Virtual EMG via facial video analysis

Giuseppe Boccignone[1], Vittorio Cuculo[12], Giuliano Grossi[1], Raffaella Lanzarotti[1], and Raffaella Migliaccio[1]

[1] PHuSe Lab - Dipartimento di Informatica, Università degli Studi di Milano
Via Comelico 39/41, Milano, Italy
giuseppe.boccignone@unimi.it
{grossi,lanzarotti}@di.unimi.it
raffaella.migliaccio@studenti.unimi.it
[2] Dipartimento di Matematica, Università degli Studi di Milano
Via Cesare Saldini 50, Milano, Italy
vittorio.cuculo@unimi.it

Abstract. In this note, we address the problem of simulating electromyographic signals arising from muscles involved in facial expressions - markedly those conveying affective information -, by relying solely on facial landmarks detected on video sequences. We propose a method that uses the framework of Gaussian Process regression to predict the facial electromyographic signal from videos where people display non-posed affective expressions. To such end, experiments have been conducted on the OPEN EmoRec II multimodal corpus.

1 Introduction

The face is the locus of a great deal of emotional expressions and researchers in different fields crossing with affective science [9] have been keen on facial electromyographic measures of muscle activity, in particular those related to the *zygomaticus major* and the *corrugator supercilii* (see Fig. 1a). The motivation for such endeavour is straightforward: the *zygomaticus major* controls the corners of the mouth (e.g., by pulling them back and up into a smile), the *corrugator supercilii* hauls the brow down and together into a frown [18]. In brief, facial electromyography is a reliable detector of the affective state, either in the continuous dimension of valence (positive versus negative affective state) [18], or to reveal the discrete emotions[16].

Electromyography measures the electrical potentials arising from skeletal muscles [27]. Facial EMG (fEMG), is based on recording the difference in electrical potential pairs of electrodes that are placed close together on the target facial muscle (Fig. 1b). Main advantages of fEMG stem from 1) the capability of intercepting even very weak affective expression and 2) the very good time resolution that allows to reliably register sudden expression changes. On the other hand, the need of placing electrodes over the face limits the applicability of this sensor to laboratory acquisition only (see again Fig. 1b). Cogently, in this case
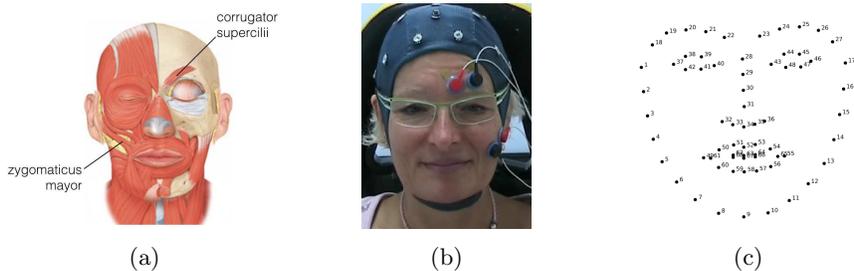
Fig. 1: (a) Anatomical location of facial muscles involved in this study. (b) Electrode placing to detect the activity of the *zygomaticus major* and the *corrugator supercilii* muscles. (c) Facial landmarks inferred by the method [8]

and more generally, the option of monitoring physiological signals via noncontact means has promise for a variety of out-of-lab applications well beyond the affective computing realm [23].

Whilst there is a number of works addressing noncontact, physiological measurements of heart rate, e.g. [23,26,30], to the best of our knowledge, this is the first attempt to estimate fEMG signals from video sequences.

We argue that, apart from the *per se* appealing issue of avoiding the obtrusiveness of fEMG, the idea of virtual fEMG derived from observing natural, non-posed facial expression, can be important for dealing with emotion understanding in a broader perspective (see Section 4, for a discussion). All things considered, this endeavour is at this stage affordable, given that in the last decade, the number of public repositories has grown larger, where behavioral data have been recorded by multiple modalities [7,29], hence providing adequate training sets and benchmarking, as will be detailed in Section 3.

In Section 2 the method we propose for the virtual fEMG generation is described; in Section 3 the experiments and the obtained results are shown and discussed. In Section 4 conclusive remarks on this preliminary study are given.

2 Method

Given a video stream $\mathbf{I}(t)$, fEMG signal generation is obtained by relying on perceived facial fiducial points, or landmarks. In a nutshell, landmarks are detected in a sparse coding framework and signal generation is obtained through Gaussian Process (GP) regression and prediction. More precisely, use the following random variables (RVs):

- **E**: a set of fEMG data over time intervals, i.e. a set of signals **e**;
- **L**: a set of landmarks **l**, over time intervals, each $\mathbf{l}^i$ being a landmark;
- **F**: a set of feature responses **f**, over time intervals, each $\mathbf{f}^i$ being a local feature response;
- $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N] \in \mathbb{R}^{D \times N}$: the matrix of observed training patches.

– $\mathbf{W} = [\mathbf{w}_1, \cdots, \mathbf{w}_L] \in \mathbb{R}^{D \times L}$: a dictionary; each column $\mathbf{w}_i$ is referred to as a basis vector or atom;
– $\mathbf{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_N] \in \mathbb{R}^{L \times N}$ the latent sparse code matrix associated to $\mathbf{W}$.

Then the proposed method can be summarised as the sampling of the virtual fEMG signal $\widetilde{\mathbf{e}} = [e(1), e(2), \cdots, e(T)]$ from the joint conditional distribution:

$$\widetilde{\mathbf{e}} \sim P(\mathbf{E}, \mathbf{L}, \mathbf{F}, \mathbf{W} \mid \mathbf{X}, \mathbf{I}). \tag{1}$$

The joint pdf can be factorised as follows:

$$P(\mathbf{E}, \mathbf{L}, \mathbf{F}, \mathbf{W} \mid \mathbf{X}, \mathbf{I}) = P(\mathbf{E} \mid \mathbf{L}) \times P(\mathbf{L} \mid \mathbf{F}) \times P(\mathbf{F} \mid \mathbf{W}, \mathbf{I}) \times P(\mathbf{W} \mid \mathbf{X}) \tag{2}$$

The method can be best explained by starting from the last factor on the r.h.s. of Eq.2. In the sparse coding framework, such term supports dictionary inference given a set of training patches:

$$\mathbf{W}^* = \arg\max_{\mathbf{W}} P(\mathbf{W} \mid \mathbf{X}) \tag{3}$$

The problem of inferring dictionary $\mathbf{W}$ can be reduced to a maximum likelihood estimation $\mathbf{W}^* = \arg\max P(\mathbf{W} \mid \mathbf{X}) \approx \arg\max P(\mathbf{X} \mid \mathbf{W})$, where the observable patch vector $\mathbf{x}_i$ is approximated as a sparse combination of basis vectors $\mathbf{w}_i$, i.e. $\mathbf{x} = \mathbf{W}\mathbf{z} + \mathbf{v}$, $\mathbf{v}$ being a residual noise vector sampled from a zero mean Gaussian distribution $\mathcal{N}(0, \sigma^2 \mathbb{I})$. The dictionary can be derived under the Olshausen and Field approximation [21], $\log P(\mathbf{X}|\mathbf{W}) \approx \sum_{i=1}^{N} \max_{\mathbf{z}_i}[\log \mathcal{N}(\mathbf{x}_i|\mathbf{W}\mathbf{z}_i, \sigma^2\mathbb{I}) + \log P(\mathbf{z}_i)]$, and turned in the minimization of the negative log-likelihood (NLL). This can be done efficiently by using either the K-SVD [3] or the R-SVD [15] algorithms as shown in [1,2,14].

The third factor represents the feature likelihood under the current observable video $\mathbf{I}$ and the inferred dictionary. The goal here is to compute feature responses

$$\mathbf{F}^* \sim P(\mathbf{F} \mid \mathbf{W}, \mathbf{I}) \tag{4}$$

at each frame in $\mathbf{I}$. Here, we adopt the Histograms of Sparse Codes (HSC) representation to sample the local response $\mathbf{f}^i$ [8].

The second factor accounts for the detection of landmarks given the observed $\mathbf{F}^*$. A part-based detection approach is adopted [8], where every facial landmark can be modeled as a part, and the locations $\mathbf{L}$ of parts of the face can be generated according to m views or poses by some similarity transformation τ, giving rise to the global model $\mathbf{L}_{k,\tau}$. The generation of $\mathbf{L}$ can be accomplished by marginalising over the set of m models, i.e., $P(\mathbf{L}|\mathbf{F}) = \sum_{k=1}^{m} \int_{\tau} P(\mathbf{L}|\mathbf{L}_{k,\tau})P(\mathbf{L}_{k,\tau}|\mathbf{F})d\tau$. The term $P(\mathbf{L}|\mathbf{L}_{k,\tau})$ accounts for dependence of $\mathbf{L}$ from the global configuration $\mathbf{L}_{k,\tau}$.

Assume that: i) the locations of the parts $\{\mathbf{l}^i\}_{i=1}^{l}$ are conditionally independent of one another and the same holds for the detector responses $\mathbf{f}^i$; ii) the relation between the transformed model landmark and the true landmark is translationally invariant, i.e., $P(\mathbf{l}_{k,\tau}^i|l_{k,\tau})$ only depends on $\Delta\mathbf{l}_{k,\tau}^i = \mathbf{l}_{k,\tau}^i - \mathbf{l}^i$. Then, the following MAP solution can be derived,

$$\mathbf{L}^* = \arg\max_{L} \sum_{k=1}^{m} \int_{\tau} \prod_{i=1}^{l} P(\Delta\mathbf{l}_{k,\tau}^i) P(\mathbf{l}^i|\mathbf{f}^i)d\tau, \tag{5}$$

where the prior $P(\Delta \mathbf{l}_{k,\tau}^i)$ accounts for the *shape* or global component of the model, and $P(\mathbf{l}^i|\mathbf{f}^i)$ for the *appearance* or local component. The latter relies on patches representing HSC responses to face landmarks.

Eventually, the first factor on the r.h.s. of Eq. 2, is the likelihood supporting the generation of the fEMG signal given the extracted landmarks. The generative model behind the conditional distribution $P(\mathbf{E} \mid \mathbf{L})$, under Gaussian assumption, assumes that a realisation of a target electromyographic signal $\mathbf{e}$ is generated by a latent function $\mathbf{g} = \{g(\mathbf{d}_n)\}$ of a suitable measurement $\mathbf{d}$ of the landmarks corrupted by additive Gaussian noise. Thus, at time (frame index) t:

$$e(t) = g(\mathbf{d}(\mathbf{l}_p(t))) + \nu(t), \quad \nu \sim \mathcal{N}(0, \sigma_e^2) \tag{6}$$

where, in our case, $\mathbf{d}(\mathbf{l}_p)$ is a vector of distances over the *pool* $\mathbf{l}_p$, a subset of the extracted landmarks $\mathbf{l}$, which is suitable to capture muscle activity. Note that the mapping function $g(\cdot)$ needs not to be linear. In other terms, the conditional distribution $P(\mathbf{E} \mid \mathbf{L})$ is defined as the marginal likelihood $P(\mathbf{E} \mid \mathbf{L}) = \int P(\mathbf{E} \mid \mathbf{g}, \mathbf{L}) P(\mathbf{g} \mid \mathbf{L}) d\mathbf{g}$, where the marginalisation over the function values g, can be performed by using a GP prior distribution over functions $P(\mathbf{g} \mid \mathbf{L}) = \mathcal{N}(\mu_g(\mathbf{L}), k(\mathbf{L}, \mathbf{L}))$, $k(\mathbf{L}, \mathbf{L})$ being the kernel function [24], i.e. in our case

$$g(\mathbf{d}(\mathbf{l}_p)) \sim \mathcal{GP}(\mu(\mathbf{d}(\mathbf{l}_p)), k(\mathbf{d}(\mathbf{l}_p), \mathbf{d}'(\mathbf{l}_p))), \tag{7}$$

and where the likelihood of the observed targets is $P(\mathbf{E} \mid \mathbf{g}, \mathbf{L}) = \mathcal{N}(\mathbf{g}, \sigma_e^2 \mathbb{I})$, from which Eq. 6 is obtained. Note that, due to analytical tractability of the Gaussian distribution, all the above computations are determined in closed form so that, prior to the prediction of the virtual fEMG signal $\widetilde{\mathbf{e}}$, parameter learning can be efficiently performed on the given dataset $\{\mathbf{L}, \mathbf{E}\}$ (see Rasmussen and Williams [24] for details).

3   Experimental work

A) Experimental setup. The experiments have been conducted on the multimodal corpus OPEN EmoRec II [25]. The dataset was designed to induce emotional responses in users involved in naturalistic-like human-computer interaction (HCI) according to two HCI-experimental settings. In the former, pictures taken from the IAPS set [17] were used to induce emotions. Stimulus sequences consisted of 10 pictures with similar ratings according to the 5 possible affective states: high valence and high arousal (HVHA), high valence and low arousal (HVLA), low valence and low arousal (LVLA), low valence and high arousal (LVHA) and neutral. In the second part of the experiment, the emotions were induced during a naturalistic-like HCI in a standardized environment. In both the experiments several data were recorded: video, audio, trigger information and physiological data, namely respiration, fEMG from *corrugator supercilii* activity, fEMG *zygomaticus major* activity, Blood Volume Pulse and Skin Conductance.

In this paper we refer to the data, videos and fEMG signals, acquired in the first experiment, that is the recording of 30 subjects, each one stimulated by 5 image sequences.

B) Landmark extraction. Given a video sequence of a facial expression, we account for Eqs. 3,4 and 5 by applying the method described in [8] to infer the locations of facial landmarks (Fig. 1c). Such method extends in a sparse coding framework Zhu and Ramanan's technique [31], which jointly performs face and landmark detection. Once landmarks $\mathbf{L}$ have been detected, an adequate pool $\mathbf{l}_p$ of landmarks should be defined in order to provide related distance measures $\mathbf{d}(\mathbf{l}_p)$ as a "proxy" to muscle activity. Figures 2 and 3 below show the landmarks involved in measuring *corrugator supercilii* and *zygomaticus major* activities, respectively.
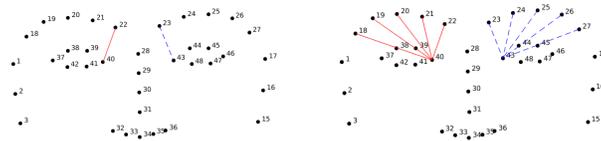

Fig. 2: Landmarks and distances accounting for the *corrugator supercilii* activity
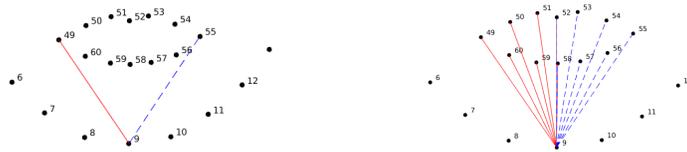

Fig. 3: Landmarks and distances accounting for the *zygomaticus major* activity

The fEMG signal captures very local muscle movements and its simulation should derive from a small subset of facial landmarks with superposition to the muscle of interest. The most natural choice would be to consider the landmarks closest to the muscle as shown in Fig. 2, (blue dashed line, left panel) for the *corrugator supercilii*, and in Fig. 3, (blue dashed line, left panel) for the *zygomaticus major*. However, landmark displacements are noisy, due to the detection method and possible occlusions caused by the sensors. We thus investigate several pools of displacements aiming at pinpointing the most suitable ones for fEMG regression.

In the case of the *corrugator supercilii*, we thus consider the symmetric distance between the inner eye corners and the inner eyebrow landmarks (Fig. 2, left panel), the two distances coupled, and more global measures obtained considering the distances between the inner eye corners and the corresponding eyebrow landmarks, both separately and all together (Fig. 2, right panel). Similarly for the *zygomaticus major*, we take into account the symmetric distance as in Fig. 3, (red line, left panel), the two punctual distances coupled, and the distances be-

tween the chin and the two halves extern lip contour landmarks, both singularly and coupled (Fig. 3, right panel).
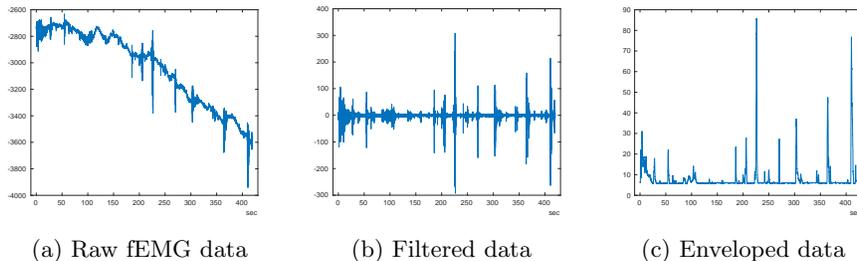

(a) Raw fEMG data   (b) Filtered data   (c) Enveloped data

Fig. 4: fEMG signal processing pipeline.

C) fEMG preprocessing. The raw data set of fEMG measurements derived from *corrugator supercilii* and *zygomaticus major* activities - which we denote $\mathbf{E}^c$ and $\mathbf{E}^z$, respectively - is a collection of 1-D signals captured at 512 Hz or more (Fig. 4a). The low frequencies are strongly influenced by artifacts such as motion potentials, eye movements, eye blinks, swallowing, and respiration, thus requiring a preliminary high-pass filtering to remove the strongest artifacts that would otherwise dominate the real facial EMG potentials. In the literature different cutoff frequencies are adopted for this purpose, ranging from 5 to 20 Hz [32,6,19], We use a 20 Hz cutoff frequency, guaranteeing artifact elimination. In addition, filtering has to be applied to remove the 50 Hz power line interference. To this aim, notch filtering is adopted (Fig. 4b). Further, when fEMG activation is addressed, the rectification and envelope are advised [5,20]. Eventually, to train the Gaussian process, the signals are down-sampled to 25 Hz so that the fEMG and the video frequencies are in correspondence (Fig. 4c).

D) $\mathcal{GP}$ model learning and fEMG prediction. Given a dataset of inputs and targets, $\{\mathbf{L}, \mathbf{E}\} = \{\mathbf{l}_n, \mathbf{e}_n \mid n = 1, \cdots, N\}$, we are interested in evaluating the mapping of S test sequences of landmarks $\mathbf{L}_{new} = \{\mathbf{l}_{new,s} \mid s = 1, \cdots, S\}$ into fEMG sequences $\mathbf{E}_{new} = \{\mathbf{e}_{new,s} \mid s = 1, \cdots, S\}$, where $\widetilde{\mathbf{e}} = \mathbf{e}_{new,s}$ is the desired virtual fEMG signal. Notice that here and in what follows, we thoroughly write $\mathbf{l}_{p,new}$ in place of actual measurements $\mathbf{d}(\mathbf{l}_{p,new})$ to simplify notation. Formally, we need to evaluate the predictive distribution $P(\mathbf{E}_{new} \mid \mathbf{L}, \mathbf{E}, \mathbf{L}_{new}) = \int P(\mathbf{E}_{new} \mid \mathbf{g}_{new}) P(\mathbf{g}_{new} \mid \mathbf{L}, \mathbf{E}, \mathbf{L}_{new}) d\mathbf{g}_{new}$, where $P(\mathbf{E}_{new} \mid \mathbf{g}_{new})$ is the likelihood given by Eq. 6. The posterior over functions $P(\mathbf{g}_{new} \mid \mathbf{L}, \mathbf{E}, \mathbf{L}_{new})$ is a Gaussian distribution $\mathcal{N}(\mu_{new}, k_{new})$, whose parameters can be written in closed form [24], namely, $\mu_{new} = k(\mathbf{L}_{new}, \mathbf{L}) \left[k(\mathbf{L}, \mathbf{L}) + \sigma_e^2 \mathbb{I} \right]^{-1}$ and $k_{new} = k(\mathbf{L}_{new}, \mathbf{L}_{new}) - k(\mathbf{L}_{new}, \mathbf{L}) \left[k(\mathbf{L}, \mathbf{L}) + \sigma_e^2 \mathbb{I} \right]^{-1} k(\mathbf{L}, \mathbf{L}_{new})$. Kernel functions and related hyperparameters are obtained from the training stage.

As to the latter, we train different models, varying the referred landmark pool, $p \in \{1, ..., 6\}$, associated with the related muscle, and exploring the GP behaviour by adopting the well-known Squared Exponential Kernel (k_{SE}), Rational Quadratic Kernel (k_{RQ}), and the Matern 3/2 kernel (k_{M32}) [24]. For each model, training and test sets are derived adopting the k-fold cross validation method, partitioning data into 10 subsets.

D) Results. The quality of the virtual fEMG, $\widetilde{\mathbf{e}}$, with respect to the original fEMG filtered signal, $\mathbf{e}$, is evaluated in terms of Mean Square Error (MSE), and by the Concordance Correlation Coefficient measures (CCC):

$$MSE(\mathbf{e}, \widetilde{\mathbf{e}}) = \frac{1}{T} \sum_{t=1}^{T} (e(t) - \tilde{e}(t))^2 \qquad CCC(\mathbf{e}, \widetilde{\mathbf{e}}) = \frac{2cov(\mathbf{e}, \widetilde{\mathbf{e}})}{\sigma_e^2 + \sigma_{\tilde{e}}^2 + (\mu_\mathbf{e} - \mu_{\widetilde{\mathbf{e}}})^2},$$

being $\mu_\mathbf{e}$ and $\mu_{\widetilde{\mathbf{e}}}$ the signal means, $\sigma_\mathbf{e}^2$ and $\sigma_{\widetilde{\mathbf{e}}}^2$ the variances, and $cov(\mathbf{e}, \widetilde{\mathbf{e}})$ the covariance.

In Table 1 we report the performances obtained in simulating the *corrugator supercilii* fEMG, adopting the different learnt models. Those concerning the virtual generation of the *zygomaticus major* fEMG are shown in Table 2.

Table 1: Performances achieved in the virtual generation of the *corrugator supercilii* fEMG, referring to different pool of landmarks ($p \in \{1...6\}$), and different kernels (k_{SE}, k_{RQ}, k_{M32}). Performances are expressed as MSE and CCC.

Pool	MSE			CCC		
	SE	RQ	M32	SE	RQ	M32
1	42.0075	31.8062	36.2262	0.9855	0.9888	0.9873
2	4.9623	4.9285	4.9343	0.9983	0.9983	0.9983
3	3.8280	3.7601	3.7185	0.9987	0.9987	0.9987
4	2.0082	1.7696	1.7273	0.9993	0.9994	0.9994
5	1.9681	1.7991	1.7070	0.9993	0.9994	0.9994
6	1.0478	0.7167	**0.6522**	0.9996	0.9997	**0.9998**

Analysing the behaviour of the models, we observe that the MSE and the CCC performances are always coherent. We can conclude that both in the simulations of the *corrugator supercilii* fEMG and of the *zygomaticus major*, best performances are achieved through the largest pool of landmark distances. This is likely to depend on the noise that characterizes landmarks localization, certainly attenuated by considering a pool of landmarks rather than punctual ones. In particular, we observe that the punctual distance $d = 1$ in the *corrugator supercilii* fEMG gives the worst performances, this because, in the considered dataset, the fEMG sensor often partial occludes the eyebrow. Also, it is worth noticing that system behaviour is robust to the use of different kernels.

Figures 5 and 6 illustrate typical fEMG signal reconstructions for both the *corrugator* and the *zygomaticus* muscles.

Table 2: Performances achieved in the virtual generation of the *zygomaticus major* fEMG. Results are organized as in Table 1

Pool	MSE			CCC		
	SE	RQ	M32	SE	RQ	M32
1	11.5408	11.5193	11.5001	0.9969	0.9969	0.9969
2	13.3317	13.2269	13.1922	0.9965	0.9965	0.9965
3	8.1977	7.8251	7.7871	0.9978	0.9979	0.9979
4	2.2541	1.5721	1.4217	0.9994	0.9996	0.9996
5	2.5401	1.2763	1.3191	0.9993	0.9997	0.9997
6	1.0401	0.6291	**0.5617**	0.9997	0.9998	**0.9999**

4   Discussion and conclusions

We have presented a method for detecting the electromyographic signal arising from muscles involved in affective, non-posed, facial expressions, which only relies on the facial landmarks detected in videos. Preliminary experiments on the OPEN EmoRec II multimodal corpus [25] have given evidence of promising results.

Clearly, one should be aware that there are limitations in the detection capability of the method. It is known that real fEMG can intercept even very weak affective expressions, even below the visible display of the expression itself [18]; however, this limit is shared by all virtual methods that attempt at simulating *in vivo* measurements from visual input.

Apart from the appealing issue of avoiding the obtrusiveness of fEMG measurement, what is to be gained by such attempt in view of the affective computing problem? All things considered, as detailed in Section 2, the landmarks we rely upon for regressing the fEMG signal are nothing but a subset of the facial landmarks we collect, the latter, in principle, providing full information - at least that available from the video sequence - to further proceed with facial expression analysis for affective computing purposes. Under the circumstances, it is worth making clear the rationale behind this study. Affective computing aims at dealing with machines that might have the ability to 1) recognize emotions, 2) express emotions, 3) "have emotions", the latter being the "hardest stuff" [22]. So far, most current research focuses on 1) and 2), with image processing and pattern recognition-based affect detection playing a prominent role[7]. The research work fostering this study pursues a different approach, centred on simulation-based affect analysis [28]. According to embodied simulation theories, understanding emotions of others is supported by running the same emotional apparatus - possibly in reverse - that is already used to generate or experience the emotion, eventually causing a "reactivation" of the corresponding mental state [13,11,12]. Indeed, an emotion is a neural reaction to a certain stimulus, realised by a complex ensemble of neural activations in the brain. The latter often are preparations for (muscular, visceral) actions (facial expressions, heart rate increase, etc.), as a consequence, the body will be modified into an "observ-
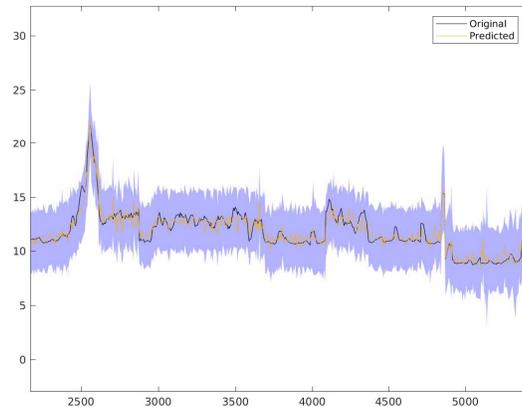
Fig. 5: Detail of fEMG reconstruction of the *corrugator supercilii* signal, using the Squared Exponential kernel and considering the 5-th landmark pool. The shaded area represents the pointwise mean plus and minus two times the standard deviation for each input value (corresponding to the 95% confidence region)

able" [10]. It is in such a broader perspective that it is particularly relevant to have available a variety of physiological signals, real or virtual, for building the latent continuous space of emotions [4]. fEMG, together with others that can be obtained by less obtrusive means (heart rate, skin conductance, respiratory rhythm, gaze scan path), is one such signal.

Acknowledgments

References

1. Adamo, A., G.Grossi, Lanzarotti, R.: Local features and sparse representation for face recognition with partial occlusions. IEEE (September 2013)
2. Adamo, A., Grossi, G., Lanzarotti, R., Lin, J.: Robust face recognition using sparse representation in lda space. Machine Vision and Applications 26(6), 837–847 (2015)
3. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: An algorithm for designing over-complete dictionaries for sparse representation 54(11), 4311–4322 (Nov 2006)
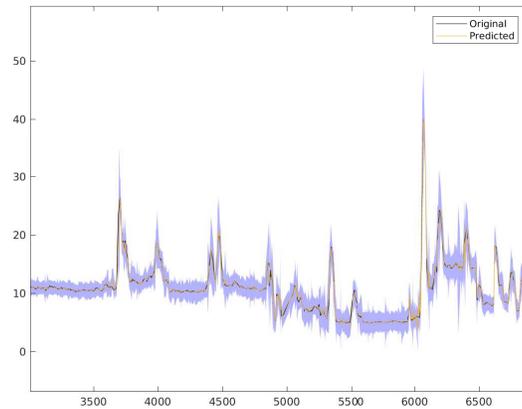4. Anderson, D.J., Adolphs, R.: A framework for studying emotions across species. Cell 157(1), 187–200 (2014)

Fig. 6: Detail of fEMG reconstruction of the *zygomaticus* signal, using the Matern 3/2 kernel and considering the 6-th landmark pool

5. Barzilay, O., Wolf, A.: A fast implementation for EMG signal linear envelope computation. J. Electromyogr. Kinesiol. 21(4), 678–682 (2011)
6. van Boxtel, A.: Optimal signal bandwidth for the recording of surface EMG of facial, jaw, oral, and neck muscles. Psychophysiology 38, 22–34 (2001)
7. Calvo, R., D'Mello, S.: Affect detection: An interdisciplinary review of models, methods, and their applications. IEEE Transactions on affective computing 1(1), 18–37 (2010)
8. Cuculo, V., Lanzarotti, R., Boccignone, G.: Using sparse coding for landmark localization in facial expressions. In: 5th European Workshop on Visual Information Processing (EUVIP). pp. 1–6 (Dec 2014)
9. Dalgleish, T., Dunn, B., Mobbs, D.: Affective neuroscience: Past, present, and future. Emotion Review 1(4), 355–368 (2009)
10. Damasio, A.R.: The feeling of what happens: Body and emotion in the making of consciousness. Houghton Mifflin Harcourt (1999)
11. Gallese, V.: The 'shared manifold' hypothesis. From mirror neurons to empathy. Journal of consciousness studies 8(5-7), 33–50 (2001)
12. Gallese, V.: The manifold nature of interpersonal relations: The quest for a common mechanism. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 358(1431), 517–528 (2003)
13. Goldman, A.I., Sripada, C.S.: Simulationist models of face-based emotion recognition. Cognition 94(3), 193–213 (2005)
14. Grossi, G., Lanzarotti, R., Lin, J.: Robust face recognition providing the identity and its reliability degree combining sparse representation and multiple features. International Journal of Pattern Recognition and Artificial Intelligence 30(10) (2016)
15. Grossi, G., Lanzarotti, R., Lin, J.: Orthogonal procrustes analysis for dictionary learning in sparse linear representation. PLoS ONE 12 (2017)
16. Hildebrandt, A., Recio, G., Sommer, W., Wilhelm, O., Ku, J.: Facial EMG Responses to Emotional Expressions Are Related to Emotion Perception Ability. PLoS ONE 9(1) (2014)

17. Lang, P.J., Bradley, M.M., Cuthbert, B.N.: International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Tech. Rep. A-8, The Center for Research in Psychophysiology, University of Florida, Gainesville, FL (2008)
18. Larsen, J., Norris, C., Cacioppo, J.: Effects of positive and negative affect on electromyography activity over zygomaticus major and corrugator supercilii. Psychophysiology 40, 776–785 (2003)
19. Lu, G., Brittain, J.s., Holland, P., Yianni, J., Green, A.L., Stein, J.F., Aziz, T.Z., Wang, S.: Removing ECG noise from surface EMG signals using adaptive filtering. Neurosci. Lett. 462, 14–19 (2009)
20. Myers, L., Lowery, M., O'Malley, M., Vaughan, C., Heneghan, C., Gibson, A.S.C., Harley, Y., Sreenivasan, R.: Rectification and non-linear pre-processing of EMG signals for cortico-muscular analysis. Journal of Neuroscience Methods 124(2), 157 – 165 (2003)
21. Olshausen, B.A., Field, D.J.: Natural image statistics and efficient coding. Network: computation in neural systems 7(2), 333–339 (1996)
22. Picard, R.W.: Affective computing. MIT press, Cambridge, MA (2000)
23. Poh, M.Z., McDuff, D.J., Picard, R.W.: Advancements in noncontact, multiparameter physiological measurements using a webcam. IEEE transactions on biomedical engineering 58(1), 7–11 (2011)
24. Rasmussen, C.E., Williams, C.K.: Gaussian processes for machine learning. The MIT Press, Cambridge, MA, USA (2006)
25. Rukavina, S., Gruss, S., Walter, S., Hoffmann, H., Traue, H.C.: OPEN EmoRec II – a multimodal corpus of human-computer interaction. International Journal of Computer, Electrical, Automation, Control and Information Engineering 9(5), 1181–1187 (2015)
26. Sun, Y., Thakor, N.: Photoplethysmography revisited: from contact to noncontact, from point to imaging. IEEE Transactions on Biomedical Engineering 63(3), 463–477 (2016)
27. Tassinary, L.G., Cacioppo, J.T., Vanman, E.J.: The skeletomotor system: Surface electromyography. In: Cacioppo, J.T., Tassinary, L.G., Berntson, G. (eds.) Handbook of Psychophysiology, chap. 12, pp. 267–300. Cambridge University Press (2012)
28. Vitale, J., Williams, M.A., Johnston, B., Boccignone, G.: Affective facial expression processing via simulation: A probabilistic model. Biologically Inspired Cognitive Architectures Journal 10, 30–41 (2014)
29. Wang, S., Ji, Q.: Video affective content analysis: a survey of state-of-the-art methods. IEEE Transactions on Affective Computing 6(4), 410–430 (2015)
30. Wu, H.Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., Freeman, W.: Eulerian video magnification for revealing subtle changes in the world. ACM Transactions on Graphics (TOG) 31(4), 65 (2012)
31. Zhu, X., Ramanan, D.: Face detection, pose estimation, and landmark localization in the wild. In: Proc. of IEEE CVPR. pp. 2879–2886 (2012)
32. Zschorlich, V.R.: Digital filtering of EMG-signals. Electromyogr. clin. Neurophysiol. 29(April), 81–86 (1989)