# Functional statistics for human emotion detection

**There is an increasing scientific interest in automatically analysing and understanding human behavior, with particular reference to the evolution of facial expressions and the recognition of the corresponding emotions. In this project we propose a technique based on Functional ANOVA to extract significant patterns of face muscles movements, in order to identify the emotions expressed by actors in recorded videos. This research is part of the BIGMATH project, a European Industrial Doctorate funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No 812912.**

Here we describe the main mathematical aspects of the work developed by Rongjiao Ji, one of the Early Stage Researchers (ESRs) enrolled in the BIGMATH PhD programme, who is jointly supervised by academic advisors, coming from the universities of Milan and Novi Sad, and by industrial advisors, coming from the company 3Lateral (Serbia, producing virtual reality for entertainement).

## Problem description

The study of human facial expressions and emotions never stops in our daily life while we communicate with others. Following the increased interest in automatic facial behavior analysis and understanding, the need of a semantic interpretation of the evolution of facial expressions and of human emotions has become of interest in recent years [1]. In this project, based on a work cooperated with the Serbian company 3Lateral, which has special expertise on building visual styles and designs in animation movies, we want to explore functional statistical instruments to identify the emotions while analyzing the expressions through recorded videos of human faces. The final aim of this research is to use this information to better and more realistically establish virtual digital characters, able to interact autonomously with real humans [2].

The data that we consider are multivariate longitudinal data, showing the evolution in time of different face muscles contraction. Functional Data Analysis (FDA) offers the possibility to analyze the entire expression evolution process over time and to gain detailed and in-depth insight into the analysis of emotion patterns. The basic idea in functional data analysis is that the measured data are noisy observations coming from a smooth function. Ramsay and Silverman [3] describe the main features of FDA, that can be used to perform exploratory, confirmatory or predictive data analysis.

In our application, Functional ANOVA can be used to determine if there are time-related differences between emotion groups by using a functional F-test [4].

> *“The final aim of this research is to use this information to better and more realistically establish virtual digital characters, able to interact autonomously with real humans”*

Our study is based on the RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song) dataset [5], which contains 24 professional actors (12 female, 12 male) to offer the performance with good quality and natural behavior under the emotions: calm, happy, sad, angry, fearful, disgusted and surprised. Also a neutral performance is available for each actor. The actors are vocalizing one lexically-matched statement in a neutral North American accent (“Kids are talking by the door”).

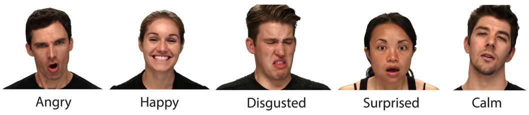


*Figure 1: Emotions represented by the actors in RAVDESS*

To avoid being lost in the difference of individual facial appearances, when analyzing the expressions and emotions, researchers mostly focus on the movements of individual facial muscles which are encoded by the Facial Action Coding System (FACS) [6]. FACS is a common standard to systematically categorize the physical expression of emotions, extracting the geometrical features of the faces and then producing temporal profiles of each facial movement. Such movements, corresponding to contraction of specific muscles of the face, are called *action units (AUs)*. As action units are independent of any interpretation, they can be used for any higher-order decision-making process including recognition of basic emotions. Following the FACS rules, OpenFace [7], an open-source software, is capable of recognizing and extracting facial action unit from facial images or videos. We applied OpenFace to extract the engagement degrees of action units for the videos in RAVDESS. The extracted action units include 17 functions for each video, taking values in $[0, 5]$, sampled in about 110 time points (which is also the number of frames in each video).

## Methods

We represent the action units evolution recorded on each video as a multivariate time series $\mathbf{Y}(t) = (Y_1(t), \dots, Y_d(t), \dots, Y_D(t)), t \in [0, T]$ containing a set of $D$ univariate longitudinal functions ($D = 17$ in our case), each defined on the finite interval $[0, T], 0 < T < +\infty$. The observation of $\mathbf{Y}$ on our sample of videos provides the set $\mathbf{Y_1}, \dots, \mathbf{Y_n}$ of multivariate curves, that we represent as multivariate functional data.

First we aligned the action units functions into a common registered internal timeline that follows the same pronunciation speed, to control the influence of the specific pronounced sentence and to detangle it from the influence of the emotions.

we then investigated if there exist patterns which could discriminate the different emotions, using a Functional ANOVA model.

Let $y_{k,g}(t)$ be the evolution of one specific action unit in the video $k \in \{1, \dots, K\}$ (in our case $K = 48$) for emotion $g \in \{1, \dots, 7\}$. We can assume that

$$y_{k,g}(t) = \mu_0(t) + \alpha_g(t) + \epsilon_{k,g}(t), \qquad (1)$$

where $\mu_0(t)$ is the grand mean function due to the pronounced sentence and to the actor, independent from all emotions. The term $\alpha_g(t)$ is the specific effect on the considered action unit of emotion $g$, while $\epsilon_{k,g}(t)$ represents the unexplained zero mean variation, specific of the $k$-th video within emotion group $g$. To be able to identify them uniquely, we require that they satisfy the constraint $\sum_{g=1}^{7} \alpha_g(t) = 0, \forall t.$

By grouping the videos representing the same emotion, we can define a $8K \times 8$ design matrix $\mathbf{Z}$ for this model, with suitable 0 and 1 entries, as described in [3, Section 9.2], and rewrite Equation 1 in matrix form: $\mathbf{y} = \mathbf{Z}\beta + \epsilon$, where $\beta = [\mu_0(t), \alpha_1(t), \dots, \alpha_7(t)]^T$.

To estimate the parameters we used a functional least squares fitting criterion.

In order to investigate which emotions are significantly influencing the change of the action units patterns, for each emotion $\tilde{g}$ and for each action unit we tested the null hypothesis $H_0 : \alpha_{\tilde{g}}(t) = 0.$

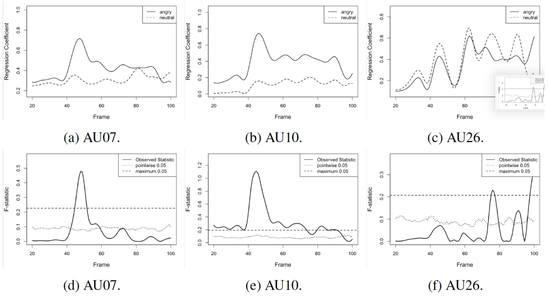
Similarly to the classical univariate ANOVA model, the statistics used to test $H_0$ is

$$FRATIO(t) = \frac{MSR(t)}{MSE(t)}$$

whose distribution under $H_0$ is estimated through a permutation test.

## Preliminary results

We applied the F-test described in the previous section to detect, for each emotion, which AUs have a mean behaviour significantly different from the neutral performance and in which time period during the videos. In Figure 2 we illustrate the results for emotion angry, as an example.



(a) AU07. (b) AU10. (c) AU26.

(d) AU07. (e) AU10. (f) AU26.

*Figure 2: The functional coefficients of action units 07 (Lid Tightener), 10 (Upper Lip Raiser) and 26 (Jaw Drop) under neutral and angry emotion and the corresponding F-test results*

The first row of Figure 1 illustrates the estimated mean $\mu_0(t)$ (neutral emotion) and the angry emotion effects for three action units. The second row displays the observed F-statistics curves together with the pointwise and maximum 95% quantile for the F-distribution in the dashed and horizontal dotted lines respectively.

Thus when the observed F-statistics is higher than the critical level lines, the emotion has a significant effect on the AU's pattern. We found in general three main situations of influence of one emotion on expression evolution: 1. locally strengthening 2. locally inhibiting 3. globally strengthening. Further, we pointed out the time zones of significant effects of the angry emotion on the action units in Figure 3, which is beneficial to understand and detect dynamically when and how the facial muscles contractions differ from the baseline.

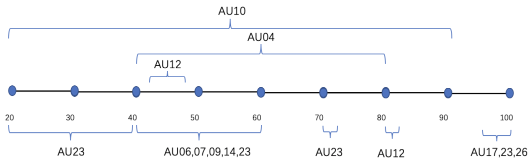


*Figure 3: Which and where AU values are affected significantly by angry emotion.*

Table 1 summarizes for each emotion of interest the related action units that show significant changes from the neutral case for our videos dataset. Similarly to the example of angry, we found that for happy and disgust emotions more action units have the globally strengthening effect on a large time range. Sad emotion sometimes affects the action units to be more constant than in neutral case. Emotion Fearful has more influence on upper half face (brows, eye lids and nose), while emotion calm is more related with the center of the face (Cheek Raiser, Lid Tightener and Lip Corner Puller). Surprised emotion is the only emotion where AU45 is significantly influenced.

| Emotions | Related Action Units |
|---|---|
| Calm | 06, 07, 10, 12, 14, 23 |
| Happy | 01, 06, 07, 10, 12, 14, 17, 23, 25, 26 |
| Sad | 04, 06, 10, 14, 17, 20, 23, 25 |
| Angry | 04, 06, 07, 09, 10, 12, 14, 17, 23, 26 |
| Fearful | 04, 09, 10, 12, 14, 15, 17, 23, 25, 26 |
| Disgust | 04, 06, 07, 09, 10, 12, 14, 17, 23, 25, 26 |
| Surprised | 06, 09, 10, 12, 14, 15, 17, 23, 25, 26, 45 |

*Table 1: Emotions with corresponding significant action units*

As a conclusion, our results can be joined in a multivariate setting and exploited to build a classifier able to automatically recognize the emotions.

> *“We found in general three main situations of influence of one emotion on expression evolution: 1. locally strengthening 2. locally inhibiting 3. globally strengthening”*

Rongjiao Ji[1], Alessandra Micheletti[1], Natasa Krklec Jerinkic[2], Zoranka Desnica[3]

[1] Università degli Studi di Milano, Italy
[2] University of Novi Sad, Serbia
[3] 3Lateral DOO, Serbia

## References

1. A.J. Fridlund. Human facial expression: An evolutionary view. *Academic Press*, 2014.
2. Rongjiao Ji, Alessandra Micheletti, Natasa Krklec Jerinkic, and Zoranka Desnica. Emotion pattern detection on facial videos using functional statistics. Technical report, arXiv - Preprint 3627227, 2021.
3. J. Ramsay and B.W. Silverman. *Functional data analysis*. Springer, 1997.
4. J. Dannenmaier, C. Kaltenbach, T. Kölle, and G. Krischak. Application of functional data analysis to explore movements: walking, running and jumping-a systematic review. *Gait & postureh*, pages 182–189, 2020.
5. S.R. Livingstone and F.A. Russo. The ryerson audio-visual database of emotional speech and song (ravdess): A dynamic, multimodal set of facial and vocal expressions in north american english. *PloS one*, 13(5):e0196391, 2018. https://smartlaboratory.org/ravdess/.
6. R. Ekman. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA, 1997.
7. B. Amos, L. Bartosz, and M. Satyanarayanan. Openface: A general-purpose face recognition library with mobile applications. Technical report, CMU-CS-16-118, CMU School of Computer Science, 2016. https://cmusatyalab.github.io/openface/.
8. Openface. *https://cmusatyalab.github.io/openface/*, 2016.