 5. The generated transitions between two affective states. From top to bottom: HVHA to HVLA, HVHA to LVHA, HVNA to LVNA, NVHA to NVLA. Note that the figures are sampled every 16th frame and are spaced linearly for visualization purposes.

As shown in Figure 5, the generated movements smoothly reflect the changes in the state of the agent. This experiment demonstrates that the model is able to generalize the expressive characterization of movement and generate movements for the combinations of the valence and arousal that do not exist in the training data.[1]

C. Extrapolation

The model can also extrapolate slightly beyond the valence and arousal levels that were intended by the actors and exist in the training data.[2] Extrapolation can be seen as a form of exaggeration, which is suggested in the computer animation guidelines as a way to improve the perception of an affective state [25]. For example, in order to make an agent look happy, the model should generate movements that are more happy than the intended levels by the human actors.

D. Validation of Expressivity

In order to assess the quality of the training data, as well as the ability of the system to communicate any given affective state of the agent, we conducted an experiment in which human participants rated the valence and arousal levels they perceived in both generated and recorded movements.

All the movement sequences (10 generated, 12 recorded) were rendered as short video clips of simple skeletal characters. In order to focus only on bodily movements and

[1] The videos of the transitions can be found at http://goo.gl/hL5kJa.

[2] The extrapolated movements are labeled as exaggerated and can be found at http://goo.gl/hL5kJa.

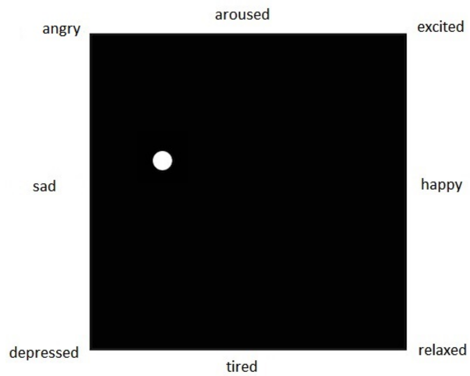

Fig. 6. The affect grid used in the experiment to collect the participants' perception of the affective state.

cues, the characters did not have a face, skin, or clothing. The length of each clip was between 10 to 25 seconds. The order of the clips were randomized for each participant. The experiment was presented using a web-based questionnaire, and the participants were instructed to watch each clip and use a 2-dimensional affect grid (Figure 6) to rate their perceived valence and arousal levels in the movements. The ratings along each dimension were mapped to the range of $[-1, +1]$ In order to avoid any bias against the computer-generated content, the participants were not told that some clips represented computer-generated movements.

Fifteen undergraduate students, in a third-year computer animation course and gathered in a classroom, participated in the experiment. Instructions were given to them on how to use the on-line questionnaire before they individually started watching the videos. There was no time limit for the students to finish the experiment, and they could watch each clip as many times as they wanted.

The means of the responses for the valence and arousal ratings are shown in Figure 7. The participants could successfully classify the arousal levels as high, neutral, and low, although they perceived these levels with less intensity compared to the instructed levels. For example, the high arousal recorded movements were rated on average as 0.44 out of 1.0 in contrast to 1.0 out of 1.0. Overall, the analysis of the responses show high inter-rater reliability (Cronbach's α = 0.89 for valence, 0.98 for arousal).

The participants could identify the neutral and low valence levels correctly, while their ratings of the high valence movements averaged near the center of the spectrum. This suggests that perhaps other cues beyond bodily movements, such as facial expression or voice, are necessary to correctly express and identify the valence level. Another possibility is that the performance of the actors did not contain enough variations along the valence dimension.

The perceptions of the participants for arousal and valence levels of the recorded and generated movements are compared using the Mann-Whitney U test. The arousal level was perceived the same between the recorded and generated

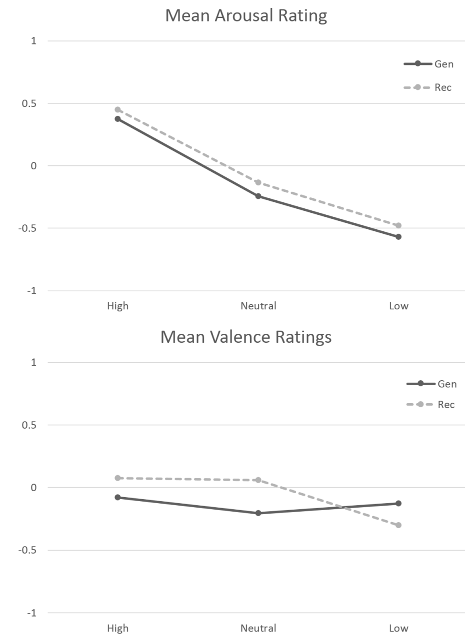

Fig. 7. The mean ratings for valence (top) and arousal (bottom) for recorded (rec) and generated (gen) movements. Size of the error bars (95% confidence interval) are on average 0.0903.

movements ($U = 14541, p < 0.227, N1 = 150, N2 = 180$). However, the valence level of the generated data was marginally perceived as less than the recorded data ($U = 15294, p < 0.038, N1 = 150, N2 = 180$). Overall, the mean rating of the valence of the generated movements is slightly lower than the recorded movements.

VI. Conclusions

We presented a generative model of affect-expressive movements, which allows controlling the emotional qualities of its output. The emotional qualities are represented and modulated by two continuous variables describing the valence and the arousal level of the agent. We applied the model on a data set of walking movements performed by two professional actors while modulating their movements based on different combinations of valence and arousal levels. The validation results show that the model can successfully express the affect along the arousal dimension. However, expressing the valence is shown to be not sufficient at the moment.

As future work, we plan to extend our model towards the following directions: (1) improve the expression of the valence dimension; (2) create a model that allows controlling the function and execution dimensions of the movement as well as its expression, extending the previous works on parametric motion graphs to support affect expression; (3) use the ground-truth labels from the experiment to train the model; and (4) use more dimension than valence and arousal to represent affect, as suggested in the literature [23].

References

[1] Michael Wooldridge, *An Introduction to MultiAgent Systems*, John Wiley & Sons, June 2009.

[2] Joseph Bates, "The role of emotion in believable agents," *Communications of the ACM*, vol. 37, no. 7, pp. 122–125, July 1994.

[3] Lucas Kovar, Michael Gleicher, and Frederic Pighin, "Motion graphs," in *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, New York, New York, USA, July 2002, p. 473, ACM Request Permissions.

[4] Alexander Shoulson, Nathan Marshak, Mubbasir Kapadia, and Norman I. Badler, "ADAPT: the agent development and prototyping testbed," in *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, New York, 2013, pp. 9–18, ACM.

[5] Diane Chi, Monica Costa, Liwei Zhao, and Norman Badler, "The EMOTE model for effort and shape," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, New York, New York, USA, 2000, pp. pp 173–182, ACM Press/Addison-Wesley Publishing Co.

[6] Jing Huang and Catherine Pelachaud, "Expressive Body Animation Pipeline for Virtual Agent," in *Intelligent Virtual Agents*, Yukiko Nakano, Michael Neff, Ana Paiva, and Marilyn Walker, Eds., pp. 355–362. Springer Berlin Heidelberg, Berlin, Heidelberg, Jan. 2012.

[7] Graham W Taylor and Geoffrey E Hinton, "Factored Conditional Restricted Boltzmann Machines for Modeling Motion Style," in *Proceedings of the th Annual International Conference on Machine Learning ICML*, Montreal, 2009.

[8] C.C. Chiu and Stacy Marsella, "A style controller for generating virtual human behaviors," in *The th International Conference on Autonomous Agents and Multiagent Systems*. May 2011, pp. 1023–1030, International Foundation for Autonomous Agents and Multiagent Systems.

[9] Albert Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament," *Current Psychology*, vol. 14, no. 4, pp. 261–292, Dec. 1996.

[10] Yingliang Ma, Helena M Paterson, and Frank E Pollick, "A motion capture library for the study of identity, gender, and emotion perception from biological motion," *Behavior Research Methods*, vol. 38, no. 1, pp. 134–141, 2006.

[11] Taesoo Kwon and Sung Yong Shin, "Motion modeling for on-line locomotion synthesis," in *The 2005 ACM SIGGRAPH/Eurographics Symposium*, New York, New York, USA, July 2005, pp. pp 29–38, ACM.

[12] Luis M Tanco and Adrian Hilton, "Realistic synthesis of novel human movements from a database of motion capture examples," in *Workshop on Human Motion*. 2000, pp. pages 137–142, IEEE.

[13] Matthew Brand and Aaron Hertzmann, "Style machines," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. 2000, pp. 183–192, ACM Press/Addison-Wesley Publishing Co.

[14] Yi Wang, Zhi-Qiang Liu, and Li-Zhu Zhou, "Learning Style-directed Dynamics of Human Motion for Automatic Motion Synthesis," in *IEEE International Conference on Systems, Man and Cybernetics, 2006. SMC '06*, Oct. 2006, pp. 4428–4433.

[15] N. Niwase, J. Yamagishi, and T. Kobayashi, "Human Walking Motion Synthesis with Desired Pace and Stride Length Based on HSMM," *IEICE TRANSACTIONS on Information and Systems*, vol. E88-D, no. 11, pp. 2492–2499, Nov. 2005.

[16] Joëlle Tilmanne and Thierry Dutoit, "Expressive gait synthesis using PCA and Gaussian modeling," in *Proceedings of the Third international conference on Motion in games*, Berlin, Heidelberg, 2010, pp. 363–374, Springer-Verlag.

[17] Jack M Wang, David J Fleet, and Aaron Hertzmann, "Multifactor Gaussian process models for style-content separation," in *Proceedings of the 24th international conference on Machine learning*, New York, NY, USA, 2007, pp. 975–982, ACM.

[18] Graham W Taylor, Geoffrey E Hinton, and Sam T Roweis, "Modeling human motion using binary latent variables," in *Advances in neural information processing system*, 2006, pp. 1345–1352.

[19] Graham William Taylor, *Composable, distributed-state models for high-dimensional time series*, University of Toronto, 2009.

[20] Robert Plutchik and Hope R Conte, *Circumplex Models of Personality and Emotions*, Amer Psychological Assn, 1997.

[21] Michelle Karg, Ali-Akbar Samadani, Rob Gorbet, Kolja Kuhnlenz, Jesse Hoey, and Dana Kulic, "Body Movements for Affective Expression: A Survey of Automatic Recognition and Generation," *Affective Computing, IEEE Transactions on*, vol. 4, no. 4, pp. 341–359, 2013.

[22] J. L Tracy and D. Randles, "Four Models of Basic Emotions: A Review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt," *Emotion Review*, vol. 3, no. 4, pp. 397–405, Oct. 2011.

[23] Johnny R J Fontaine, Klaus R Scherer, Etienne B Roesch, and Phoebe C Ellsworth, "The World of Emotions is not Two-Dimensional," *Psychological Science*, vol. 18, no. 12, pp. 1050–1057, Dec. 2007.

[24] F Sebastian Grassia, "Practical Parameterization of Rotations Using the Exponential Map," *Journal of Graphics Tools*, vol. 3, no. 3, pp. 29–48, 1998.

[25] John Lasseter and John Lasseter, *Principles of traditional animation applied to 3D computer animation*, vol. 21, ACM, Aug. 1987.