# Imitation as a Communication Tool for Online Facial Expression Learning and Recognition

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



*We are interested in understanding how babies learn to recognize facial expressions without having a teaching signal allowing to associate a facial expression to a given abstract label (i.e the name of the facial expression 'sadness', 'happiness'...). Our starting point was a mathematical model showing that if the baby uses a sensory motor architecture for the recognition of the facial expression then the parents must imitate the baby facial expression to allow the on-line learning. In this paper, a first series of robotics experiments showing that a simple neural network model can control the robot head and learn on-line to recognize the facial expressions (the human partner imitates the robot prototypical facial expressions) is presented. We emphasize the importance of the emotions as a mechanism to ensure the dynamical coupling between individuals allowing to learn more complex tasks.*

## I. Introduction

Since several years, the subject of Human/Robot interactions is became an important area of research. Yet, the proposed architectures use mainly an ad hoc engineering strategy allowing to show some impressive results but even if learning technics are used most of them use a-priori information. In the case of complex interactions, we believe the behavior must be understood in a developmental perspective to avoid the symbol grounding problem [10] (a human expert must provide knowledge to the system). We can obtain really autonomous systems as the result of the interaction between human and robot. Understanding how emotional interactions with a social partner can bootstrap increasingly complex behaviors, which is important both for robotics application and understanding the human development. Gathering information through emotional interaction seems to be a fast and efficient way to trigger learning. This is especially evident in early stages of human cognitive development, but also evident in other primates [23]. The emotion can be provided by a variety of modalities of emotional expressions, such as facial expressions, sound, gestures, etc. We choose to explore the facial expressions since they are an excellent way to communicate important information in ambiguous situations [3] but also because we can show that learning to recognize facial expression can be autonomous and very fast [2] which was not evident at first. For this purpose, we were interested in understanding how babies learn to recognize facial expressions without having a teaching signal allowing to associate for instance the vision of an "happy face" with their own internal emotional state of happiness [7].

Our starting point was motivated by the question of how a "naive" system can learn to respond correctly to other's expressions during a natural interaction. "Natural" here means that the interaction should be the less constrained as possible, without explicit reward or ad-hoc detection mechanism or formated teaching technique. In this case, a good inspiration is given by the baby-mother interaction, where the newborn or the very young baby, has a set of expressions linked with his/her own emotions. Yet, the link with the expressions of others still needs to be built. How does the link between his own emotions and the expression of others can emerge from non-verbal interactions?

Using the cognitive system algebra [8], we showed a simple sensory-motor architecture based on a classical conditioning paradigm could learn online to recognize facial expressions if and only if we suppose that the robot or the baby produces first facial expressions according to his/her internal emotional state and that next the parents imitate the facial expression of their robot/baby allowing in return the robot/baby to associate these expressions with his/her internal state [20]. Imitation is used as a communication tool instead of learning tool: the caregiver communicates with the robot through imitation. Psychological experiments [18] have shown that humans "reproduce" involuntary the facial expression of our robot face. This low level resonance to the facial expression of the other could be a bootstrap for the robot learning ("empathy" for the robot head).

Using a minimal robotic set-up (Figure 1), is interesting first to avoid the problems linked to the uncanny valley [16] and next to test which are the really important features for the recognition of a given facial expression. The robot is considered as a baby and the human partner as a parent. Originally, the robot knows nothing about the environment but it starts to learn as it interacts with the environment. Using a physical device instead of a virtual face brings several difficulties but induces a visible "pleasure" linked to the "presence" of the robot for the human partner. The robot head is also very useful because the control of the gaze direction (pan/tilt camera) that can be used both as a active perception and communication tool.

In this paper, we summarize first our formal model for the online learning the facial expressions. Next the implementation of this theoretical model without a face detection will be presented and the constraints due to the online learning will be studied.

## II. Material & Method: On line learning of facial expression recognition, an interactive model

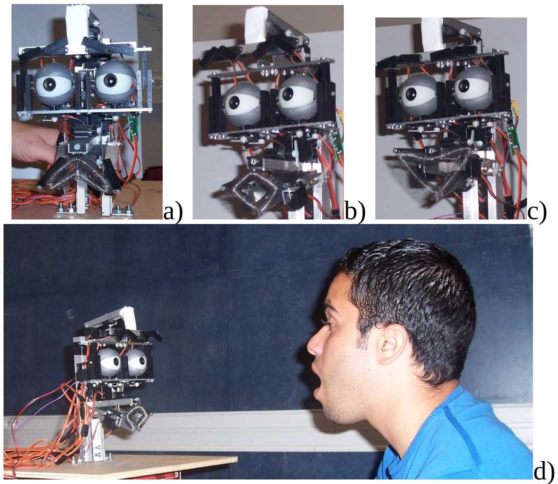


Fig. 1. Examples of robot facial expressions: a) sadness, b) surprise, c) happiness. d) Example of a typical human / robot interaction game (here the human imitating the robot).

A very simple robotic expressive head was developed as a tool for researches in cognitive sciences involving both psychological experiments and computational modelling[19]. The head was designed to be a minimal system allowing to display some prototypical basic emotions [5]. In this work, we will only use: happiness, sadness, hunger and surprise (Figure 1). The validity of this choice could be discussed but, for our purpose, all we need is a small set of emotions that can be associated to internal signals that should be present in the human or animal brain.

Our robot head is composed of 13 servo motors which are controlled by a mini SSC3 servomotor controller card allowing to maintain the servo motors in a given position (control in position) and control the different parts of the face. 4 motors control the eyebrows (bending), 1 motor controls the forehead (to move up and move down), 5 motors control the mouth (opening and bending). At last, 3 motors control the orientation of the 2 cameras located in the robot "eyes" : 1 motor controls the vertical plane (pan movement) and 2 motors control the horizontal plane (1 servos for each camera and independent tilt movement). The robot head has been programmed to display the 4 facial expressions plus a neutral pattern. Each of the four facial expressions have been controlled by FACS experts [5]. The program controlling the robot head is able to reproduce prototypical facial expressions, in other words, all the servo will move in parallel, each unit executing the position command given by the controller. This results in a dynamic and homogenous process where all the parts of the face change to form a given expression. One change of facial expression is achieved in approximately 200-400 ms depending of the distance in the joint space between two particular facial expressions. Thanks to the servo dynamics, the robot head is able to produce a infinity of facial expressions. In this paper, we want test our model with simply 5 prototypical facial expressions.

To test our paradigm, we propose to develop a neural network architecture and to adopt the following experimental protocol: 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 during 2s. The human subject is asked to mimic the robot head. After this first phase lasting between 2 to 3min according to the subject "patience". The generator of random emotional states is stopped. If the N.N has learned correctly, the robot must be able to mimic the facial expression of the human partener.

The computional architecture (Figure 2) 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.

## III. Facial expression recognition

### A. Model

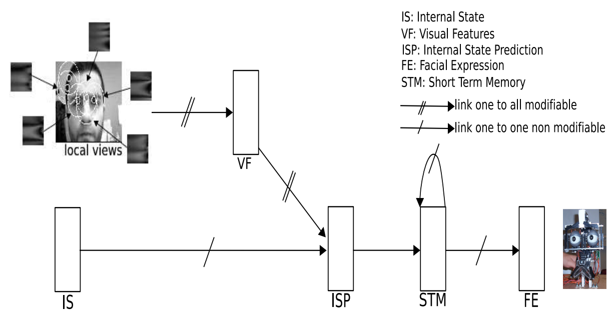


Fig. 2. The global architecture to recognize facial expression and imitate. A visual processing allows to extract sequentially the local views. The $VF$ (Visual features: local view recognition) group learns the local views. The $ISP$ (internal state prediction)learns the association between $IS$ (internal state) and $VF$. $STM$ is a short term memory in order to obtain more robustness. Each group of neuron $IS$, $ISP$, $STM$ and $FE$ contains 5 neurons corresponding to the 4 facial expressions plus the neutral face.

Our initial approach followed classical algorithms: (1) face localization using for instance [22] or [25], 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). Moreover, the robot head cannot be really autonomous because of the offline learning of the face/non face. Surprisingly, an online learning of the face/non face recognition is not as easy as the online learning of the facial expressions in the case of our mimicing paradigm since we do not have a "simple" internal signal to trigger a speficic face/non face reaction of the human partner. In the perspective of an autonomous learning avoiding any ad hoc framing mechanism appeared as an important feature. Our solution

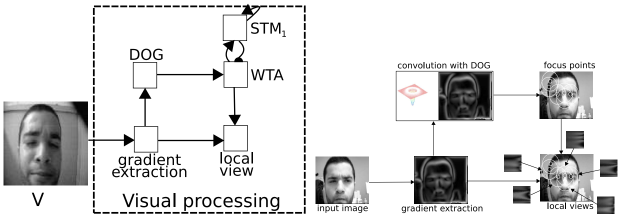


Fig. 3. Visual processing: This visual system is based on a sequential exploration of the image focus points. The input image (256x192 pixels) is performed the gradient extraction, convolution with a Difference Of Gaussian (DOG) providing the focus points, the focus points extraction, local views extraction around each focus points.

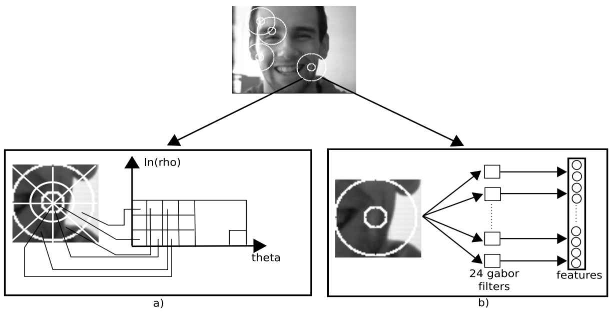


Fig. 4. Visual features: a) The local polar transform increases the robustness of the extracted local views to small rotations and scale variations (log polar transform centered on the focus point is performed to obtain an image more robust to small rotations and distance variations and his radius is 20 pixels). b) gabor filters are performed to obtain an image more robust to rotations and distance variations (the gabor filters are 60x60), the features extract for each convolution with a gabor filter are the mean and the standard deviation.

uses a visual system independent from face framing. The visual system is based on a sequential exploration of the image focus points (Figure 3). The focus points are the result of a DOG filter convolved with the gradient of the input image. This process allows the system to focus more on the corners and end of lines in the image for example eyebrows, corners of the lips, but also distractors (hair, background). Its main advantages over the SIFT (Scale Invariant Feature Transform) [15] method are its computational speed and a fewer extracted focus points (the intensity of the point is directly its level of interest). One after the other, the most

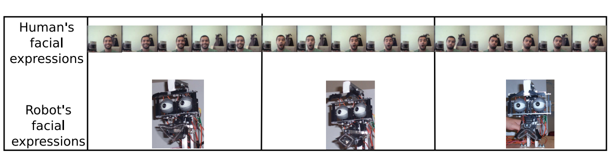


Fig. 5. The robot is able to recognize the facial expressions when the human's partner is at a distance of 2 m.

active focus points of the same image are used to compute local views: either a log polar [1] transform centered on the focus point is performed to obtain an image more robust to small rotations and distance variations and his radius is 20 pixels, and gabor filters are performed (robust to rotations and distance variations) (Figure 4). The features extract for the convolution between the gabor filter and the focus point are the mean and the standard deviation. This collection of local views is learned by the recrutement of new neurons in the visual features ($VF$) group using a k-means variant allowing online learning and real time functions [12]:

$$VF_j = net_j.H_{max(\gamma,\overline{net}+\sigma_{net})}(net_j) \quad (1)$$

$$net_j = 1 - \frac{1}{N}\sum_{i=1}^{N}|W_{ij} - I_i| \quad (2)$$

$VF_j$ is the activity of neuron $j$ in the group $VF$. $I$ is a visual input. $H_\theta(x)$ is the Heaviside function [2]. $\gamma$ is the vigilance (threshold of recognition, if the prototype recognition is below $\gamma$ then a new neuron is recruited). $\overline{net}$ is the average of the output, $\sigma_{net}$ 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:

$$\Delta W_{ij} = {\delta_j}^k(a_j(t)I_i + \epsilon(I_i - W_{ij})(1 - VF_j)) \quad (3)$$

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 $\varepsilon$ 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 $\varepsilon(I_i - W_{ij})(1 - VF_j)$ 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. If $\varepsilon$ is chosen too small then it will have a small impact. Conversely, if $\varepsilon$ is too big, the previously learned prototypes can be unlearned. Thanks to this learning rule, the neurons in the $VF$ 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 (no 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 these distractors can be learned on $VF$. Nevertheless, the architecture will tend to learn and reinforce only the expressive features of the face (Figure 2). In our face to face situation, the distractors are present for all the facial expressions so their correlation with an emotional state tends toward zero.

The internal state prediction ($ISP$) associates the activity of $VF$ with the current $IS$ (internal state) of the robot (simple conditioning mechanism using the Least Mean Square

[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:

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

($LMS$) rule [26]):

$$\Delta w_{ij} = \epsilon . VF_i . (IS_j - ISP_j) \quad (4)$$

$STM$ is Short Term Memory used to sum and filter on a short period ($N$ iterations) the emotional states $ISP_i(t)$ associated with each explored local view:

$$STM_i(t+1) = \frac{1}{N} . ISP_i(t+1) + \frac{N-1}{N} STM_i(t) \quad (5)$$

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

Arbitrary, a limited amount of time is fixed for the visual exploration of one image. The system succeeds to analyse 10 local views on each image. It is a quite small number of points but since the system usually succeeds to take 3 to 4 relevant points on the face (mouth, eyebrow).Yet, it is enough in most cases and it allows to maintain real time interaction (3 to 5 images/second) in order to test our model.

$FE$ triggers the facial expression of the robot, the $FE_i$ highest activity triggers the $i^{th}$ facial expression thanks to a WTA.

### B. Experiment results

After learning, the associations between the view recognition ($VF$) and the emotional state ($ISP$) are strong enough to bypass the low level reflex activity coming from the internal state $IS$. In this case, the facial expression $FE$ 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). The robot head can imitate the human's facial expression and the focus points are associated to each facial expression i.e these focus points vote for the recognition of a given facial expression. Each facial expression is mainly characterized by a specific set of focal points corresponding to local areas on the face which are relevant for the recognition of that expression. For example, some local view around the mouth (lip) characterize the "happyness" facial expressions, some others around the eyebrows characterize the anger facial expression. After learning of the N.N, Figure 5 shows that the robot recognizes the facial expressions even when the interaction distance is important (2m of distance). In this case, we can see the system learns to discriminate background informations (distractors in the image) from the visual features on the face, really relevant for our interaction game (local views associated to an emotional content).

Figure 6 shows that the interaction with the robot head during 2 min can be enough in order to learn the facial expressions before the robot can imitate the human partner. This incremental learning is robust although the number of human partners increases and that the expressivity between the humans (for example the sadness facial expression) is very different. Figure 7 shows that the model can generalize to people who were not present during the learning phase. A possible explanation for the bad result concerning sadness is

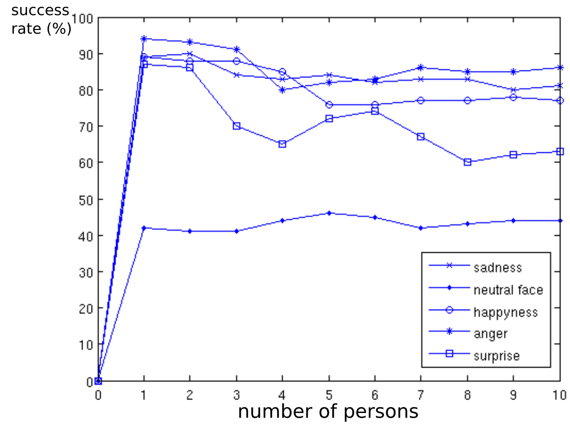


Fig. 6. The success rate of each facial expression (sadness, neutral face, happyness, anger, surprise). These results are obtained during the natural interaction with the robot head. 10 persons interacted with the robot head (32 images by facial expression by person). During the learning phase (2 minutes), these humans imitate the robot, then the robot imitates them. In order to build statistics, each image was annotated with the response of the robot head. The annotated images were analyzed and the correct correspondance was checked by a human. On line robot performances are far better but more difficult to analyze.

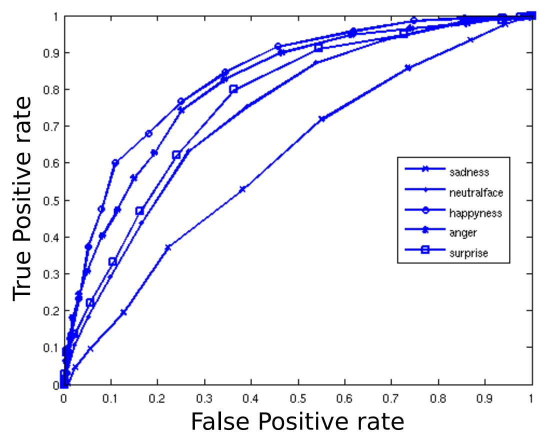


Fig. 7. Generalisation to new faces: After 20 persons interacted with the robot head (learning phase), the robot had to imitate new persons never seen. The false positive rate and true positive rate of each facial expression (sadness, neutral face, happiness, anger, surprise) with the visual process fusion (log polar transform and gabor filters). Here, we don't use a $WTA$, but a threshold function is used to enable all neurons above the threshold. A true positive is a correctly categorized positive instance and false positive is a negative instance which is categorized positive.

that the people have difficulties to display sadness without a context. Each partner imitating the robot displays the sadness in a different way. Nevertheless, the on line learning can involve problems because the human reaction time to the robot facial expressions is not immediate (Figure 8a). First, 150 ms are required to recognize an object [24], hence the minimal duration to recognize the facial expression for a human is not negligible. The minimal period $T$ of an interaction loop is the sum of $t_1$ the delay for the robot to perform a facial expression plus $t_2$ the delay for the human to recognize the facial expression plus $t_3$ the delay 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 800ms for expert subjects and 1.6 s for a novice subject. This time lag can pertubate the learning because if

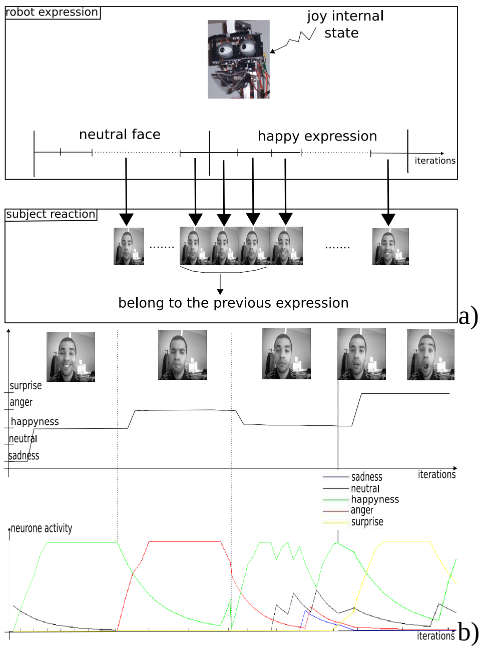


Fig. 8. a) phase shifting between the human facial expression and the robot facial expression during an imitation game (the human imitating the robot). b) Temporal activity of the neurons associated to the triggering of the different facial expressions when the robot imitates the human (after learning).

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. Figure 8.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.

In this section, we showed that the robot head is able to learn and recognize autonomously facial expressions if during the learning the robot head does facial expressions and the human partner mimicks it.

## IV. Discussion & Conclusion

Many existing researches focus on the building of a robust system to recognize the facial expressions but they are not interested in understanding how this learning could be performed autonomously. Some methods are based on the Principal Component Analysis (PCA) for example the LLE (Locally Linear Embedding) [14]. Neuronal methods have also been developed for facial expression recognition. In Franco and Treves[6] network uses a multi layer network using a classical supervised learning rule. Others methods are based on face models which try to match the face (appearance model[1]). Yu[29] uses a support vector machine (SVM) to categorize the facial expressions. Wiskott[27] uses Gabor wavelets to code the face features as 'jets'. All these technics used an offline learning and try to introduce a lot of a priori to improve the performances of the system. Moreover, all these methods need to access the whole learning database thus they can't be accepted for a realistic model of the baby learning.

These methods have better results (above 80%) but they use databases without "noise" (database clean) where the face are framed (only the face in the image), the facial expressions are checked by human experts and the problems of the brightness are controlled. The question about how a robot can learn the facial expressions without supervision is not essential for them. Moreover, our model has abilities of adaptation thanks to the neural network and the on line learning . The "database" is built through emotional interactions, as a consequence the robot can start to reproduce the facial expressions even if the database is incomplete (incremental learning).

Breazeal[4] designed Kismet, a robot head that can recognize human's facial expressions. Thanks to an interaction game between the human and the robot, kismet learns to mimic the human's facial expressions. In this work, there is a strong a priori about what is a human face. Important focus points such as the eyes, the eye brows, the nose, the mouth, ..., are pre-specified and thus expected. These strong expectations lead to a lack autonomy because the robot must have a specific knowledge (what is a human face) in order to learn the facial expressions. On the contrary, in our model, facial expressions can be learned without any prior knowledge about what is a face. Moreover, facial expressions recognition, instead of needing a face model to be usable, can bootstrap face/non-face discrimination. Others robot heads as Einstein's robot [28] explores the process of self-guided learning of realistic facial expression production by a robotic head (31 degrees of freedom). Facial motor parameters were learned using feedback from real-time facial expression recognition from video. Their work interested to how learning to make the facial expressions (fit very well with our theoretical framework and will be useful for motor control of more complex robot head).

Our robot learns thanks to the interaction with a human partner, so several difficulties occur. First, the on-line learning can involve problems because the human reaction time can be long. This point is crucial in order to improve the results. In classical image processing system, this problem is avoided because the learning database is labelled by human experts. Moreover, some human partners are not expressive therefore the robot has difficulties to categorize the facial expressions. What is interesting is that the robot makes mistakes but if a person checks the facial expressions that these human partners do then this person make the same mistakes.

Our theorical model [20] 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 had no real interaction capability during the learning phase since this phase was completely predefined. 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 [13]. In previous works [9], we tested simpler and faster architectures using the whole image. They could correspond to the short thalamo-amygdala pathway [21], [13] implied in rapid emotional reactions. In conclusion, this work suggests the baby/parents

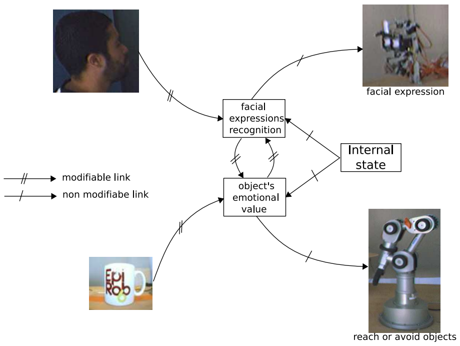


Fig. 9. Experimental set-up for social referencing. We rely upon the use of a robotic head which is able to recognize facial expressions. A robotic arm will reach the positive object and avert the negative object as a result of the interaction with a human partner.

system is an autopoietic social system [17] 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 as the social referencing[4]. Figure 9 presents new experiments in which a robotic arm learns to reach positive objects or avoid negative objects as a result of the emotional interaction with a human partner. The emotional interaction provides an emotional value to the objects (the objects have a meaning: a dangereous object or a interested object). This work emphasizes that the recognition of others agents and objects can be built through 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.


## VIDEO

http://www.etis.ensea.fr/˜sofibouc/social_referencing_v2.avi
http://www.etis.ensea.fr/˜sofibouc/feelix_interaction_emotionnel.avi

## REFERENCES


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

[2] S. Boucenna, P. Gaussier, and P. Andry. What should be taught first: the emotional expression or the face? *epirob*, 2008.

[3] S. Boucenna, P. Gaussier, L. Hafemeister, and K. Bard. Towards a new social referencing paradigm. *epirob09*, pages 201–202, 2009.

[4] Cynthia Breazeal, Daphna Buchsbaum, Jesse Gray, David Gatenby, and Bruce Blumberg. Learning from and about others: Towards using imitation to bootstrap the social understanding of others by robots. *Artificial Life*, 11(1-2):31–62, 2005.

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

[6] L. Franco and A. Treves. 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, 2001.

[7] J. Watson G. Gergely. 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, 1999.

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

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

[10] S. Harnad. The symbol grounding problem. *Physica D*, 42:335–346, 1990.

[11] C. Hasson, S. Boucenna, P. Gaussier, and L. Hafemeister. Using emotional interactions for visual navigation task learning. *International conference on Kansei engineering and emotion researchKEER2010*, pages 1578–1587, 2010.

[12] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24:881–892, 2002.

[13] J.E. LeDoux. *The Emotional Brain*. Simon & Schuster, New York, 1996.

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

[15] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2:91–110, 2004.

[16] M. Masahiro. The uncanny valey. *Energy*, 7:33–35, 1970.

[17] H.R. Mataruna and F.J. Varela. *Autopoiesis and Cognition: the realization of the living*. Reidel, Dordrecht, 1980.

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

[19] Jacqueline Nadel, Maud Simon, Pierre Canet, Robert Soussignan, P. Blancard, Lola Canamero, and Philippe Gaussier. Human responses to an expressive robot. In *Epirob 06*, 2006.

[20] J. Nadel P. Gaussier, K. Prepin. 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, 2004.

[21] J.W. Papez. A proposed mechanism of emotion. *Archives of Neurology and Psychiatry*, 1937.

[22] A.K Jain R.L Hsu, M. Abdel-Mottaleb. Face detection in color images. *Pattern Analysis and Machine Intelligence, IEEE Transactions*, 24:696–706, 2002.

[23] C.L. Russell, K.A. Bard, and L.B. Adamson. Social referencing by young chimpanzees (pan troglodytes). *journal of comparative psychology*, 111(2):185–193, 1997.

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

[25] Paul Viola and Michael Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57:137–154, 2004.

[26] Bernard Widrow and Marcian E. Hoff. Adaptive switching circuits. In *IRE WESCON*, pages 96–104, New York, 1960. Convention Record.

[27] L. Wiskott. Phantom faces for face analysis. *Pattern Recognition*, 30:586–191, 1991.

[28] Tingfan Wu, Nicholas J. Butko, Paul Ruvulo, Marian S. Bartlett, and Javier R. Movellan. Learning to make facial expressions. *International Conference on Development and Learning*, 0:1–6, 2009.

[29] J. Yu and B. Bhanu. Evolutionnary feature synthesis for facial expression recognition. *Pattern Recognition Letters*, 2006.


[4]the ability to recognize, understand, respond to and alter behavior in response to the emotional expressions of a social partner