A Machine Learning Approach to Automate Facial Expressions from Physical Activity

Tarik Boukhalfi
École de technologie supérieure
Montreal, Canada
tarik.boukhalfi.1@ens.etsmtl.ca

Christian Desrosiers
École de technologie supérieure
Montreal, Canada
christian.desrosiers@etsmtl.ca

Eric Paquette
École de technologie supérieure
Montreal, Canada
eric.paquette@etsmtl.ca

ABSTRACT

We propose a novel approach based on machine learning to simulate facial expressions related to physical activity. Because of the various factors they involve, such as psychological and biomechanical, facial expressions are complex to model. While facial performance capture provides the best results, it is costly and difficult to use for real-time interaction during intense physical activity. A number of methods exist to automate facial animation related to speech or emotion, but there are no methods to automate facial expressions related to physical activity. This leads to unrealistic 3D characters, especially when performing intense physical activity. This research highlights the link between physical activity and facial expression, and to propose a data-driven approach providing realistic facial expressions, while leaving creative control. First, biological, mechanical, and facial expression data are captured. This information is then used to train regression trees and support vector machine (SVM) models, which predict facial expressions of virtual characters from their 3D motion. The proposed approach can be used with real-time, pre-recorded or key-framed animations, making it suitable for video games and movies as well.

Keywords

Facial Animation - Biomechanics - Physical Activity - Machine Learning.

1 INTRODUCTION

Facial animation remains one of most tricky, time-consuming, and costly aspects of 3D animation. Facial expressions are difficult to model because of the numerous factors underlying them: emotions (joy, sadness, etc.), mouth movements (speech, deep breath, etc.), eye and eyelid movements (blinking, gaze direction, etc.) and physiological (fatigue, pain, etc.).

Different approaches for automating facial expressions related to emotion or speech exist, but none are available to automate expressions related to physical activity. In the visual effects and computer animation communities, facial animations are most often key-framed or motion-captured. Even though this is a relatively long and costly procedure, it is understandable for main characters. For secondary characters, such as a crowd, the facial animation related to physical activity will often be disregarded. In video games, although characters often have to provide significant physical exertion, facial animation related to this component is somewhat neglected. It is sometimes present during cinematic sequences, but it suffers from a crude approximation during gameplay, and it is often simply overlooked. According to discussions we have had with video game companies, current approaches for gameplay facial animations related to physical activity rely on ad hoc techniques based on linear functions and thresholds. Such approaches are far from the complexity of human facial animations, in relation to physical activity. In this paper, we propose a novel approach based on machine learning to simulate facial expressions related to physical activity, in order to improve the realism of 3D characters. The approach is based on the analysis of motion capture data acquired from real exercise sessions. Given the captured animations and physiological data, specific machine learning techniques are selected to enable the synthesis of facial expressions corresponding to physical activity. The main contributions of the proposed approach can be summarized as:

- A machine learning framework to derive facial expressions from physical activity;
- An approach to link mechanical, physiological, and facial measurements;
- An analysis of the most effective way to compute energy values for machine learning purposes;
- A set of empirical rules relating physical activity to specific facial expressions;
- A normalization procedure to make better use of heart rate and blend shape data.

With this machine learning framework and captured data, we are able to synthesize realistic facial expressions. The approach can be used for real-time as well as off-line facial animation. Furthermore, the method allows for control over stylization, as key-framed data could be used instead of captured data. It enables control over expressiveness, as the animator can adjust various parameters that have an impact on the facial expression of the character. Finally, the models developed in this work also provide metabolic data that could be used for purposes other than facial animation.

2 RELATED WORK

Previous works are categorized in four topics: the description of facial expressions, the synthesis of speech-related expressions, the synthesis of emotion-related expressions, and the description of facial expressions related to physical activity.

2.1 Facial Expression Coding

Several objective and systematic approaches to encode facial expressions have been proposed. Although facial expressions are due to a wide range of factors, only facial changes due to emotions, intentions or social communication are taken into account [17]. Various coding systems have been developed mainly for psychological studies, including FACES (Facial Expression Coding System) [14] and FACS (Facial Action Coding System) [8], which are presented in a survey paper [19]. FACS is an anatomically-based expression space grouping together facial muscle groups as AUs (Action Units), whose combination can be used to form any possible expression [26]. The MPEG-4 standard proposes a similar approach using FAPs (Facial Action Parameters), which has been used in various research projects. In the proposed approach, the facial expressions are built using blend shapes that correspond to facial muscle groups similar to the FACS approach.

The automation of the coding process generally relies on video tracking software or motion capture systems that require complicated setup. In recent years, novel techniques emerged using depth cameras such as *FaceShift* [4] or *Brekel ProFace* [6]. Other software use simple webcams, such as *Mixamo Face Plus* [22], *di-o-matic Maskarad* or *Emotient* software [18]. While most of the available software extract a set of facial features or blend shapes, *Emotient* software extracts Ekman's facial expressions and a set of Action Units.

2.2 Speech-Related Facial Expressions

Animation synthesis is generally done by analyzing an audio input, extracting phonemes, and then animating the 3D face model's visemes (phoneme's visual counterpart) [15]. Different approaches have also been developed to enhance realism in animation, such as blending speech-driven animation into emotion-driven animation and using anatomically-based structures [16]. Other works [31] have focused on improving the visual behavior related to speech. Some works use machine learning methods such as SVMs [33] or neural networks [7, 21]. All of these methods help to achieve more realistic results in facial expressions related to speech, and substantially reduce manual animation time. They give good results, given that there is a single input (the audio) that captures all of the required information to adjust the facial animation.

When considering physical activities, a character's motion involves several limbs, as well as potential and kinetic energy, torques, etc., which results in a broader set of inputs. Furthermore, simulating speech-related animation from audio is synchronized to one input signal, while the facial expression of the character's motion might be the result of both its instantaneous motion and the movements or activities performed by the character in the past few minutes. Finally, the character's motion triggers facial expressions that will influence parts of the face that are not related to speech, such as the region around the eyes. For these reasons, works dealing with speech cannot fully solve the problem of generating the facial animation related to physical activity.

2.3 Emotion-Related Facial Expressions

Researchers in psychology studied emotions and came up with classifications based on a limited number of emotions. To further simplify the relationships between emotions, they can be represented in simpler 2D expression spaces [24, 28]. Computer graphics researchers have taken advantage of such approaches and proposed different two dimensional emotion layouts that allow a meaningful blending between emotions [25]. Other approaches have relied on coding systems such as FACS to provide realistic transitions between emotion expressions [1]. While these works provide interesting approaches for the transition and blending of facial expressions, they work when the emotion is already known, and when a set of face poses is provided. Animating the right combination of emotions through time remains a complex problem. It is similar to the challenge involved in this work: developing an approach that can predict the facial expression from observations and models describing how a human subject reacts in different circumstances.

2.4 Physical Activity-Related Facial Expressions

Even though 3D characters often perform intense physical activities, we could not find any research addressing the automatic and realistic facial animation related to physical activity. Outside of the computer graphics field, the work of McKenzie [20] describes the facial expressions related to substantial effort, exhaustion,

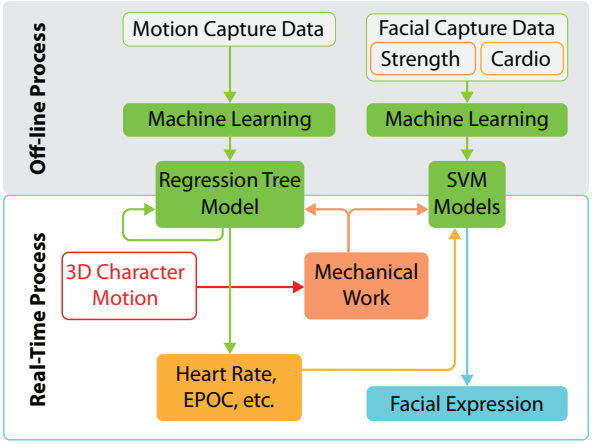

Figure 1: Machine learning facial expressions

and fatigue. In the computer graphics field, facial animation literature related to physical activity is found mostly in art books [9, 11]. These contain numerous facial expressions, some of which are related to physical activity. While these works provide useful information for manually animating expressions, they are not useful for the automatic facial animation from the character's motion. Zordan's work [34] target the modeling and control of 3D models in relation to respiration.

Numerous works address facial animation in the context of speech and emotion, but they are not adapted to the synthesis of physical activity expressions. Based on machine learning and captured data, the proposed approach derives a model, to animate facial expressions.

3 FACIAL EXPRESSION SYNTHESIS

Fig. 1 shows an overview of the proposed approach. The first step is to acquire real-life motion capture data, providing information on the facial expressions observed under various types of exercise. These data are used to train machine learning models, which are then used to generate realistic facial expressions.

3.1 Data Acquisition

Various capture sessions were conducted in order to gather the information required to develop a model that gives realistic results. During these sessions, information describing the type of activities and physical state (heart rate and facial expressions) were recorded.

A full-body motion capture was done in a large room where 15 participants of different age (20 to 46 years old) and training level (0 to 7.5) exercised freely without training devices. The resolution of the motion capture cameras did not allow facial capture along with the full-body capture. Furthermore, using training devices would have led to markers occlusion. For these reasons, the capture was split in two sessions: full-body and facial. The goal of the full-body session was to provide data to establish the relationship between the motion and physiological measures. The participants were asked to alternate between exercises of low and high levels of intensity, and to slow down to ensure that a large range of data was acquired. While participants were training, both full-body motion and heart rate data were recorded. The software used to record the heart rate provided an estimation of other metabolic indicators, such as metabolic energy consumption, breathing rate and EPOC (Excess Post Oxygen Consumption).

The second capture session was done at a fitness center, where 17 participants from the same age groups and training level as the previous session, were asked to exercise on either a cardiovascular training machine or a mixture of strength training machines and free weights. The goal of this capture session was to establish the relationship between motion, heart rate, and facial expression. Again, this capture involved exercising at different levels of intensity and a slow-down period. Using this procedure, the data collected for each exercise included repetitions for the same participant as well as for different participants. Facial expressions were filmed, while heart rate data were recorded following the same procedures and with the same material as during the first session. Together with the height and weight of the participants, the specific loads used with the strength training machines and free weights allowed for a good approximation of the involved work and forces.

3.2 Biomechanical Model

One of the key inputs to both the off-line and online phases of the proposed approach is the mechanical work resulting from the motion. Different methods were evaluated to approximate the work: potential energy, translational kinetic energy, and rotational kinetic energy. Different ways of evaluating the mechanical energies were tested: using the center of mass of the whole character, using the lower/upper body, and computing these values for each limb. Potential energy, translational and rotational kinetic energies were used.

Tests were conducted to find an approach that would be efficient to compute, while providing good results for both the learning and synthesis phases. While separate inputs for each limb intuitively seemed to provide better knowledge about the type of exercise and effort, they resulted in noisy facial animations with blend shape weights that changed too rapidly compared to the real data. An explanation for this phenomenon is that even though there are several captures, the amount of input data is still too small to correctly capture the intricate interrelations between specific limbs and facial expressions. Ultimately, what provided the best result was using the sum of the mechanical work (potential, translational kinetic and rotational kinetic) for all limbs.

3.3 Machine Learning Facial Expressions

To get a better understanding of the underlying mechanisms and relations between the exercises and the facial expressions, a preliminary analysis of the data was conducted. As the relations between the metabolic, mechanical and facial parameters are too complex to model using simple polynomial equations, it made sense to use machine learning. Given the type of captured data and the kind of predictions required, regression techniques were the most appropriate. Several models were trained using different sets of features as input, and the quality of these models was evaluated. Likewise, appropriate model parameters were selected using cross-validation. The data flow, learning approaches and models are presented in Fig. 1.

3.3.1 Metabolic Parameters Prediction

To predict the heart rate from the character's motion, various learning techniques were tested with different combinations of features as input. The heart rate increases or decreases depending on the intensity of the exercise: for each person, there is a certain threshold in exercise intensity that results in an increase or decrease in the demand for oxygen. To model this behavior, regression trees were found to give an accuracy comparable to more complex models such as SVM. Furthermore, this technique was also selected for its ability to provide a human-interpretable model, which can be used to get more artistic control on the final result.

Since the range of input values affects learning techniques, and as the range of heart rates varies among the participants, the data were transformed to the $[0,1]$ interval resulting in the normalized heart rate (NHR):

$$\text{NHR} = \frac{\text{current heart rate} - \text{resting heart rate}}{\text{maximum heart rate} - \text{resting heart rate}}$$

The maximum and resting heart rates are found in standard training charts based on the age and training level.

A regression tree model was built using the UserClassifier in the *Weka* software [12]. Several combinations of inputs were tested (mechanical work as input and heart rate as output, mechanical work and heart rate difference as input and heart rate difference as output, etc.). Among the tested models, the one providing the best results was to predict the difference in heart rate using the current heart rate and the instantaneous mechanical work. By using the last predicted NHR in the subsequent prediction, the model considers the temporal information and the accumulated fatigue.

While a model trained using the data from a single participant could accurately predict the heart rate of this participant (correlation coefficient of 0.88 and root-mean-square error – RMSE – of 19%, see Fig. 2(a, c)), combining the data from every participant in a single model resulted in significant errors (correlation coefficient of 0.21 and RMSE of 79%, see Fig. 2(b, d)). To improve the results, simpler models were derived.

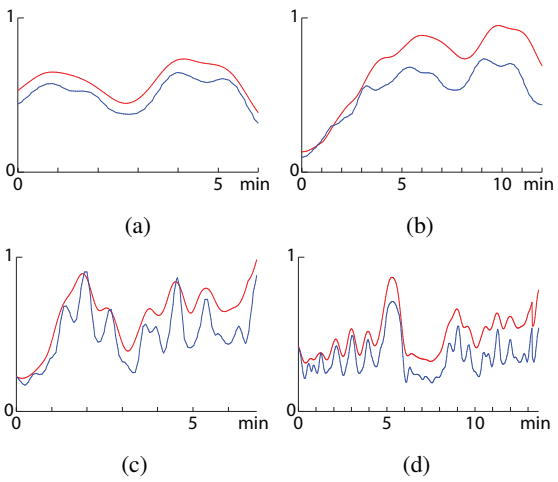

Figure 2: Comparison between real (blue) and predicted (red) NHR as a function of time (in minutes). (a-b) cardiovascular and (c-d) strength training exercises. (a,c) the same and (b,d) different participants.

Using appropriate regularization parameters and tree pruning yielded simple regression trees with only two leaves. By analyzing these simpler models for each participant, a common structure emerged: all regression trees had the same test at the root node, comparing the mechanical work w to a threshold value t_{root}, which broke the predictions into increases or decreases in the heart rate. Moreover, the main difference between these personalized models was the threshold value used in the test, this value depended largely on the fitness level of the participant. Based on these observations, a linear regression between the training level and t_{root} was used to improve the model. The resulting model based on the linear regression and regression tree is as follows:

$$t_{root} = 7.13 + 0.42 \times \text{training level}$$

$$\Delta(\text{nhr}) = \begin{cases} c_{inc} \times (w - t_{root}) \times (1 - \text{nhr})^2 & t_{root} < w \\ c_{dec} \times (w - t_{root}) \times \text{nhr}^2 & w \leq t_{root} \end{cases}$$

$$c_{inc} = 0.0056 - 0.00043 \times \text{training level}$$

$$c_{dec} = 0.0009 + 0.00025 \times \text{training level}$$

The threshold t_{root} determines when the heart rate starts to increase while c_{inc} and c_{dec} are the factors of increase or decrease. These values were obtained by calculating a linear regression between the individual values obtained in the regression tree of each participant.

This model provided a prediction that was almost as good as the one for separate participants (correlation coefficient of 0.87, RMSE of 24%). Furthermore, the training level can be used to control the response level of characters to various types of exercises.

The normalized heart rate is used to approximate the oxygen consumption (VO_2) and respiration rate. Both the VO_2 and respiration rate have to be normalized relatively to their minimal and maximal values. Given the normalized values, the VO_2 and respiration rate are proportional to the normalized heart rate [27]. With the VO_2 estimation, EPOC can be approximated [10]. These estimations, together with the mechanical work and mechanical power, are then used to predict the facial animations as will be described in the next section.

3.3.2 Predicting Expression Components

To animate a virtual character, the four weights corresponding to the blend shapes associated with the basic components identified in the preliminary analysis (see Section 4.1.1) should be obtained from the movement of the character. Compared to the metabolic parameters, the facial expressions in our capture data exhibited more sudden and frequent changes. Because of this behavior, regression trees did not provide adequate results to predict blend shape weights.

Instead, we opted for SVM regression, which provided better prediction results and had already been used successfully for facial animation [32]. Tests were conducted with multiple participants, for multiple exercises as well as for single participant and single exercise (see Fig. 3). For a single participant and exercise, the prediction of the participant's blend shapes corresponding to exercises not used to train the model was accurate (Fig. 3 (a), (b)). Compared to what was observed for the metabolic parameter, the prediction from strength training exercises (Fig. 3 (b), (c), and (d)) lines up quite well with the real data, while the prediction for cardiovascular exercises (Fig. 3(b)) follows the general trend of the curve, but presents variations of smaller amplitude due to the regularization of the model.

For multiple participants, training with a single exercise enabled a good prediction of the same exercise for a participant not used in the training data (Fig. 3(c)). Nevertheless, generalization across all participants and exercises was relatively poor. Again, the data had to be normalized, but this time with respect to the expressiveness of the participant. This can be seen in Fig. 3(c), as the curves are well aligned, but the blend shape weights are on a different scale. Some of the participants could endure incredible exertion with a relatively neutral expression while others depicted pronounced expressions. The expressiveness could not be linked to any of the parameters collected about the participants (age, training level, etc.). It still can be computed for each participants by finding the maximal weight for each blend shape, resulting in a four-dimensional expressiveness vector. The captured blend shape weights of the participants were then normalized. This expressiveness vector allows to control the facial expressions by increasing or reducing the expressiveness values.

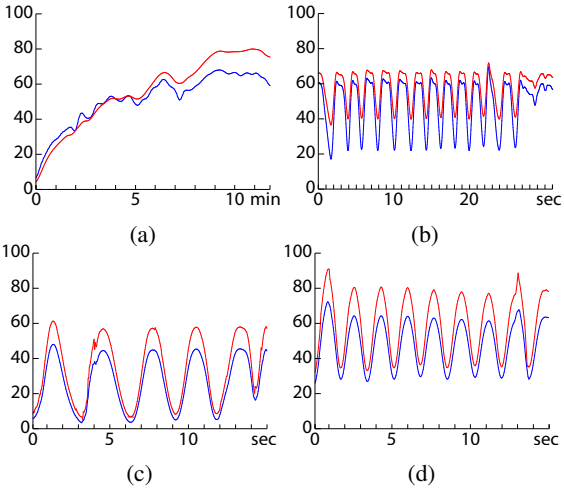

Figure 3: In each case, the acquired data (blue) was not used to train the model for the prediction (red) of the blend shape weight. In (a), (b), and (c), the prediction and SVM model are for the same exercise, while in (d) we tested the prediction for an exercise not used to train the SVM model. The prediction works for (a) cardio and (b) strength training exercises. The motion of a participant not used to construct the SVM model is used for the prediction in (c). In (d), the prediction is for an exercise not used to train the SVM model.

Given these normalized blend shape weights, the whole dataset could be learned using an SVM. To select the best-suited set of inputs and model parameters, several combinations were evaluated on a test data set and using cross validation. The combination that gave the best results was mechanical work, mechanical power, normalized heart rate and EPOC as inputs, and blend shape weight as output. As the predicted heart rate is used as an input for the next prediction (see Fig. 1), the models consider the temporal information and do not only model the correlation at a single-frame level.

With respect to the selection of the SVM parameters, the radial basis function (RBF) kernel was selected, and several combinations of parameters were tested: regularization parameter and gamma of the RBF varying independently from 10^{-10} to 10^{10}. The values of these parameters that produced the best results were different from one blend shape to the other. The regularization parameter ranged from 10^3 to 10^6 while the gamma of the RBF ranged from 10^{-5} to 10^{-2}.

Since different areas of the face reacted in different ways depending on the physical activity and the participant, the predictions use four SVM models. Although these are independent, the predictions were consistent in all of our tests. The training RMSE of the models was in the $[17\%, 26\%]$ range. The final models enabled a good generalization of the captured data, which indicates that our method could be used to generate realistic facial animations on other types of movements.

(a) $\frac{\text{Mouth}}{\text{Exertion/EPOC}}$ (b) $\frac{\text{Mouth}}{\text{Breathing/EPOC}}$ (c) $\frac{\text{Eyes}}{\text{Exertion/EPOC}}$ (d) $\frac{\text{Eyes}}{\text{Fatigue}}$ (e) $\frac{\text{Eyebrows}}{\text{Exertion}}$ (f) $\frac{\text{Eyebrows}}{\text{Fatigue}}$

Figure 4: The physical effort resulted in a broad range of facial expressions

(a) (b) (c)

Figure 5: Results for five minutes of running at medium speed with different training levels: (a) 10, (b) 5, (c) 0

4 RESULTS

In this section, the approach and its experimental evaluation are discussed in the context of the expected use of the models. The accompanying video presents results for different types of exercises. The results present believable expressions based on the physical activity of the character, and these expressions improve the realism of the 3D character (see Fig. 5).

The learned regression tree and SVM models were used to animate facial expressions of a 3D head model. Using the LibSVM implementation to perform the predictions showed that the approach can easily achieve real-time frame rates. On an average computer, 60 to 600 FPS were obtained depending on the render settings and mesh complexity. The prototype can be used either by triggering pre-recorded animation clips or by using a live input from a Kinect device. The user can specify the age, height, weight, fitness level and expressiveness vector of the 3D character, and can also add weights lifted in each hand as well as on the back. The facial expressions change relatively to the motion of the character and the specified parameters. The user can change the character parameters and get a real-time feedback. Fig. 5 shows different results achieved with different training levels and Fig. 6 shows different results obtained by changing the expressiveness vector.

4.1 Discussion

4.1.1 Observations on the Captured Data

A qualitative analysis of the captured data (full-body capture, video and heart rate) was conducted. While these observations helped us in the selection of the appropriate machine learning methods, they should also benefit artists in animating 3D characters. Two types of relations were discovered: general relations that hold for most facial expressions, and specific relations that apply to particular expressions.

The first general rule is that the intensity of expressions is proportional to the displaced mass and inversely proportional to the mass of the muscle. The expression related to the instantaneous exertion is proportional to the mechanical power. The evolution of the expression intensity is proportional to the change in heart rate and metabolic energy. Finally, the recovery time is proportional to the effort intensity and inversely proportional to the training level.

Rules specific to individual components of facial expression were also observed. It was determined that the facial expression features associated with physical activity were concentrated around the eyes and the mouth (see Fig. 4). Regarding the expressions related to the mouth, the stretching is induced by two factors: instantaneous physical exertion and fatigue level. The mouth remains closed at the beginning of the training session. After a certain time, it starts to open, and the opening is linked to the respiration rate and the fatigue level.

Other observations were related to the region of the eyes. Eye squinting is mainly induced by instantaneous physical exertion until a certain fatigue level. At a higher level of fatigue, the eyes tend to relax in connection with the fatigue level with a remaining constant squinting value. The behavior of the eyebrows is a combination of a downward movement related to the physical exertion and an upward movement related to fatigue.

Finally, some observations were made with respect to both the breathing and the swallowing. The frequency of occasional breathing movements related to loud and quick expiration is induced by two factors: fatigue level and respiration rate. The frequency of the occasional gulping is proportional to the fatigue and to the regular respiration rate. These observations helped in defining blend shape selection greatly inspired by the muscular groups of the human face, as described in FACS [8]. The model consisted of a neutral face and four blend shapes that can be used in various expressions linked to physical activity (see Fig. 7).

4.1.2 Manual Blend Shapes Recovery

To simplify the capture sessions, only a video recording of the face was used for the facial expression capture.

(a) (b) (c) (d) (e)

Figure 6: Results obtained with different expressiveness vectors: (a-b) softened expressions, (c) predicted expression, and (d-e) exaggerated expressions. The red lines are added as guides to help in comparing the expressions

(a) Neutral (b) Mouth Stretch (c) Mouth Opening (d) Eyes Squinting (e) Eyebrows

Figure 7: The four blend shapes used in our implementation.

Different techniques were evaluated for retrieving facial animation data, but it was found that manual animation provided more reliable results. To recover objective values that can be used with machine learning approaches, a virtual character's face was key-frame animated to match the expression of the participants. To ensure the results are reproducible, blend shapes were key-framed, one at a time, and always in the same order. Furthermore, to measure how perceptually meaningful the values were, three different people independently adjusted the blend shape weights for a selection of eighteen representative poses. Even though the blend shape weights were not identical, the error remained limited to 11% on average and was considered to be quite sufficient for the purposes of this work.

4.1.3 Limitations

As shown in Fig. 7, the blend shape model used in this work is sufficient, but it does not cover the whole range of expressions. Since each blend shape is predicted separately, there could be inconsistencies in the face of the character. Resolving such inconsistencies and providing a better correspondence could be achieved through a constrained weight propagation [23]. While the mesh deformation used in this paper is based on blend shapes, the models could be learned with the use of other control mechanisms, such as bone systems.

The generated facial expressions are generalizations of the observed data. They correspond to mean values and sometimes lack expressiveness (see Fig. 8). The models sometimes output results that deviate significantly from the observations. As they happen quite infrequently, they can be easily filtered out.

The metabolic prediction model uses its last prediction as input. It is thus subject to error accumulation and could diverge from the observed values over time. Approaches to steer the values back to the observed range should be used to solve such problems.

(a) (b) (c)

Figure 8: Comparison between real and predicted facial expression from one participant to another: (a) generated, (b-c) different participants doing the same exercise.

5 CONCLUSION

By analyzing two sets of captured data, this paper reveals several important observations about what triggers specific facial expressions. A combination of two machine learning techniques was used in order to automatically synthesize some metabolic parameters as well as the facial animation of a 3D character. While being automatic, this approach provides meaningful parameters that animators can change to deliver realistic and compelling facial animations that automatically adjust to the motion of the character. Furthermore, the metabolic parameters provided by the approach could also be helpful in animating other aspects of the character, such as breathing and sweating. Finally, the approach can be used for real-time applications as well as

off-line high quality rendering. The approach provides more realistic characters while reducing the burden of capturing or hand animating the facial expressions resulting from physical activity.

As the manual blend shape animation was a time consuming process, the method was limited to four blend shapes. Automating this process by using novel techniques [5, 29, 30] would allow for a larger number of blend shapes or Ekman's AU's by using the *Emotient* software for even more realistic results. Like other methods described in Section 2, the proposed approach addresses a single aspect of facial animation (only from physical activity). A future work would be to provide a framework that allows mixing different types of expressions through various methods [2, 3, 13]. The proposed approach is deterministic in nature: given the same control parameters and motions, it will result in the same facial animation. An interesting future research would be to incorporate the probabilistic and stochastic nature of human reactions into the models.

6 ACKNOWLEDGMENTS

This work was funded by the GRAND NCE.

7 REFERENCES

[1] A. Arya and A. DiPaola, S. Parush. Perceptually valid facial expressions for character-based applications. *Intl. Journal of Computer Games Technology*, 2009.

[2] T. Beeler, F. Hahn, D. Bradley, B. Bickel, P. Beardsley, C. Gotsman, R. W Sumner, and M. Gross. High-quality passive facial performance capture using anchor frames. *ACM Trans. Graph.*, 30(4):75, 2011.

[3] A. H. Bermano, D. Bradley, T. Beeler, F. Zund, D. Nowrouzezahrai, I. Baran, O. Sorkine-Hornung, H. Pfister, R. W Sumner, B. Bickel, and M. Gross. Facial performance enhancement using dynamic shape space analysis. *ACM Transactions on Graphics (TOG)*, 33(2):13:1–13:12, 2014.

[4] S. Bouaziz and M. Pauly. Dynamic 2d/3d registration for the kinect. In *ACM SIGGRAPH 2013 Courses*.

[5] C. Cao, Y. Weng, S. Zhou, Y. Tong, and K. Zhou. Facewarehouse: A 3d facial expression database for visual computing. *Visualization and Comp. Graph, IEEE Trans. on*, 20(3):413–425, March 2014.

[6] J Chow, K Ang, D Lichti, and W Teskey. Performance analysis of a low-cost triangulation-based 3d camera: Microsoft kinect system. In *Intl. Society for Photogrammetry and Remote Sensing Congress (ISPRS)*, volume 39, pages 175–180, 2012.

[7] P. Eisert, S. Chaudhuri, and B. Girod. Speech driven synthesis of talking head sequences. In *3D Image Analysis and Synthesis*, pages 51–56, 1997.

[8] P. Ekman and W.V. Friesen. *Facial action coding system: investigator's guide*. Consulting Psychologists Press, 1978.

[9] G. Faigin. *The Artist's Complete Guide to Facial Expression*. Watson-Guptill Publications, 1990.

[10] Firstbeat Technologies. Indirect epoc prediction method based on heart rate measurement. White paper.

[11] E. Goldfinger. *Human Anatomy for Artists: The Elements of Form*. Oxford University Press, 1991.

[12] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten. The weka data mining software: An update. *SIGKDD Explorations*, 11:10–18, November 2009.

[13] M. Kapadia, I. Chiang, T. Thomas, N. I. Badler, and J. T. Kider, Jr. Efficient motion retrieval in large motion databases. In *Proc. of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, pages 19–28, 2013.

[14] A.M. Kring and D.M. Sloan. The facial expression coding system (faces): Development, validation, and utility. *Psychological Assessment*, 19(2):210–224, June 2007.

[15] S. Kshirsagar and N. Magnenat-Thalmann. Visyllable based speech animation. *Computer Graphics Forum*, 22(3):631–639, 2003.

[16] S. Kshirsagar, T. Molet, and N. Magnenat-Thalmann. Principal components of expressive speech animation. In *Proc. of Computer Graphics International 2001*, pages 38–44, 2001.

[17] S.Z. Li, A.K. Jain, Y.-L. Tian, T. Kanade, and J.F. Cohn. Facial expression analysis. In *Handbook of Face Recognition*, pages 247–275. Springer New York, 2005.

[18] M. Malmir, D. Forster, K. Youngstrom, L. Morrison, and J. R. Movellan. Home alone: Social robots for digital ethnography of toddler behavior. In *Computer Vision Workshops (ICCVW), 2013 IEEE Intl. Conf. on*, pages 762–768, 2013.

[19] I.B. Mauss and M.D. Robinson. Measures of emotion: A review. *Cognition & Emotion*, 23(2):209–237, 2009.

[20] R.T. McKenzie. The facial expression of violent effort, breathlessness, and fatigue. *Journal of anatomy and physiology*, 40:51–56, October 1905.

[21] W. Zhen P. Hong and T.S. Huang. Real-time speech-driven face animation with expressions using neural networks. *Neural Networks, IEEE Transactions on*, 13(4):916–927, Jul 2002.

[22] C. Piña, E. Gambaretto, and S. Corazza. Live real-time animation leveraging machine learning and game engine technology. In *ACM SIGGRAPH 2014 Talks*.

[23] L. Qing and D. Zhigang. Orthogonal-blendshape-based editing system for facial motion capture data. *Computer Graphics and Applications, IEEE*, 28(6):76–82, 2008.

[24] J.A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

[25] Z. Ruttkay, H. Noot, and P. Ten Hagen. Emotion disc and emotion squares: Tools to explore the facial expression space. *Computer Graphics Forum*, 22(1):49–53, 2003.

[26] M. Sagar. Facial performance capture and expressive translation for king kong. In *ACM SIGGRAPH 2006 Sketches*, 2006.

[27] D.P. Swain and B.C. Leutholtz. Heart rate reserve is equivalent to %vo2 reserve, not to %vo2max. *Med Sci Sports Exerc*, 29:410–4, March 1997.

[28] R.E. Thayer. *The biopsychology of mood and arousal*. Oxford University Press, 1989.

[29] T. Weise, S. Bouaziz, H. Li, and M. Pauly. Kinect-based facial animation. In *SIGGRAPH Asia 2011 Emerging Technologies*.

[30] T. Weise, S. Bouaziz, H. Li, and M. Pauly. Realtime performance-based facial animation. *ACM Trans. Graph.*, 30(4):77:1–77:10, July 2011.

[31] Y. Xu, A. W. Feng, S. Marsella, and A. Shapiro. A practical and configurable lip sync method for games. In *Proc. of Motion on Games*, MIG '13, pages 109:131–109:140, 2013.

[32] P. Faloutsos Y. Cao, W.C. Tien and F. Pighin. Expressive speech-driven facial animation. *ACM Trans. Graph.*, 24:1283–1302, October 2005.

[33] Y.Q. Xu E. Chang H.Y. Shum Y. Li, F. Yu. Speech-driven cartoon animation with emotions. In *Proc. of ACM Intl. Conf. on Multimedia*, MULTIMEDIA '01, pages 365–371, 2001.

[34] V. B. Zordan, B. Celly, B. Chiu, and P. C. DiLorenzo. Breathe easy: Model and control of human respiration for computer animation. *Graph. Models*, 68(2):113–132, March 2006.