

Connectionist and Ethological Approaches for Discovering Salient Facial Movements Features in Human Gender Recognition

Sébastien Guérif, Younès Bennani

University of Paris 13, CNRS UMR 7030 - LIPN, F-93430 Villetaneuse
{sebastien.guerif, younes.bennani}@lipn.univ-paris13.fr

Claude Baudoin

University of Paris 13, CNRS UMR 7153 - LEEC, F-93430 Villetaneuse
claude.baudoin@leec.univ-paris13.fr

Abstract. *Individual Facial movements signal various social information to other persons, like the gender of the sender. We used an ethological and a connectionist approaches in order to detect these movements and their characteristics in men and in women. Behavioural results indicate both qualitative and quantitative differences between men and women. The connectionist approach involves similar and complementary conclusions. The ethological study has been focused on the main movement differences as well as did the connectionist one but this last approach showed important differences between men and women in motionless events. These pilot results leads to a re-examination of behavioural events and a checking of lateralization of movements correlated with the gender.*

Keywords. *Facial movements, gender recognition, unsupervised learning, clustering, self-organizing maps*

1. Introduction

Social life in human groups involves constant regulatory processes like social categorization of interacting partners. One type of social category among the most obvious is the gender. Various body parts are used in signaling gender and the face is an important area as it has been demonstrated in previous studies [2, 3, 4]. Several authors showed the role of facial movements in gender categorization [9, 13]. Curiously only few studies concern the production of facial movements and their temporal organization [1, 8, 14]. Moreover the use of complex experimental systems for facial recording induces unnatural situations. Our study concerns an experimental sit-

uation with young adult subjects confronted to a cognitive task without direct interacting partners but with a female experimenter welcoming them before testing and video recording their behaviour from a contiguous room. This situation was not social but the context was social. Our aims were (i) to constitute a database allowing further comparisons between men and women, (ii) to code facial movements using an objective method, (iii) to detect and to characterize the temporal organization of the facial movements, (iv) to use the same data base for studying salient facial features with a connectionist approach, (v) to compare emerging results from ethological and connectionist approaches.

The rest of this paper is organized as follows. After, a brief presentation of the protocol used to collect the data, we present both ethological and connectionist approaches. Then, experiments and results are presented and discussed. Finally, we conclude and we give some point that will we developed in further research.

2. Collection of data base

Our purpose was to obtain a video recording from women and men in a standardized situation that permits expression of various facial movements: labial movement related to verbal answer, emotional reaction, etc. The experimental situation was a cognitive task realized in an indirect social context (reception followed by task instructions, filmed by a video camera operated by a female experimenter in the next room).

A total of 20 students (11 women and 9 men) from the University of Paris XIII volunteered to participate in the study. All were naive to the true

purpose of the study and were not paid for their participation.

Subjects were received by an experimenter and then left alone in a room where they followed instructions given by a laptop screen. The task consisted in looking at pictures and saying whether it was ambiguous/normal or not. No time limit had been imposed and the experiments lasted between 1.5 and 4.5 minutes (mean duration was 2.75 minutes). Some subjects were set aside because of particular situations (important movements of the body or the head, wearing glasses or a beard, etc.) Only 5 subjects of each gender were selected for the remaining part of the study.

With a view to standardizing the database we chose 3 sequences of 3 seconds centred on an easy to locate verbal answer from the subjects. Thus 3 sequences with a similar context are available per subject. We defined 36 face points involved in the facial movements that were easy to identify [8,14]. Figure 1 indicates the positions of the face points considered.

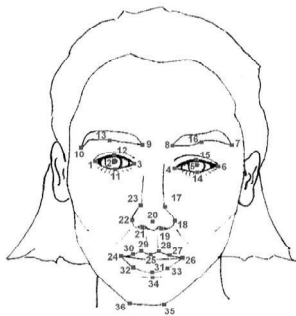


Figure 1: Position of the face points

The sequences were sampled at 13 images per second, and an operator recorded the 36 face points coordinates. This selection was repeated at least twice and the mean position was retained to reduce errors due to tiredness of the operator. The coordinates of the points were relative to the subjects face. Actually, the x axis is the line between points 3 and 4, and the y axis the orthogonal line crossing through point 20.

3. Ethological approach

For a given facial point (fig. 1), a salient movement was defined as the distance from origin which was higher than the mean distance calculated during a 3 seconds period (39 images) majored by the standard error. The number of move-

ments was studied comparatively between men and women. Then we detected how distance variations from origin of the 36 facial points occurred during the 3 seconds periods for men and for women using the Magnussons THEME 5.0 software (<http://www.noldus.com>). This software allowed to detect T-patterns of facial movements, which are defined as repetitive real time organized behavioural structures [11]. Only some results are presented below.

We observed a higher mean number of movements produced per 3 seconds period by men comparatively to women ($n = 86$ vs. 69 , $p < .05$, exact permutation test) as well as a tendency to present a higher number of T-patterns in men (on a basis of 100 movements, men produced 66 T-patterns vs. 46 in women, $p = .055$, exact permutation test). T-patterns involved on average 4 different facial points in men and 3 in women ($p = .079$). Moreover we discovered qualitative differences in the T-pattern composition linked to gender: men produced simple patterns involving temporal left eyebrow and left nostril, and women produced simple patterns involving internal and median parts of the right eyebrow.

Our pilot results indicated that man and woman facial movements were quantitatively and, at least for some of them, qualitatively different during a cognitive task performed in a social context.

4. Connectionist approach

Two different approaches are available to exploit our dataset: classification and clustering. The first one falls into supervised learning and builds a classifier. The latter approach detects groups of similar observations, called clusters. Our purpose is to determine whether the intrinsic structure of the data space is related to the gender of subjects or not. So, our interest has been focused on unsupervised learning approaches and Self-Organizing Maps (SOM) [10] were chosen to carry out our analysis. On one hand, SOM provides a convenient way to visualize the structure of our data [15]. On the other hand, the different clusters can be labeled according the gender of grouped observations and then be used as a classifier whose evaluation may give us some interesting information. First SOM are briefly introduced, then our methodology is explained and finally, the experimental results obtained using the Matlab somtool-

box [16] are given and discussed.

4.1. Connectionist model : Self-Organizing Maps

SOM was introduced by Pr. Teuvo Kohonen in the early 80's as a convenient clustering and visualization tool. High-dimensional data are projected on a low dimension discrete space, called the topological map, preserving the local topology of the initial space; thus, the observations which are close to each other are projected on a localized area. A map should be viewed as a set of neurons (or units), organized according to a grid that defines their neighbourhood relationships. Each neuron is associated to one point of the observations' space: its profile.

Self-Organizing Maps (SOM) implement a particular form of competitive artificial neural networks; when an observation is recognized, activation of an output cell competition layer leads to inhibit activation of other neurons and reinforce itself. It is said that it follows the so called Winner Takes All rule. Actually, neurons are specialized in the recognition of one kind of observations. The learning is unsupervised because neither the classes nor their number is fixed a priori.

A SOM consists in a two dimensional layer of neurons which are connected to n inputs according to n exciting connections of respective weights w and to their neighbors with inhibiting links.

The training set is used to organize these maps under topological constraints of the input space. Thus, a mapping between the input space and the network space is constructed; closed observations in the input space would activate two closed units of the SOM.

An optimal spatial organization is determined by the SOM from the received information, and when the dimension of the input space is lower than three, both position of weights vectors and direct neighbourhood relations between cells can be represented visually.

4.2. Learning algorithm

Connectionist learning is often presented as a minimization of a risk function (cost function). In our case, it will be carried out by the minimization of the distance between the input samples and the map prototypes (referents), weighted by a neighbourhood function h_{ij} . To do that, we use a gradient algorithm for optimization. The criterion to be

minimized is defined by:

$$R_{SOM} = \frac{1}{N} \sum_{k=1}^N \sum_{j=1}^M h_{jNN(x^{(k)})} \left\| \omega_j - x^{(k)} \right\|^2 \quad (1)$$

N represents the number of learning samples, M the number of neurons in the map, $NN(x^{(k)})$ is the neuron having the closest referent to the input form $x^{(k)}$, and h the neighbourhood function.

The weights of all the neurons are updated until stabilization according to the following adaptation rules: If $\omega_j \in V_{NN(x^{(k)})}$ then adjust the weights using:

$$\omega_j(t+1) = \omega_j(t) - \varepsilon(t) h_{jNN(x^{(k)})} \left(\omega_j - x^{(k)} \right) \quad (2)$$

4.3. Labelling the map

Training of the self-organizing map is totally unsupervised; and actually, it does not make use of the data labels (namely female or male). Therefore, at the end of the training phase we only had a topological map based on the transformed coordinate data without any additional information. Nevertheless, it should be emphasized that the map defined a partition of the dataset which can be used to assign each neuron a label. Actually, each neuron is labelled using the most represented gender associated with that neuron. As such, the labelling is very sensitive to small changes in gender distribution. Therefore, to increase robustness of the labelling, a chi-square test was used to check whether the distribution of that part is significantly different from that of the whole dataset. Therefore, it should be emphasized that some neurons remained unlabeled.

4.4. SOM segmentation

We segment the SOM using the K-means algorithm. It is another clustering method. It consists in choosing arbitrarily a partition. Then, the samples are treated one by one. If one of them becomes closer to the center of another class, it is moved into this new class. We calculate the centers of new classes and we reallocate the samples to the partitions. We repeat this procedure until having a stable partition.

The criterion to be minimized in this case, is defined by:

$$R_{K-means} = \frac{1}{C} \sum_{k=1}^C \sum_{x \in Q_k} \|x - c_k\|^2 \quad (3)$$

where C represents the number of clusters, Q_k is the cluster k , c_k is the center of the cluster Q_k or the referent.

The basic algorithm requires fixing K , the number of clusters wished. However, there is an algorithm to calculate the best value for K assuring an optimal clustering. It is based principally on the minimization of Davies-Bouldin index, defined as follows :

$$I_{DB} = \frac{1}{C} \sum_{k=1}^C \max_{k \neq l} \left\{ \frac{S_c(Q_k) + S_c(Q_l)}{d_{ce}(Q_k, Q_l)} \right\} \quad (4)$$

where $S_c(Q_k) = \frac{\sum_i \|x_i - c_k\|}{|Q_k|}$ is the intracluster dispersion of cluster Q_k and $d_{ce}(Q_k, Q_l) = \|c_k - c_l\|$ is the distance (centroid linkage) between the center of clusters k and l . This clustering procedure aims to find internally compact spherical clusters which are widely separated.

There are several methods to segment the SOMs [17]. Usually, they are based on the visual observations and the manual assignment of the map cells to the clusters. Several methods use the K-means algorithm with given ranges for K value. Our work is based on the approach of Davies-Bouldin index minimization [5].

4.5. Statistical measure for cluster characterization

In the sequel, the word cluster refers to a group of neurons that share the same label and which define a contiguous area on the map. The test-value, proposed in [12] was used to identify dimensions that were relevant for each cluster. Intuitively, it indicates how different a cluster is from the whole population according to the feature considered. Thus, the more different is the feature from the whole population the more relevant it is to describe that cluster. It is defined by

$$t_k = \frac{(\mu - \mu_k)}{\sigma_k} \quad (5)$$

where, μ is the mean of the whole dataset and, μ_k and σ_k are respectively the mean and standard deviation of the class k . Therefore, to interpret subsequences seemed to us more natural than to interpret the dynamic covariance matrices. So rather than directly use subsets of the covariance matrices, we used subsets of the corresponding subsequences. Thus we are able to quantify the relative importance of each point, at each step in time, for the different clusters.

5. Data pre-processing

Analysis was focused on facial motion; therefore, the gradients of the coordinate points were computed. Then, to eliminate the structural cue to individuals with a larger face who have a longer shift, the gradients were normalized. Thereafter, sequences of movements were resampled using a sliding window to improve robustness to the time lag of the selected video recording. Nevertheless, it introduced an additional parameter that had to be chosen carefully, namely the width of the temporal window. The observations then had too many dimensions to be used. So, the dynamic covariance matrix of each sub-sequence was computed according the following expression [18, 19]:

$$\Sigma_d = \frac{x_{(1)}x_{(1)}^T + \sum_{i=2}^W (x_{(i)} - \bar{x}_{(i)}) (x_{(i)} - \bar{x}_{(i)})^T}{W} \quad (6)$$

with $\bar{x}_{(i)} = \frac{1}{i} \sum_{j=1}^i x_{(j)}$. Thus, the dimension of the data only depends only on the number of face points considered.

6. Experiments and results

Our objective was to verify whether facial movements are related to the subject gender or not. Thus, it appeared relevant to select the parameter value that involved the best separation between the two classes. A cross-validation was adopted to evaluate values from 1 to 38 and each evaluation was repeated 5 times. The SOM that were trained with the dynamic covariance matrices from nine of the ten subjects was labelled. Then, the labelling of the map was evaluated by comparing the label from the remaining data with their best matching unit label.

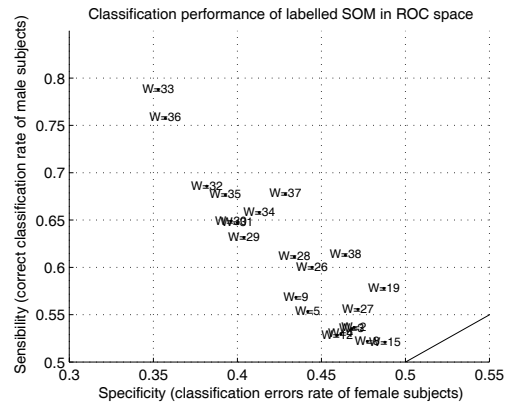


Figure 2: SOM based classifiers performance

| Point | Important Move | Motionless |
|-------|----------------|------------|
| 1 | 0.06 | 0.10 |
| 3 | 0.07 | 0.08 |
| 6 | 0.07 | 0.09 |
| 13 | 0.09 | 0.09 |
| 21 | 0.08 | 0.10 |

Table 1: Significant test values for Female

Receiver Operating Characteristics (ROC) graphs are a useful technique for visualizing, organizing and selecting classifiers based on their performance [7]. Thus, performances of SOM based classifiers are given in the ROC space. For convenience, only ones with more than 50% correct classification rate of both gender have been represented on figure 2. The nearest point $W=33$ from the upper left corner corresponds to a 33 time units sliding window. So this value has been retained for the remaining exploratory analysis of our data.

Figure 3 shows the distribution of subsequences gender over the final map and the segmentation obtained using the Davies-Bouldin index. On the left hand side, dark and light grey represents respectively female and male neurons, while black colour stands for unlabelled neurons. On the right hand side, 1 and 2 respectively stands for male and female, and the number between parenthesis indicates the number of hits.

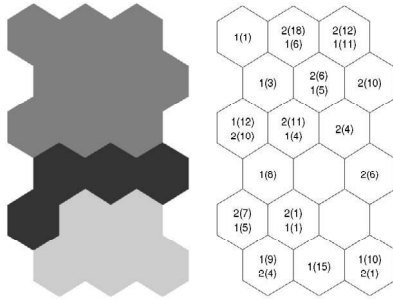


Figure 3: Final map

The tables 1 and 2, and the figures 4 and 5 show the significant test values for each points considered at each time step. Columns on the right, indicate the significance of the corresponding points for the whole subsequences. A visual inspection of the test values indicates that male produce more structured movement than female.

A deeper analysis of the test values emphasized that female cluster (respectively male cluster) is characterized by more structured movements of points 1, 3, 6, and 13 (respectively 8, 17 and 33). It should be highlighted that points 1, 3 and 13 are from the right part of the face whereas points 8, 17

| Point | Important Move | Motionless |
|-------|----------------|------------|
| 8 | 0.14 | 0.23 |
| 17 | 0.14 | 0.19 |
| 21 | 0.14 | 0.21 |
| 33 | 0.14 | 0.28 |

Table 2: Significant test values for Male

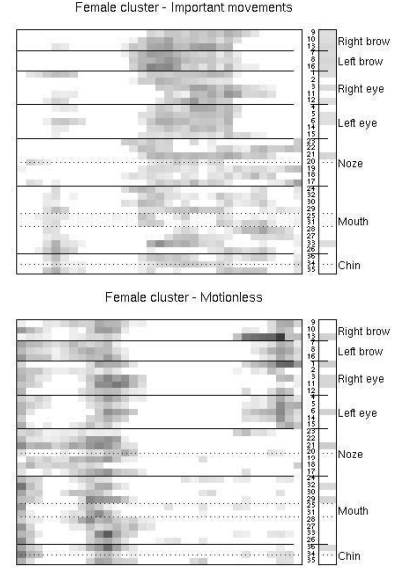


Figure 4: Test values for Female

and 33 are on the left part.

7. General discussion

The two approaches presented above had involved similar conclusions. On one hand, male facial movements appear more structured than female ones. On the other hand, points the more implicated in movements seems to differ from one gender to an other. Our results led us to hypothesise that the lateralization of facial movements should be an important feature to discriminate ones gender. Anyway, experiments should be repeated with a larger sample of population and with subjects from more different culture to confirm our hypothesis.

8. Conclusion and further research

In this paper, we have presented results from a pilot study with both an ethological and a connectionist approaches which had involved similar and complementary conclusions. Moreover, we chose a quite simple connectionist model for this first study, nevertheless, more elaborated connectionist model have been developed to integrate the temporal dimension of our data [6, 18, 19] and should

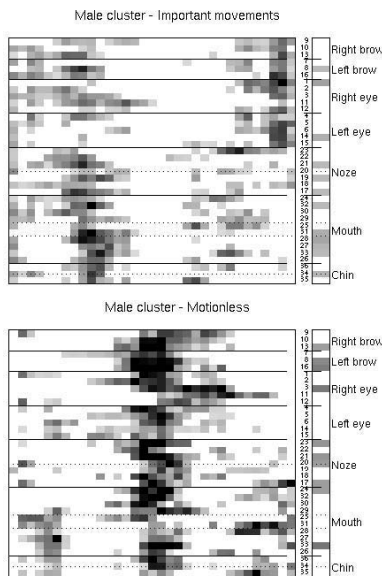


Figure 5: Test values for Male

be considered in future work.

9. References

- [1] Anolli L, Duncan S, Magnusson M, Riva G, editors. *The hidden structure of interaction: from neurons to culture patterns*. IOS Press; 2005.
- [2] Berry D. S. Child and adult sensitivity to gender information in patterns of facial motion. *Ecological Psychology* 1991; 3(4):349-366.
- [3] Bruce V, Burton AM, Hanna E, and al. Sex discrimination : how do we tell the difference between male and female faces ? *Perception* 1993, 22(2):131-152.
- [4] Bruce V, Young A. *In the Eye of the Beholder: The Science of Face Perception*. Oxford: Oxford University Press; 1998.
- [5] Davies DL, Bouldin DW. A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI 1979, 1(2):224-227.
- [6] Euliano N. *Temporal Self-Organization for Neural Networks*. PhD Thesis, University of Florida, USA; 1998.
- [7] Fawcett T. ROC Graphs: Notes and Practical Considerations for Data Mining Researchers. HP Labs Tech Report HPL-2003-4; 2003.
- [8] Giovanoli P, Tzou C-H J, Ploner M, Frey M. Three-dimensional video analysis of facial movements in healthy volunteers. *British Journal of Plastic Surgery* 2003, 56(7): 644-652.
- [9] Hill H, Johnston A. Categorizing sex and identity from the biological motion of faces. *Current Biology* 2001, 11(11):880-885.
- [10] Kohonen T. *Self-Organizing Maps*, Third Extended Edition. Berlin, Heidelberg, New York: Springer; 2001.
- [11] Magnusson MS. Discovering hidden time patterns in behaviour: T-patterns and their detection. *Behavior research methods, instruments and computers : a journal of the Psychonomic Society, Inc.* 2000, 32(1):93-110.
- [12] Morineau A. Note sur la caractérisation statistique d'une classe et les valeurs-tests. *Bulletin technique n 2*, p.20-27. Centre international de statistique et d'informatique appliquées, Saint-Mandé, France; 1984.
- [13] Thornton IM, Kourtzi Z. A matching advantage for dynamic human faces. *Perception* 2002, 31(1):113-132.
- [14] Tzou C-H J, Giovanoli P, Ploner M, Frey M. Are there ethnic differences of facial movements between Europeans and Asians? *British Journal of Plastic Surgery* 2005, 58(2):186-195.
- [15] Vesanto J. *SOM-Based Data Visualization Methods*. Intelligent Data Analysis 1999, 3(2):111-126.
- [16] Vesanto J, Himberg J, Alhoniemi E, Parhankangas J. *Self-Organizing Map in Matlab: the SOM Toolbox*. In: *Proceedings of the Matlab DSL Conference*; Espoo, Finland; 1999. p. 35-40.
- [17] Vesanto J, Alhoniemi E. Clustering of the Self-Organizing Map. *IEEE Transactions on Neural Networks* 2000, 11(3):586-600.
- [18] Zehraoui F, Bennani Y. M-SOM: Matricial Self Organizing Map for sequences clustering and classification. In: *Proceeding of the International Joint Conference on Neural Network, IJCNN'04*; Budapest, Hungary. 2004.
- [19] Zehraoui F, Bennani Y. M-SOM-ART: Growing Self Organizing Map for Sequences Clustering and Classification. In: *Proceedings of the 16th European Conference on Artificial Intelligence, ECAI'2004*; 2004 Aug 22-27; Valencia, Spain; 2004. p. 564-570.