

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/334170560>

OpenFACS : an open source FACS-based 3D face animation system

Preprint · August 2019

CITATIONS

0

READS

510

2 authors:



Vittorio Cuculo

University of Milan

28 PUBLICATIONS 54 CITATIONS

[SEE PROFILE](#)



Alessandro D'Amelio

University of Milan

23 PUBLICATIONS 32 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Facial signals [View project](#)



Eye movements [View project](#)

OpenFACS : an open source FACS-based 3D face animation system

Vittorio Cuculo and Alessandro D'Amelio

PHuSe Lab - Dipartimento di Informatica
University of Milan, Milano, Italy
`{vittorio.cuculo,alessandro.damelio}@unimi.it`

Abstract. We present OpenFACS, an open source FACS-based 3D face animation system. OpenFACS is a software that allows the simulation of realistic facial expressions through the manipulation of specific action units as defined in the Facial Action Coding System. OpenFACS has been developed together with an API which is suitable to generate real-time dynamic facial expressions for a three-dimensional character. It can be easily embedded in existing systems without any prior experience in computer graphics. In this note, we discuss the adopted face model, the implemented architecture and provide additional details of model dynamics. Finally, a validation experiment is proposed to assess the effectiveness of the model.

Keywords: Facial expression · FACS · Emotion · 3D facial animation · HCI

1 Introduction

The chief purpose of systems that realise natural interactions is to remove the mediation between human and machine typical of classic interfaces. Among the main modalities of interaction there are speech, gestures, gaze and facial expressions (see [24] for a thorough review). The latter is particularly relevant because it plays a fundamental role in non-verbal communication between human beings. In particular, as to the human-computer interaction (HCI) aspects, the ability to recognize and synthesize facial expressions allows the machines to gain significant communication skills, on the one hand by interpreting emotions relying on the face of a subject; on the other hand by translating their communicative intent through an output, such as movement, sound response or colour change [7]. The latter skill is the one specifically addressed by the system presented here, which aims at providing the scientific community with a tool that can be easily used and integrated into other systems.

2 Related works

A significant body of work concerning facial animation has been reported since its early stages [19]. In recent years, thanks also to the entertainment industry, this



Fig. 1. Visualization of the OpenFACS 3D face animation system with the actor showing a neutral facial expression.

field of research has undergone a boost in the technologies developed, reaching high levels of realism [14]. However, the demand for specific skills in the computer graphics field or, alternatively, the high cost of third-party software, creates a barrier to the effective usability of an animated 3D model by researchers not directly involved in the computer graphics field but who still require a precise and realistic visible facial response for emotion research [20, 8, 4, 5].

In terms of facial movements coding systems two are the prominent approaches: MPEG-4 and the Facial Action Coding System (FACS). The former [18] identifies a set of Face Animation Parameters (FAP), each corresponding to the displacement of a subset of 84 Feature Points (FP) of the face. These displacements are measured in FAP Units, defined as the distance between the fiducial points of the face. A similar approach, but with different motivation, has been realized by Paul Ekman [10] in the Facial Action Coding System (FACS). As the name suggests, the work presented here is based on the FACS, and related works that share the same system will be taken in consideration below. In 2009 a 3D facial animation system named FACE! [23] has been presented. This system was able to reproduce 66 action units from FACS as single activations or combined together. In the same year, Alfred [3] provided a virtual face with 23 facial controls (AUs) connected with a slider-based GUI, a gamepad, and a data glove. They concluded that the use of a gamepad for facial expression generation can be promising, reducing the production time without causing a loss of quality. A few years later, FACSGen 2.0 [15] provides a new animation software for creating facial expressions adopting 35 single AUs. The software has been evaluated by four FACS-certified coders resulting in good to excellent classification levels

for all AUs. HapFACS 3.0 [1] was one of the few free software, providing an API based on the Haptek 3D-character platform. It has been developed to address the needs of researchers working on 3D speaking characters and facial expression generation. Among the most recent software based on FACS, FACSHuman [12] represents a suite of plugins for MakeHuman 3D tool. These should permit the creation of complex facial expressions manipulating the intensity of all known action units.

All the above tools, when still available, are not free to use, nor cross-platform or easily integrable within a pre-existing system. The only exception could be represented by FACSHuman but, at the time of writing, no technical details or software are still provided.

3 Theoretical background

As mentioned in the introduction, facial expressions play a very important role in social interaction and their analysis has always represented a complex challenge [24, 6]. They have been under study since 1872 when Charles Darwin published “The Expression of the Emotions in Man and Animals” [9], positing his thesis on the universality of emotions as a result of the evolutionary process, and considering facial expressions as a residue of behaviour, according to the principle of the “serviceable habits”. From Darwin’s work also stems the view addressing the communicative function of emotions (*emotions as expressions*). Researchers have expanded on Darwin’s evolutionary framework toward other forms of emotional expression. The most notable ones are from Tomkins [22], proposing that there is a limited number of pan-cultural basic emotions, such as surprise, interest, joy, rage, fear, disgust, shame, and anguish, and by Ekman and Friesen [11]. In particular, Ekman’s facial action coding system (FACS, [10]) influenced considerable research that tackles the affect detection problem developing systems that identify the basic emotions through facial expressions (and in particular extracting facial action units). According to the FACS, the emotional manifestations occur through the activation of a series of facial muscles which are described by 66 action units (AU). This encoding system allows the realisation of about 7000 expressions that can be found on a human face by the combination of such atoms. Each AU is identified by a number (AU1, AU2, AU4 ...) and correspond to the activation of a single facial muscle (e.g. *Zygomatic Major* for AU12). Intensities of FACS are expressed by letters from A (minimal intensity) to E (maximal intensity) postponed to the action unit number (e.g. AU1A is the lowest representation of the Inner Brow Raiser action unit).

4 OpenFACS system

The system presented here is an open-source, cross-platform, stand alone software. It relies on a 3D face model where FACS AUs are employed as a reference

for creating specific muscle activations, that can be manipulated through a specialized API. OpenFACS software, including examples of usage and Python interface with the API are freely available at <https://github.com/phuselab/openFACS>.

4.1 Model

The 3D model adopted in OpenFACS is instantiated by exploiting the free software Daz3D¹. It consists of approximately 10000 vertices and 17000 triangles (ref. Fig. 2). The handling of its parts relies on the so-called morph targets (also known as blend shapes), that describe the translation of a set of vertices in the 3D space to a new target position. In the proposed system, each of the 18 considered action units is implemented by the contribution of one or more morph targets. Table 1 summarizes the correspondence between the considered action units and the respective morph targets.

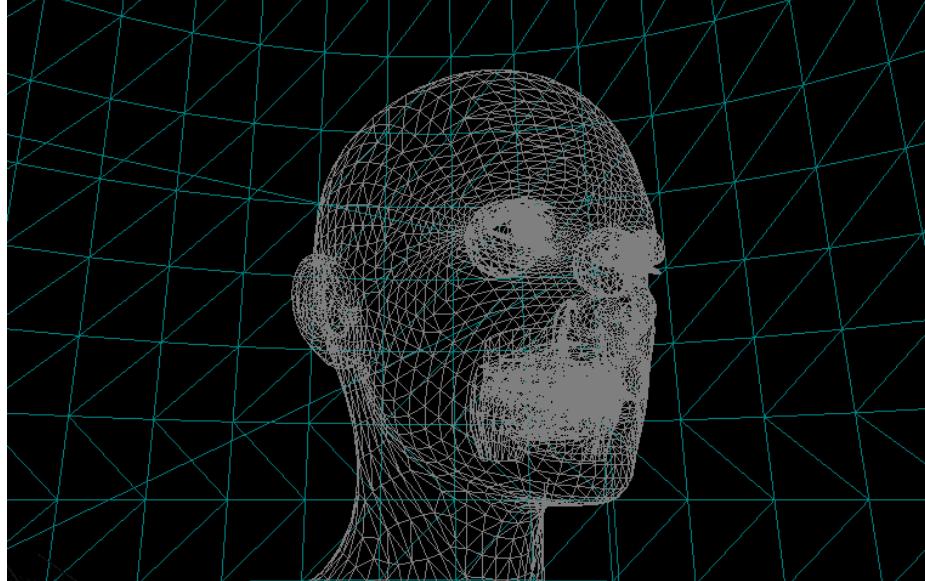


Fig. 2. Wireframe visualisation of the adopted 3D face model consisting of ~ 10000 vertices and 17000 triangles.

To reproduce the FACS standard, the intensities for each of the action units are expressed with a value in the range [0, 5], where 0 corresponds to the absence of activation and 1 to 5 follows the A to E encoding. The value of muscle activation speed ranges from 0 to 1 and directly affects the linear interpolation speed from the current to the target configuration (e.g.: speed equal to 0.25

¹ <https://www.daz3d.com/>

Table 1. List of considered FACS action units and corresponding model morph targets.

Action Unit	Description	Morph target
1	Inner Brow Raiser	head_CTRLBrowInnerUp
2	Outer Brow Raiser	head_CTRLBrowOuterUp
4	Brow Lowerer	head_PHMBrowSqueeze
5	Upper Lid Raiser	head_CTRLEyeLidsTopUp
6	Cheek Raiser	head_PHMSSmileFullFace head_PHMMouthSmile head_CTRLEyesSquint
7	Lid Tightener	head_CTRLEyesSquint
9	Nose Wrinkler	head_PHMNoseWrinkle
10	Upper Lip Raiser	head_CTRLLipTopUp
12	Lip Corner Puller	head_PHMMouthSmile
14	Dimpler	head_PHMCheeksBaloonPucker
15	Lip Corner Depressor	head_PHMMouthFrown
17	Chin Raiser	head_CTRLLipBottomOut head_PHMMouthOpen
20	Lip Stretcher	head_PHMLipsPucker
23	Lip Tightener	head_CTRLMouthNarrow
25	Lips parter	head_PHMLipsPart
26	Jaw Drop	head_PHMMouthOpenWide
28	Lip Suck	head_CTRLLipTopDown head_CTRLLipBottomIn
45	Blink	head_CTRLEyeLidsTopDown head_CTRLEyeLidsBottomUp

means that every tick it goes 25% of the way to the target). In Figure 3 are shown the prototypical facial expressions of six basic emotions: anger, disgust, fear, happiness, sadness and surprise. These were obtained following the FACS [10], for instance, the surprise is made up by the combination of AU1, AU2, AU5 and AU26 (all with intensities equal to C).



Fig. 3. Examples of six different facial expressions resulting from the combination of a specific set of action units, reported in brackets.

4.2 Architecture

The 3D model presented above is imported and managed by the source-available game engine Unreal Engine². In such environment, an UDP based API server has been developed. Such API permits to an external software, even remotely, to communicate action units intensity values as well as the speed of muscle activation. This information must be serialized following the JSON data-interchange format and exchanged via classical client-server pattern. This implementation choice paves the way to a cross-platform and language-independent embedding into external systems. The engine, in real-time, takes care to realise the desired facial movement.

A sequence diagram of the implemented architecture can be found in Figure 4, it includes the *ServerListener* that implements the UDP server, the *JSONParser* receives and interprets the JSON messages, the *AUInfos* keeps the information about action unit intensities and speed, while the *HumanMesh* is the only who can operate on the 3D model.

² <http://www.unrealengine.com/>

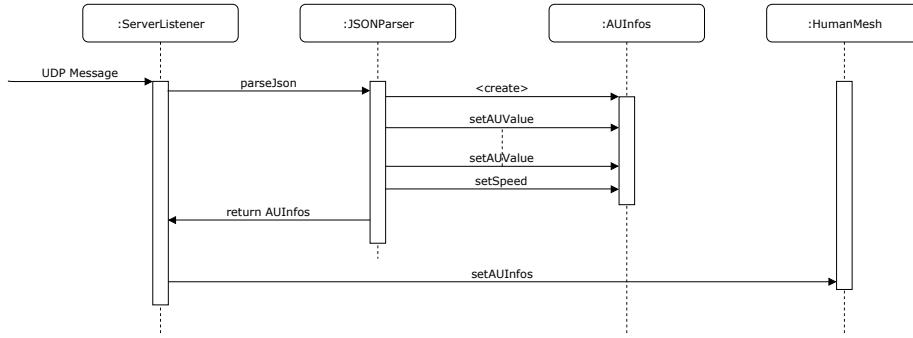


Fig. 4. Sequence diagram of the system to set a new facial expression.

4.3 Additional details

In order to increase the realism of the simulation, in addition to a natural background and a clearer and realistic lighting, some basic automatic movements were added, unrelated to the action units. Bearing in mind that movements could amplify the effect of the so called *uncanny valley* [17, 16], we modelled very slight movements of mouth corners, eyelids, neck and eye blink as described below.

Eye blink It has been shown that the average blink rate varies between 12 and 19 blinks per minute [13]. Here it is determined by sampling from a Normal distribution with $\mu = 6.0$ and $\sigma = 2.0$. The sampled value represents the waiting time (in seconds) between two consecutive blinks. The blink duration, in this case, is constant and set to 360 milliseconds following the findings of Schifmann [21] who claims that blink duration lies between 100 and 400 milliseconds.

Mouth corners Each mouth corner is handled by a specific morph target, namely *PHMMouthNarrowR* and *PHMMouthNarrowL*. The position (m) of such morph targets assume values in range $[-1, 1]$. At each tick, the probability that a corner is moved is equal to 0.8. In this case, when the morph target position is $m = 0$, it assumes probability 0.5 to be increased or decreased. The step done at each tick (t) is equal to $\Delta_m = 0.02$. A movement in any direction makes that direction less likely in the next step, in other words the probability that $m_{t+1} = m_t + \Delta_m$ is equal to

$$p = 1 - (0.5 + 0.5 * m) \quad (1)$$

This choice allows to keep constrained mouth corner movements reducing the possibility of abrupt changes.

Eyelids Eyelids movement results in a constant and slight vibration of the interested vertices. The algorithm behind this movement is the same as the one

described for mouth corners. Differently from Equation 1, the probability that $m_{t+1} = m_t + \Delta_m$ is equal to $p = 1 - (0.8 + 0.2 * m)$, thus obtaining a lower probability of these movements.

Neck Neck movement is realised through the manipulation of two different morph targets: *CTRLNeckHeadTwist* and *CTRLNeckHeadSide*. Even in this case the two morph targets are governed by the approach described for the previous movements, where Eq. 1 becomes $p = 1 - (0.6 + 0.4 * m)$, and $\Delta_m = 0.001$, since the neck movements require to be much more contained.

5 Validation

To characterise the behaviour of proposed simulation model, in terms of its expressive abilities, we conceived an experimental setup where an “expert” evaluates the unfolding of the facial dynamics of human \mathcal{H} and artificial \mathcal{A} expressers. In this perspective, a human expert (e.g., a FACS certified psychologist) would compare AUs’ behaviour of \mathcal{H} and \mathcal{A} while expressing the same emotion.

We use in the role of a “synthetic expert” a freely available AU detector [2]. The inputs to the detector are the original frame sequence of \mathcal{H} ’s facial actions and \mathcal{A} ’s output. The AU detector provides, at each frame, the activation level of the following AUs ($N_{AU} = 12$): $AU_k, k = 1, 2, 5, 9, 12, 14, 15, 17, 20, 23, 25, 26$.

For each of the six basic emotions (anger, disgust, fear, happiness, sadness and surprise), we run the AU detector over either an original \mathcal{H} ’s sequence, excerpted from the classic Cohn-Kanade dataset, and its synthetic reproduction, i.e., \mathcal{A} ’s actual expression sequence resulting from the prototypical activation of facial action units as conceived in [10]. It is worth mentioning that no explicit expression design has been carried out, conversely the generated facial expression is the result of simultaneous activation of those AUs that, according to [10], are responsible for displaying the emotion at hand.

To quantify the expressive abilities of OpenFACS we compared the AU activations of \mathcal{H} and \mathcal{A} in terms of their two dimensional correlation coefficient. This was done for each of the six emotion categories. No specific tuning was adopted for optimizing detector performance, namely, it was used as black box AU expert.

Figure 5 illustrates the results achieved in terms of time-varying AU activation maps (each row denoting a single AU activation in time, brighter colour corresponding to higher activation). It can be noted at a glance that human’s patterns of activation, for each expression, are similar to artificial’s patterns. This is also confirmed by the fairly strong correlation (anger: 0.83, fear: 0.73, happiness: 0.89, sadness: 0.74, surprise: 0.86) achieved in all the basic emotions but one (disgust: 0.40).

In the latter case, although a positive correlation exists it is not as strong as in the others. This result is mostly due to the lack of wrinkles around the nose area in the OpenFACS model, that probably brings the AU expert to miss the activation of some AUs, e.g. AU9 and AU10.

Interestingly enough, associated AUs can be activated as surrogates, for instance AU14 (Dimpler, that forms in the cheeks when one smiles) as consequence of AU12 activation. This effect can be easily noticed for the “happy” expression.

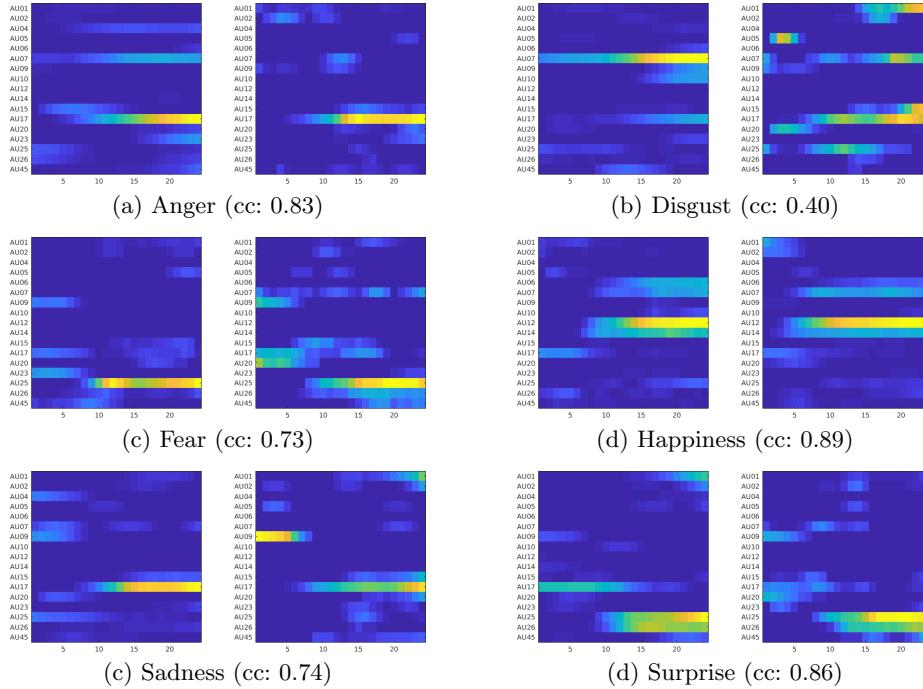


Fig. 5. AU activation maps for the expression associated with each of the six basic emotions performed by the human actor (left) and OpenFACS model (right). Each row represents the activation over time of a single AU (brighter colours for higher activations). In brackets the 2D correlation coefficient between human and simulated AU activation maps.

6 Conclusions and future works

In this paper we presented a novel 3D face animation system that relies on the Facial Action Coding System for the facial movements. This framework is intended to be used by researchers not directly involved in the computer graphics field but who still require a precise and realistic visible facial expression simulation. OpenFACS would overcome the lack of a free and open-source system for such purpose. The software can be easily embedded in other systems, providing a simple API.

The model has been evaluated in terms of its expressive abilities by means of quantitative comparison with the unfolding of humans' facial dynamics, for each of the six basic emotions. Obtained results proved the effectiveness of the proposed model.

In a future work, in addition to the objective evaluation already done, OpenFACS could be validated by a set of FACS-certified coders, in order to consolidate the operational definitions of the implemented AUs. Moreover, OpenFACS could benefit from the implementation of phonemes actions that may be used to simulate speech, increasing its communication skills.

7 Acknowledgements

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Quadro P6000 GPU used for this research.

References

1. Amini, R., Lisetti, C., Ruiz, G.: HapFACS 3.0: FACS-Based Facial Expression Generator for 3D Speaking Virtual Characters. *IEEE Transactions on Affective Computing* **6**(4), 348–360 (Oct 2015). <https://doi.org/10.1109/TAFFC.2015.2432794>
2. Baltrušaitis, T., Mahmoud, M., Robinson, P.: Cross-dataset learning and person-specific normalisation for automatic action unit detection. In: 11th IEEE Int. Conf. and Workshops on Automatic Face and Gesture Recognition. vol. 6, pp. 1–6. IEEE (2015)
3. Bee, N., Falk, B., André, E.: Simplified facial animation control utilizing novel input devices: A comparative study. In: Proceedings of the 14th international conference on Intelligent user interfaces. pp. 197–206. ACM (2009)
4. Boccignone, G., Bodini, M., Cuculo, V., Grossi, G.: Predictive sampling of facial expression dynamics driven by a latent action space. In: Proceedings of the 14th International Conference on Signal-Image Technology Internet-Based Systems (SITIS), Las Palmas de Gran Canaria, Spain. pp. 26–29 (2018)
5. Boccignone, G., Conte, D., Cuculo, V., D'Amelio, A., Grossi, G., Lanzarotti, R.: Deep construction of an affective latent space via multimodal enactment. *IEEE Transactions on Cognitive and Developmental Systems* **10**(4), 865–880 (Dec 2018). <https://doi.org/10.1109/TCDS.2017.2788820>
6. Ceruti, C., Cuculo, V., D'Amelio, A., Grossi, G., Lanzarotti, R.: Taking the hidden route: Deep mapping of affect via 3d neural networks. In: International Conference on Image Analysis and Processing. pp. 189–196. Springer (2017)
7. Cuculo, V., Lanzarotti, R., Boccignone, G.: The color of smiling: computational synesthesia of facial expressions. In: International Conference on Image Analysis and Processing. pp. 203–214. Springer (2015)
8. D'Amelio, A., Cuculo, V., Grossi, G., Lanzarotti, R., Lin, J.: A note on modelling a somatic motor space for affective facial expressions. In: International Conference on Image Analysis and Processing. pp. 181–188. Springer (2017)
9. Darwin, C.: *The Expression of the Emotions in Man and Animals*. John Murray (1872)
10. Ekman, P.: *Facial action coding system (facs). A human face* (2002)

11. Ekman, P., Friesen, W.V.: Constants across cultures in the face and emotion. *Journal of personality and social psychology* **17**(2), 124 (1971)
12. Gilbert, M., Demarchi, S., Urdapilleta, I.: FACSHuman a Software to Create Experimental Material by Modeling 3D Facial Expression. In: Proceedings of the 18th International Conference on Intelligent Virtual Agents - IVA '18. pp. 333–334. ACM Press, New York, New York, USA (2018)
13. Karson, C.N., Berman, K.F., Donnelly, E.F., Mendelson, W.B., Kleinman, J.E., Wyatt, R.J.: Speaking, thinking, and blinking. *Psychiatry research* **5**(3), 243–246 (1981)
14. Klehm, O., Rousselle, F., Papas, M., Bradley, D., Hery, C., Bickel, B., Jarosz, W., Beeler, T.: Recent advances in facial appearance capture. In: Computer Graphics Forum. pp. 709–733. Wiley Online Library (2015)
15. Krumhuber, E.G., Tamarit, L., Roesch, E.B., Scherer, K.R.: Facsgen 2.0 animation software: Generating three-dimensional facs-valid facial expressions for emotion research. *Emotion* **12**(2), 351 (2012)
16. MacDorman, K.F., Ishiguro, H.: The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies* **7**(3), 297–337 (2006)
17. Mori, M.: The uncanny valley. *Energy* **7**(4), 33–35 (1970)
18. Pandzic, I.S., Forchheimer, R.: MPEG-4 facial animation: the standard, implementation and applications. John Wiley & Sons (2003)
19. Parke, F.I., Waters, K.: Computer facial animation. AK Peters/CRC Press (2008)
20. Roesch, E.B., Sander, D., Mumenthaler, C., Kerzel, D., Scherer, K.R.: Psychophysics of emotion: The quest for emotional attention. *Journal of Vision* **10**(3), 4–4 (2010)
21. Schiffman, H.R.: Sensation and perception: An integrated approach. John Wiley & Sons (1990)
22. Tomkins, S.: Affect, imagery, consciousness, vol. 1. Springer, New York (1962)
23. Villagrasa, S., Sánchez, S., Others: FACE! 3D Facial Animation System based on FACS. IV Iberoamerican Symposium in Computer Graphics pp. 203–209 (2009). <https://doi.org/10.1002/9780470682531.pat0170>
24. Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F., Schroeder, M.: Bridging the gap between social animal and unsocial machine: A survey of social signal processing. *IEEE Transactions on Affective Computing* **3**(1), 69–87 (2012)