Emotionally Expressive Motion Controller for Virtual Character Locomotion Animations

Diogo Silva
INESC-ID & Instituto Superior Técnico
Universidade de Lisboa
Lisbon, Portugal
diogo.goncalves.silva@tecnico.ulisboa.pt

Pedro A. Santos
INESC-ID & Instituto Superior Técnico
Universidade de Lisboa
Lisbon, Portugal
pedro.santos@tecnico.ulisboa.pt

João Dias
Faculty of Science and Technology
University of Algarve
Faro, Portugal
jmdias@ualg.pt

***Abstract*—Style and emotional expressiveness are essential aspects of virtual character computer animation. For a virtual character to display different emotions, motion capture data conveying each desired style has to be recorded, even if the baseline motion is the same. Animators then have to refine and conjoin each recording in order to create the final animations making it a timely and costly process. Although there have been efforts made into the automatic generation of motions, the problem persists that, for each new desired emotion, reference data displaying said emotion has to be readily available and a new motion has to be learned from scratch. By combining Machine Learning with Emotion Analysis - in particular Laban Movement Analysis and the Pleasure, Arousal, Dominance Emotional State Model - we have developed a system that is capable of not only identifying the perceived emotion of locomotion animations but that also allows users to alter the character's expressed emotion in real time and without the need of additional data.**

***Index Terms*—computer animation, machine learning, sentiment analysis, motion synthesis**

I. Introduction

Enabling a virtual character to convey different emotions is paramount to making that character feel realistic, believable and impactful. Animators are tasked, not only with creating the character's baseline animation, but also tweaking its body language to make it able to express different emotional states. The problem then lies in the fact that, should animators want their character to convey different emotions, they would need to record actors portraying the same motion in all desired emotions. For example, if an animator wants a character to walk sadly and happily, they need to gather mocap data of the same walking animation but with the actor conveying these emotions. They then need to generate an entirely new animation for each emotion, be it through manual computer animation, or through automatic motion learning systems [4], [6]. This process gets repeated for every different emotion.

We propose a novel solution in the form of a tool, capable of both emotional discernment and motion generation through the combination of Machine Learning (ML), the Pleasure, Arousal, Dominance Emotional Model (PAD) [5] and Laban Movement Analysis (LMA) [2]. The developed Emotionally Expressive Motion Controller (EEMC) framework, shown in Figure 1 focuses on locomotive motions and allows users to edit the virtual character's expressed style and emotion in real-time, any number of times, without slowing down or stopping the animation and without the need for any additional mocap data or motion learning training. Moreover, our system works not only with Kinematic mocap data but also automatically generated Physics-Enabled Policy based controllers [4].

This work was partially supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference FCT: UIDB/50021/2020 and Project SLICE reference: PTDC/CCICOM/30787/2017

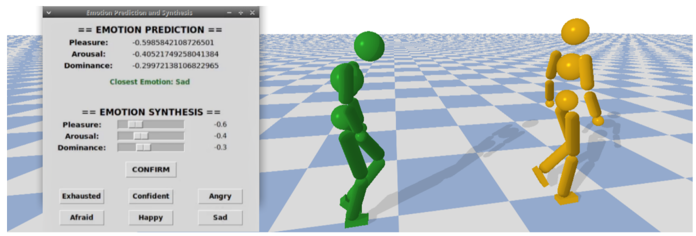

Fig. 1. A baseline motion (right) and a physics-enabled policy-controlled character (left) whose movement has been altered to showcase "Sadness".

II. Emotionally Expressive Motion Controller

The EEMC system can be subdivided into several core sub modules. Figure 2 illustrates the system's overall architecture. At the core lies a character controller loaded with the baseline animation. This controller can either be Kinematic, driven directly by mocap, or a learned Policy-Based Physics-Enabled. A dataset of 78551 LMA Feature sets ,labelled according to their emotion's PAD coordinates [3], was used to train our models. Each feature set was composed of 25 LMA Features from the Body, Shape and Effort LMA Categories such as "Max Hand Distance", "Feet Speed" and "Body Volume".

A. Emotional Discernment

To classify the motion's perceived emotion a set of three Gradient Tree Boosting Regressors were trained to map our set of 25 LMA Features into each PAD coordinates. The LMA Feature dataset was first standardized and then split into a train/validation (80%) and test set (20%). All features extracted from the same animation were either placed into the train or test set. Hyper parameter tuning was done for each regressor using Random Search 10-Fold Cross Validation. The final models managed to a Mean Absolute Error (MAE) of 0.02, 0.06 and 0.03 using the Test set for the Pleasure,

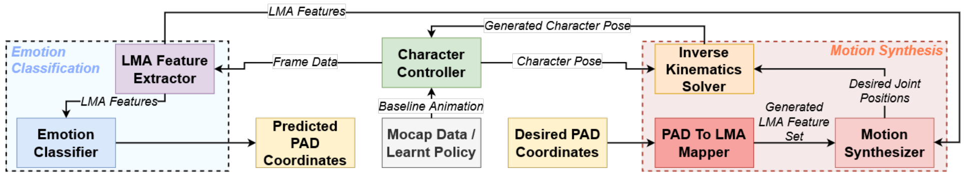

Fig. 2. Emotional Classification uses LMA Features as input to a set of Gradient Tree Boosting Regressors to output the animation's current PAD coordinates. Motion Synthesis receives new PAD coordinates, generates LMA Features, and uses Heuristic Rules and Inverse Kinematics to output a new pose.

Arousal and Dominance coordinates, which corresponds to an error of under 5% of the total value spectrum ($[-1.0, 1.0]$). Using the trained predictors it is then possible to identify a given motion's perceived emotion in real time. During an animation's playtime LMA Features are extracted at every keyframe. After a list of 10 LMA Feature sets has been stored a new multithreaded process is started. This process uses the predictors to compute the PAD coordinates for each of the sets. Each coordinate's predictions is then averaged and output.

B. Motion Synthesis

The first step of motion synthesis is generating a new set of LMA Features for the new desired PAD coordinates. To do so, an *Autoencoder* was created to convert the 25 LMA Features into a 5 dimensional Latent Feature space and vice-versa. This was done to decrease the overall complexity of the PAD-LMA mapping problem [9]. A set of 5 Gradient Tree Boosting regressors was then trained to map PAD coordinates into each of these features. After training, this methodology achieved an overall MAE of 0.19 between the predicted coordinates of the generated LMA Feature set and the original ones. Given a new set of desired PAD coordinates it is possible to synthesize and apply motion changes to the character in real time. A set of 6 Heuristic Rules was designed, each responsible for tweaking a core joint - Hips, Chest, Hands, Elbows, Feet and Neck. Changing upper body joints was the main focus as these tend to have the most impact on the conveyed emotion [1]. Each of these rules works by taking into account the current position or rotation of the joint its trying to change and one or more associated coefficients. Whenever a new set of PAD coordinates is provided, new values for our set of LMA Features are created. These generated LMA Features. together with the animation's recorded LMA features, are utilized to compute the coefficients used in our heuristic rules. Each rule is associated with a different subset of LMA Features and its associated coefficients are computed by finding the value that minimizes the distance between the corresponding subset of recorded and generated LMA features. All coefficients are initialized at 1.0 and are minimized using Powell's method [7]. Afterwards, the system then synthesizes the changes to the pose necessary to convey the desired emotion. These changes are then handed to an Inverse Kinematics Solver to compute the character's new pose at any given frame.

III. Results & Conclusion

Two sets of user tests were conducted to validate the EEMC system's synthesis emotional quality, when compared to reference mocap animations. Each test counted with the collaboration of 40 anonymous paid participants. The first task - "Emotional Identification" - had participants view a clip of an animation - be it synthesized or not - and select from a list of preset emotions, which they thought was most indicative of the character's emotion. The second task - "Primed Emotional Agreement" - was used because certain emotions have intrinsic ambiguity when lacking context which might influence the results of first test [8]. This test showed participants a clip and told them what emotion the character was trying to express. They were then asked to rate how much they agreed with that statement.

Through these tasks we were able to ascertain that, although there was statistically significant differences between the answers provided depending on whether the clip was synthesized or not ($p > 0.05$) for some emotions, most of them managed to blend seamlessly with the reference clips. Overall, generated emotions managed to be easily identified by the participants on the first task, and rate highly on the agreeing scale on the second, therefore proving the effectiveness of the EEMC system for altering a baseline motion into conveying different emotions.

References

[1] Aristidou, A., Zeng, Q., Stavrakis, E., Yin, K., Cohen-Or, D., Chrysanthou, Y., Chen, B.: Emotion control of unstructured dance movements. In: Proceedings of the ACM SIGGRAPH/Eurographics symposium on computer animation. pp. 1–10 (2017)

[2] Groff, E.: Laban movement analysis: Charting the ineffable domain of human movement. Journal of Physical Education, Recreation & Dance **66**(2), 27–30 (1995)

[3] Hoffmann, H., Scheck, A., Schuster, T., Walter, S., Limbrecht, K., Traue, H.C., Kessler, H.: Mapping discrete emotions into the dimensional space: An empirical approach. 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC) pp. 3316–3320 (2012)

[4] Ma, L.K., Yang, Z., Tong, X., Guo, B., Yin, K.: Learning and exploring motor skills with spacetime bounds. In: Computer Graphics Forum. vol. 40, pp. 251–263. Wiley Online Library (2021)

[5] Mehrabian, A.: Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. Current Psychology **14**(4), 261–292 (1996)

[6] Peng, X.B., Abbeel, P., Levine, S., van de Panne, M.: Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. ACM Transactions on Graphics (TOG) **37**(4), 1–14 (2018)

[7] Powell, M.J.: An efficient method for finding the minimum of a function of several variables without calculating derivatives. The computer journal **7**(2), 155–162 (1964)

[8] Reynolds, R., Novotny, E., Lee, J., Roth, D., Bente, G.: Ambiguous bodies: The role of displayed arousal in emotion [mis] perception. Journal of Nonverbal Behavior **43**(4), 529–548 (2019)

[9] Wang, Y., Yao, H., Zhao, S.: Auto-encoder based dimensionality reduction. Neurocomputing **184**, 232–242 (2016)