Real-time simulation of virtual humans' emotional facial expressions, harnessing autonomic physiological and musculoskeletal control

Tisserand, Yvain; Aylett, Ruth; Mortillaro, Marcello; Rudrauf, David

Real-time simulation of virtual humans' emotional facial expressions, harnessing autonomic physiological and musculoskeletal control

Yvain Tisserand
University Of Geneva
Geneva, Switzerland
yvain.tisserand@unige.ch

Ruth Aylett
Heriot-Watt University
Edinburgh, United Kingdom
r.s.aylett@hw.ac.uk

Marcello Mortillaro
University Of Geneva
Geneva, Switzerland
marcello.mortillaro@unige.ch

David Rudrauf
University Of Geneva
Geneva, Switzerland
david.rudrauf@unige.ch

ABSTRACT

Emotion expressions and communication involve involuntary and voluntary processes that may not always operate consistently. Expressive behaviour involving physiological changes may carry information about hidden aspects of emotional experience important for accurate emotion recognition and social cognition. In this paper, we describe a principled approach to simulating Virtual Human physiological facial features related to the autonomic nervous system (ANS). Our approach is based on typical synergies within the two branches of the ANS. It covers both parasympathetic tone and sympathetic tone, and their impact on skin tone, pupil diameter and sweat. We also present the triggering of tears. We discuss the implementation of the approach as part of a 3D toolkit. This work is aimed at supporting the development of affective features for real-time intelligent artificial agents, and the study of the perception of mixed emotion and emotion regulation. We demonstrate how varying ANS parameters impacts facial behaviour, contrasting emotionally consistent vs inconsistent musculoskeletal and ANS-related features.

KEYWORDS

Virtual Human, facial communication, facial expression, physiological control, musculoskeletal control, autonomic nervous system, emotions

ACM Reference Format:
Yvain Tisserand, Ruth Aylett, Marcello Mortillaro, and David Rudrauf. 2020. Real-time simulation of virtual humans' emotional facial expressions, harnessing autonomic physiological and musculoskeletal control. In *IVA '20: Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents (IVA '20), October 20–22, 2020, Virtual Event, Scotland Uk*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3383652.3423904

IVA '20, October 20–22, 2020, Virtual Event, Scotland Uk

ACM ISBN 978-1-4503-7586-3/20/09.
https://doi.org/10.1145/3383652.3423904

1 INTRODUCTION

Much recent work has gone into expressive behaviour, and particularly expressive facial behaviour, for embodied social agents. This forms an important communication channel which in human-human interaction is involved in the generation of social signals that both regulate interaction and also support the attribution of affective state to others.

Simple approaches, whether based on cognitive appraisal and other psychologically-inspired [29] models, or on annotation of a conversation stream [36] may generate a specific expression for a single modelled affective state[3]. This suffers from being a static approach (using graphical melding to move between expressions), and single-level, using linguistically-labelled categorical emotions (often relying heavily on the idea of a small set of basic emotions). It also assumes emotional transparency, a direct mapping between modelled state and expressive behaviour, though we know that this is frequently not the case in human-human interaction [3]. This research, both for non-verbal social communication on the agent side and emotion recognition on the user side, has largely focused on musculoskeletal features, for instance through Action Units (AU) control [17].

Dynamic models for expressive behaviour require a more direct actuation, i.e. control of expressive modalities, and support the fluid composition of different aspects of affective experiences and their expressive consequences [30]. In particular, they support multi-stage appraisal architectures [34] in which multiple processes run on different time-scales. They also support the expressive consequences of physiological responses associated with emotion (heart rate, skin conductance, etc.). Though these aspects of affective experiences have been widely researched [26], much less work has been carried out on their expressive correlates on the face: pupil dilation, skin surface blood perfusion, sweating, etc.

These responses relate to the activity of the autonomic nervous system (ANS), through its two branches: the parasympathetic and sympathetic. Each branch has widespread effects over multiple organs and effectors that operate in synergy, and generally act in opposite direction with respect to the other branch. The parasympathetic tone, typically associated with relaxed states, increases redness of skin and eyes, reduces pupil diameter, and sometimes produces tears. The sympathetic tone, typically associated with tense states, increases paleness of skin and eyes, produces increase of pupil diameter, and also sweat. The ability to manipulate these parameters in a principled manner in intelligent virtual agents would open new avenues for research.

One of these would be the study of the relative impact of musculoskeletal vs ANS-related expressive features, both on emotion recognition and on inferences about emotional states in others. ANS-related expressive features could be used to reinforce musculoskeletal affective cues and increase overall expressiveness. However, these features are also highly significant for the study of emotion regulation, mixed emotions and social cognition and communication. Where affective expressive behaviour is being used as a social signal, ANS-related expressive features may provide clues to underlying affective experiences.

It is known that while we can often compose our face in social situations at the level of our musculoskeletal effectors, it is harder to voluntarily control physiological responses. For instance, we may maintain a somewhat neutral expression at a musculoskeletal level while experiencing strong emotional states, but those may nevertheless manifest, for instance, through increased sympathetic tone, making us pale and sweating with wider pupils.

A few studies have endeavored to develop intelligent virtual agents integrating such physiological expressive features. As we will see below, these vary from the application of ad hoc graphical effects examining the impact upon users, through more purely graphical work focusing on the realism of the effects, to direct translation from categorical emotions. However, our approach leverages known synergies within the branches of the ANS, and is not tied to specific computational models of appraisal and emotion processing, and can thus offer a wider applicability than competing approaches.

Another major contribution of our tool is the improved graphical aspect and control interface, offering an advanced tool for research in psychological science and a richer interface for artificial agent models. We present an implementation of a virtual humans' (VH) facial expression system, introducing the Geneva Virtual Humans toolkit, with real-time rendering, combining physical and graphical solutions. The implementation offers a principled integration of musculoskeletal features based on AUs and physiological features based on ANS parameters. The interface allows the manipulations of these features, in both a manual and automatic manner.

To illustrate the tool, we show examples of facial expressions, contrasting emotionally consistent vs inconsistent musculoskeletal and ANS-related features. Consistent expressions are operationalized as musculoskeletal features expressing positive emotional states with high parasympathetic tone, and negative states with high sympathetic tone. Conversely, inconsistent expressions are operationalized as musculoskeletal features expressing positive, or alternatively negative emotional states, with high sympathetic, alternatively high parasympathetic tones. We discuss perspectives for research and current limitations as well as remaining challenges for the development of this emerging field and its tools.

2 RELATED WORK

Our understanding of the mechanisms and cues involved in the perception and inference of others' affective and intentional states through the processing of facial expressions remains limited [7]. There are a number of issues.

First, there are inherent ambiguities and limited degrees of freedom in facial expressions. Some visible movements can be produced by more than one muscle. For example, naïve perceivers, do not have the terms to clearly distinguish between all the different forms of smiling. This increases the complexity of studying what people are perceiving and to what extent they are using each signal in their inferences. Second, recent work shows there is cultural variation in the processing of facial expressions [23], for example depending on which parts of the face observers focus most. There are also contextual effects, and even within the same person we can expect some variability across times and situations.

Third, much work has been based on databases generated in laboratory contexts[33] and this may reduce the ecological validity of the stimuli/expressions. As a consequence, the composition of facial expressions may not reflect genuine and sufficiently intense emotional states. Fourth, musculoskeletal mechanisms can - to a great extent - be controlled in a voluntary manner, and are frequently used intentionally as social signals. For this reason, facial expressions may or may not reflect actual emotional states [3], and indeed may run counter to them: for example, smiling in a social context when one is actually sad [30], or engaging in deliberate deception as in various games [31].

However, the way in which musculoskeletal mechanisms control facial configurations in the generation of expressions of prototypical or basic emotions [15, 16], has been extensively studied. Much of this work has been based on the concept of facial action units, with an interest in both static and dynamic parameters [8] [12]). This work typically relates AU activation to categorical emotions, sometimes generated by a cognitive appraisal model, and creates complex and masked expressions via combinations of action units [4] [30]. Dynamic approaches are often based on a multi-stage component appraisal mechanism [34] in which different stages work on different time-intervals and therefore modify action units temporally. [10].

A limitation of this approach is that facial features are manipulated based on assumed relationships between categorical emotions and expressive features. This creates a close coupling between the affective architecture being used and expressive outputs. This is all the more the case in work that depends on a small set of *primitive emotions*. It limits the possibility of exploring the role of facial parameters in a more systematic manner, and in connection with other affective models.

However, this limitation may not apply to the same extent to ANS-related parameters. Control parameters based upon sympathetic vs parasympathetic ANS tone can build upon background knowledge at a more basic level, free of strong assumptions about the affective architecture, while providing a principled manner to control physiological facial features, which act in synergy.

Beyond musculoskeletal mechanisms, the notion that ANS parameters - difficult to control voluntarily - can in general reflect genuine affective states, is certainly not new, having been at the core of affective science for more than a century [26]. However, the specific role of physiological mechanisms related to the ANS in the facial expression of emotion and related facial cues supporting emotion recognition, has been largely overlooked and understudied.

The ANS, which is modulated in a complex manner by the central nervous system, has two main branches, sympathetic and parasympathetic. They influence the activity of numerous organs in a systematic manner, through smooth muscles and glands, including facial features, and play a role in affective states and emotions [25, 26]. The sympathetic branch has been related to "fight and flight" or arousing accelerating responses, and the parasymapthetic branch to "relax and digest" or break responses. Negative emotions such as fear and anxiety, but also more neutral emotions such as surprise, are associated with increased sympathetic tone, and positive emotions such as joy and relief, and negative emotions such as sadness, are associated with increased parasympathetic tone. Emotions such as anger (which involves a combination of both fight responses and appraisals of coping and control) as well as embarrassment, are associated with a mixture of increased sympathetic and parasympathetic tone.

The two branches may typically act in antagonistic ways, but their relative balance determines the states of effectors. The parasympathetic tone increases digestive system activity, constricts lungs' bronchioles, relaxes bladder sphincter, decreases heart rate and increases flushing, stimulates the production of saliva and tears, constricts pupils and controls the eye lens for close vision. The sympathetic tone decreases digestive activity, triggers epinephrine and glucose release in the blood stream (related to the stress response), dilates bronchioles (for increased intakes of oxygen), constricts bladder sphincter, increases heart rate and visceral blood vessel constriction, inhibits saliva and tear production, dilates pupils and controls the eye lens for far vision, and produces sweat.

These processes affect facial physiology [5, 6, 14, 27, 38]. Increase in sympathetic tone, when related to negative emotional states calling for avoidance behaviors (e.g. fear), can be expected to increase pupil diameter, render the face paler, and produce sweat. Conversely, an increase in parasympathetic tone, for instance related to emotional states calling for approach behaviors (e.g. a state of joy and well-being), can be expected to decrease pupil diameter, and render the face more red (due to influx of skin surface blood, as in flushing).

A small number of recent studies using virtual agents have examined the impact of expressive wrinkles as well as sweating and variations of pupil diameter on human interaction partners. [13] implements wrinkles, blushing, sweating, tears, and changes in respiration in real-time using the graphics pipeline. Rather than parametrising using ANS levels however, as in this work, it maps categorical emotions onto these expressive modalities and studies their impact. Statistically significant positive effects were shown on the perception of surprise, sadness, anger, shame, and fear in one study, and on perception of excitement, pain, relief, boredom, anger, fear, panic, disgust, and startle in a second. This underlines the utility of making such expressive effects generically available.

[28] studied the impact of expressive wrinkles and pupil diameter variation on user perceptions, finding a positive impact for the former, but not the latter. This work did not attempt a generic graphical model. [9] also focused on expressive wrinkles, and [42] examined graphical implementation methods for these and other expressive effects using mpeg-4 and focusing on low bit-rates and plausible dynamics. [1] applies a particle physics approach to the generation of sweat and tears but is not concerned with an interface to affective states. As with other virtual human animation issues, it is not clear that purely physically-based modelling produces the most graphically-convincing effects, especially considering the computational effort required [13].

In the next sections, we describe and illustrate an implementation of these effects in Unity3d, as part of our current development of the Geneva Virtual Humans toolkit, which is aimed at real-time rendering and control of virtual humans for affective science and artificial agent modeling. We integrate ANS parameters, with an explicit control of sympathetic and parasympathetic tones, along with musculoskeletal parameters, as parts of a body model, focusing on facial expressions, and extending standard approaches.

We designed flexible interfaces for the dynamical control of facial expressions, which can then be used for commands from high-level cognitive models of affect, creating a behavioural model. This can sit at the level of existing behaviour planning, translating virtual human actions into animation outputs. Though not part of this work, our approach could for example be used to extend the Behaviour Markup Language (BML)[24] already used in other work as a standardised interface between modelled affect and facial animation. Because purely physical modeling aiming at accuracy can be unsatisfying in terms of user experience and believability, and too computationally heavy for smooth real-time applications, we opted for a combination of simplified physical and advanced graphical modeling.

3 METHODOLOGY

The proposed facial control system comprises three major components: the physiological control of the face, the musculoskeletal control and the appraisal mapping. The latter combines physiological and musculoskeletal controls, based on appraisal theory. In the remainder of this section, we elaborate upon the three components, focusing more on the physiological control.

3.1 Physiological control

An innovative approach to control the physiological states of a Virtual Human (VH) is proposed, which addresses both sympathetic and parasympathetic physiological responses. We implement dedicated eye and skin shaders with algorithms based on advanced rendering techniques, such as Sub Surface Scattering (SSS) and tension map computation. A fully integrated pipeline is proposed to allow realistic rendering of facial expressions. Compared to other methods used to animate the face of VH [37], the present method is more generic and can be used without requiring any complex integration process, specialised digital artists or special equipment. Our work is divided into four sub-modules: skin reaction, eye control, sweat reaction and emotional tears generation. More details about the different modules are provided below.

3.1.1 Skin tone. Skin tone varies in paleness and redness depending on the state of the ANS. A three-dimensional simulation of blood flow would be a principled approach here; yet this is a computationally expensive process, still unsuitable for real-time applications. We therefore propose a method where surface shaders are used to represent the blood flow variations. More precisely, the diffuse textures that come with the 3D model of the virtual character are combined with the thickness maps representing the distance between the skin and the blood vessels. The change of tone between paleness and redness is controlled by an external continuous parameter, with values between 0 (pale-toned skin) and 1 (red-toned skin). The value 0.5 corresponds to neutral skin tone. According to the ANS parameters, we obtain a specific change of skin tone, using the following formula:

$$Skintone = (sympathetic/2) - (parasympathetic/2) + 0.5; \quad (1)$$

To achieve a consistency between the face and the neck skin tone colour, the change in tone is applied both to the face and the body material.

3.1.2 Skin sweat. As a result of intense emotional states or of physical activity, human skin sweats to regulate its temperature [22]. For this reason, it is important to include sweating as part of VH physiological control. A 3D physically-based fluid simulation is computationally costly for skin meshes, as it requires collision detection in every frame for each triangle of the mesh of a VH's face. To simulate sweat in real-time applications, we propose an alternative method, based on 2D physical simulations. More precisely, image sprites corresponding to the sweat droplets are instantiated at run-time, at specific predefined spawn regions of the face (e.g. frontal, temporal). The sweat droplets are equipped with a physical 2D collider and physics simulation to approximate the contact between the skin and the other sweat droplets, based on physical laws, such as gravity and friction. To avoid unrealistic movement of the sweat droplets, several bound colliders are placed in specific locations of the face (e.g. nose, mouth) creating a physical path for the downward movement of a sweat droplet. A virtual camera is placed, rendering the 2D physical simulation system into a texture which is added to the smoothness layer of the shader's input, to produce specular reflection. To avoid multiple camera renderings and loading several textures separately, consuming extra memory, the simulation information is packed into the green channel of a RGB 2D texture. This packed render texture is also used to store the information about the lacrimal simulation system (see Section 3.1.4).

3.1.3 Eye control. Important information is also provided by human eyes in response to emotional states. For instance, pupil dilation, redness of eye sclera and cornea' wetness all give information about a person's emotions. Below, we describe the method developed to control these parameters.

For the control of the virtual eye, we developed a dedicated shader which supports the ANS control, in terms of pupil diameter and eyes' redness. The redness of the eyes is generated by computing a linear colour interpolation between the neutral sclera albedo colour and a redness colour, defined based on visual feedback.

As part of the eye shader, pupil diameter can be adjusted using a normalized value from 0 (contracted pupil) to 1 (dilated pupil). The value 0.5 corresponds to a neutral pupil diameter. To approximate the real sensitivity of the pupils to light intensity, in addition to the manual control of the pupils based on the ANS control, a function to simulate the pupillary light reflex (PLR), controlled by the Central Nervous System (CNS) has been integrated.

Figure 1: Tear rendering

3.1.4 Lacrimal system control. We propose a dynamic lacrimal system control, using a defined adjustable tear generation frequency. Based on that frequency, a tear is generated. Its spawn location is defined by a simplification of the human tear drainage system. More precisely, if during the periodic time, the VH does not blink, the tear is "collected" in the tear line - the area between the eye and the lower eyelid. Due to the virtual accumulation of the lacrimal liquid, a tear is instantiated at the lacrimal caruncle position. If during the periodic time, the VH blinks, then a tear is instantiated at the center of the lower eyelid, approximating the result produced due to the force exerted by the eyelid's mechanical movement towards the stored lacrimal liquid propelling the escape of a tear from the tear line. Similarly to the method followed to simulate skin sweat (Section 3.1.2), tears are simulated using 2D physics. The proposed solution is computationally cost-effective and offers realistic results (Figure 1). As the tear drops are affected by physics simulation, they follow different paths and their movement depends on the head's tilting. The tear element consists of a 2D sprite representing the tear droplet and a trail renderer simulating tears' trace. As already mentioned in Section 3.1.2, the 2D simulation output is packed into a render texture (red channel) and is added to the smoothness layer of the shader. Additionally, to create a volume illusion, making the tear droplets more realistic, the shader enhances the intensity of the normal map at the specific area.

Finally, to approximate the "misty-eyed" effect, a tear line mesh with reflective material representing the accumulation of tears between the eye and the lower eyelid is added to the virtual character. Reflecting the emotional situation, the activation of tears triggers the increase of the visibility of the tear line by gradually adjusting the index of reflection of the material used in the tear line.

3.1.5 Integration into sympathetic and parasympathetic parameters. The sympathetic and parasympathetic components can be controlled both independently and as part of the overall ANS control system.

The parasympathetic and sympathetic values, ranges from 0 (inactive) to 100 (high parasympathetic/sympathetic). High sympathetic tone implies increased pupil diameter, paleness of skin, paleness of the eyes and skin sweating. High parasympathetic tone implies decreased pupil diameter, redness of skin, redness of eyes (see Figure 2,3).

Based on the aforementioned characteristics, the algorithm controls the physiological aspects, applying the necessary changes to the eyes, the face and the body materials using linear interpolation, through high level parameters.

Figure 2: Effects of ANS parameters on facial features. Musculoskeletal AU parameters are all at 0, corresponding to a typical neutral face. Manipulation of ANS parameters, extrema examples. Left tier: "high sympathetic", parasympathetic tone = 0, sympathetic tone = 100 (max); Middle tier: "neutral:, parasympathetic tone = 0, sympathetic tone = 0; Right tier: "high parasympathetic", parasympathetic tone = 100 (max), sympathetic tone = 0

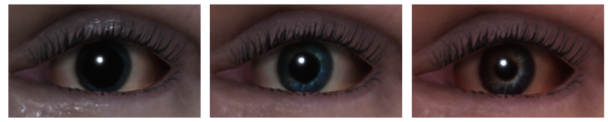

Figure 3: Effects of ANS parameters: focus on the eyes. Magnified version of Figure 2, with emphasis on the eyes.

Parasympathetic and sympathetic effects may occur in combination. As a first approximation, a simple linear blending abstraction layer is used to fuse the parasympathetic and sympathetic components, based on known typical synergies (see Section 2).

High parasympathetic effects combined with high sympathetic effects cancel each other at the level of skin tone and pupil diameter, as the two ANS branches act on these organic targets in opposite ways, but not at the level of sweat intensity, which is only modulated by sympathetic influences. If the character has a high parasympathetic value (100) combined with a high sympathetic value (100) this would produce a neutral skin tone (0.5), while the sweat intensity would be at maximum (1).

The lacrimal control remains an independent feature and can be enabled at will, as emotional tears are highly contextual and not systematically produced when parasympathetic tone increases..

Although motivated based on physiological studies, the grouping of different physiological parameters such as skin tone, sweating, and pupil dilatation, according to the two branches of the ANS, is a simplification, and a first approximation calling for future more realistic nuances. Thus in order to also allow investigators to studies the effects of those parameters independently, we also enabled the possibility of setting their values independently in the interface.

3.2 Musculoskeletal control

Several level of abstractions are used to control the VH musculoskeletal system. To drive facial expressions based on appraisal theory, a first mapping of facial vertex animation to Action Unit system (AU) is achieved, using the Facial Action Coding System (FACS). Then, based on the generated AUs, a mapping to emotion theory and appraisal level is proposed. An oculomotor system is integrated to control eye gaze. Finally, to increase the realism of the facial animations, a dynamic wrinkles system is proposed based on tension maps. A description of the different elements of the musculoskeletal system is given hereunder.

3.2.1 Action Units mapping. To control the Virtual Human using Facial Action Coding System (FACS)[18], we define an abstraction level, enabling manipulation of the facial vertex positions based on the defined AUs. To animate the 3D mesh representing a VH face, three main methods exist: vertex displacement using blendshapes, vertex displacement based on joint control and a combination of both. To have an optimal control of the facial expressions, an abstraction layer, able to handle blendshapes and control the joints has been developed. AUs can be generated by selecting and blending blendshapes and joint displacements. For every AU, the user can select one or more blendshapes and joints, then define a weight for each one, corresponding to the intensity of the transformation. The user can then control the activation of the AU using a high-level parameter with ranging value from 0 to 100. As some AUs share common blendshapes/joints with a different weight, it is important to define a method to fuse blendshapes and avoid potential conflicts. For this, a linear blending method is followed to create realistic and responsive results. Overall, thirty main AUs are accessible to the user, generated in collaboration with a certified FACS coder. We acknowledge that Ekman's AU-based methods, although quite broadly used, are not necessarily the ultimate solution for modeling the complexity of facial expressions, including at the musculoskeletal level, even though no alternate solution is readily available. By adding physiological parameters (see above), we do not solve the intrinsic limitations of AU-based methods, but we contribute to opening richer and more naturalistic venues for research.

3.2.2 Oculomotor system. The oculomotor system is based on a state of the art method [32] and includes the convergence and fixation of the eyes as well as the control of blinking. In more detail, eye convergence is computed based on a defined 3D point of interest; the direction of each eye globe is set accordingly. To enhance the naturalistic look of eye movement, the fixation is not static; rather, it drifts around the target point, generating micro saccadic movements. The fixation percentage, the eye speed movement and the blinking frequency can be configured, through high-level parameters.

3.2.3 Dynamic wrinkles system. To improve the realism of facial expressions, a system is developed to simulate expressive wrinkles. Various methods are proposed in the literature, for instance based on reconstructed data [21] and blendshapes intensity [2], or based on Finite Element Method (FEM) [19, 41]. The proposed tool is based on computing tension and relaxation of the skin, using vertex positions at runtime. The relaxed state (no AU activated) for each triangle of the face is represented by its edge magnitude. At every frame, edge magnitudes are re-computed and compared with those of the relax state (Equation 2).

$$|Triangle_{abc}| = |a - b| + |b - c| + |c - a| \tag{2}$$

$$tension = |T_1| - |T_0| \tag{3}$$

Figure 4: Wrinkles system based on tension map: no simulation (left), simulated wrinkles (center), tension visualisation (right)

An edge magnitude increase denotes skin tension, while a decrease denotes relaxation. The obtained tension values are encoded into a tension map that is stored using the vertex colour. The skin tension is encoded as a red value, while the relaxation is encoded as a blue value (Figure 4).

On the shader side, vertex colour is used to apply deformations on the VH face. More precisely, the intensity of the red channel is used to define movement in the Z axis, while movement in the -Z axis is set based on the intensity of the blue channel.

To approximate the perturbing effect in the VH skin produced by the expressive wrinkles, a depth offset method is followed. Contrary to other methods, such as a tesselation function that modifies the vertex positions, using a depth offset, reduces the computational expense of wrinkling, making it ideal for real-time applications.

3.3 Expression control

The ANS parameters values (sympathetic, parasympathetic) can be set by the user, if the expression requires a physiological control. Thirty Action Units (AUs) have been integrated, allowing the user to choose one or a combination of different AUs, to produce a mapping corresponding to a desired facial expression. The selected AUs can be bundled and weighted using a high-level intensity parameter to produce facial expressions. For instance, the proposed system allows to control AUs together to produce expressions corresponding to standard basic emotions and appraisal dimensions. For the release of the tool as a demonstration of the feature, a preset mapping for basic emotions is proposed based on EMFACS (Emotion FACS [20]). The basic emotion contains the following elements: Happiness / Joy, Sadness, Surprise, Fear, Anger, Disgust, Contempt. We also integrated appraisal dimensions based on Scherer's theory [34], even though other approaches could be used. More specifically, we applied the mapping of specific appraisal dimensions onto AUs as proposed by [35]. This work systematically studied the empirical relationships between target appraisal dimensions and AU activations: Novelty, Uncertainty, Intrinsic pleasantness / unpleasantness, Goal conduciveness / obstructiveness, Coping potential high / low, Norm compatibility. As we develop the tool, we also intend to incorporate bundles of expressive features covering a broader range of emotional states (see for instance, [11]). Note however that users are already able to generate any bundles of expressive features to map emotional states of interest using the current version of the software.

4 IMPLEMENTATION

To demonstrate the capability of the proposed method, an interactive tool is implemented through which the user can manipulate the ANS parameters, the AUs system and the expressions control to produce visual material. A dedicated key-framing system is integrated into the animation maker of the tool to allow the dynamic generation of facial animations.

4.1 Character creation

To generate characters, commercial software is used. The software supports the creation of characters based on template models, by using morphological, gender and age parameters. For the proposed tool, the characters are created based on 3D scanned bodies, offering a high level of realism. The software automatically generates the main maps used in our skin and eye shader (defuse, normal, SSS). It is noteworthy that the proposed method supports the integration of several VHs, regardless of their skin color, gender and age.

The proposed method is not dependent on the software used to generate the characters, thus can be applied to any rigged 3D character. However, several adjustments may be required based on the character's features (e.g. mesh structure, UV mapping, blend-shapes/joints availability).

4.2 Technical aspect

The application has been implemented using the Unity3D game engine [39] with the High Definition Render Pipeline (HDRP). The shaders have been designed using Shader Graph, a node-based system supporting shaders creation. The skin and the eye shader use advanced rendering techniques, such as subsurface scattering, skin micro details and cavity maps.

External libraries have been used to facilitate the implementation of the tool. More precisely, an Inverse Kinematics (IK) system, is used to control the character's limbs and an external library is used to control the gaze of the VH. The application is designed to offer high rendering quality at a stable framerate of 60 frames per second, at the maximum quality options available on HDRP (tested on i7-8700K/16Go/GTX 1080).

4.2.1 Authoring Tool for Facial Animations. To generate facial animations in a user-friendly manner, a key-framing system has been developed (Figure 5). The system creates animation dynamics by controlling AUs and physiological parameters.

The user can manipulate the different parameters, visualize the effect on character's face at run-time, create key-frames and control motion dynamics by modulating the animation curve. The animations can be previewed and exported into an animation file that can be later loaded and triggered in real-time.

4.2.2 Facial Animations Rendering. A standalone application is freely accessible, allowing users to interact with the options that the proposed method offers trough the Graphical User Interface. Users can render the facial expressions in high resolution photos and videos (4K). The obtained visual material is available together with a file listing the parameters selected by the user.

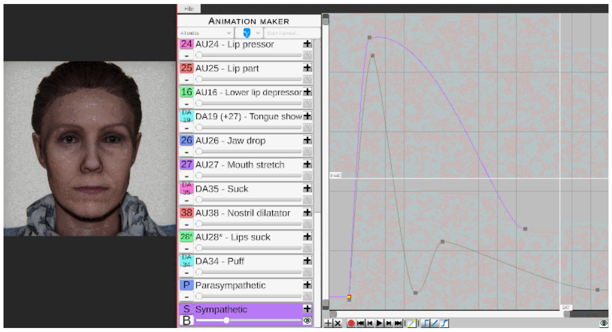

Figure 5: Screenshot of the Graphical User Interface of the key-frame based animation tool.

5 EXAMPLE OF CONSISTENT VERSUS INCONSISTENT MUSCULOSKELETAL AND ANS FEATURES IN BASIC EMOTION EXPRESSIONS

To illustrate the potential interest of our approach for the generation of stimuli for the study of emotion recognition and affective state attribution, we parametrised a female Virtual Human with ANS responses that were consistent versus inconsistent with two basic emotions (happiness and fear) expressed through musculoskeletal channels. We show extreme examples (Figure 6). Such stimuli featuring expected associations versus unexpected dissociations between musculoskelettal and physiological features can be used to assess the extent to which these associations or combinations are valid. For instance, the hypothesis is that participants would show better performance in emotion recognition for consistent than for inconsistent stimuli. An important issue will be to assess the extent to which our virtual humans and the manipulated features may appear uncanny. While we aim at increasing expressive realism through the addition of physiological features, their manipulation may in some case induce Uncanny Valley effects, e.g. in particular for extreme values of those parameters as used in the figures presented in this contribution for illustrative purpose. Generally, we hypothesize that, within a reasonable range, the addition of such physiological features will tend to reduce the uncanniness of the virtual humans, by generating a more nuanced and lively experience. We could test this hypothesis by assessing uncanniness for stimuli with versus without physiological features added for the same musculoskeletal expressions. In any case, several evaluation studies will be required to fine tune the gain of the different parameters. These studies will require to jointly assess emotion perception and uncanniness, in order to calibrate the values of expressive control parameters jointly maximizing emotion perception performance and minimizing uncanniness.

6 LIMITATIONS AND PERSPECTIVES

The toolkit introduced in this article is a first version of a project in development. It has several limitations at this point. These include: control mechanisms and animations that are limited to facial expressions; the lack of full automation for integrating databases of virtual humans with our Unity rig; a simplistic model of ANS effects; and limited user interface features. Empirical studies also need to be performed to validate the tool and refine its calibration. In this article, we concentrated on presenting the development of the tool and its capabilities and relevance for research, notably in the generation of stimuli for empirical research. We are working on developing multiple elements of our toolkit. This includes full body control of : posture, head movements, locomotion, and repertory of other actions, but also ANS-related effects on breathing, as well as fusion with libraries of animations. We are also working on developing an automated pipeline to import a range of body models (with different age, gender, ethnicity, identity), in order to offer a package including a large database of different bodies. We plan improved user interfaces and an API for external control. Notably, we plan to develop an API and adapt our application to make it compatible with the Behavioral Markup Framework (BML) [40]. We will also develop more advanced models for the ANS control layer. This will operate both at the level of subparameters involved in the contextual modulation of organic effects of the sympathetic and parasympathetic branches, based on neurophysiological knowledge and models, and at the level of their graphical and behavioral outputs (blushing and goosebumps). We intend to make the tool available to the community from the outset as it is, and to release new versions regularly as we progress in our development.

Figure 6: ANS-mediated effects consistent versus inconsistent with the musculoskeletal expression of happiness and fear. Non-zero AUs for happiness were: AU6 = 100 (max) and AU12 = 100. Non-zero AUs for fear were: AU1 = 75, AU2 = 75, AU4 = 75, AU5 = 75, AU7 = 75, AU20 = 38, AU26 = 75. Consistent versus inconsistent ANS parameters for happiness were: parasympathetic tone = 100, sympathetic tone = 0, versus parasympathetic tone = 0, sympathetic tone = 100. Consistent versus inconsistent ANS parameters for fear were: parasympathetic tone = 0, sympathetic tone = 100, versus parasympathetic tone = 100, sympathetic tone = 0.

7 CONCLUSIONS

The tool we have introduced is the first systematic approach to controlling the physiological state of the face for virtual humans. We introduced the Geneva Virtual Humans toolkit, including ANS features, specifically for real-time rendering of high-quality graphical effects, compatible with Virtual Reality systems. Our tool can also be used to generate static and dynamic parametric stimuli databases (pictures and videos) that can be presented on 2D displays, for a range of experiments in psychology and neuroscience. It is also designed to provide virtual bodies for artificial agent models, within Unity3D and outside (with a planned API for integration into BML in future releases). The integration of ANS-related control parameters in virtual humans could increase realism and believability in affective behaviors of virtual humans, and open new venues for affective science and artificial agent modeling.

ACKNOWLEDGMENTS

This research has been funded by the Swiss National Science Foundation (FNS 205121_188753): "Implementation of artificial affective agents capable of complex emotion and social perspective taking in virtual reality for behavioral science". Project funding in Mathematics, Natural sciences and Engineering (division II).

REFERENCES

[1] Mohammed Hazim Alkawaz, Ahmad Hoirul Basori, Dzulkifli Mohamad, and Farhan Mohamed. 2014. Realistic facial expression of virtual human based on color, sweat, and tears effects. *The Scientific World Journal* 2014 (2014).
[2] Mohammed Hazim Alkawaz, Dzulkifli Mohamad, Ahmad Hoirul Basori, and Tanzila Saba. 2015. Blend shape interpolation and FACS for realistic avatar. *3D Research* 6, 1 (2015), 6.
[3] Ruth Aylett, Christopher Ritter, Mei Yii Lim, Frank Broz, Peter McKenna, Ingo Keller, and Gnanathusharan Rajendran. 2019. An architecture for emotional facial expressions as social signals. *IEEE Transactions on Affective Computing* (2019).
[4] Elisabetta Bevacqua, Maurizio Mancini, Radoslaw Niewiadomski, and Catherine Pelachaud. 2007. An expressive ECA showing complex emotions. In *Proceedings of the AISB annual convention, Newcastle, UK*. 208–216.
[5] Stella Y Botelho. 1964. Tears and the lacrimal gland. *Scientific American* 211, 4 (1964), 78–87.
[6] Margaret M Bradley, Rosemarie G Sapigao, and Peter J Lang. 2017. Sympathetic ANS modulation of pupil diameter in emotional scene perception: Effects of hedonic content, brightness, and contrast. *Psychophysiology* 54, 10 (2017), 1419–1435.
[7] Manuel G. Calvo and Lauri Nummenmaa. 2016. Perceptual and affective mechanisms in facial expression recognition: An integrative review. *Cognition and Emotion* (2016). https://doi.org/10.1080/02699931.2015.1049124
[8] Jeffrey F Cohn and Paul Ekman. 2005. Measuring facial action. *The new handbook of methods in nonverbal behavior research* (2005), 9–64.
[9] Matthieu Courgeon, Stéphanie Buisine, and Jean-Claude Martin. 2009. Impact of expressive wrinkles on perception of a virtual character's facial expressions of emotions. In *International Workshop on Intelligent Virtual Agents*. Springer, 201–214.
[10] Matthieu Courgeon, Céline Clavel, and Jean-Claude Martin. 2009. Appraising emotional events during a real-time interactive game. In *Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots*. 1–5.
[11] Alan S Cowen and Dacher Keltner. 2020. What the face displays: Mapping 28 emotions conveyed by naturalistic expression. *American Psychologist* 75, 3 (2020), 349.
[12] William A. Cunningham, Kristen A. Dunfield, and Paul E. Stillman. 2013. Emotional states from affective dynamics. *Emotion Review* (2013). https://doi.org/10.1177/1754073913489749
[13] Celso M. De Melo, Patrick Kenny, and Jonathan Gratch. 2010. Influence of autonomic signals on perception of emotions in embodied agents. *Applied Artificial Intelligence* (2010). https://doi.org/10.1080/08839514.2010.492159
[14] Peter D Drummond. 1995. Mechanisms of physiological gustatory sweating and flushing in the face. *Journal of the autonomic nervous system* 52, 2-3 (1995), 117–124.
[15] Paul Ekman. 1992. Facial Expressions of Emotion: New Findings, New Questions. *Psychological Science* (1992). https://doi.org/10.1111/j.1467-9280.1992.tb00253.x
[16] Paul Ekman. 1993. Facial expression and emotion. *American Psychologist* (1993). https://doi.org/10.1037/0003-066X.48.4.384
[17] Paul Ekman and Wallace V. Friesen. 1976. Measuring facial movement. *Environmental Psychology and Nonverbal Behavior* (1976). https://doi.org/10.1007/BF01115465
[18] Paul Ekman and Wallace V Friesen. 1978. *Facial action coding system: Investigator's guide*. Consulting Psychologists Press.
[19] Cormac Flynn and Brendan AO McCormack. 2010. Simulating the wrinkling and aging of skin with a multi-layer finite element model. *Journal of biomechanics* 43, 3 (2010), 442–448.
[20] Wallace V Friesen, Paul Ekman, et al. 1983. EMFACS-7: Emotional facial action coding system. *Unpublished manuscript, University of California at San Francisco* 2, 36 (1983), 1.
[21] Haoda Huang, Jinxiang Chai, Xin Tong, and Hsiang-Tao Wu. 2011. Leveraging motion capture and 3D scanning for high-fidelity facial performance acquisition. In *ACM SIGGRAPH 2011 papers*. 1–10.
[22] Nina G Jablonski. 2008. *Skin: A natural history*. Univ of California Press.
[23] Rachael E Jack, Caroline Blais, Christoph Scheepers, Philippe G Schyns, and Roberto Caldara. 2009. Cultural confusions show that facial expressions are not universal. *Current biology* 19, 18 (2009), 1543–1548.
[24] Stefan Kopp, Brigitte Krenn, Stacy Marsella, Andrew N Marshall, Catherine Pelachaud, Hannes Pirker, Kristinn R Thórisson, and Hannes Vilhjálmsson. 2006. Towards a common framework for multimodal generation: The behavior markup language. In *International workshop on intelligent virtual agents*. Springer, 205–217.
[25] Sylvia D Kreibig. 2010. Autonomic nervous system activity in emotion: A review. *Biological psychology* 84, 3 (2010), 394–421.
[26] Robert W. Levenson. 2014. The autonomic nervous system and emotion. *Emotion Review* (2014). https://doi.org/10.1177/1754073913512003
[27] David H McDougal and Paul D Gamlin. 2011. Autonomic control of the eye. *Comprehensive Physiology* 5, 1 (2011), 439–473.
[28] Anne-Sophie Milcent, Erik Geslin, Abdelmajid Kadri, and Simon Richir. 2019. Expressive Virtual Human: Impact of expressive wrinkles and pupillary size on emotion recognition. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*. 215–217.
[29] Radosław Niewiadomski, Sylwia Julia Hyniewska, and Catherine Pelachaud. To appear. Computational models of expressive behaviors for a virtual agent. (To appear).
[30] Magalie Ochs, Radosław Niewiadomski, and Catherine Pelachaud. 2010. How a virtual agent should smile?. In *International Conference on Intelligent Virtual Agents*. Springer, 427–440.
[31] Matthias Rehm and Elisabeth André. 2005. Catch me if you can: exploring lying agents in social settings. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*. 937–944.
[32] Kerstin Ruhland, Christopher E Peters, Sean Andrist, Jeremy B Badler, Norman I Badler, Michael Gleicher, Bilge Mutlu, and Rachel McDonnell. 2015. A review of eye gaze in virtual agents, social robotics and hci: Behaviour generation, user interaction and perception. In *Computer graphics forum*, Vol. 34. Wiley Online Library, 299–326.
[33] Kr Scherer and T Bänziger. 2010. On the use of actor portrayals in research on emotional expression. *Blueprint for affective computing: A sourcebook* (2010).
[34] Klaus R. Scherer. 2005. Appraisal Theory. In *Handbook of Cognition and Emotion*. https://doi.org/10.1002/0470013494.ch30
[35] Klaus R Scherer, Marcello Mortillaro, Irene Rotondi, Ilaria Sergi, and Stéphanie Trznadel. 2018. Appraisal-driven facial actions as building blocks for emotion inference. *Journal of personality and social psychology* 114, 3 (2018), 358.
[36] Marc Schröder. 2010. The SEMAINE API: towards a standards-based framework for building emotion-oriented systems. *Advances in human-computer interaction* 2010 (2010).
[37] Mike Seymour, Chris Evans, and Kim Libreri. 2017. Meet mike: epic avatars. In *ACM SIGGRAPH 2017 VR Village*. 1–2.
[38] Stuart R Steinhauer, Greg J Siegle, Ruth Condray, and Misha Pless. 2004. Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *International journal of psychophysiology* 52, 1 (2004), 77–86.
[39] Unity 3D. 2020. *Unity 3D*. https://unity.com/
[40] Hannes Vilhjálmsson, Nathan Cantelmo, Justine Cassell, Nicolas E Chafai, Michael Kipp, Stefan Kopp, Maurizio Mancini, Stacy Marsella, Andrew N Marshall, Catherine Pelachaud, et al. 2007. The behavior markup language: Recent developments and challenges. In *International Workshop on Intelligent Virtual Agents*. Springer, 99–111.
[41] Yin Wu, Prem Kalra, and Nadia Magnenat Thalmann. 1996. Simulation of static and dynamic wrinkles of skin. In *Proceedings Computer Animation'96*. IEEE, 90–97.
[42] Yongmian Zhang, Qiang Ji, Zhiwei Zhu, and Beifang Yi. 2008. Dynamic facial expression analysis and synthesis with MPEG-4 facial animation parameters. *IEEE Transactions on circuits and systems for video technology* 18, 10 (2008), 1383–1396.