Berthouze, L., Prince, C. G., Littman, M., Kozima, H., and Balkenius, C. (2007).
Proceedings of the Seventh International Conference on Epigenetic Robotics: Modeling
Cognitive Development in Robotic Systems. *Lund University Cognitive Studies, 135*.

Emotional interactions as a way to structure learning

P. Gaussier[1,2], S. Boucenna[2], J. Nadel[3]
[1] Member of the Institut Universitaire de France,
[2]Neuro-cybernetic team, Image and Signal processing Lab.,
Cergy Pontoise University / CNRS UMR 8051 / ENSEA
[3] LVAP Hopital de la Pitié Salpétrière Paris, France

Since several years, we are interested in understanding how babies learn to recognize facial expressions without having a teaching signal allowing to associate for instance an "happy face" with their own internal emotional state of happiness (Gergely and Watson, 1999). Using the cognitive system algebra (Gaussier, 2001), we showed a simple sensori-motor architecture using a classical conditioning paradigm could solve the task if and only if we suppose that the baby produces first facial expressions according to his/her internal emotional state and that next the parents imitate the facial expression of their baby allowing in return the baby to associate these expressions with his/her internal state (Gaussier et al., 2004). If the adult facial expression are not synchronized with the baby facial expression, the task cannot be learned. Psychologigal experiments (Nadel et al., 2006) have shown that humans "reproduce" involuntary the facial expression of our robot face. This low level resonance to the facial expression of the other can be a bootstrap for the baby learning.

In this work, the first goal was to show our theoretical model can control a real robot. Next, we wanted to verify in dynamical conditions if humans naturally enter in an emotional resonance with the robotic head, allowing the robot to perform an online learning without any explicit supervision (or predefined communication format). In a first series of experiments, we thought that the use of an ad hoc mechanism to focus on the face in order to obtain an information invariant to translation and scale could simplify the problem (we have tried numerous algorithms more or less biologically plausible based on color detection, Hough or Haar transforms...). A single neuron (or perceptron) was able to learn offline (using a database of face/non face examples) to reject the non face examples if the human was always facing the robot head. As others, we verified the performances were limited by the capability of the face detection mechanism to focus always in the same way on the face. Whatever, at the end, it was disappointing to obtain an autonomous learning of the facial expression recognition with a system using a supervised recognition of the human face! Moreover, it was clear the face recognition alone cannot be learned autonomously since we were unable to propose a criteria to bootstrap the autonomous learning of the faces / non face discrimination. Hence, we decided to sup-

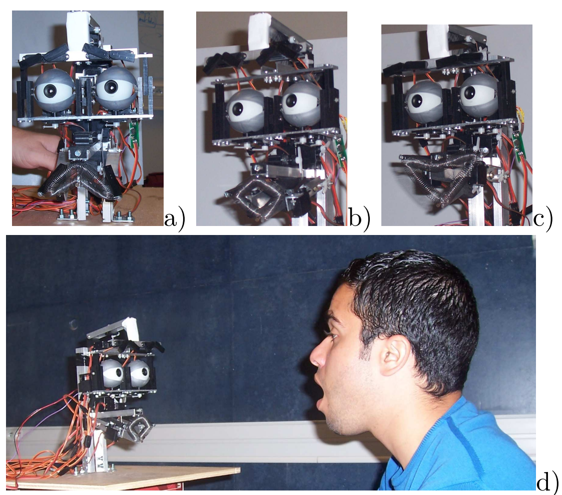

Figure 1: Examples of robot facial expressions: a) sadness, b) surprise, c) happiness. d) Example of a typical human / robot interaction game (here the human imitating the robot).

press this first step of face detection since it was not necessary in the mathematical model. In order to obtain a limited number of local views to be learned and analyzed a Difference of Gaussian (DoG) filter is applied on the gradient image to determine stable focus points associated to angular or curved areas (these points remain stable from small perspective or scale variations). An inhibition of return allows to explore sequentially the image around each focus point. To increase the analysis speed, movement information is used to amplify the activity of the focus points in the moving parts of the image (presumably at the position of the human partner face). Local views are next obtained after a log polar transform of the input image centered on each focus point (this increases the robustness of the extracted local views to small rotations and scale variations). Finally, for each emotional state, a winner takes all mechanism is used to store and recognize the explored local views. In a first phase of the experiment, the robot produces random facial expressions and analyzes the images grabbed from its CCD camera. Neuronal activities correlated with the robot facial expression (its motor command) are reinforced through a classical conditioning mechanism. The only difference here is that the unconditional stimulus is the signal generated by the robot itself to trigger a particular facial expression. As a result, the robot learns to discriminate the features associated to a given facial expression from the distractor in the image (objects in the background but also non relevant parts of the human body). In typical conditions, 2 to 4 views associated to different parts of the face are found (these views are mainly centered on the lips corners or the eyebrows). Learning and performances are tested with 4 basic facial expressions (happiness, sadness, surprise, anger) plus a "neutral" face to avoid the direct transition from one facial expression to the next. The learning phase lasts 10 min, each facial expression being displayed by the robot during 5 s, followed by 5 s of a neutral face and so on. The instantaneous performances of the robot is not very high (from 40 to 80 % according to the facial expression). The robot mistakes correspond to images we had difficulties to categorize ourselves. Yet, the introduction of a short term memory (STM) at the motor level (for the triggering of the facial expression) allows to obtain quite convincing results at the price of a small hysteresis in the decision process. As a result, the human partner has really the feeling to be imitated by the robot head. The STM filters efficiently the views incorrectly recognized since their categorization is usually different from one image to the next at the opposite to the well recognized expressions. The really interesting part of this work is that the classical procedure that suppose to localize first the face and next to recognize its expression can be avoided (see for instance (Ogino et al., 2006) for the use of an intuitive parenting procedure similar to the one we proposed but using first a face tracking system). It was a big challenge for us since we were at first unable to propose an autonomous procedure to learn the face / non face discrimination. In our architecture, the autonomous recognition of face / non face discrimination results from the facial expression recognition. A human face is recognized as such because his/her local views were associated to emotion recognition and not the opposite.

In conclusion, we postulate that in a social environment, even if some low level features can be used to help the baby to focus more on faces than other objects (because of the lack of mobility of the infant making the interaction mostly face to face and because of the pre-existence of some low level visual filters adapted to facial stimuli), the emotion communication may be used as a way to shape and trigger more and more complex learning when no explicit reinforcement signal can be used. We believe the double function of emotions as a meta controller and a communication system (Canamero and Gaussier, 2005) are deeply intricate and shape the infant cognitive development.

Acknowledgments

The authors thank 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 network of Excellence HUMAINE IST-507422, and also the European project "FEELIX Growing" IST-045169. P . Gaussier thanks also the Institut Unisersitaire de France for its support.

References

Canamero, L. and Gaussier, P. (2005). Emotion understanding: robots as tools and models. In Nadel, J. and Muir, D., (Eds.), *Emotional Development*, pages 235–258. New York: Oxfort University Press.

Gaussier, P. (2001). Toward a cognitive system algebra: A perception/action perspective. In *European Workshop on Learning Robots (EWLR)*, pages 88–100, http://www-etis.ensea.fr/~neurocyber/EWRL2001_gaussier.pdf.

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

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

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

Ogino, M., Watanabe, A., and Asada, M. (2006). Mapping from facial expression to internal state based on intuitive parenting. In *Epirob 06*.