Using the Interaction Rhythm as a Natural Reinforcement Signal for Social Robots: A Matter of Belief

Antoine Hiolle[1] Lola Cañamero[1] Pierre Andry[2]
Arnaud Blanchard[2] Philippe Gaussier[2]

[1]Adaptive Systems Research Group
School of Computer Science
University of Hertfordshire
{a.hiolle, l.canamero}@herts.ac.uk
[2]ETIS,ENSEA
Universite de Cergy-Pontoise
CNRS
{andry, gaussier, blanchard@ensea.fr}

Abstract. In this paper, we present the results of a pilot study of a human robot interaction experiment where the rhythm of the interaction is used as a reinforcement signal to learn sensorimotor associations. The algorithm uses breaks and variations in the rhythm at which the human is producing actions. The concept is based on the hypothesis that a constant rhythm is an intrinsic property of a positive interaction whereas a break reflects a negative event. Subjects from various backgrounds interacted with a NAO robot where they had to teach the robot to mirror their actions by learning the correct sensorimotor associations. The results show that in order for the rhythm to be a useful reinforcement signal, the subjects have to be convinced that the robot is an agent with which they can act naturally, using their voice and facial expressions as cues to help it understand the correct behaviour to learn. When the subjects do behave naturally, the rhythm and its variations truly reflects how well the interaction is going and helps the robot learn efficiently. These results mean that non-expert users can interact naturally and fruitfully with an autonomous robot if the interaction is believed to be natural, without any technical knowledge of the cognitive capacities of the robot.

1 Introduction

The question of how to have robot able to be useful and adaptive in our socially situated environment is of growing interest. Indeed, in a not so far future, human will have to interact daily with robots in various settings. During these interactions, robots will have to gain information from humans, and humans will have to learn from robots. Concerning the non-verbal aspects of these interactions, the field of developmental robotics has been trying to develop and study algorithms and architectures as generalisable as possible, in order for these systems to be as

minimal as possible, be that on the lower motor level [1] or at the motivational level [2]. The underlying principle of the field is to try and model phases and phenomena from the development of children and animals to understand and take advantage from the adaptivity and efficiency we observe in them.

Within this framework, the authors have been interested in defining and testing how and when learning from a human partner can be achieved with the minimum amount of prior knowledge from the robot, as a young infant has to do in the early years. During this period, the most important partner the infant has is its mother, or primary caregiver. As described by Bowlby in [3], the infant uses the attachment figure, often the mother, as a secure base to explore and learn from its experiences in unknown situations. However, the question is how does the mother elicit these positive responses and promotes healthier cognitive and socio-emotional development. One hypothesis is that the mother's sensitivity, as described in [4], or consistency in the mother's behaviour and responses to stimuli is crucial. The positive emotions and mutual delight that mother promotes a healthier development for the infant, and deep engagement from the mother [5]. Additionally, within the still-face paradigm[6], where a caretaker would produce a neutral expression after a few minutes of interaction which in turn would produce a significant fall in the infant's positive responses. Other frameworks like the Double Video paradigm, for instance [7], measured the same responses when the synchrony of the interaction was altered by introducing time delays in the mother's response. This would indicate that synchrony and timing during a mother-infant non-verbal interaction is a strong indicator of the infant pleasure and emotional responses.

Defining the notion of sensitivity in the context of human-robot interactions is far from trivial. The experiment discussed in this paper is based on our previous work [8][9][10], where was raised the question of how important the consistency of the behaviour of the human to the stability and accuracy of learned sensorimotor associations. We attempt to unify these notions in order to build a general reinforcement signal that could be used by a robot in a large number of settings, which in turn would help humans interact with robots without any knowledge of how the robot cognitive system is designed.

We here present an experiment where a NAO robot has to learn, without any prior knowledge, the correct sensorimotor associations in a "mirroring game". The actions of the human are mediated by a pink ball, and the robot uses the rhythm at which the human is performing a new action in order to reinforce the correct action to perform. This experiment is an extension of the work presented in [11], where the same algorithm was used in a human-computer interaction, and [12] where the setup was extended to work on an AIBO robot and then a NAO robot.

The results showed that the rhythm could be used as a reinforcement signal for the robot to learn the correct associations, even more so when the principle was explained, since they would explore all possibilities they know of. But the non-expert subjects, who are not used to interact with robots and do not know the rule, did not significantly manage to teach the robot the associations.

If the rhythm of the interaction is, as hypothesized, an intrinsic component of a natural, surely something was missing for the non-expert user. The modified version of this experiment presented here aims at discovering what was missing in the interaction for these subjects to succeed.

2 Architecture and experimental setup

Fig. 1: Experimental setup. The human partner is in front of the robot, moving a pink ball between the four different positions in the visual field. The robot learns the proper response to mirror the actions of the human partner.

In our setup, the robot is trying to learn to mirror the actions of the human partner, following the position of a pink ball in its visual field, as in Fig 1. The robot has to learn the four different sensorimotor associations, corresponding to the four possible positions (left arm up when the ball is in the top right of the visual field,left arm down when the ball is in the bottom right, and respectively for the left side). The learning algorithm itself functions as follows and the main components are depicted in Fig.2.

The robot has access to four different perceptions (ball in the top-left part of the visual field, ball top-right, ball bottom-left and ball bottom-right) to which it will associate an action. Every time an action is performed, the rhythm prediction component will reset and peak after a time corresponding to the last gap learned between two actions. To summarise, the robot learns in one-shot the time elapsed between two different perceptions, and expects the next action to

be performed after this precise duration. The reinforcement signal is calculated as the difference between the duration expected and the duration observed. The only prediction made here is the occurrence of the next action from the partner, which for the robot is an change in the current perception.
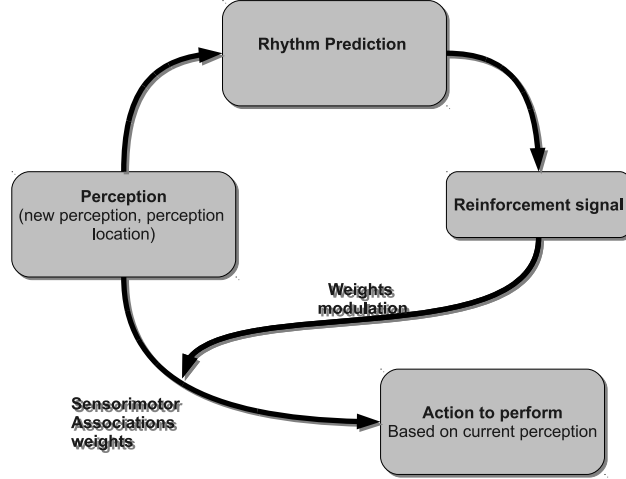

Fig. 2: Abstract representation of the robot architecture to learn the sensorimotor associations

The reinforcement signal $\mathbf{R(t)}$ from the rhythm prediction module varies as a Gaussian centred on the time t, which is the time of the next predicted event (see [11] for more details). $\mathbf{R(t)}$ is then used to change the weights between two fully connected layers of neurons (the perception and the action to be performed). The synapses have a weight W_{ij} and a probability p_{ij} associated to them, and the rule used is the **Probalistic Learning rule**. Using this rule, a fully connected neural network (perceptions connected to potential actions) behaves as follows:

$$\Delta p_{ij} = (\epsilon + \alpha \times R(t)) \times C_{ij} \tag{1}$$
$$p_{ij}(t) = H(p_{ij} + \Delta p_{ij}) \tag{2}$$

With ϵ the learning speed, α the reinforcement factor and C_{ij} the average of the past activation of unit i. Then, if a random draw $Rand$ is higher than the confidence, $Rand > p_{ij}$:

$$W_{ij} = 1 - W_{ij} \tag{3}$$
$$p_{ij} = 1 - p_{ij} \tag{4}$$

Using *Rand* promotes an exploratory behaviour when the confidence is low, and a more exploitative behaviour when it is high.

3 Experimental Design

The aim of the experiment is to assess if and how the human subjects are able to teach 4 different sensorimotor associations without any explicit feedback or reinforcement signal being used by the robot, and without the human having a prior knowledge of the signal used by the robot. If the humans are successful in that task, this will show that the rhythm is potentially an intrinsic component of non-verbal interaction that can help identify successful interaction and allow a robot to learn **without any specific reward**.

In order to keep the subjects engaged and to make the robot's behaviour appear slightly more life like, we introduced a slow balancing movement on the robot. Its torso would lean closer to the human partner and then slowly back away with a low frequency modulated by the rewards obtained over time. Moreover, in the architecture, a notion of **well-being** has been added in order to control the expressions of the robot. We therefore decouple the reward used to learn the sensorimotor associations and the overall behaviour of the robot. The robot will express happiness when the **well-being** is high, and sadness when it is low, and boredom when it is low and the perceptions stay too stable when the human is always repeting the same action over and over.

Finally, one major change was in the protocol of the experiment. Regardless of the background of the subject, they would all hear the same guidelines which are as follows: *You will be asked to use the pink ball to teach the robot to mirror your actions. The robot is able to hear your voice, but does not understand words. The robot is able to see your face and what you are expressing. The robot will only respond to movements (a change of the position of the ball in its visual field). The LEDs in the eyes of the robots will reflect the quadrant where the robot perceives the ball, and are turned off when it cannot see the ball any more. Try and act as though you were teaching this to a 6 to 15 month old infant, who is able to process voices and faces.* As a monitoring feature, we also reflected the expected rhythm in the LEDs of the robot, as an indicator for the experimenter to see how the system was performing. The LEDs woud turn to a brighter white the closer to the predicted action we got, and then fade the longer after the predicted action. The guidelines were modified to provide the human with potential existing and natural rewards (tone of voice and facial expressions) they can use without having to be trained. This was also believed to raise their confidence in the overall capabilities of the robot.

We conducted a study with 10 subjects, with ages ranging from 23 to 60 years old, and with various backgrounds. The robot used was the same which was used previously, the NAO robot (Fig. 1). The interaction would last typically ten minutes, and was ended by the experimenter. We recorded the value of the rhythm used by the human, the movements performed by the robot, and the rewards the system identified. Using this we have enough data to know what

rhythm a particular subject used in a particular situation(for instance which one was used when the robot was right, when it was wrong, when it was wrong for a long period of time, etc.). We also asked the subjects to fill in a questionnaire, designed to rate how the body postures of the robot helped the users and how they were identified. The questionnaire also asked the subject what kind of cues they thought the robot was using to learn, choosing from 4 choices: facial expressions, tone of the voice, rhythm of the humans action, the repetitiveness of the action(explained as repeting the same action over and over consecutively), or free choice. The subject could select any number of these, if they thought the robot actively used them.
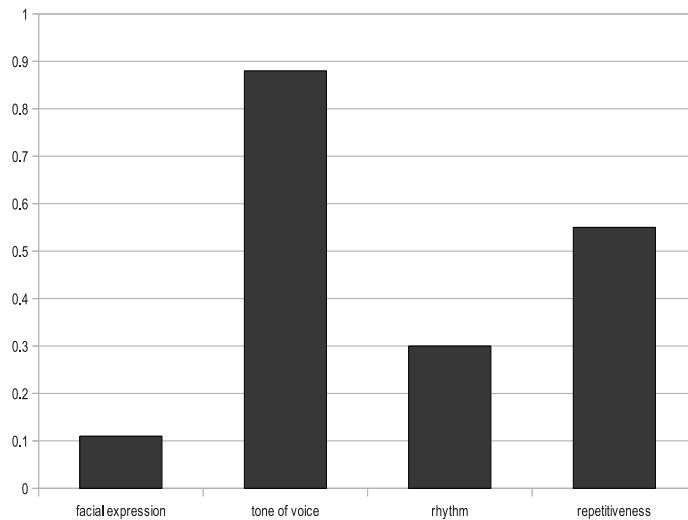
4   Results and Discussion

Fig. 3: Subjects rating on what signal is used by the robot to learn the correct associations. Although the tone of voice was not used by the robot, almost 90 percent of the subjects described it as a cue used by the robot to learn.

4.1   General Observations

Firstly, the subjects seemed far more engaged and showed less frustration during the experiment then when the expressivity was not displayed and when they did not know that the robot could use the voice and the face as available social signals. Only two of them seemed to want to stop the interaction after more

than 7 minutes. All the other subjetcs were disappointed to stop the experiment after ten minutes, since they managed to teach the robot to mirror their actions successfully several times. Since this setup still allows the system to interpret false negatives, the robot was forgetting the correct actions and then the subjects were keen on trying to make the robot learn again. Every subject successfully managed to teach the robot at least 3 correct associations, which is far more than in the previous trials, even when the subjetcs were told that the rhythm was used by the robot to learn. After a first phase of success, where the robot had learned successfully, the subject would accidentally cross another area of the visual field, changing the rhythm, leading the robot to forget an association, which would disturb the human, leading to further mistakes. This is one of the major problems of this setup. The system is prone to pick up false negative really easily.

4.2   Objective and subjective reinforcement signal

During a successful period of the interaction, the subjects would typically encourage with a "Yes" or "Very good", or any other short positive phrases when the robot was performing the correct action, using an exaggerated tone, as parents or adults use with young infants. Alternatively, they would also use "No" or "Not that" in the case of a wrong action being performed. But they did use a different rhythm consequently to a good or a wrong action of the robot, which leads to a meaningful reinforcement of the associations.

As can be seen on figure 3, when asked what signal they thought the robot used, 90% of the subjects picked the tone of voice as the main cue to learn the cordrect associations. Although most of them during the experiment were expressing the fact that at first their voice did not have any impact on the behaviour of the robot, when they managed to teach one or more associations to the robot, they would go on using their voice as a signal for the robot to use. Moreover, it suggest that the whole behaviour of the robot, was believable and consistent with what the subject would expect from an agent using natural cues in order to be taught a simple task, producing in turn a smoother and relaxed behaviour from the subjects, using implicitly the rhythm. Also, when the subjects were asked what they thought the white LEDs were reflecting, none of them identified them as an expectation of an event, or anything related to the rhythm. This suggests that this cue was not biasing their behaviour and that their focus was mainly on the task at hand. Finally, when asked about the relevance and coherence of NAO's body postures, the subjects expressed that some of them were slightly confusing (the happy and bored postures), but the sad posture, displayed when too many negative rewards were experienced, was rated as good, but more compared to frustration or disappointment.

5   Conclusions and Perspectives

The results of this experiment are promising as they indicate that the rhythm of the interaction can be used as a reinforcer to learn new sensorimotor associ-

ations without prior knowledge to the robot or the human. This system seems to be even more efficient when the human partner is unaware of the underlying mechanism and behaves more naturally, using the usual cues people would use with a young infant such as the voice and the facial expression. We are argue that this was made possible by making the robot more life-like, never really as still and predictable as a usual machine would be, but mainly by pushing the the subjects to believe that the robot could process information as a young infant would, namely the tone prosody of the voice and facial expressions of the partner. Therefore, if research goes on towards making robots use synchrony and rhythm, as young infants use during non-verbal interactions, the success will depend on the belief and confidence of the human in the capabilities of the robot. Once the human partner believes he can act naturally without any guidelines and coaching from expert users.

These results are also consistent with the ones found in [10], where subjects rated the behaviour of an Aibo robot having two sets of parameters rendering it more or less needy and reactive towards events and interactions with humans. The more reactive, moving and consistent the behaviour, the easier it was for the human to interact with the robot. From a long term perspective, if the results found here are confirmed by a more extensive and broader study, the rhythm or other synchrony related measures could be used by a robot to evaluate the potential of a human partner in becoming a reliable teacher or learning partner. Indeed, in the future if robots meet an increasing number of humans, and will need themselves assistance, knowing which partner is the most useful should be a clear advantage. In comparison with humans, who from birth do not choose their human partner in term of socio-cognitive development, the robot could actually choose carefully its primary caregiver(s).

In the future, we plan on confirming these results with a broader set of subjects, in age, technological and cultural background. We would then be interested to see with what kind of different interactions the rhythm can be used. Finally, we would work at extending the architecture to allow different "natural" reinforcement signals to be used by the robot. We would then be able to study the possibility of conflicts between these signal and develop a system which would try and cope with the possible contradictions.

Acknowledgements

References

1. L. Berthouze and M. Lungarella. Motor skill acquisition under environmental perturbations: On the necessity of alternate freezing and freeing of degrees of freedom.

Adaptive Behavior, 12(1):47–63, 2004.

2. P-Y. Oudeyer, F. Kaplan, and V. Hafner. Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, 2(11), 2007.

3. J. Bowlby. *Attachment and loss*, volume 1:Attachment. New York : Basics Books, 1969.

4. Marianne S. De Wolf and Marinus H. van IJzendoorn. Sensitivity and attachment: A meta-analysis on parental antecedents of infant attachment. *Child Development*, 68(4):571–591, August 1997.

5. E. Tronick. *The neurobehavioral and social-emotional development of infants and children*. WW Norton and Company, 2007.

6. J. Nadel, R. Soussignan, P. Canet, G. Libert, and P. Grardin. Two-month-old infants of depressed mothers show mild, delayed and persistent change in emotional state after non-contingent interaction. *Infant Behavior and Development*, 28:418–425, 2005.

7. J. Nadel, K. Prepin, , and M. Okanda. Experiencing contigency and agency : first step toward self-understanding. *Interaction Studies*, 2:447–462, 2005.

8. A. Hiolle and L. Cañamero. Why should you care? an arousal-based model of exploratory behavior for autonomous robots. In S. Bullock, J. Noble, R. Watson, and M. A. Bedau, editors, *Artificial Life XI: Proceedings of the Eleventh International Conference on the Simulation and Synthesis of Living Systems*, pages 242–248. MIT Press, Cambridge, MA, 2008.

9. A. Hiolle and L. Cañamero. Developing sensorimotor associations through attachment bonds. In C. Prince, C. Balkenius, L. Berthouze, H. Kozima, and M. Littman, editors, *Proc. 7th Intl. Wksp. on Epigenetic Robotics*, pages 45–52. Lund University Cognitive Studies, 2007.

10. A. Hiolle, A. Bard, K., and L. Cañamero. Assessing human responses to different robot attachment profiles. Proceedings of the 18th International Symposium on Robot and Human Interactive Communication, . In *Proceedings of the 18th International Symposium on Robot and Human Interactive Communication*, pages 251–257, 2009.

11. P. Andry, Ph. Gaussier, S. Moga, J.-P. Banquet, and J. Nadel. Learning and communication in imitation: An autonomous robot perspective. *IEEE Transactions on Man, Systems and Cybernetics, Part A: Systems and humans*, 31(5):431–442, 2001.

12. P. Andry, N. Garnault, and P. Gaussier. Using the interaction rhythm to build an internal reinforcement signal: a tool for intuitive hri. In C. Prince, C. Balkenius, L. Berthouze, H. Kozima, and M. Littman, editors, *Proceedings of the Ninth Int. Conf. on Epigenetic Robotics*. Lund University Cognitive Studies, 2009.