What should be taught first: the emotional expression or the face?

S. Boucenna, P. Gaussier, P. Andry

What should be taught first: the emotional expression or the face?

S. Boucenna[1], **P. Gaussier**[1,2], **P. Andry**[1]
[1]ETIS, CNRS UMR 8051, ENSEA, Univ Cergy-Pontoise, [2]IUF,
{boucenna,gaussier,andry}@ensea.fr

We are interested in knowing how a robot head can learn to recognize facial expressions without supervision. Our starting point is a mathematical model showing that a sensory-motor architecture is able to express its emotions succeedes to recognize on-line the facial expression of a caregiver if this latter naturally tends to imitate or to resonate with the system. Interestingly, our works also show that, learning autonomously to recognize a face/non face is more complex than to recognize a facial expression. We propose an architecture using the interaction rhythm to allow first a robust learning of the facial expression without a face tracking and next to perform the learning of the face/non face recognition. Finally we emphasize the importance of the emotions as a mechanism to ensure the dynamical coupling between individuals allowing to learn more and more complex tasks.

1. Introduction

The main goal of this research is to understand the role of an emotional system in the development of autonomous agents. By "emotional system", we mean the system allowing to manage **and** to express theses emotions. Emotional changes, lead to particular facial expressions (angry, sadness, etc...) potentially inducing caregivers to interact and help the agent to solve the task. In this paper, we are studying a Neural Network (NN) model allowing an expressive robot head to engage "natural" emotional interactions in order to learn on line to recognize facial expressions of human partners.

The recognition of facial expressions is a well known issue, and classical solutions have shown impressive results. Some methods are based on the Principal Component Analysis (PCA). For example, the LLE (Locally Linear Embedding) in (Liang et al., 2005) performs a dimension reduction of the inputs vectors. Neuronal methods have also been developed for facial expression recognition. Franco and Treves (Franco and Treves, 2001), use a multi layer network with a classical supervised learning rule. The designer has to determine the number of neurons associated to the different expressions according to their complexity. Other methods are based on face models, which try to match the face (see for instance the appearance model in (B. Abboud, 2004)). (Yu and Bhanu, 2006) use a support vector machine (SVM) to categorize the facial expressions. (Wiskott, 1991) uses Gabor wavelets to code the face features as 'jets'(labeled graph where the nodes are 'jets'). While all these technics show a high level of recognition performances, it is important to notice that they are not easy to adapt in the frame of autonomous systems. Most of the time, the proposed solutions use ad hoc engineering strategies that need a strong control of the experimental conditions (strong supervision, accurate face detection in order to process the expression recognition). They use off line learning algorithms with the need to access to the whole learning database. Moreover, the related statistical models do not take into account the interaction dynamics between the human and the robot, and can not be accepted for realistic models of emotional expression recognition development. In the case of complex and unconstrained interactions, our starting point is the fact that babies learn to recognize facial expressions without explicit teaching signal or strong supervision (G. Gergely, 1999). Using the cognitive system algebra (Gaussier, 2001), we showed that a simple sensory-motor architecture based on a classical conditioning paradigm (Schmajuk, 1991, Balkenius and Moren, 2000) can learn on line to recognize facial expressions under the following condition (P. Gaussier, 2004).

In this paper, we summarize first the conditions derived from our formal model for the facial expressions on line learning. Next, we present the implementation of this theoretical model on a robotic head (fig. 1), with the constraints of on line learning, that will allow us to underline three original findings : first, the classical procedure that suppose to first localize the face and then to recognize its expression can be avoided : human face is recognized as such because his/her local views were associated to emotion recognition and not the opposite. Second, the dynamics of the human-robot interaction, brings important and non explicit

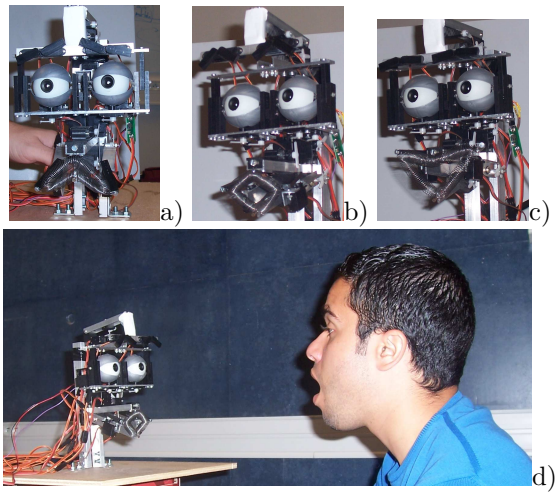


Figure 1: The robotic expressive head designed for developmental psychology and computational modeling studies (Nadel et al., 2006a). Examples expressions: a) sadness, b) surprise, c) happiness. d) Setup of a typical human / robot interaction game (here the human imitating the robot).

signals, such as the interaction rhythm, that helps the system to recognize face/non face.

2. On line learning of facial expression recognition: an interactive model

We consider a single system composed of two agents interacting in a neutral environment (see Fig.2). One

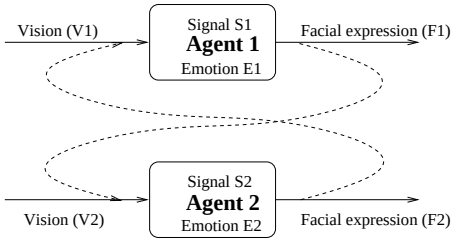


Figure 2: The bidirectional dynamical system studied. Both agents face each other. Agent 1 will be considered as a newborn and agent 2 as an adult mimicking the newborn facial expressions. Both agents are driven by internal signals which can induce the feeling of particular emotions.

agent is supposed to be an adult with perfect emotion recognition capabilities and reproduction capabilities. The second agent is considered as a new born without any previous learning on the social role of emotions. Formally, the 'baby' agent can be described as follow (Fig.3). We suppose our agents receive a visual signal (V_i vision of agent i). It can be learned and recognized in R_i group, R_i being the result of, for example an

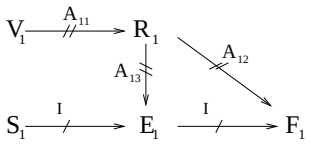


Figure 3: Schematic representation of an agent that can display and recognize "emotions". Arrows with one stroke represent "one to one" reflex connections. Arrows with labels and 2 parallel strokes represent "one to all" modifiable connections.

unsupervised pattern matching such as a WTA, an ART network, a Kohonen map. Hence, the vision of a face displaying a particular expression should trigger the activation of a corresponding node in R_i:

$$R_i = c(A_{i1}.V_i) \tag{1}$$

Where c is a competitive mechanism. A_{i1} represents the weights of the neurons in the recognition group of the agent i allowing a direct pattern matching. Our agents are also affected by the perception of their internal milieu (hunger, fear etc.). S_i the internal signals linked to physiological inputs such as fear, hunger... The recognition of a particular internal state will be called an emotional state E_i. We suppose also E_i depends on the visual recognition R_i of the visual signal V_i. At last, the agents can express a motor command F_i corresponding to a facial expression. If one agent can act as an adult, it must have the ability to "feel" the emotion recognized on someone else's face (empathy). At least, one connection between the visual recognition and the neuron group representing its emotional state must exist. In order to display an emotional state, we must also suppose there is a connection from the internal signals to the facial expression control. The connection can be direct or through another group devoted to the representation of emotions. For sake of homogeneity, we will suppose that the internal signal activates through an unconditional link the emotion recognition group which activates through an unconditional connection the display of a facial expression (hence it is equivalent to a direct activation of F_i by S_i - see (Gaussier, 2001) for a formal analysis of this kind of properties). Hence, the sum of both flows of informations is:

$$E_i = c\,(I.S_i + A_{13}.R_i) \tag{2}$$

At last, we can also suppose the teacher agent can display a facial expression without "feeling" it (just by a mimicking behavior obtain form the recognition of the other facial expression). The motor output of the teacher facial expression then depends on both facial expression recognition and the will to express a particular emotion:

$$F_i = c\,(I.E_i + A_{12}.R_i) \tag{3}$$

In a previous paper, we have studied the minimal conditions allowing the building of a global behavioral attractor (learning to imitate and to understand facial expression). Fig.4 represents the complete system with both agents in interaction. After simplifications

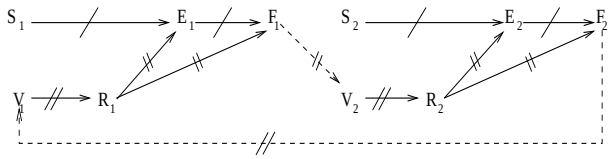


Figure 4: Schematic representation of the global network representing the interactions between 2 identical emotional agents. The dashed links represent the connections from the display of a facial expression to the other agent vision system (effect of the environment).

(P. Gaussier, 2004), we finally obtained the network shown Fig.5a.

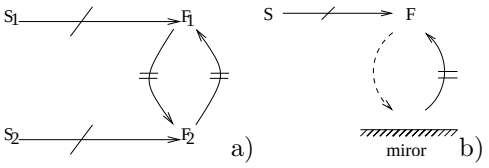


Figure 5: a) Final simplification of the network representing the interaction between our 2 identical emotional agents. b) Minimal architecture allowing the agent to learn "internal state"-"facial expression" associations.

It is simple on Fig.5 to see the condition of the learning stability. If both agents display their internal emotional state S_1 and S_2, the learning is impossible if we suppose both agents have independent emotional states (there is no correlation between S_1 and S_2). The learning cannot stabilized and then the simplification rules to obtain this Fig.5 cannot apply. If we suppose that there is no way to control the internal state of the baby, the only solution is to suppose that the second agent mimics or resonate (Nadel et al., 2006b) to the facial expressions of the baby thus allowing an explicit correlation (the parent is no more than a mirror).
If this condition is verified the system can learn, the agent 1 (baby) learns to associate the visual recognition of the tested facial expressions to its own internal feeling (E_1). The agent learns how to connect the felt but unseen movements of self with the seen but unfelt movements of the other.
To test this model, we propose to develop a neural network architecture and to adopt the following experimental protocol: The facial expressions of the robotic head calibrated by FACS experts (Ekman and Friesen, 1978), without any instruction the human subject resonates with the facial expressions of the robot head (Nadel et al., 2006b). In a first phase of interaction, the robot produces a random facial expression (sadness, happy, anger, surprised) plus the neutral face during 2s, then returns to a neutral face to avoid human misinterpretations of the robot facial expression (The same procedure is used in psychological experiment) during 2s. The human subject is asked to mimic the robot head. After this first phase lasting between 5 to 10 min according to the subject "patience". The generator of random emotional states is stopped. If the N.N has learned correctly, the robot must mimic the facial expression of the human partener.
The computional architecture (see fig.6) is close to the therorical model, this architecture allows to recognize the visual features of the people interacting with the robot head and to learn if these features are correlated with its own facial expression. Moreover, another sub network learns to predict the rhythm of the interaction allowing to detect if an interacting agent (a human) faces the robot head. The justifications of the architecture will be provided in the next sections.

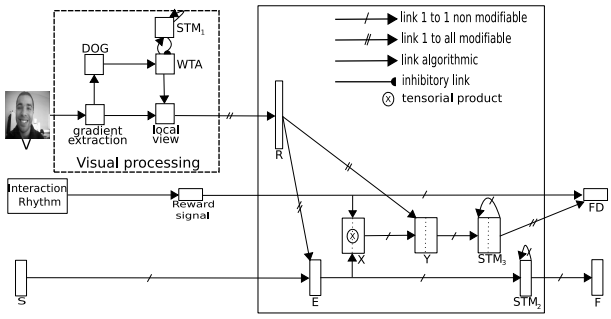


Figure 6: The global architecture to recognize facial expression, imitate and to recognize face/non face. A visual processing allows to extract sequentially the local views. The R group (local view recognition) learns the local views (each group of neuron S, E, STM_2 and F contains 5 neurons corresponding to the 4 facial expressions plus the neutral face). A tensorial product is performed between E (emotional state) and a reward signal to select the neuron which must learn. The FD group (Face detection) group learns the correlation between the tensorial product and the reward signal (its activity corresponds to the recognition of a face). Y learns the correlation between a local view and a facial expression on a specific neuron activated if a reward linked to the interaction has been obtained or not.

3. Facial expression recognition

Our initial approach followed classical algorithms: (1) face localization using for instance (R.L Hsu, 2002)

or (Viola and Jones, 2004), then (2) face framing, and (3) facial expression recognition of the normalized image. In this case the quality of the results is highly dependant on the accuracy on the frame of the face (the generalization capability of the N.N can be affected).

In the perspective of embedding the process, we had to avoid any ad hoc framing mechanism. Our solution introduces a visual system independent from face framing. The visual system is based on a sequential exploration of the image focus points (Fig.7). The focus points are the result of a DOG filter convolved with the gradient of the input image (we used this technics both for visual place and object recognition (Giovannangeli et al., 2006)). This process allows the system to focus more on the corners and end of lines in the image (eyebrows, corners of the lips, etc). Its main advantages over the SIFT method are its computational speed and the few number of needed focus points.

One by one, the most active focus points of the same image are used to compute local views (a log polar [1] transform centered on the focus point and his ray is 20 pixels).

This collection of local views is learned by R:

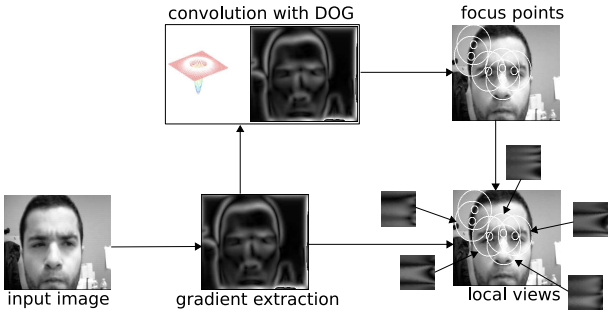


Figure 7: Visual processing: with the input image (100x100 pixels) is performed the gradient extraction, convolution with a Difference Of Gaussian (DOG) providing the focus points, the focus points extraction, local views extraction around each focus points.

$$R_j = net_j . H_{max(\gamma, \overline{net}+\sigma_{net})}(net_j)$$

$$net_j = 1 - \frac{1}{N}\sum_{i=1}^{N} |W_{ij} - I_i|$$

R_j is the activity of neuron j in the group R. $H_\theta(x)$ is the Heaviside function [2]. γ is the vigilance (threshold of recognition, if the prototype recognition is below γ then a new neuron is recruited). $\overline{net}$ is the average of the output, σ_{net} is the standard deviation. The learning rule allows both one shot learning and long term averaging. The modification of the weights is computed as follow:

$$\Delta W_{ij} = {\delta_j}^k (a_j(t) I_i + \epsilon (I_i - W_{ij})(1 - R_j))$$

with $k = ArgMax(a_j)$, $a_j(t) = 1$ only when a new neuron is recruited otherwise $a_j(t) = 0$. ${\delta_j}^k$ is the Kronecker symbol [3] and ε is the constant in order to average the prototypes. When a new neuron is recruited, the weights are modified to match the input (term $a_j(t) I_i$). The other part of the learning rule $\varepsilon(I_i - W_{ij})(1 - R_j)$ averages the already learned prototypes (if the neuron was previously recruited). The more the input will be close to the weights, the less the weights are modified. Conversely the less the inputs will be close to the weights, the more they are averaged. The quality of the results depends on the ε value as shown in Fig.8. If ε is chosen too small then it will have a small impact. Conversely, if ε is too big, the previously learned prototypes can be unlearned. Thanks to this learning rule, the neurons in the R group learn to average prototypes of face features (for instance, a mean lip for an happy face).

Of course, there is no constraint on the selection of the local views (absence of any framing mechanism). This means that numerous distractors can be present (local views in the background, or inexpressive parts of the head). It also means that any of this distractors can be learned by R. Nevertheless, the complete architecture will tend to learn and reinforce only the expressive features of the face (see Fig.6). The robot will extract important information from the dynamics of the interaction, thus to know when the face is present and when a facial expression is showed. In such face to face situation, the distractors are present for all the facial expressions so their correlation with an emotional state is null. E associates the activity of R with the current internal state S of the robot (simple conditioning mechanism using the Least Mean Square (LMS) rule (Widrow and Hoff, 1960)). STM_2 is Short Term Memory used to sum and filter on a short period (N iterations) the emotional states $E_i(t)$ associated with each explored local view:

$$STM_{2,i}(t+1) = \frac{1}{N} . E_i(t+1) + STM_{2,i}(t)$$

i is the indice of the neurons, for instance E_i corresponds to the i^{th} emotional state ($0 < i \leq 5$).

[1] The local polar transform increases the robustness of the extracted local views to small rotations and scale variations

[2] Heaviside function:

$$H_\theta(x) = \begin{cases} 1 & \text{if } \theta < x \\ 0 & \text{otherwise} \end{cases}$$

[3] Kronecker function:

$${\delta_j}^k = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{otherwise} \end{cases}$$

coefficient ε	sadness	neutral	happy	anger	surprised	total
1	49%	6%	92%	64%	42%	51%
0.1	51%	4%	92%	66%	49%	52%
0.01	51%	34%	94%	67%	49%	59%
0.001	49%	51%	79%	70%	57%	61%
0.0001	49%	51%	78%	69%	56%	60%

Figure 8: Effect of the ε local view averaging on the global performances of the system (with γ=0.92).

Generally, the 10 most activive focus points are selected. A majority of focus point (around 7/10) belongs to the face (mouth, eyebrow) but there are distractors belonging to the background.

F triggers the facial expression of the robot, the F_i highest activity triggers the i^{th} facial expression thanks to a WTA. For more robustness, F used also a short term memory(it gives more importance at the present than at the past) with $\beta = 1$ and $\alpha < 1$ (usually $\alpha = 0.8$):

$$F_i(t+1) = \beta.STM_{2,i}(t+1) + \alpha.F_i(t)$$

After learning, the associations between R the view recognition and E the emotional state are strong enough to bypass the low level reflex activity coming from the internal state S (see section 2.). In this case, the facial expression F will result from the temporal integration of the emotional state associated to the different visual features analyzed by the system (features will have an emotional value if they are correlated with the robot facial expression, basically the expressive features of the human head). Hence the robot head will begin to imitate the facial expression of the human partner.

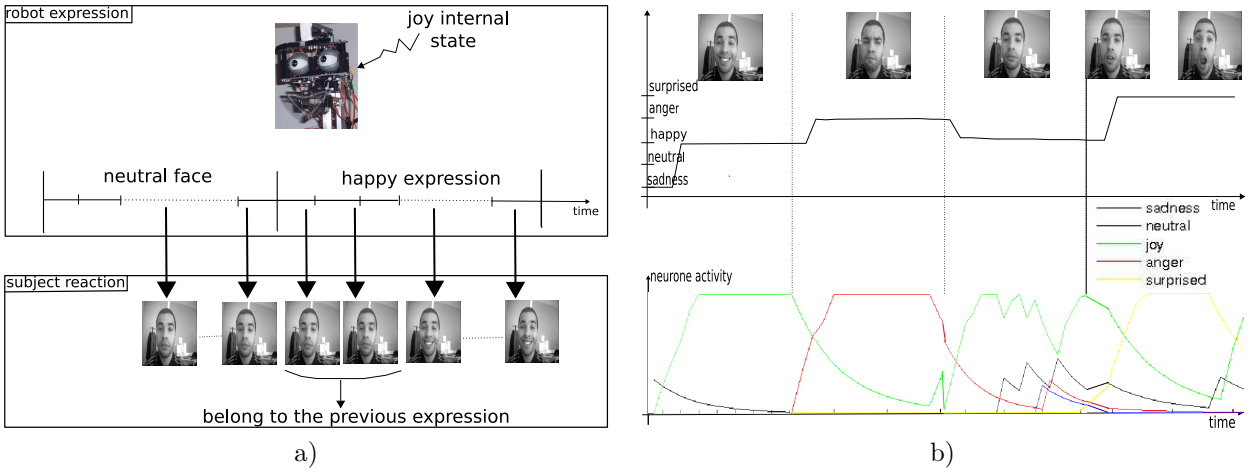
The results of Fig.8 show that our architecture is able to recognize the facial expression without the face detection. The theoretical model has been successfully translated in a computational model. Nevertheless, the on line learning can involve problems because the human reaction time is not immediate (see Fig.9a). First, 150 ms are required to recognize an object (Thorpe et al., 1996), hence the minimal duration to recognize the facial expression for a human is 150ms. The minimal period T of an interaction loop is the sum of t_1 the time for the robot to perform a facial expression plus t_2 the time for the human to recognize the facial expression plus t_3 the time for the human subject to mimic the recognized expression ($T = t_1 + t_2 + t_3$). When the robot is only an automata producing facial expressions, we measure a minimal period T around 800 ms for expert subjects (a person knowing the robot) and 1.6 s for a novice subject (a person who never interacted with the robot). This time lag can pertubate the learning because if the robot learns the first images which are still associated to the human previous facial expression then the previous expression is unlearned. The presentation time of a given expression must be long enough to neglect the first images.

Fig.9.b shows the neural activity during the test phase. In this figure, we can see that the robot reacts correctly for the different facial expressions excepted the neutral face.

4. Face/non face recognition thanks to facial expression recognition

The goal of this section is the face/non face recognition without an external supervision. Since, the robot is able to recognize a facial expression without the face/non face recognition, we will show that the facial expressions recognition can be a bootstrap to recognize the face/non face. To perform this task, the robot uses the interaction rhythm (reinforcing signal) and the facial expression recognition. We will see the importance of the interaction rhythm and we will describe the global system for the face/non face recognition. This system has no real interaction capability during the learning phase since this phase is completely predetermined (the robot continues to trigger randomly the facial expression even if the subject is gone). In first time, we introduce the prediction of the interaction rhythm to solve this problem. Psychologists underline the importance of the synchrony during the interaction between the mother and the baby. For instance, babies are extremely sensitive to the interaction rhythm with their mother(Devouche and Gratier, 2001). A social interaction rupture involves negative feelings (agitation, tears ...). However, a rhythmic interaction between the baby and her mother involves positif feelings and smiles. These works show the importance of the interaction rhythm. In our case (following (Andry et al., 2001)), the rhythm is used as a reward signal:

- A rhythmic interaction is equivalent to a positive reward: The robot head and the subject produce a coherent action at each instant.
- Conversely, an interaction rupture means a nega-



a) b)

Figure 9: a) phase shifting between the human facial expression and the robot facial expression during an imitation game (the human imitating the robot). b) Temporal activity of the neurons associated to the triggering of the different facial expressions when the robot imitates the human (after learning).

tive reward.

When a subject displays a facial expression, he/she performs whole face or body motions. If the subject imitates the robot then his/her movement peaks have a frequency depending on the frequency of changes in the robot facial expressions (in our case this frequency is constant since the robot facial expression changes after 4s). Hence, the robot can predict the interaction rhythm either using a prediction of the timing between 2 visual peaks (stable frequency of interaction of the human partner) or using the prediction of the delay between the triggering of its facial expression and the motions perceived by its CCD cameras (reaction of the supposed human partner to the robot expression). So the robot can predict the interaction rhythm and it can measure the prediction error. If the error is important, there is a novelty (subject is not in the rhythm). Otherwise, the prediction error is small which involves a good interaction between the subject and the robot (see Fig.10b)). The details of the neural network used for the rhythm prediction, were presented in (Andry et al., 2001, Banquet et al., 1997). The Neural Network uses three groups of neurons, each group having a different functionality. A Derivation Group (DG) receives the input signal. The Temporal Group (TG) is a battery of neurons (15 neurons) with different temporal activities. The Prediction Group (PG) learns the conditioning between DG (the present) and TG (the past) informations. This model (see Fig.10 a) is grounded on the following rule: a PG neuron can learn and also predict the delay between two events from DG.

The interaction rhythm provides an interesting reinforcement signal to learn to recognize an interacting partner, in our case a human and more specifically his/her face because of the short interaction distance (the robot sees the human face and not really the other part of his/her body).

A tensorial product is performed between the reward signal (built according to the interaction rhythm, one neuron is used if the interaction rhythm is coherent, otherwise another neuron is used) and E (5 neurons), to build the X group of neurons which is a matrix of 10 neurons (5 lines and 2 columns). A simple conditioning mechanism using the LMS rule is used to associate the activity of the neurons in the recognition of the local views R with the current X activity, the group Y learns this conditioning (if $X_{i,j}$ is activated then $Y_{i,j}$ must learn). After learning, the associations between R activity and Y activity are strong enough to bypass the low level activity coming from $X_{i,j}$. Next, a STM_3 (10 neurons) is used to accumulate the focus points of an image:

$$STM_{3,(i,j)}(t+1) = \frac{1}{N} Y_{(i,j)}(t+1) + STM_{3,(i,j)}(t)$$

(i,j) is the indice of the neurons ($0 < i \leq 5$ and $0 < j \leq 2$) and N is the number of focus points.

When the system has learned, The STM_3 matrix tends to activate more the first column when there is a face than the second column, inversely if it is not a face. It remains to learn the conditioning to associate the activity of the neurons in the group STM_3 and the reward signal provided by the interaction rhythm. A simple LMS (FD Face Detection) is used to perform this task. The group of neuron FD (2 neurons) is able to recognize the face after the learning.

The first results linked to this on-line learning of the face are very positive. When the face detection is learned and tested on the same subject, the system success rate on that subject tends toward 100%.

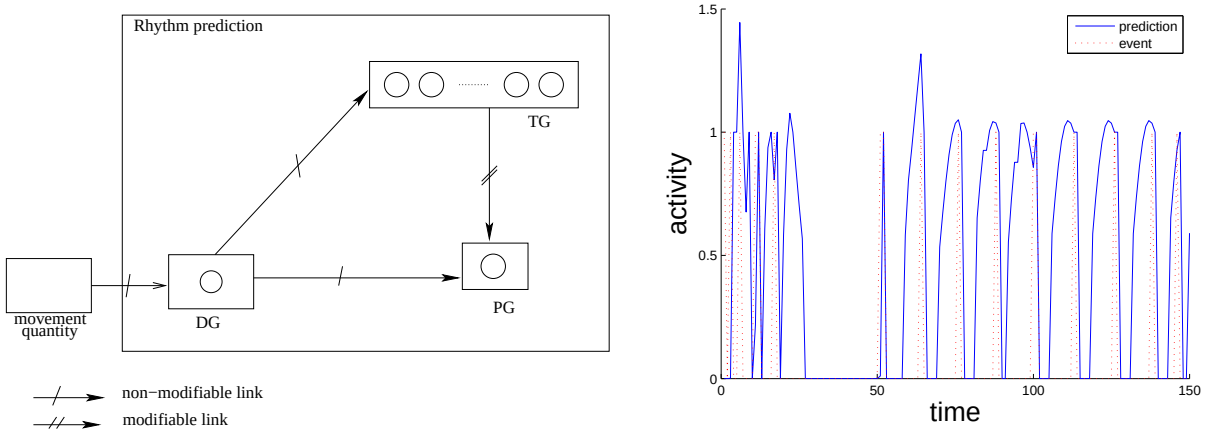


Figure 10: a)The model for the prediction of the interaction rhythm between the subject and the robot. b)Activity of neuron *PG* for the rhythm prediction, when the robot performs the facial expressions and the human imitates the robot head.

However, when the face detection is learned with a single subject and tested on 4 other subjects, the system success rate scales between 29% (for people with beard) and 90%. It is not so bad if we consider that the learning was performed during only 10 minutes (in real time) with a single subject. This show the generalization capabilities of our visual system and justify the choice of DOG filter to focus the robot attention on particular visual features. Now, when the face detection is learned on 3 subjects (1500 images) and the tests are performed on 5 others different subjects (5000 images), the system success rate tends toward 70% for the face detection. The performances should be improved after the interactions with more and more people. The goal here was to show that the emotional interactions can structure the learning. Thanks to the emotional interactions (the on-line facial expression recognition), the neural network is able to perform a face/non face recognition. The emotional interaction is a bootstrap to learn to recognize what is a human face.

5. Conclusion

The theorical model has allowed us to show that in order to learn on line to recognize the facial expressions, the learner must produce facial expressions first and be mimicked by his/her caregiver.

The system proposed in (Gaussier et al., 2007) had no real interaction capability during the learning phase since this phase was completely predefined. The idea used in this paper is to introduce the prediction of the interaction rhythm as a way to build an internal reinforcement signal allowing to change the robot behavior. Interestingly, the reward can also be used to detect if the robot is interacting with a partner or not. Since, in our case, the interacting agent is a human, it was easy to derivate a neural network for the face/non face discrimination as a particular case linked to the properties of the visual signal. If the human partner is near the robot head then a face/non face discrimination can be learned. For longer distances, one can imagine a human/thing discrimination could be performed. We have shown there is no need to find first the face and to recognize next the facial expression. The recognition of local views associated or not to a given emotional state is sufficient to "recognize" the facial expression of the human partner. The attentional strategy (using focus points) presented in this paper corresponds to a sequential and time consuming analysis of the image. It could be seen as a simple implementation of the thalamo-cortico-amygdala pathway in the mammal brain (LeDoux, 1996). In previous works (Gaussier et al., 2007), we tested simpler and faster architectures using the whole image. They could correspond to the short thalamo-amygdala pathway (Papez, 1937, LeDoux, 1996) implied in rapid emotional reactions. In future works, we will try to verify the idea emerging from the present work that the thalamo-cortico-amygdala network may be a way to control the learning of the thalamo-amygdala network allowing both a quick recognition of the facial expressions and their precise labelling.

Finally, in the proposed architecture, the emotional interaction can be seen as a way to structure learning (the emotional interaction is a bootstrap for the face/non face discrimination). Our approach could be to generalized to the learning of more complex tasks involving other kinds of movements since we have

shown in (Gaussier et al., 1998, Andry et al., 2001, Andry et al., 2002) that a simple sensory-motor system is sufficient to trigger low level imitations.

In conclusion, this work suggests the baby/parents system is an autopoietic social system (Mataruna and Varela, 1980) in which the emotional signal and the empathy are important elements of the network to maintain the interaction and to allow the learning of more and more complex skills. Future works using our robotics head will try to test this hypothesis in more dynamical situations involving human/robot and robot/robot interactions.

Acknowledgments

The authors thank J. Nadel, M. Simon and R. Soussignan for their help to calibrate the robot facial expressions and P. Canet for the design of the robot head. Many thanks also to L. Canamero for the interesting discussions on emotion modelling. This study is part of the European project "FEELIX Growing" IST-045169 and also the French Region Ile de France. P. Gaussier thanks also the Institut Unisersitaire de France for its support.

References

Andry, P., Gaussier, P., Moga, S., Banquet, J., and Nadel, J. (2001). Learning and communication in imitation: An autonomous robot perspective. *IEEE transactions on Systems, Man and Cybernetics, Part A*, 31(5):431–444.

Andry, P., P.Gaussier, and Nadel, J. (2002). From sensorimotor coordination to low level imitation. In *Second international workshop on epigenetic robotics*, pages 7–15.

B. Abboud, F. Davoine, M. D. (2004). Facial expression recognition and synthesis based on an appearance model. *Signal Processing: Image Communication*, 19:723–740.

Balkenius, C. and Moren, J. (2000). Emotional learning: a computational model of the amygdala. *Cybernetics and Systems*, 6(32):611–636.

Banquet, J., Gaussier, P., Dreher, J. C., Joulain, C., Revel, A., and Günther, W. (1997). Space-time, order, and hierarchy in fronto-hippocampal system: A neural basis of personality. In Matthews, G., (Ed.), *Cognitive Science Perspectives on Personality and Emotion*, volume 124, pages 123–189, Amsterdam. North Holland.

Devouche, E. and Gratier, M. (2001). Microanalyse du rythme dans les échanges vocaux et gestuels entre la mère et son bébé de 10 semaines. *Devenir*, 13:55–82.

Ekman, P. and Friesen, W. (1978). Facial action coding system: A technique for the measurement of facial movement. *Consulting Psychologists Press, Palo Alto, California*.

Franco, L. and Treves, A. (2001). A neural network facial expression recognition system using unsupervised local processing. *2nd international symposium on image and signal processing and analysis. Cognitive neuroscience*, 2:628–632.

G. Gergely, J. W. (1999). Early socio-emotional development: contingency perception and the social-biofeedback model. *In P. Rochat, (Ed.), Early Social Cognition: Understanding Others in the First Months of Life*, pages 101–136.

Gaussier, P. (2001). Toward a cognitive system algebra: A perception/action perspective. *In European Workshop on Learning Robots (EWRL).*, pages 88–100.

Gaussier, P., Boucenna, S., and Nadel, J. (2007). Emotional interactions as a way to structure learning. *epirob*, pages 193–194.

Gaussier, P., Moga, S., Quoy, M., and Banquet, J. (1998). From perception-action loops to imitation processes: a bottom-up approach of learning by imitation. *Applied Artificial Intelligence*, 12(7-8):701–727.

Giovannangeli, C., Gaussier, P., and Banquet, J.-P. (2006). Robustness of visual place cells in dynamic indoor and outdoor environment. *International Journal of Advanced Robotic Systems*, 3(2):115–124.

LeDoux, J. (1996). *The Emotional Brain*. Simon & Schuster, New York.

Liang, D., Yang, J., Zheng, Z., and Chang, Y. (2005). A facial expression recognition system based on supervised locally linear embedding. *Pattern recognition Letter.*, 26:2374–2389.

Mataruna, H. and Varela, F. (1980). *Autopoiesis and Cognition: the realization of the living*. Reidel, Dordrecht.

Nadel, J., Simon, M., Canet, P., Soussignan, R., Blancard, P., Canamero, L., and Gaussier, P. (2006a). Human responses to an expressive robot. In *Epirob 06*.

Nadel, J., Simon, M., Canet, P., Soussignan, R., Blanchard, P., Canamero, L., and Gaussier, P. (2006b). Human responses to an expressive robot. *In Epirob 06*.

P. Gaussier, K. Prepin, J. N. (2004). Toward a cognitive system algebra: Application to facial expression learning and imitation. *In Embodied Artificial Intelligence, F. Iida, R. Pfeiter, L. Steels and Y. Kuniyoshi (Eds.) published by LNCS/LNAI series of Springer*, pages 243–258.

Papez, J. (1937). A proposed mechanism of emotion. *Archives of Neurology and Psychiatry*.

R.L Hsu, M. Abdel-Mottaleb, A. J. (2002). Face detection in color images. *Pattern Analysis and Machine Intelligence, IEEE Transactions*, 24:696–706.

Schmajuk, N. (1991). A neural network approach to hippocampal function in classical conditioning. *Behavioral Neuroscience*, 105(1):82–110.

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381:520–522.

Viola, P. and Jones, M. (2004). Robust real-time face detection. *International Journal of Computer Vision*, 57:137–154.

Widrow, B. and Hoff, M. E. (1960). Adaptive switching circuits. In *IRE WESCON*, pages 96–104, New York. Convention Record.

Wiskott, L. (1991). Phantom faces for face analysis. *Pattern Recognition*, 30:586–191.

Yu, J. and Bhanu, B. (2006). Evolutionnary feature synthesis for facial expression recognition. *Pattern Recognition Letters*.