

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

S. Boucenna¹, P. Gaussier^{1,2}, P. Andry¹

¹ETIS, CNRS UMR 8051, ENSEA, Univ Cergy-Pontoise, ²IUF,
 {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 and to recognize the facial expression of a caregiver online if this latter naturally tends to imitate the system or to resonate with it. Interestingly, our work also shows that learning autonomously the face/non face discrimination is more complex than recognizing a facial expression. We propose an architecture using the interaction rhythm to allow first a robust learning of the facial expression without face tracking and next to perform the learning of the face/non face discrimination. Finally we emphasize the importance of emotions as a mechanism ensuring the dynamical coupling between individuals in order to learn more and more complex behaviors.

Using the cognitive system algebra (Gaussier et al., 2004), we showed that a simple sensory-motor architecture based on a classical conditioning paradigm (Schmajuk, 1991, Balkenius and Moren, 2000) can learn to recognize facial expressions online. Furthermore, the dynamics of the human-robot interaction bring important non explicit signals, such as the interaction rhythm that helps the system to perform the face/non face discrimination.

We describe here a neural network architecture which is able to learn the facial expressions and then the face/non face discrimination. We adopt the following experimental protocol: the facial expressions of the robotic head are calibrated by FACS experts (Ekman and Friesen, 1978). In the first phase of interaction, the robot produces a random facial expression during 2s (among the following: sadness, happiness, anger, surprise), then returns to a neutral face during 2s to avoid human misinterpretations of the robot facial expression (same procedure as in psychological experiments). The human subject is explicitly asked to mimic the robot head (even without any instruction, the human subject resonates with the facial expressions

of the robot head (Nadel et al., 2006)). This first phase lasts between 5 and 10 min depending on the subject "patience". Then, in the second phase, the random emotional states generator is stopped. After the N.N has learned, the robot mimics the human partner facial expressions.

This architecture (see fig.1) allows to recognize the subjects visual features and to learn if these features are correlated with its own facial expressions. Moreover, another sub network learns to predict the interaction rhythm allowing to detect if an interacting agent (a human) faces the robot head. In this case, the facial expression recog-

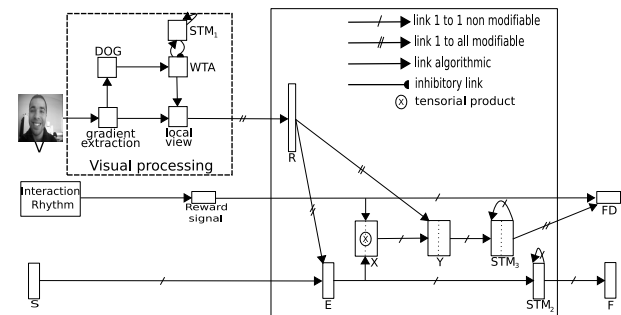


Figure 1: The global architecture is able to recognize and imitate a facial expression and to perform a face/non face discrimination. The following groups, S , E , STM_2 and F , each contain 5 neurons corresponding to the 4 facial expressions plus the neutral face. A visual processing allows to sequentially extract the local views. The local view recognition (group R) learns the local views. A tensorial product (group X) is performed between the emotional state (group E) and a reward signal in order to select which neurons must learn. The group Y learns the conditioning between R and X . The face detection (group FD) learns the conditioning between the Y Short Term Memory (group STM_3) and the reward signal. The activity of this group corresponds to the recognition of a face.

dition can be a bootstrap to discriminate a face from a non face. To perform the face/non face discrimination task, the robot uses the interaction

rhythm (reinforcing signal) and the facial expression recognition. Psychologists underline the importance of the synchrony during the interaction between the mother and the baby (Devouche and Gratier, 2001). If a rhythmic interaction between the baby and his mother involves positive feelings and smiles (positive reward), a social interaction rupture involves negative feelings (negative reward). In our case (following (Andry et al., 2001)), the rhythm is used as a reward signal. It provides an interesting reinforcement signal to learn to recognize an interacting partner.

When the face detection is learned on 3 subjects (1500 images) and tested on 5 others subjects (5000 images), the success rate of face detection tends toward 70% (with 90% on the subjects used for learning).

The idea used in this paper is to introduce the interaction rhythm prediction as a way to build an internal reinforcement signal allowing to influence the robot behavior. Interestingly, the reward can also be used to detect if the robot is actually interacting with a partner or not. 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 whole human body discrimination could be performed. We have shown that there is no need to find first the face and then to recognize the facial expression. The recognition of local views associated to a given emotional state is sufficient to "recognize" the human partner facial expression. The attentional strategy (using focus points) allows a sequential image analysis. In previous works (Gaussier et al., 2007), we tested simpler and faster architectures using the whole image instead of local views. These architectures could correspond to the short thalamo-amygdala pathway implied in rapid emotional reactions (Papez, 1937, LeDoux, 1996). In this paper, the architecture (see fig.1) could be seen as a simple implementation of the thalamo-cortico-amygdala pathway in the mammalian brain (LeDoux, 1996). 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 thus allowing both a quick facial expressions recognition 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).

This work suggests the baby/parents system is an autopoietic social system (Mataruna and Varela, 1980). The emotional signals are important elements in order to maintain the interaction and to allow the learning of more and more complex skills.

Acknowledgments

The authors thank: J. Nadel, M. Simon and R. Soussignan and P. Canet for the design and calibration of the robot head. L. Canamero for the interesting discussions on emotion modelling. This study is part of the European project "FEELIX Growing" IST-045169, the French Region Ile de France "DIGITEO" and the Institut Universitaire de France.

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.
- Balkenius, C. and Moren, J. (2000). Emotional learning: a computational model of the amygdala. *Cybernetics and Systems*, 6(32):611–636.
- 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*.
- Gaussier, P., Boucenna, S., and Nadel, J. (2007). Emotional interactions as a way to structure learning. *epirob*, pages 193–194.
- Gaussier, P., Prepin, K., and Nadel, J. (2004). Toward a cognitive system 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.
- LeDoux, J. (1996). *The Emotional Brain*. Simon & Schuster, New York.
- Mataruna, H. and Varela, F. (1980). *Autopoiesis and Cognition: the realization of the living*. Reidel, Dordrecht.
- Nadel, J., Simon, M., Canet, P., Soussignan, R., Blanchard, P., Canamero, L., and Gaussier, P. (2006). Human responses to an expressive robot. In *Epirob 06*.
- Papez, J. (1937). A proposed mechanism of emotion. *Archives of Neurology and Psychiatry*.
- Schmajuk, N. (1991). A neural network approach to hippocampal function in classical conditioning. *Behavioral Neuroscience*, 105(1):82–110.