Effect of facial cues on identification

Radoslaw Niewiadomski — Telecom ParisTech
Jing Huang — Telecom ParisTech
Catherine Pelachaud — LTCI-CNRS
{niewiado, jing.huang, pelachaud}@telecom-paristech.fr

Abstract

In this work we study factors that may influence the identification of single facial actions (e.g., frown or cheek raising), namely their intensity, dynamics and the application of wrinkles. Results of our evaluation study show that single facial actions are better identified when they are dynamic and with higher intensity. On the other hand, intense expressions of single facial actions are perceived less natural and less realistic. In this paper we also describe our wrinkles model. From our evaluation study, we found that wrinkles did not improve significantly the identification of facial actions.

Keywords: facial animation and expressions, action units, wrinkles

1. Introduction

Several studies (e.g. [1-4]) focus on the perception of synthesized facial expressions of the virtual agent (VA). They mainly focus on the stereotypical full-blown expressions of emotions. However recent research e.g. [5] shows that expressions of emotion are rather a sequence of facial actions than a full-blown single-shot displays. The later ones occur rarely in the real-life interactions. Single facial actions are ambiguous as they can be the components of different facial expressions. For example the action *lips-corners-up* (i.e. AU 12) occurs in the expression of joy as well as of embarrassment [5]. They also are more subtle (consequently more difficult to perceive). While the perception of a full-blown stereotypical expression can be done from a subset of its visible features (e.g. frown in anger, see [6]) it is not the case for single facial actions which are unitary features. Thus, to understand the perception of synthetic emotional expressions and generate more realistic ones, it is more appropriate to focus on single facial actions rather than on stereotypical expressions. For our perceptive test we have chosen some frequently appearing facial actions corresponding to certain Action Units (AUs [7]). We have tested 3 factors that may influence their perception: dynamic displays (vs static image), their (muscular) intensity, and the role of wrinkles. We have measured their identification rate, their realism and their naturalness.

The previous studies focus on the identification of synthetic expressions of basic emotions. Katsyri and Sams [2] showed that synthetic dynamic expressions were identified better than static ones only for expressions whose static displays were not-distinctive. In a similar study of Noel et al. [3] the effect of the dynamics was however not observed. In Bartneck and Reichenbach's work [4] the higher intensity expressions were better recognized. Finally, in Courgeon et al. [1] the application of wrinkles increases the agent's expressivity but does not improve the recognition. All these studies used the stereotypical expressions of the basic emotions. However these stereotypical expressions can be easily identified even without wrinkles or only from the images. This may not be the case for more "unitary" actions. These actions may be more easily confounded with each other. This made us decide to study single facial actions rather than full-blown expressions.

2. Model of wrinkles

Lately there are two main approaches that have been used to generate wrinkles effect: texture and geometry based methods [1] [8]. Due to the computation cost, we have chosen to apply the texture based technique for our model. This method, such as the bump mapping technique, was first introduced by Blinn [9]. The author proposed to modify the surface normal vector before the lighting computation, thus the bent normal can give the visually satisfying result without changing the surface geometry.

For the purpose of this study we generate wrinkles using Screen Space Bump Wrinkle approach. It extends Blinn's idea [9] and is based on GPU Bump Mapping approach [10]. We simulate wrinkle effect by performing the computation of perturbed Normal vector only in screen space with Pixel Shader (OpenGL/ DirectX). So the complexity only depends on the number of pixels, not on the number of vertices of the facial model. The perturbed normal vector depends only on the input surface normal and on the height value in screen space. We do not need to apply transformation between tangent space and world space. The bent normal vector is computed by composing the changes of local surface gradient of x and y directions in screen space for each pixel.

We define 12 groups of wrinkles in textures, which correspond to different AUs. In runtime, when an action unit is to be activated on the face mesh, the GPU receives its intensity value and computes the corresponding wrinkles. The final result is the composition of all active wrinkles (see Figures 1 and 2).

3. Experiment setup

We chose the following 6 facial actions: 3 for the upper face and 3 for the lower face. They are: frown (AU4), oblique eyebrows (AU1+4), raised eyebrows (AU1+2), cheeks raise (AU6), lip straightly extended (AU20) and lip corners pulled up (AU12). These actions are particularly important in communicating emotions by VA. Some may occur in several expressions (e.g. frown), others have a particular role (e.g. AU6 in the spontaneous expression of joy). The remaining actions were intentionally chosen to create confusions e.g. oblique and raised eyebrows.

In the experiment we used 128 different stimuli. Two different agents were used: one female and the other male. All 6 facial actions were displayed with and without wrinkles and with low and high intensity (see Figures 1 and 2). Finally they were shown by: image in the expression at its apex (static condition) and in animation (dynamic condition). This gives 96 stimuli. The remaining 32 stimuli presented other actions such as wide opening of the eyes or no action at all. They were not considered in the results. All the animations were validated by a certificated FACS expert.

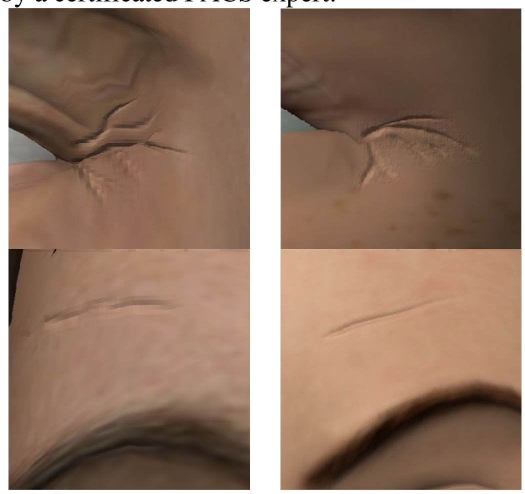

Figure 1. Wrinkles of two agents (a female on the right, and a male on the left) corresponding to action AU6 (cheek raise) and AU1+2 (raised eyebrows).

The test was placed on the web and it was organized as a set of web pages. Each web page displayed one stimulus at a times. Participants could see the animations only once and they had to answer all the questions before seeing the next animation. Each participant could evaluate maximum 32 stimuli (16 upper, 16 lower-face, 16 animations and 16 images). They were however told to stop the experiment wherever they wanted. This explains the different number of answers for each stimulus.

In the experiment we ask participants to identify stimuli using 5 point Likert scale from "totally disagree" to "totally agree". For the upper facial actions there are 3 separate scales labelled "raised eyebrow in half circle", "frowning" and "raised eyebrow in oblique". For the lower face expressions there are 3 other scales: "cheek raise", "lip extension - corner up", "lip extension - straight". In both cases participants were also asked to express their opinion on the naturalness and realism using two separate 5 point Likert scales. Three hypotheses were tested:

- H1. High intensity facial actions are better identified but they are less natural and less realistic,
- H2. Facial actions displayed by an animation are better identified than the ones presented by static images,
- H3. Facial actions with wrinkles are better indentified, or, at least they decrease confusions.

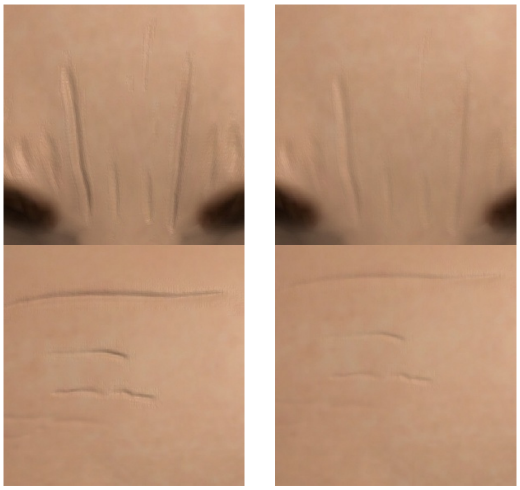

Figure 2: Wrinkles resulting of eyebrow movements with high and low intensity (AU4 and AU 1+2).

	high intensity			***low intensity***		
AU	***EXP***	***NAT***	***RE***	***EXP***	***NAT***	***RE***
1 + 2	**4,23**	3,40	3,32	**3,18**	3,46	3,33
4	**3,84**	3,75	3,69	**3,31**	3,9	3,82
1 + 4	**3,03**	3,28	3,23	**2,64**	3,33	3,35
6	**3,14**	**3,03**	**3,03**	**2,46**	**3,33**	**3,34**
20	**4,33**	**2,30**	**2,34**	**4,02**	**3,21**	**3,16**
12	**4,21**	**2,67**	**2,72**	**3,77**	**3,29**	**3,36**

Table 1: [1] The effect of intensity (significant values in bold).

4. Results

We collected 2092 answers in total from 92 participants (aged 19-51). The Likert scale usually does not satisfy the condition of the normal distribution. For this reason we applied to the data non parametric Mann-Whistley test. The global Mann-Whistley test shows the effect of intensity and dynamics on naturalness and realism. At the same time the effect of wrinkles was not observed. Next, in order to compare the identification rates we considered each facial action separately. Five out of six actions (see Table 1) were better identified on the more intense expressions than on the low intense ones (M-W test, $p<0.05$). Similar tendency was observed for AU1+4 (M-W test, $p=0.065$). At the same time 3 intense actions (AU12, 20 and 6) were perceived less realistic while 2 of them less natural (M-W test, $p<0.05$). Similar tendency was observed for the third action, AU6, (M-W test, $p=0.065$).

Three out of six actions (AU4, AU6 and AU12) were better identified from the animations than from the images (M-W test, $p<0.05$). Similar tendency was observed for AU1+4 (M-W test, $p=0.088$). 4 actions were perceived more natural and 2 more realistic when presented as animations (see Table 2).

	animation			***static image***		
AU	***EXP***	***NAT***	***REA***	***EXP***	***NAT***	***REA***
1 + 2	3,82	**3,68**	**3,53**	3,55	**3,11**	**3,07**
4	**3,86**	3,79	3,74	**3,21**	3,87	3,77
1 + 4	**3,00**	**3,48**	3,41	**2,63**	**3,10**	3,15
6	**3,46**	3,24	3,24	**1,89**	3,10	3,10
20	4,20	**3,07**	**3,06**	4,15	**2,34**	**2,35**
12	**4,22**	**3,14**	**3,18**	**3,70**	**2,80**	**2,87**

Table 2: The effect of presentation (significant values in bold).

Finally, only one action, AU4, (M-W, test $p<0.05$) was significantly better indentified with wrinkles. To check if the confusion rate is influenced by the wrinkles or not, we evaluated the number of "first choices": For each facial action we calculated the percentage of answers in which the corresponding label received the higher score than the remaining facial labels.

Wrinkles:	**4**	**1+4**	**1+2**	**20**	**12**	**6**
with	0.73	0.29	0.57	0.56	0.74	0.47
without	0.5	0.47	0.68	0.52	0.77	0.45

Table 3: "First choice" percentage.

5. Discussion

In this experiment we tested 3 hypotheses. The hypothesis H1 concerning the intensity of AUs was confirmed. The more intense AUs are the better indentified. At the same time they are perceived less natural and less realistic. Consequently, when creating facial animations,

[1] The second column presents the mean of the responses to the appropriate question.

animators need to balance between the readiness and naturalness of synthesized facial expressions. In our study this relation was particularly visible for the lower face actions.

Regarding the hypothesis H2, as expected, the animations were better indentified than the images. They were also perceived more natural (4 out 6) and more realistical (3 out 6). It also confirms that the dynamics of an expression is the most important cue allowing identification.

Finally the effect of wrinkles was limited (Hypothesis 3). Only for one action the result was significantly higher. In the case of ambiguous actions wrinkles may even increase the confusion. Indeed for a high intensity expressions the correct identification of an action AU1+4 was significantly lower (and significantly higher for a wrong label) in the wrinkles condition. Thus, synthetic wrinkles may influence facial action identification both positively or negatively and consequently, they are not very helpful in this task.

Six facial actions were considered in our evaluation. Below we discuss them in details.

AU 4. The Frown is the only action that is better identified with wrinkles, from the videos and with higher intensity. **Oblique (AU1+2) and raised eyebrows (AU1+4).** We chose these two actions as they can be easily confounded. They use the same unitary action, namely AU1 (inner raise eyebrow). Indeed both received the smallest values in the "first choice" test. It seems that the wrinkles did not help in distinguishing between these two actions. **AU6** was recognized much better in the intense and dynamic conditions. Indeed this action is particularly difficult to be perceived from the static images. Nevertheless, similarly to other lower face actions the intense AU6 was perceived less natural. **AU12 and AU20.** The last two actions can be confused (especially with low intensity). Indeed they were better recognized in the intense and dynamic conditions rather than in the image condition. On the other hand wrinkles did not improve the identification of these two actions.

This work studies the role of the intensity, the dynamics and the wrinkles in the perception of the synthesized facial displays. The important contribution of this research is that instead of using stereotypical expressions (that rarely occur) we analyze single facial actions i.e. unitary components of more "real-life"-like expressions. Secondly, more intense expressions are perceived less natural and less realistic. Last but not least the role of wrinkles is ambivalent.

Acknowledgements

This research is supported by the FP7 - IP VERVE Project and ANR CECIL.

References

[1] M. Courgeon, S. Buisine, J.-C. Martin. Impact of expressive wrinkles on perception of a virtual character's facial expressions of emotions. In: Proceedings of IVA'09, pages 201–214, 2009

[2] J. Katsyri, M. Sams. The effect of dynamics on identifying basic emotions from synthetic and natural faces. *Int. J. Hum.-Comput. Stud.* 66, 4:233-242, 2008

[3] S. Noël, S. Dumoulin, T. Whalen, J. Stewart. Recognizing emotions on static and animated avatar faces. *HAVE 2006*, pages 99-104, Ottawa, Canada, 2006

[4] C. Bartneck, J. Reichenbach. Subtle emotional expressions of synthetic characters. *Int. J. Hum.-Comput. Stud.* 62, 2: 179-192, 2005

[5] D. Keltner. Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. Journal of Personality and Social Psychology, 68:441–454, 1995

[6] R. Niewiadomski, S. Hyniewska, C. Pelachaud. Constraint-Based Model for Synthesis of Multimodal Sequential Expressions of Emotions. IEEE Transactions on Affective Computing, 2, 3:134-146, 2011

[7] P. Ekman, W. Friesen. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, 1978.

[8] C. Larboulette, M.-P. Cani, Real-time dynamic wrinkles. In CGI'04, pages 522–521, IEEE, 2004

[9] J.F. Blinn. Simulation of wrinkled surfaces. In: *Proceedings of* SIGGRAPH '78, pages 286-292, ACM, NY, USA, 1978

[10] M.S. Mikkelnsen. Bump mapping unparametrized surfaces on the GPU. Journal of Graphics, GPU, and Game Tools 15, 1: 49–61, 2010