2019 18th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)

Investigating Emotion Style in Human Faces and Avatars

Julia Kubiak Melgare, Soraia Raupp Musse
School of Technology, PUCRS
Porto Alegre, Brazil
Email: julia.melgare@acad.pucrs.br, soraia.musse@pucrs.br

Nathan Rainieri Schneider, Rossana Baptista Queiroz
School of Technology, Unisinos
Porto Alegre, Brazil
Email: nathanm@unisinos.br, rossanaqueiroz@unisinos.br

***Abstract*—This paper describes a computational study regarding the way real humans manifest their facial expressions and emotions and how other people perceive that when applied to virtual humans. We propose a new metric for measuring individuals' emotion style where subjects were recorded while expressing the six basic emotions (happiness, fear, disgust, anger, surprise and sadness). With this metric, we were able to group the subjects into four different clusters and provide evidence that shows a visual correlation between the groups and the video footage. After applying the styles in virtual humans, a survey was also applied to lay people in order to understand how emotion style is perceived and identified by the general public. The survey not only indicated that people are in fact able to perceive an individual's emotion style regardless of facial geometry, but also showed that there seems to be a particular style that is considered more sympathetic and approachable when compared to others.**

***Index Terms*—Motion style; emotion; facial expression; virtual agents.**

I. Introduction

Motion style is a topic that includes two main concepts: content and style [1]. Content represents the task that is performed with an objective purpose, e.g. walking, jumping, throwing a ball, etc. Style describes how each individual performs the defined action, i.e. how a person walks, jumps or throws a ball; and that style might differ from another individual's manner of executing such actions. In our study, the concept of emotion style would be a person's unique way of expressing an emotion and how that style differs from other individuals' even when expressing the same emotion. This type of research is relevant in the context of interfaces for entertainment applications as games and gamified products, where the user's emotion can be relevant for the included narrative.

In this work, with the goal of investigating the extraction of facial emotion styles, we considered the acted content to be all six basic emotions according to Ekman's universal emotions [2], i.e. happiness, fear, disgust, anger, surprise and sadness; while style would be the variations observed within the expression of these emotions. We performed an experiment with 11 different subjects who were asked to express the same emotions in their own style. We then measured their individual facial expressions based on the Facial Action Coding System (FACS) [3] in order to later compare them quantitatively. We opted to use this small amount of subjects because we also wanted to create their avatars in a non-fully automated process. For a more qualitative analysis, a public survey was also applied to lay people so that their responses can give us clues about how people understand some expressions and empathize with them. With the future application of this study in the entertainment and virtual agent area in mind, we also conducted an experiment where we attempted to replicate a person's emotion style to a 3D virtual avatar and tested if that style could still be perceived and remained the one most similar to the original when compared amongst other 3D model expressions.

The main contribution for this work is the proposal of a metric system for motion style analysis in human emotions. In order to evaluate this measurement, the aforementioned experiment with the 11 subjects was performed. Also, we began a preliminary study to investigate if people are capable of recognizing the emotion style of an individual if we replicate it to another, using expression transfer techniques. In this paper, we show a brief summary of some studies that have been developed on the motion style area in Section II and we explain the proposed experiments and methodologies in Section III. In Section IV we analyze and interpret the results obtained with the defined experiments, and ponder on what would be some other applications for the study of emotion style. Finally, in Section V we discuss future aspects of this research.

II. Related Work

Studies that have been made on the subject of motion style focus either on the expression style of individuals regarding the human face (focusing more on visemes and facial expressions and not on emotions per se) or on the whole body movement of human beings.

Learning the motion style of people for purposes of 3D character animation has been turning into an exciting research field since the past few years. Facial motion style is a novel research field in the computer facial animation area, where usually the goal is to learn the motion style of an actor and either synthesize novel motions or transform existing motions as if they were performed by the original actor. However, most of the work in the literature addresses body motion [1], [4]–[6] instead of the facial motion style [7], [8]. Ma *et al.* [7] introduced a Constraint-based Gaussian Process model (CGP) that

2159-6662/19/$31.00 ©2019 IEEE
DOI 10.1109/SBGames.2019.00025

can effectively learn the editing style from a small set of facial editing pairs, which can be applied to automate the editing of the remaining facial animation frames or transfer editing styles between different animation sequences, being able to reduce the manual efforts necessary in most of current facial animation editing practices, e.g., sculpting a facial keyframe every several frames. Cao *et al.* [8] proposed an unsupervised learning technique based on *Independent Component Analysis* (ICA) for editing speech related facial motions, which provides a meaningful parameterization of the original data, making it suitable for editing.

Braun *et al.* [9] proposes the learning of facial motion style and its use in avatars using monocular cameras as input. The result of this work is a structure named Persona, which provides a set of facial animation parameters. The inspiration for this term are the masks used in ancient Greek theater. Their methodology to obtain the Persona includes three main steps:

1) first, they analyze the tracking results from a markerless real-time facial motion capture system (e.g. Faceware Live[1] software to provide the tracking of feature points on the face) using a single video segment;
2) as the second step, it classifies action units and emotions on each frame of the input video;
3) then, it saves and organizes the new face expression in the Persona structure and it builds 3D control masks.

The Persona data (control masks) were applied to synthetic faces using a methodology for expression transfer proposed by Queiroz *et al.* [10]. They developed a methodology for transferring facial expressions between different 3D facial models, being able to separate an individual's expression style from their respective facial geometry and apply that style to a completely different face. They accomplish that through three main steps: Rigging, Expression Transfer and Animation. In so doing, they managed to develop a semiautomatic rigging process that produces robust rigged faces with little user intervention. Their proposed methodology of expression transfer has also been shown to work with both low and high polygonal 3D mesh topologies and different representation languages (realistic and cartoon faces) as well. Although Braun and Queiroz worked on that area, they did not provide a metric to estimate emotion style as we propose in this work.

Wang et al. [11] have done something similar by extracting an individual's expression style through 3D-scanning their face while performing various facial expressions, thus being able to not only transfer expressions between two different faces, but also extrapolating a new previously unknown expression style given two different individuals. They achieved that by representing an individual's expression style through a vector, containing information of their facial geometry and motion style. This allowed them to morph two styles together in order to generate new ones, while also being able to generate new different face geometries. In order to successfully transfer expressions between different faces, they proposed a method for extracting a new vector containing only an individual's expression style, ignoring their facial geometry. By selecting a default facial model to be applied and normalizing these vectors in various ways in order to not be affected by other factors such as head size or head motion, they were able to transfer expression style between 3D representations of actors regardless of facial geometry.

[1]http://facewaretech.com/products/software/realtime-live/

As an example of work on motion style in the entire human body, Okamoto et al. [12], through a learning from observation (LFO) paradigm, presented a humanoid robot which extracts the different person-specific styles of humans performing a physical action such as tossing rings to a goal. By decomposing such action into different parameters, e.g. stance, wrist angle, etc., the robot is able to imitate a person's style when executing it. Torresani et al. [13] proposes a system that applies style to movements that previously had only the content, being useful for animations where different characters can have different styles. They do not, however, use this method on facial expressions and emotions but focus only on body movements like walking, running, etc.

There are also some recent works focused on transferring and learning of speech motion style. Taylor *et al.*'s work [14] uses a deep learning approach to analyze video and sound inputs and automatically generates speech animation that synchronizes to input speech applied to different 3D characters. We also can find some works that learn the speech motion style from videos, such as Zakharov *et al.* [15]'s work. They first perform lengthy meta-learning on a large dataset of videos, and after that are capable of transfer speech motion from a few-shot (even one photo) source data to another person's image. Machine learning approaches have been widely explored with this purpose, generating convincing results. The main con of these approaches is related do the nature of the technique which does not allow to decode "easily" (or meaningfully) the model (weights and network topology) learned by the method.

III. Method

This section describes the proposed methodology in this work. Our goal is to i) obtain footage of different people expressing the same six emotions, ii) extract quantitative information from this footage in order to study different facial styles and lastly, iii) propose a way to measure the emotion style and to analyze it quantitatively.

A. Obtaining footage

At first, we attempted to record the subjects reacting to videos that elicit specific emotions in order to obtain genuine expressions. However, even if we show a sad sequence, for example, some people react in a different way, i.e. talking about how the sequence makes them feel sad instead of expressing it through facial expressions. Given that, we decided to ask to the subjects to try to express some specific emotion on their own way in front of a camera, as commonly used in the area [6].

For each subject, 6 videos were recorded, one for each emotion, using a 720p/30fps camera. We began recording the subject in a neutral expression and then asked them to

Fig. 1. A graphic representation of the emotions used in the experiment [3].

express a specific emotion and attempt to hold it for 5 seconds. Figure 1 shows a guideline that was presented to subjects who wished to see a reference of what the expressions should look like, however we informed the subjects that they should be spontaneous and not attempt to mimic the pictures, doing facial expressions on their own way. We used a small group of 11 subjects, male and female and around the same age group (ranging from 18 to 29 years old). We used only 11 subjects in this study because we intend to manually transfer their motion style to virtual humans in order to replicate the subjects' facial expressions. The subjects were lay people, not actors.

B. Extracting data from footage

In order to obtain quantitative data from the recorded videos, we utilized Openface [16], a free open source face recognition software with deep neural networks which features, among other functionalities, the detection of up to 17 action units [3] (facial expressions) on photos and videos. Table I lists all of the action units that the software is able to detect. In our work, we consider all of them except the last one (AU45), as we concluded that it is not relevant when comparing the expression of emotions, since it represents blinking, which doesn't affect the perception of any of the six basic emotions used in this work, according to Ekman's EMFACS [17] . Using Openface, we are able to obtain the intensity of each action unit on each frame of each video. Then, for each subject and each emotion, we extracted the frame which had the most average AU intensity to be processed in the next step. Indeed, such frames were visually evaluated in order to verify if they really represent the expected emotion.

C. Defining and comparing emotion styles

In order to quantitatively define a facial emotion style, we decided to firstly assume a reference for each expressed emotion. We use the standard emotion style of Ekman-trained professionals [3] as such reference, since these individuals are trained to express emotions in the most standard and comprehensible style as possible. So, in order to determine the motion style of other people, when expressing the six basic emotions in their own spontaneous way, we compare it to the Ekman style and observe the difference among them. The photo-set of the Ekman-trained expressions used in our work can be seen in Figure 2.

TABLE I
TABLE PRESENTING ALL OF THE 17 FACIAL EXPRESSIONS DETECTED BY OPENFACE AND THEIR RESPECTIVE CODE (AU) ON THE FACS

Inner Brow Raiser	AU1
Brow Raiser	AU2
Brow Furrow	AU4
Eye Widen	AU5
Cheek Raiser	AU6
Lid Tighten	AU7
Nose Wrinkler	AU9
Upper Lip Raiser	AU10
Lip Corner Puller	AU12
Dimpler	AU14
Lip Corner Depressor	AU15
Chin Raiser	AU17
Lip Stretcher	AU20
Lip Tightener	AU23
Lips Part	AU25
Jaw Drop	AU26
Blink	AU45

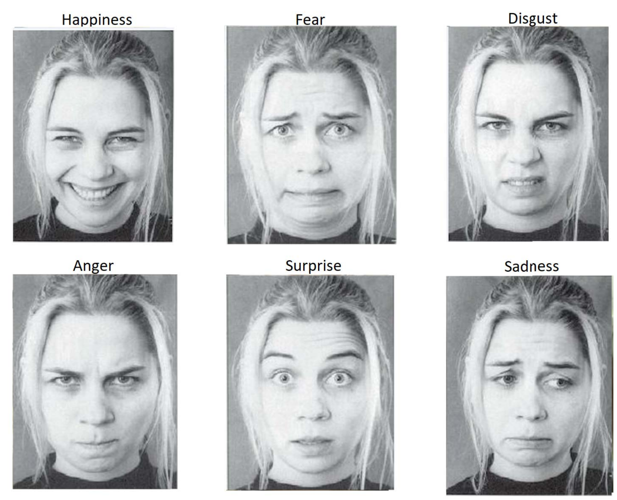


Fig. 2. Set of images of an Ekman trained professional expressing all six basic emotions.

We propose one distance metric to compute differences between AUs intensities obtained using OpenFace from the Ekman references and the subjects' faces. The proposed metric D uses the Euclidean distance between the Ekman-trained individual's 16-point vector, where each point corresponds to the intensity of an AU captured by the software while expressing one of the six basic emotions, and a subject's 16-point AUs vector. So, given the Ekman-trained subject's AU vector $\vec{E}$, and one of our eleven subjects' i vector $\vec{S}_i$, D_i represents how different subject i is from the Ekman reference when discretized in 16 AUs:

$$D_i = \frac{\sqrt{(\vec{E}_{AU1} - \vec{S}_{AU1,i})^2} + ... + \sqrt{(\vec{E}_{AU26} - \vec{S}_{AU26,i})^2}}{16}. \quad (1)$$

Having a subject's D_i for each emotion e, i.e. Happiness (H), Fear (F), Disgust (D), Anger (A), Surprise (S) and Sadness (SA), we can say that a person's overall distance to the Ekman reference is the mean of their emotion computed distances, as seen in Equations 2:

$$\bar{D}_i = \frac{D_{H,i} + D_{F,i} + D_{D,i} + D_{A,i} + D_{S,i} + D_{SA,i}}{6}. \quad (2)$$

In addition to the features vector of each subject $\vec{S}_i$ (action units $AU1_i, ..., AU26_i$), it also comprises distance values, i.e. $D_{H,i}, .., D_{SA,i}$, where D_i represents the differences from the reference for the 6 emotions computed using the distance metric.

In order to show some qualitative assessment of the D metric, Figure 3 illustrates the happiness as performed by subject S1, which presents the smallest distance among the 11 tested subjects, when compared to the Ekman reference.

Fig. 3. Visual comparison between the Ekman trained professional and S1.

Once we are interested in finding emotion styles, we compute the difference among the subjects' styles as well. We propose to use the Modified Hausdorff Distance [18] among computed values of D_i and D_j for each pair of subjects (i, j) in order to group them in clusters of similar style among all emotions. We used the very known method of k-means clustering algorithm [19]. Results are presented and discussed later in Section IV.

D. Perception of emotion styles

Alongside the numerical analysis, a survey was also applied to lay people in order to understand if there is a particular emotion style which people find more sympathetic, and if there is, how close it is to the Ekman reference. In order to do this, we decided to present all the different styles we found for happiness among our 11 subjects as well as the Ekman style in the survey questions, which will be described later in this section. In order to avoid that subjects answering the survey be influenced by the real faces' empathy, we created an arbitrary 3D model using FaceGen[2] to generate a realistic face and Blender[3] to create the blendshapes set corresponding to the AUs presented in Table I. Figure 4 shows two subjects (S1 and S2) expressing happiness and the respective representation of their style in our 3D model.

(a) S1 expressing happiness

(b) 3D model expressing S1's style

(c) S2 expressing happiness

(d) 3D model expressing S2's style

Fig. 4. Two subjects and their respective style represented on our 3D model.

The survey has 4 questions and all of them show images of our 3D model expressing all the different styles of happiness extracted from the 11 subjects as well as the Ekman happiness. Figure 5 show all 11 subjects that took part in our study.

The last question also featured the original photo of $S1$ (Figure 4a), i.e. the subject who presented the smaller D when compared to Ekman reference). The intention is to validate whether people could perceive the same emotion style in a photo of a real person as well as in a virtual human. Following are the presented survey questions:

1) Which of these expressions do you consider to be the most sympathetic?
2) Which one of these ways of expressing happiness do you identify with the most?
3) Which one of these ways of expressing happiness would you consider to be best suited for a virtual character?
4) Observing the following photo (Figure 4a was presented), which image is the one most similar to the original photo?

Question 1 had the intent to know which style people found the most approachable, and see if it is among the ones most similar to the Ekman reference. Question 2 was meant for the person to try to identify the alternative most similar to their style, as it not necessarily would be the most sympathetic.

[2]https://facegen.com/
[3]https://www.blender.org/

(a) S1 (b) S2 (c) S3 (d) S4 (e) S5 (f) S6 (g) S7

(h) S8 (i) S9 (j) S10 (k) S11

Fig. 5. Eleven subjects and their respective happiness manifestation.

Question 3 was to see which style people would find more appropriate for a virtual character, going from the hypothesis that more expressive or exaggerated styles are more preferable in a virtual environment over those more subtle [20]. Lastly, question 4 was meant to test whether people could perceive and identify the same style in real life as well as in a virtual human, thus we hoped to have the image of the 3D model expressing S1's style (Figure 4b) as the answer most voted in this question. The results obtained in this survey will be shown in the next section.

IV. Experimental Results

This section discusses some obtained results. After calculating the Ekman distance for each emotion of each subject and their overall Ekman distances (Equation 2), we were able to make some observations. Figure 6 shows a histogram of all Ekman distances calculated using Equation 1 for each emotion of each subject (66 values). In the histogram, we can see that most part of the obtained distance values D are in the interval $[6; 20]$.

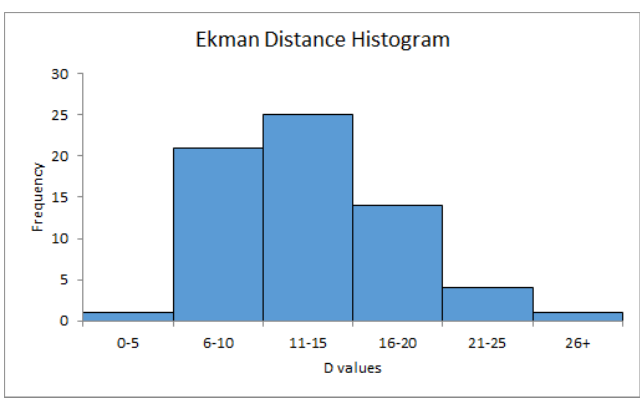


Fig. 6. Ekman distance histogram, calculated using $ED2$ and then OED for each subject.

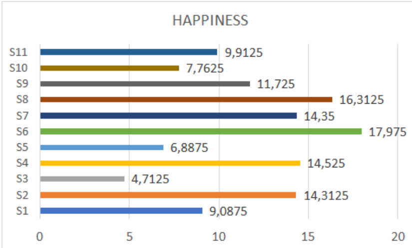
Figure 7 shows graphs of all the Ekman distances (calculated using Equation 1) for each subject, grouped by emotion. In the graphs, we can see that the emotions Sadness (Figure 7f) and Anger (Figures 7d) represent the emotions where subjects manifest themselves more similarly to the Ekman style (average distances=10.74 and 11.55 and median values =$10,43$ and 9.63, respectively). On the other hand, emotions such as disgust (Figure 7c) and surprise (Figure 7e) presented the higher distances from the Ekman reference (average distances=15.46 and 15.03 and median values =15.43 and 12.85, respectively).

As an example of what would be a high value for D, we tried calculating the distance between the Ekman-trained professional expressing surprise and one of our subjects, S3, expressing happiness. The obtained D when using Equation 1 was 29.475, which would sit in the last block of the histogram (right side of the histogram of Figure 6). In order to provide a visual assessment we include Figure 8 that shows the pictures related to such analysis.

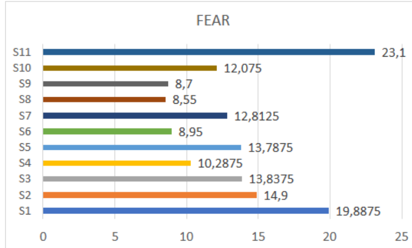
As for the k-means clustering, we decided that four clusters ($k = 4$) were the best number of groups after analyzing the results obtained using 2, 4 and 6 as values for k. The data input are the 6 distance values from each subject, so eleven 6-feature vectors in total. Tables II, III and IV were used in order to determine the best k to be used in our dataset. As can be seen in such tables, $k = 4$ represents the best guess once that in $k = 2$ we had the maximum distance to the centroid in cluster 1 be greater than the distance from both centroids. In addition, Table IV presented two unitary clusters, also indicating that $k = 4$ is the best option for our analysis.

The results of the clustering (using $k = 4$) can be seen in Table V, while Figure 9 shows the plotting of the D distances for one representative of each cluster (subject closest to the cluster centroid): S7 for cluster 0, S2 for cluster 1, S8 for cluster 2 and S1 for cluster 3. These charts also serve for graphically representing a subject's emotion style in a simplified manner.

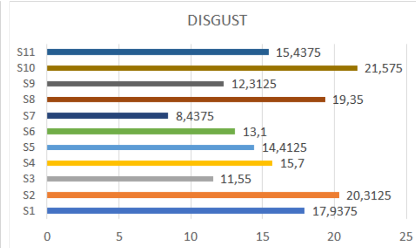
The centroid for each cluster can indicate the mean Ekman distance of each of the six emotions for a member of that group. In other words, it classifies subjects using the criteria of having each emotion be more or less similar to the default



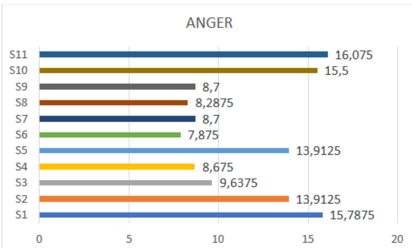
(a) D for Happiness



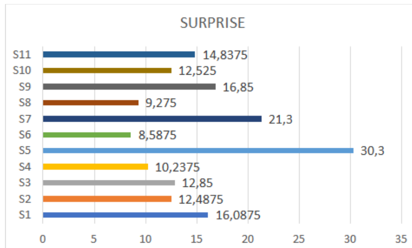
(b) D for Fear



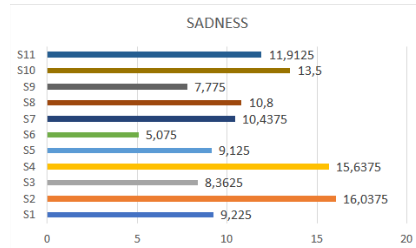
(c) D for Disgust



(d) D for Anger



(e) D for Surprise



(f) D for Sadness

Fig. 7. The calculated Ekman distance for each subject for all of the six emotions

Fig. 8. Visual comparison between the Ekman-trained professional expressing surprise and subject S3 expressing happiness.

TABLE II
TABLE PRESENTING EACH CLUSTER'S MAXIMUM DISTANCE TO IT'S CENTROID AND THE DISTANCE BETWEEN EACH CLUSTER'S CENTROID WHEN USING $k = 2$.

Cluster ID	Max Distance to Centroid	C0	C1
0	10.3157	0	11.7444
1	12.1604	11.7444	0

TABLE III
TABLE PRESENTING EACH CLUSTER'S MAXIMUM DISTANCE TO IT'S CENTROID AND THE DISTANCE BETWEEN EACH CLUSTER'S CENTROID WHEN USING $k = 4$

Cluster ID	Max Distance to Centroid	C0	C1	C2	C3
0	9.8089	0	14.8558	16.3408	12.2996
1	6.4006	14.8558	0	11.2555	10.7454
2	4.3418	16.3408	11.2555	0	16.1098
3	8.4126	12.2996	10.7454	16.1098	0

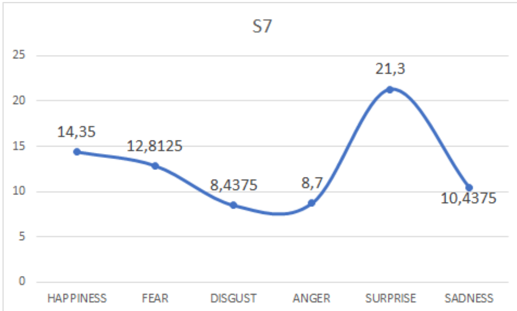
TABLE IV
TABLE PRESENTING EACH CLUSTER'S MAXIMUM DISTANCE TO IT'S CENTROID AND THE DISTANCE BETWEEN EACH CLUSTER'S CENTROID WHEN USING $k = 6$

Cluster ID	Max Distance to Centroid	C0	C1	C2	C3	C4	C5
0	0	0	13.9047	18.3381	12.6189	10.971	13.9721
1	6.4769	13.9047	0	24.1463	17.0754	11.8858	11.7702
2	0	18.3381	24.1463	0	17.2539	14.7134	20.2254
3	2.5553	12.6189	17.0754	17.2539	0	15.3341	10.6047
4	4.0531	10.971	11.8858	14.7134	15.3341	0	15.3116
5	3.9191	13.9721	11.7702	20.2254	10.6047	15.3116	0

TABLE V
TABLE PRESENTING THE RESULTS OF THE k-MEANS CLUSTERING EXPERIMENT USING $k = 4$

Cluster ID	Cluster Centroid	Members
0	(10.9875, 11.7666, 11.7208, 10.4375, 22.8166, 9.1125)	S5, S7 and S9
1	(12.2, 12.4208, 19.1958, 12.6958, 11.75, 15.0583)	S2, S4 and S10
2	(17.1437, 8.75, 16.225, 8.0812, 8.9312, 7.9375)	S6 and S8
3	(7.90416, 18.9418, 14.9749, 13.8333, 14.5916, 9.8333)	S1, S3 and S11

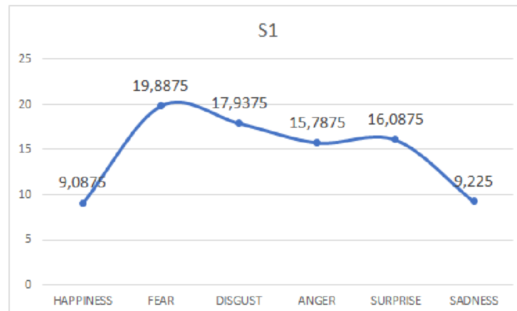
manner of expressing it. For example, subjects in cluster 0 would be expected to have a way of expressing surprise that is different of the standard, while members on cluster 3 would be expected to have their happiness and sadness more Ekman-like and their fear less Ekman-like. Being more similar to Ekman means the subject activated the same facial expressions (or AUs) that the Ekman-trained professional did when expressing one of the emotions, and that the intensity for said facial expressions was also close to that of the professional. On the other hand, being less similar to Ekman means the subject could have either been inexpressive, i.e having more subtle expressions, or over-expressive, i.e being more exaggerated or over-the-top. It could also simply mean that the subject



(a) S7's Emotion Style plot (subject closest to the cluster 0 centroid)

(b) S2's Emotion Style plot (subject closest to the cluster 1 centroid)

(c) S8's Emotion Style plot (subject closest to the cluster 2 centroid)



(d) S1's Emotion Style plot (subject closest to the cluster 3 centroid)

Fig. 9. Some examples of the plotting of D values for each subject for all emotions.

expressed that emotion in an unexpected way, e.g smiling when expressing an emotion that is not happiness. From a more technical standpoint, it would mean that the subject either did not activate the expected AUs, or that the intensity for them was either too low or too high.

The general common characteristics found between members of each cluster were as follows:

- All members of cluster 0 expressed Surprise in a a different way, when compared to Ekman reference, resulting in a high value for their $ED_{Surprise}$;
- Members of cluster 1 were generally the less expressive subjects (with respect to obtained numbers, but also qualitatively) and present very similar ways of expressing most part of the emotions;
- Members of cluster 2 were subjects that expressed happiness in a seemingly more "exagerated manner (as the recorded expressions were not spontaneous), but were more visually convincing for the rest of the emotions; and finally
- Members of cluster 3, in contrast to cluster 1, were the ones who generally had the most exaggerated expressions.

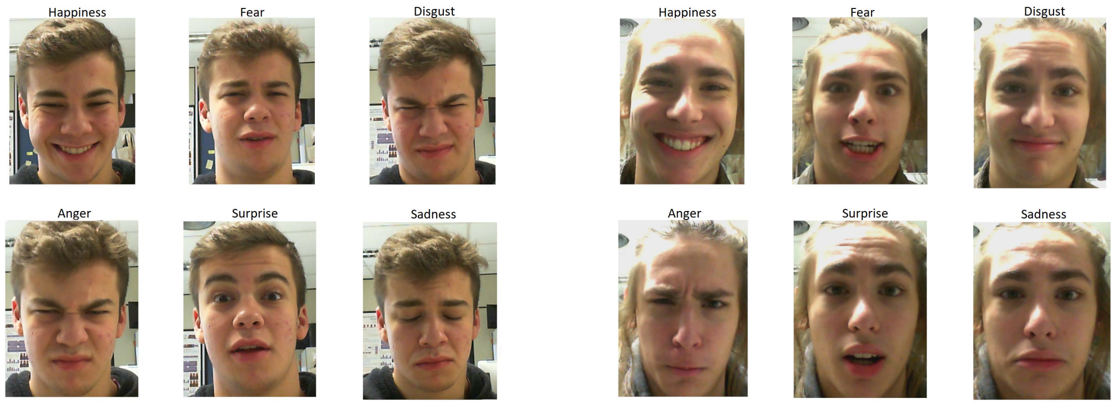
Regarding to the visual similarity between members of the same cluster, a comparison has been made between cluster 3 and 1, which is illustrated by Figures 10 and 11. In this example, members of cluster 3 seem to generally be more expressive than the members of cluster 1. This phenomenon can be seen as both S1 and S3 expressed happiness through an open smile, while subjects S2 and S4 had their mouths closed and only pulled the corner of their lips. It is interesting to note that this observation matches with the visual inspection of Figure 3, once the Ekman trained professional expressed happiness through an open smile as well.

In addition, we can also note the difference between two clusters on the surprise emotion, where both members of cluster 3 had their mouths opened and eyes widened, while in cluster 1 the subjects remained with their mouths closed, only widening their eyes. In the rest of the emotions we can also notice that the expressions of subjects in cluster 3 were more exaggerated than those of cluster 1. Thus, this example shows that we can visually perceive not only the similarities between members of the same group, but also general differences between members that do not belong to the same cluster.

A. Survey Results

The conducted survey (Section III-D) obtained 217 answers with people ranging from 18 to 80 years old, 60% female and 40% male; 43.3% of the population of the survey completed their graduate studies, 27,6% had completed their undergraduate studies and the other 28.6% had completed high school.

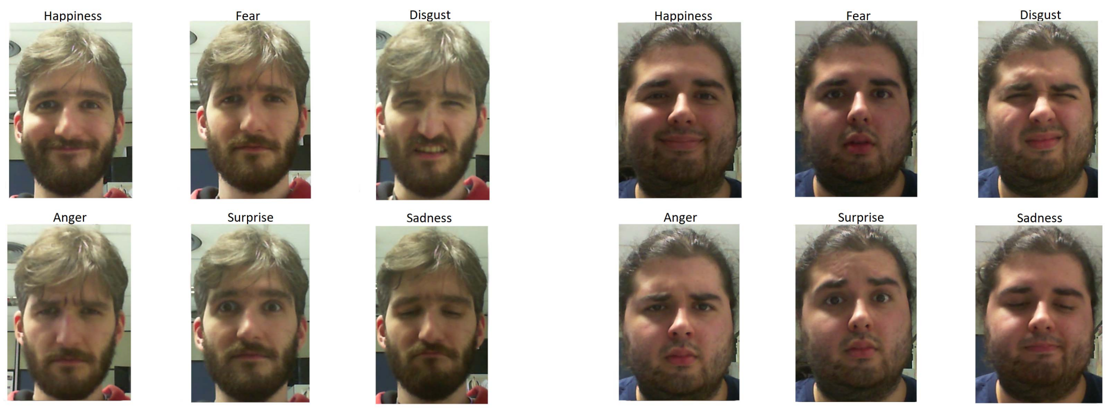
Question 1 was asked twice on the survey, each with different images to choose from in order to present all the



(a) S1's six basic emotions

(b) S3's six basic emotions

Fig. 10. Visual comparison between S1 and S3's 6 basic emotions. S1 and S3 are grouped in cluster 3.



(a) S2's six basic emotions

(b) S4's six basic emotions

Fig. 11. Visual comparison between S2 and S4's 6 basic emotions. S2 and S4 are grouped in cluster 1.

12 possible images (11 subjects and Ekman reference). Since question 1 asked for people to choose which expression they find the most sympathetic, our hypothesis was to select the image of the 3D model containing the Ekman style, which is considered to be the most default smile. The results are illustrated in Figure 13 and it shows that S1's style was the most preferred. When asking the same question but with other options, subject S11 was the one most chosen. We know that S1 and S11 belong to the same group (cluster 3) and have similar styles, so it could mean that people with the style observed in cluster 3 seem more sympathetic. The fact that S1's style is more exaggerated could justify why it was chosen over Ekman, however the fact that S2 had more votes than Ekman contradicts this assumption, as S2 can be considered less expressive in the visual inspection.

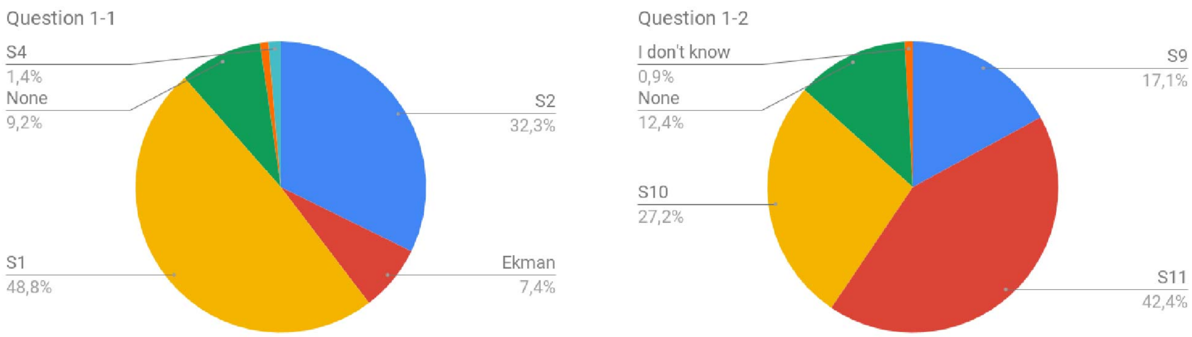
Question 2 asked subjects to choose the style that they identified with themselves the most. Twelve options are presented, being the same as the ones presented in Questions 1 (1-1 and 1-2). The results can be seen in Figure 14 and again, the option most chosen was S1 and the second most chosen was S10, excluding those who voted for None. It makes sense that the styles people found to be the most sympathetic were also the one they identify with the most. With this, we can notice that the majority of the population surveyed considers themselves to be more expressive and exaggerated when expressing happiness.

Question 3 asked to the subjects to choose the expression that they would find more appropriate for a virtual character in a virtual environment and were presented with the same previous 12 options. Figure 15 shows that S1 was again the most chosen, and S11 was the second most chosen. This time, we can see that the two most voted options belonged to the same cluster, which indicates that the style observed in cluster 3 is also the one that seems more appropriate for virtual humans. Since the members of cluster 3 have the most expressive and exaggerated style, our hypothesis that the general public prefers more expressive virtual humans holds.

Lastly, Question 4 presented a photo of S1 expressing

(a) S1 (b) S2 (c) Ekman (d) S10 (e) S11

Fig. 12. Our 3D model expressing S1, Ekman, S10 and S11's styles.



(a) (b)

Fig. 13. Answer results for Question 1.

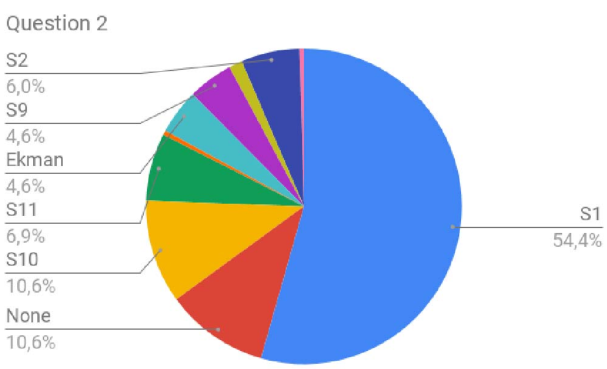


Fig. 14. Answer results for Question 2.

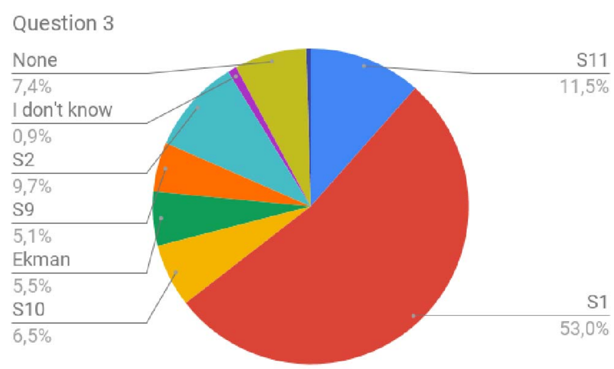


Fig. 15. Answer results for Question 3.

happiness (Figure 4a) and some virtual human faces in order to point our the best correspondence. We included the virtual human presenting S1's style, Ekman's style and also the style of the three subjects most similar to S1 (calculated using the same metric that was applied to the clustering, the modified Hausdorff distance). The subjects most similar to S1 were S11, S3 and S10 (ordered by similarity) and some of them are illustrated in Figure 12. The images were displayed in a random order so as to not influence the answers. Observing Figure 16 we can see that 94% of the surveyed selected S1, the correct answer. Another interesting observation is that S10, the only alternative available who does not belong to the same cluster as S1 and also the least similar to S1, according to our metrics, got 0 votes. With the large majority of the surveyed selecting the correct style corresponding to the original photo of S1, we can say that in this experiment people were able to identify the same style even when transferred to a 3D virtual human.

V. Conclusion

In this paper, we have shown a proposed metric for measuring an individual's emotion style using the Ekman style as reference. We also described this metric used to compare and

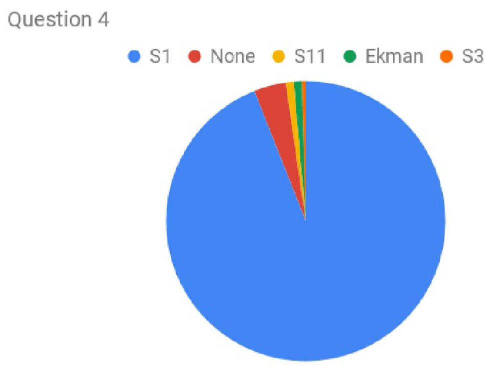


Fig. 16. Answer results for Question 4.

cluster individual motion styles. Our method present evidences that indicate that individuals emotion styles can be computed, analyzed and even used to group similar facial expressions. The results obtained in the survey have shown that people were able to perceive an individual's emotion style even when it was transferred to a different face in a 3D medium and also helped to show that the clustering process made sense.

For future studies, we would like to better evaluate these emotion style metrics with a larger dataset than the one used in this work, investigating styles of emotions in different age groups and cultures. We understand that lay people do not have the same ability to convey emotion non-spontaneously as trained actors do, so either using a dataset with emotions conveyed by actors or emotions conveyed spontaneously will be considered. In addition, we could study whether individuals who present disorders that impair their ability to express themselves emotionally, such as autistics, have a different style of emotion from those without these characteristics, with the intention of helping others to better understand what these individuals are feeling. We would also like to eventually do another survey similar to the one presented, but with a more realistic 3D model capable of performing more subtle expressions. Lastly, we also plan to create a framework for extracting an individual's emotion style and transfer it to a 3D model without the need of human intervention and also be able to extrapolate how an individual would express a previously unknown emotion that is not among the six basic ones, based on their extracted emotion style. We believe our style extraction method could help make video-game characters more emphatic and unique, as well as supporting narratives where the user emotion can be included as a part of it.

REFERENCES

[1] A. Elgammal and C.-S. Lee, "Separating style and content on a non-linear manifold," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 1, June 2004, pp. I–I.

[2] P. Ekman, "An argument for basic emotions," 2013.

[3] P. Ekman and W. V. Friesen, "Facial action coding system (facs)," 01 2002.

[4] G. Liu, Z. Pan, and Z. Lin, "Style subspaces for character animation," *Comput. Animat. Virtual Worlds*, vol. 19, no. 3-4, pp. 199–209, Sep. 2008. [Online]. Available: http://dx.doi.org/10.1002/cav.v19:3/4

[5] G. W. Taylor and G. E. Hinton, "Factored conditional restricted boltzmann machines for modeling motion style," in *Proceedings of the 26th Annual International Conference on Machine Learning*, ser. ICML '09. New York, NY, USA: ACM, 2009, pp. 1025–1032. [Online]. Available: http://doi.acm.org/10.1145/1553374.1553505

[6] E. S. Chuang, H. Deshpande, and C. Bregler, "Facial expression space learning," in *Proceedings of Computer Graphics and Applications*. IEEE, 2002, pp. 68–76.

[7] X. Ma, B. H. Le, and Z. Deng, "Style learning and transferring for facial animation editing," in *Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. ACM, 2009, pp. 123–132.

[8] Y. Cao, P. Faloutsos, and F. Pighin, "Unsupervised learning for speech motion editing," in *Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*. Eurographics Association, 2003, pp. 225–231.

[9] A. Braun, R. Queiroz, W. Lee, B. Feijo, and S. R. Musse, "Persona: A method for facial analysis in video and application in entertainment," *Comput. Entertain.*, vol. 16, no. 3, pp. 4:1–4:19, Sep. 2018. [Online]. Available: http://doi.acm.org/10.1145/3236495

[10] R. B. Queiroz, A. Braun, and S. R. Musse, "An adaptive methodology for facial expression transfer," in *2015 14th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*, Nov 2015, pp. 11–23.

[11] Y. Wang, X. Huang, C.-S. Lee, S. Zhang, Z. Li, D. Samaras, D. Metaxas, A. Elgammal, and P. Huang, "High resolution acquisition, learning and transfer of dynamic 3-d facial expressions," in *Computer Graphics Forum*, vol. 23, no. 3. Wiley Online Library, 2004, pp. 677–686.

[12] T. Okamoto, T. Shiratori, M. Glisson, K. Yamane, S. Kudoh, and K. Ikeuchi, "Extraction of person-specific motion style based on a task model and imitation by humanoid robot," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2014, pp. 1347–1354.

[13] L. Torresani, P. Hackney, and C. Bregler, "Learning motion style synthesis from perceptual observations," in *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. C. Platt, and T. Hoffman, Eds. MIT Press, 2007, pp. 1393–1400. [Online]. Available: http://papers.nips.cc/paper/3109-learning-motion-style-synthesis-from-perceptual-observations.pdf

[14] S. Taylor, T. Kim, Y. Yue, M. Mahler, J. Krahe, A. G. Rodriguez, J. Hodgins, and I. Matthews, "A deep learning approach for generalized speech animation," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 93:1–93:11, Jul. 2017. [Online]. Available: http://doi.acm.org/10.1145/3072959.3073699

[15] E. Zakharov, A. Shysheya, E. Burkov, and V. S. Lempitsky, "Few-shot adversarial learning of realistic neural talking head models," *CoRR*, vol. abs/1905.08233, 2019. [Online]. Available: http://arxiv.org/abs/1905.08233

[16] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "Openface: A general-purpose face recognition library with mobile applications," CMU-CS-16-118, CMU School of Computer Science, Tech. Rep., 2016.

[17] W. V. Friesen, P. Ekman *et al.*, "Emfacs-7: Emotional facial action coding system," *Unpublished manuscript, University of California at San Francisco*, vol. 2, no. 36, p. 1, 1983.

[18] M. . Dubuisson and A. K. Jain, "A modified hausdorff distance for object matching," in *Proceedings of 12th International Conference on Pattern Recognition*, vol. 1, Oct 1994, pp. 566–568 vol.1.

[19] K. Teknomo, "K-means clustering tutorial," *Medicine*, vol. 100, no. 4, p. 3, 2006.

[20] J. Gratch, J. Rickel, E. Andre, J. Cassell, E. Petajan, and N. Badler, "Creating interactive virtual humans: some assembly required," *IEEE Intelligent Systems*, vol. 17, no. 4, pp. 54–63, July 2002.