12

Using the rhythm of non-verbal human-robot interaction as a signal for learning

Pierre Andry, Arnaud Blanchard, Philippe Gaussier

***Abstract*—Human robot interaction is a key issue in order to build robots for everyone. The difficulty for people to understand how robots work and how they must be controlled will be one of the mains limit for broad robotics. In this paper, we study a new way of interacting with robots without needing to understand how robots work or to give them explicit instructions. This work is based on psychological data showing that synchronization and rhythm are very important features for pleasant interaction. We propose a biologically inspired architecture using rhythm detection to build an internal reward for learning. After showing the results of keyboard interactions, we present and discuss the results of real human-robots (Aibo and Nao) interactions. We show that our minimalist control architecture allows the discovery and learning of arbitrary sensorimotor associations games with expert users. With non-expert users, we show that using only the rhythm information is not sufficient for learning all the associations due to the different strategies used by the human. Nevertheless, this last experiment shows that the rhythm is still allowing the discovery of sub-sets of associations, being one of the promising signal of tomorrow social applications.**

***Index Terms*—Human-robot-interaction, Autonomous robotics, Self-supervised learning, artificial neural networks, rhythm detection and prediction.**

I. Introduction

UNDERSTANDING non-verbal communication is crucial for building really adaptive and interactive robots. Young infants, even at pre-verbal stages take turn, naturally switch roles and maintain bi-directional interaction without the use of explicit codes or declarative "procedures" [1]. Among the implicit but important signals of interaction, synchrony and rhythm appear to be fundamental mechanisms in early communication among humans [2], [3].

Our long term goal is to understand the emergence of turn taking or role switching between a human and a robot (or even between two robots). Understanding what the good dynamical properties of turn taking are would be an important progress towards maintaining for free robust and long lasting interaction (without explicit programming or explicit signals). It could lead to the learning of long and complex sequences of behaviors. Moreover, this work is linked to the fundamental question of how a "natural" interaction can emerge from few basic principles [4]. Taking inspiration from numerous studies in developmental psychology on how newborns and young pre-verbal infants react in interpersonal relations (see next section for a review), this work focuses on the role of rhythm in human-robot face to face interaction. Our main goal in this paper is to show that an implicit signal such as the rhythm of the interaction can be used to build internal rewards and enhance the learning of an interacting robot. To do so, we propose a simple methodology illustrating how many good associations can be learned thanks to the conjunction of a rhythm detector with different reinforcement learning rules: is a robot detecting cadence changes able to improve its responses and converge toward a satisfying interaction with a caregiver ? We present an artificial Neural Network (NN) control architecture inspired from properties of the hippocampus and cerebellum. This model allows on-line learning and the prediction of the rhythm of the sensorimotor flow of the robot. As a result, the robot is able, from its own dynamics, to detect changes in the interaction rhythm, that is to say changes in the cadence of the gesture production of the human-robot system. The architecture uses the rhythm prediction to compute an internal reward reinforcing the robot's set of sensorimotor associations. Therefore, a stable and rhythmic interaction should indicate that the robot's behavior is correct (generating internal positive rewards, strengthening sensorimotor rules, and allowing to respond correctly). On the other hand, an interaction full of breaks (the time for the caregiver to express his disagreement or even to interrupt the interaction) should lead the robot to change its motor responses due to the internal negative reward and converge toward new ones corresponding to the human's expectancies. Rhythm prediction plays the role of a control mechanism allowing the building of an internal reward, which is used to guide the learning of sensorimotor associations.

Finally we will discuss the importance of the learning rule coupled to the rhythm prediction mechanisms and highlight how rule properties can affect the quality of the interaction, and the learning of the task.

ETIS, CNRS UMR 8051, University Cergy-Pontoise, ENSEA, 6avenue du ponceau, F-95000

Institut Universitaire de France (IUF)

II. Interdisciplinary motivation

In the last decade, an important effort has been made to make robots more attractive to human beings. For instance, a major focus of interest has been put on the expressiveness and the appearance of robots [5], [6], [7] . Targeting the facilitation of human-machine communication, theses approaches have neglected the importance of the dynamics when two agents interact. In our opinion, intuitive communication (verbal or not) refers to the ability to "take turn" and to respond coherently to the stimulations of others without needing the presence of an expert. In summary, it means being able to detect the crucial signals of the interaction and use them to adapt one's dynamics to the other's behavior. Obviously, such an issue questions the sense of "self" and "others", and the position of

such notions in the frame of building an interactive robot. In the present work, we do not want to define an a priori boundary between the robot and the external world. We follow a bottom-up approach, testing how the combination of a few low-level mechanisms can allow the emergence of cooperative behaviors without any notion of self. From an epigenetic perspective this combination of low-level mechanisms can then appear as an interesting hypothesis in the discussion of the emergence of self and agency in artificial and natural systems. Among these bootstrapping mechanisms, we will emphasize novelty detection (and more precisely rhythm detection) as a way of building an internal reward structuring the learning of any interacting agent.

In developmental robotics, a lot of research focus on how artificial systems can detect novelty in their surrounding environment in order to learn and generalize new skills, i.e. how robots can, autonomously, develop and ground their own learning experience [8]. The notion of environment itself plays a crucial role, with numerous researches focusing on the embodiment or the surrounding physical environment while a few deal with the adaptation to the social environment of the robot. One of the seminal works concerns the role of prediction in the building of the relationships between the environment and the robot. In [9], the authors showed how multiple internal predictors playing the role of experts can help a mobile robot segment the outside world according to the confidence and the scale of the prediction of each expert. The segmentation of the external world in different categories (and therefore the convergence of the experts predictions) is made on-line, while the robot is exploring the environment. Another good example of an active exploration and learning of the physical environment can be given by the humanoid robots Cog and Obrero. In [10], Cog produces random arm movement to roll and push small objects, and use the induced optical movement to discriminate the objects from the background (see also [11] for the same experiment using sound with the robot Obrero). This is used to visually segment the objects and to learn some of their properties (roll vs non-roll, etc.). These experiments are examples in which the robot's progress is directly linked to its activity and the changes it induces. Also in the frame of the exploration of the physical surrounding environment, authors have proposed to introduce the notion of intrinsic motivation in artificial systems [12], [13]. They propose that self-motivated experiments of a prediction-based controller allow the robot to discover and split its physical environment into a finite number of states. The different areas of interest are segmented and labeled according to the consequence of the experiments on the learning progress of the robot [14]. This mechanism allows the robot to seek the situations where novelty will be the most "interesting" by providing the maximal learning progress (fully predictable and totally un-predictable situations progressively becoming repellers). Such an "artificial curiosity" is a good example of how the building of an internal reward value can drive the development of the robot in its physical environment. Finally, in preliminary experiments Pitti et al. [15] have started to link internal dynamics with the dynamics of a human manipulating the robot. They highlight how synchronization allows to integrate sensing and acting in an embodied robot, and then facilitates the learning of simple sensorimotor patterns. The authors show that the synchronization between input (sensing) and output (action) is a necessary condition for information transmission between the different parts of the architecture and therefore for learning new patterns.

Reviewing these studies, one important question we may ask is: can the same principles of novelty detection and intrinsic motivation be applied to a social environment, for example in the case of an autonomous robot exchanging with a caregiver? The first works started to show that behaviors such as imitation could allow autonomous robots to get for free some adaptation to the physical environment thanks to simple "social" behaviors [16]. Following this idea, we have started to consider low level imitation (imitation of meaningless gestures) as a proto-communicational behavior obtained as an emergent property of a simple perception-action homeostat based on perception ambiguity [17], [18]. Tending to balance vision and proprioception (direct motor feedback) the controller acts to correct errors induced by an elementary visual system (unexpected movement detection). An imitative behavior emerges without notion or detection of the experimenter. Based on the same principles, [19] highlights how a system naturally looking for the initial equilibrium of its first perceptions (based on the imprinting paradigm) can alternate an exploration of the physical environment and a return to the caregiver for learning grounding. More recently, researches have started to investigate the stability of human-robot face-to-face interactions, using synchrony and anti-synchrony as a link between stable internal dynamics and bi-directional phase locking with the caregiver [20], [4], [21], [22]. According to developmental psychology, the ability to synchronize, to detect the synchrony, and the sensitivity to the timing appear to play an important role in early communication among infants from birth on. Works have shown that babies are endowed with key sensorimotor skills exploited during interpersonal relations [3]. For example, the young infants show sensitiveness to timing with the ability to imitate, exchange and preserve the temporality and rhythm of sequences of sounds during bi-directional interaction with their mother [23]. Neonates (between 2 and 48 hours after birth) are able to anticipate the causality between two events, and also to express a negative expression if the second event does not follow the first one [24], [25] . Gergely and Watson have unified these findings in a model of contingency detection (DCM) extending the notion of causality to the social events [26]. According to the DCM model, infants are able to discriminate the causality provoked by their own actions (direct feedback inducing a perfect synchrony with their own motor production) for example when seeing and/or feeling the movement of their own body, from the contingencies of social interactions (due to the regular but non perfect synchrony between the emission of the stimulus by the baby and the response of the caregiver). From this point, contingency detection have turn to be one of the key mechanism of the development of self and social cognition [27]. In the same line, Prince et Al. are interested in understanding the development of infant skills with a model of contingencies detection based on the inter-modal perception of

audio-visual synchrony [28]. An extended verson of this low-level model is also tested by Rolf et al. in order to bind words or sounds with the vision of object [29].

In parallel, a lot of studies have been aimed at studying how infants detect changes in the other's responses during face-to-face interactions. Since 1978, the *still-face* paradigm, introduced by Tronick et al has been widely used and studied (see [30] for a review) especially during pre-verbal interactions. A still-face consists in the production of a neutral, still-face of the caregiver after a few minutes of interaction. Interestingly, this sudden break of the response and the timing of the interaction induce a fall of the infant's positive responses. The same responses where also measured with the more accurate Double Video paradigm allowing to shift the timing of the interaction using a dual display and recording system [2]. In this second paradigm, the content of the caregiver's responses remains the same as in a normal interaction, but a more or less long decay can be introduced in the display of the mother's responses. Using this dispositive during natural, bi-directional mother-babies interactions, Nadel and Prepin [2] highlight the importance of the timing of the response and show how synchrony and rhythm are fundamental parameters of pre-verbal communication. Breaking the timing results in violating the infant's expectations, and produces a strong increasing of negative responses.

Consequently, in the frame of a face-to-face human-robot interaction composed of simple gestures (for example an imitation game), our working hypothesis will be the following:

- a constant rhythm should naturally emerge if the interaction goes well (that is to say, if the robot's responses correspond to the human's expectancies).
- Conversely, if the robot produces the wrong behavior or the wrong responses, we suppose that the human may introduce more breaks in the interaction, for example to take the time to restart the game, manifest his disagreement(even if the robot is not able to process any signal concerning this dissatisfaction), or simply withdrawing the interaction.

Therefore, we will present in the next section our model for rhythm learning and prediction, and the internal reward extraction mechanism based on the rhythm detection.

III. Model

Our model is an artificial Neural Network divided in two main parts (Fig.1). The first part (the *Rhythm detection and prediction* layer in Fig.1, up) learns on-line and very quickly (one-shoot learning) the rhythm of the $Input$-$Output$ activity and computes a reward value R according to the accuracy of the rhythm prediction. The second part (the *sensorimotor learning* layer in Fig.1, bottom) is a reinforcement learning mechanism allowing to change the $Input$-$Output$ associations according to R values updated during the interaction.

A. Rhythm detection and prediction

The role of this network is to learn and predict the timing of the changing sensory-motor events. The network is inspired by the functions of two brain structures involved in memory and timing learning: the cerebellum and the hippocampus (see [31] for the neurobiological model we have proposed). It processes the sum of information from the *sensorimotor* pathway (integrating vision processing and the link between vision and motor output, see section III-B), stimulated each time the human makes a gesture in front of the robot. Consequently, the robot can have information about the whole dynamics of the interactions by monitoring only the flow of its own sensorimotor activity, without having any notion of the other. A single S neuron is summing the activities coming from the *sensorimotor* pathway. When S is connected to the $Input$ vector, the result is an activity representing the total intensity of the human stimulations. In all the experiments, $Input$ responses are the result of a competition applied to the visual detection. Therefore, S gives the undifferentiated activity of the successive inputs activated.

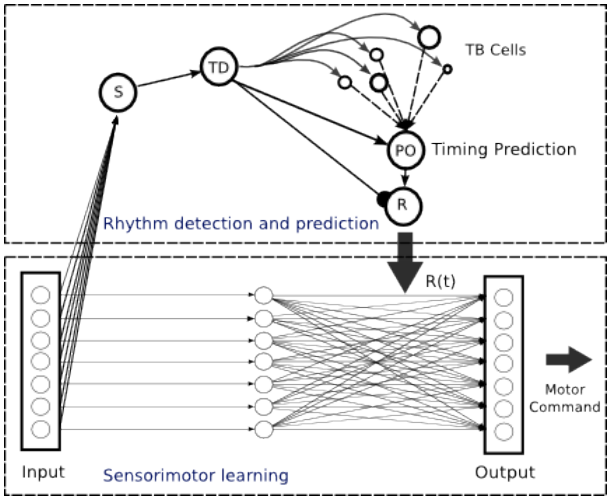

Fig. 1. Model for rhythm detection and prediction. The interaction dynamics is read through the sum (S) of the input stimulations ($Input$). $Input$-$Output$ associations are explored and learned thanks to R based on the difference between the prediction of the next stimulation (PO) and the effective detection of the current stimulation (TD)

The core of the timing prediction is composed of three groups of neurons (Fig. 1): Time Derivation (TD) group, Time Base (TB) group and Prediction Output (PO) group. A single TD neuron fires at the beginning of new actions (it detects S deviation). Therefore, TD detects the successive changes of activation of the $Input$ group. TB group decomposes the time elapsed between two TD spikes, with j cells responding at different timing and with different time span. It simulates the activities $Act^{\text{TB}}(t)$ of cells of different sizes to the same TD stimulation (such cells are known to be found in the Dentate Gyrus of the hippocampus [31]):

$$Act_j^{\text{TB}}(t) = \frac{m_0}{m_j} \cdot \exp - \frac{((t-\tau)-m_j)^2}{2 \cdot \sigma_j} \qquad (1)$$

where j is the number of the cell of TB, m_j and σ_j are the time constant and the standard deviation associated to the jth cell. τ is the instant of the activation of TB by one TD spike.

One PO cell learns the association between TB cells states (the time trace of the previous $Input$ stimulation) and the new TD activation (the detection of the current $Input$ stimulation).

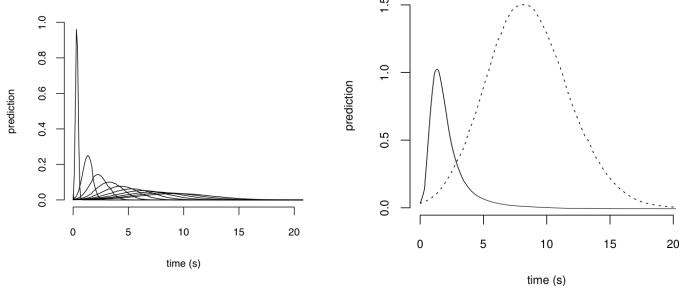

Fig. 2. Left: Activity of a group of TB cells measuring the time elapsed between two TD spikes. Here TB is stimulated at $t = \tau = 0$ and is composed of 12 cells. Each cell has a distinct set of parameters m_j and σ_j allowing the cells activities to overlap. TB plays the role of a short-term memory (here for 20 seconds) of one TD spike. Right: example of the shape of two PO predictions. The solid line is the prediction activity for a timing of 1,7 s. The dotted line is the prediction activity for a timing of 9s.

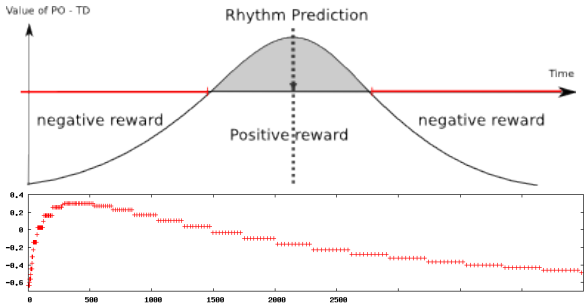

Fig. 3. Building the reinforcement $R(t)$. Up: illustration of the principle. The analogical activity of the prediction group PO peaks at the next TD period. By computing the difference between PO and TD when TD fires, we obtain the difference between the prediction and the effective rhythm. If the experiment maintain a constant rhythm, then PO will predict correctly and the TD-PO difference will be positive. At the opposite the more the experimenter changes the timing of its response, the more TD-PO will be negative. Bottom: Experimental record of the TD-PO values. A nominal rhythm is learned at period $= 450ms$, and the TD-PO response is recorded for each test period ranging from 0 to 5000 ms.

The strength $W_{\text{PO}}^{\text{TB}(j)}$ between PO neuron and TB battery is modified according to:

$$W_{\text{PO}}^{\text{TB}(j)} = \begin{cases} \frac{Act_j^{\text{TB}}}{\sum_j \left(Act_j^{\text{TB}}\right)^2} & \text{if } \ Act^{\text{TD}} \neq 0 \\ \text{unmodified} & \text{otherwise} \end{cases} \quad (2)$$

The potential of PO is the sum of the information coming from TD and the delayed activity in TB.

$$Pot^{PO} = \sum_j W_{po}^{TB(j)} \cdot Act_j^{TB} \quad (3)$$

$$Act^{PO} = f_{PO}\left(Pot^{PO}\right) \quad (4)$$

$$f_{PO} = \exp \frac{x^2}{2 \cdot \sigma^2} \quad (5)$$

After one learning (at least the presentation of two $Input$ stimulations), TB cells activation leads to the growing activity of PO until a maximum corresponding to the stimuli period (Fig. 2). The learning is done on-line, meaning that repetitive activations of the different units of $Input$ will lead to repetitive predictions of the timing of the next input by PO. The prediction itself is not linked to a particular $Input$ unit, but only to the precise timing at which the next $Input$ unit should be observed. At last, comparing PO and TD activities gives the success value of the rhythm predictor. This comparison is used to build the reward R for sensorimotor associations learning:

$$R(t) = PO(t) - \alpha \cdot TD(t) \quad (6)$$

with $PO(t)$ the value of the activity of PO in $[0, 1.5]$ at the instant t, $TD(t)$ the value of the activity of TD in $[0, 1]$, t the instant of TD spike, and $\alpha = 0.75$ in all experiments. The value of $R(t)$ is calculated when TD spikes, on the basis of PO state at time t .

For example, let us suppose that $Input$ is activated with a period t_{ref}. PO learns to predict the timing with a maximal activity at t_{ref}, and the value of $R(t)$ is maximal while the frequency of $Input$ activation stays unchanged. If, for some reason, the input frequency is changed to a period t_{new}, then the shape of PO ("bell" curve) will induce a decrease of $R(t)$ proportional to the shift between t_{new} and t_{ref} (see Fig. 3). If t_{new} is too different from the previous t_{ref}, a negative reward is produced (rhythm break detection) at the first occurrence of the new TD spikes. After the emission of the negative reward, t_{new} will be learned by PO (on-line learning), being the new reference of the interaction rhythm and the new period predicted. Letting PO learn on-line the period allows to obtain a system which suits to the changes of timing (see Fig. 4), without negative reward if the frequency is slowly changing and sliding during the interaction. Moreover, eq. 1 and 2 allow a one shot learning of the timing between two stimulations. It means that the rhythm is learned once the human and the robot have exchanged two gestures (approximately 3 to 5 seconds of interaction, depending on the human intrinsic period of interaction t_{ref}). In the frame of a human robot interaction, the rhythms can not be learned faster. During the interaction, the rhythm prediction is constantly updated according to the flow of the Input/Output information, at each new gesture. Of course, the rhythm prediction can be updated or progressively modified to take into account past value (averaging with a sliding window).

Such properties are welcome in the frame of a real human-robot interaction where the rhythm can slowly vary without negative impact until a strong break arises. When a negative reward is emitted (case of a strong rhythm break), PO is temporally inhibited until the next TD activation. This modulation of PO prevents the learning of the duration of the rhythm break itself and having two consecutive negative rewards.

One of the main interests of this NN algorithm is that PO prediction respects the Weber-Fechner law [31]: the standard deviation of PO activity is inversely proportional to the period of the timing learned. This property is particularly appropriated to the detection of various human-robot interaction rhythms, where for example a delay of 100ms must not have the same consequence if it happens in a interaction that is carried at 0.5 Hz *vs* 5Hz . The global shape of the reinforcement is

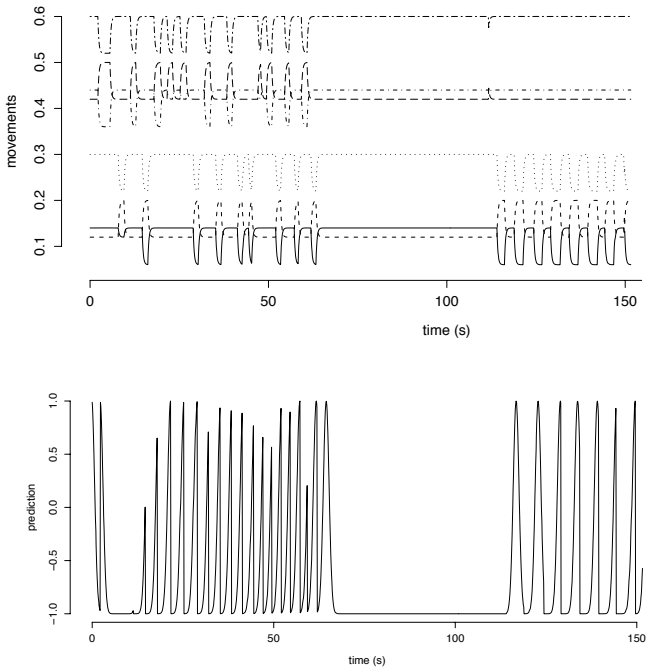

Fig. 4. Sample of a rhythmic interaction. Up: the user produces constant movements activating motor responses of the robot, from $t = 0$ to $t = 65$ at 0.3 Hz and from $t = 115$ to $t = 150$ at 0.15 Hz. Bottom: PO activity of showing the rhythm prediction. The neuron has adapted on-line the prediction of the rhythm

dependent of the PO activity, itself directly dependent of the incoming energy given by TB cells (eq 3). TB parameters play a direct role in the accuracy and the shape of PO activity. In the above experiments, the number of cells j was tested between 10 and 30 according to the precision of PO needed (j does not affect the global shape of PO activity but the resolution of the curve). m_j defines the center of each curve on the time course since the activation of the battery (at a relative time τ). It is important to maintain an overlap of the activity of each cell of TB, otherwise a strict separation would result in noisy and erroneous predictions (in this case the maximum of PO would not correspond to the learned timing). Practically, the value of m_j is proportional to j multiplied by a constant, and shifts the center of each curve along the time course. To ensure the overlap of multiples cells and the production of a "bell" curve of PO, σ_j is also linked to j. More precisely, σ_j is inversely proportional to the value of $j \cdot m_j \cdot constant$. This ensures that the first cells of TB have high standard deviation, resulting in a "thin" curve, while the last cells have a smaller deviation, resulting in larger curves. This variation of thickness of TB curves induce the variation of thickness of the PO curve at the origin of the Weber-Fechner property (see Fig. 2).

B. Sensorimotor Learning

The $Input$-$Output$ link (Fig.1, bottom) is the sensorimotor pathway of the architecture. $Input$ processes visual information from the CCD camera. The visual space of the robot is split into different areas stimulated by the gestures of the human. For seek of simplicity, the actions of the human are filtered by a color detector, and the human is asked to handle a pink ball (see Fig. 5).

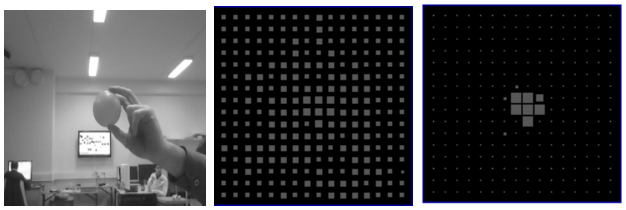

Fig. 5. Elementary Visual system. Left: image captured by the robot. Middle: detection of the red component. Right: detection of the pink color before competition at the basis of the $Input$ activity.

$Output$ triggers the robot's motor response. The activation of one $Output$ neuron induces the movement of the robot's effector toward a region of the working space. The number of $Input$ and $Output$ neurons is the same, and the division of the working space is the same on each layer. For example, $Input$ first neuron is activated when the upper right area of the visual space of the robot is stimulated, and the trigger of the first $Output$ neuron will induce the robot to raise the right arm upside.

In the frame of our experiments, $Input$ and $output$ are competitive layers forcing the activation of a sole $Input$-$Output$ association at a time. Both groups also have activity thresholds, allowing the robot to be still (no motor command) if no relevant input is provided (no significant visual activity). From these properties, we obtain an elementary but reliable basis for a non-verbal human robot interaction. Each time the human caregiver makes a move (stimulating $Input$ neurons) the robot will trigger a motor response. $Input$-$Output$ links are initially random, and in the frame of our experiments, the robot will have to discover and learn the right $Input$-$Output$ associations corresponding to the caregiver expectancies. Therefore, finding the good responses corresponds to discovering the right sensorimotor associations among the possible ones. To do so, a learning rule using reward R computed by the *Rhythm detection and prediction* Layer (section III-A) is introduced.

The rule learning $Input$-$Output$ associations is directly inspired from the *Associative Search Element* (*ASE*) from Barto, Sutton and Anderson [32]. *ASE* uses a "*generate and test*" search process in order to find the correct outputs. To simplify the equations, the letters I and O stands for $Input$ and $Output$ groups.

$$O_i = H(\sum_j W_{ij} * I_j + noise) \tag{7}$$

With H the Heaviside function and $noise$ the amount of additional random noise. At the beginning of the experiment the W_{ij} have the same value and the $noise$ value is driving the selection of O_i neurons. It ensures the initial exploration of the search space before any learning. When an output is tested, the correction of the weight is made according to the variation of $R(t)$ and $O_i(t-1)$:

$$\Delta W_{ij}(t) = \alpha \cdot \Delta R(t) \cdot \Delta O_i(t-1) \cdot O_i(t-1) \cdot I_j(t-1) \tag{8}$$

with $\Delta x(t) = x(t) - x(t-1)$ and W_{ij} the weight between $Input_j$ and $Output_j$ updated at time t. For the Input I_j activated at $t-1$, the network tests the output O_i at $t-1$ and obtain the reward R at t. This learning rule results in a modification of the $Input$-$Output$ link according to the variations of $R(t)$ and $Output(t)$, Fig. 6.

$I_i(t)$	$\Delta O_i(t-1)$	$\Delta R(t)$		$\Delta W_{ij}(t)$
0	.	.		0
.	0	.		0
.	.	0		0
1	↗	↗		↗
1	↗	↘		↘

Fig. 6. Variation of W_{ij} according to eq. 8. As shown in eq. 7, O is binary, which explain the limited number of cases: only the weights with $O_i(t-1) = 1$ are updated

Therefore, constant on-line updates of the rhythm prediction produce a new R value at each interaction step (that is to say at each new human-robot exchange of gesture), and to change the $Input$-$output$ weights accordingly. Doing so, it is important to notice that the convergence time is strongly dependent of the interaction cadency, the value of R being updated at each exchange according to each new rhythm prediction.

IV. Experiments

In this section, we describe three kind of experiments:

1) A simple human-machine interaction based on key-sound associations. The experimenter has to teach the computer to associate sounds with keystrokes. In this experiment, the human is expert, and the goal is to demonstrate the convergence of the model with a growing number of $Input$-$Output$ associations
2) A human-robot interaction with a Sony Aibo robot, testing gestures-gestures associations. The human has to teach the robot to respond with the right gesture among 4 possible ones, according to 4 human gestures. In this experiment we tested expert and non-expert users in the frame of very limited instructions.
3) A human-robot interaction with a small humanoid Aldebaran Robotics Nao robot. The subjects are involved in the same gesture experiment as experiment 2, and we also tested two groups of subjects, experts *vs* non-experts. Taking into account the results of experiment 2, we propose to enhance the learning rule to take into account a delayed reinforcement and we introduce a confidence value to enhance the stability of the robot's associations testing.

In all the experiments, we have defined expert and non-expert subjects as follow:

- An expert user is a subject that knows that the rhythm of the interaction has an effect on the robot learning. It concerns people from the lab aware of this research, and also peoples naive to robotics that are instructed before the experiment that they can change the timing of the interaction or break the rhythm in order to change the robot's behavior.
- A non-expert user is a subject that is not aware of the robot's functioning, and that has for only instruction the one given to each experiment, without additional information.

A. Human-machine interaction: the sound experiment

In order to test the feasibility of our model, we have established a human-computer interaction setup.

1) Setup: The setup is designed as a game where the expert must teach the computer to produce a given note when he hits a given key. As presented in section III, the architecture must learn to associate key strokes with sounds, through the interaction (see Fig. 7).

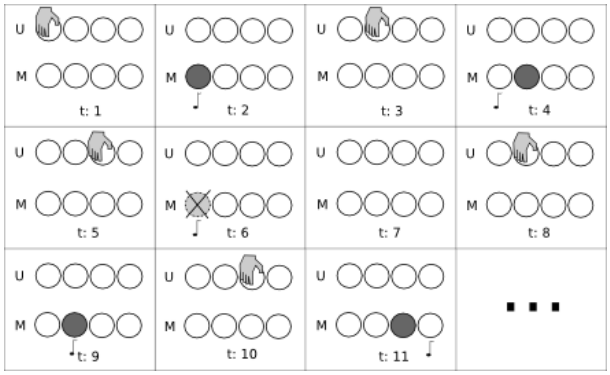

Fig. 7. Principle of the human-computer interaction. The user (U) pushes a keyboard key and expects a given note in response. From t_1 to t_4 the machine (M) responds correctly, so the user continues the interaction and tests new keys. At t_6 the note emitted is incorrect, and the user interrupt shortly the interaction (or at least change noticeably the timing) to restart the interaction, or for example express his disagree before going on.

2) Results: We have tested the network with sizes of $Input$ and $Output$ groups going from 2 neurons to 10 neurons. In all tests, the number of neurons in $Input$ and $Output$ is equal, meaning that the search rule had to discover the right set of n associations in a set of n^2 possible associations. Figure 8 shows the average time for discovering the right set of associations according to n (Fig. 8, up) and the progress of the on-line learning during one experiment, with $n = 10$ (Fig. 8, bottom).

The system succeeded in learning the set of correct associations. Obviously, we see in Figure 8, up that the convergence time of the architecture is strongly linked to the combinatory of the exploration space before discovering the right associations. For example the experiment presented in Figure 8, bottom took approximately 7min of interactions before fully converging. The network is running permanently, but it is important to notice that the learning rule can only be triggered after each update of the interaction cycle. Therefore, the time taken to converge is not a limitation of the algorithm, but a limitation imposed by the user's cadence taking the time to interact every 2 or 3 seconds in the frame of such a keyboard-human game. If we take for example the sequence of 10 elements, it means that the interaction has to go through the test of the 100 possible input-output associations. With an interaction cycle of 3 s, this gives us a "theoretical" total interaction duration of 300 seconds, without taking into account the additional time taken

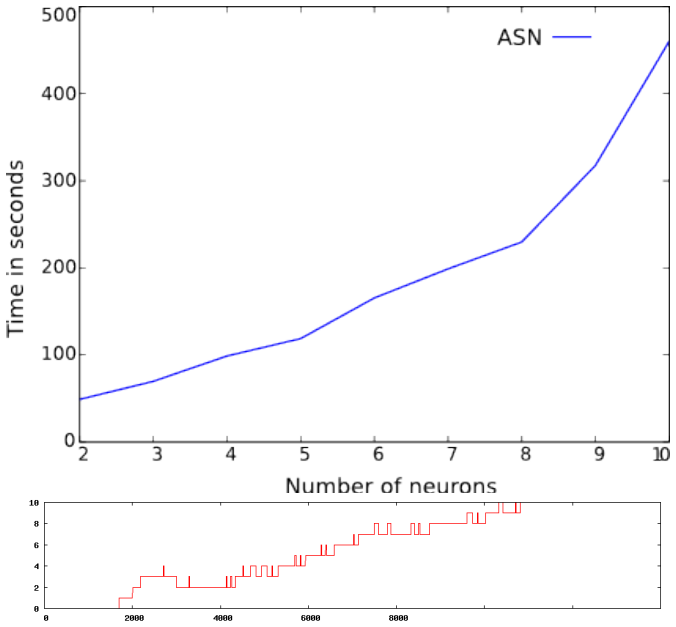

Fig. 8. Mean convergence time for discovering and learning n associations in a set of n^2 possible ones through a keyboard-sound interaction and using the rhythm of the interaction and a classical reinforcement rule. Up: mean time in seconds to converge to the right set of associations according to n. To each value of n corresponds the mean value of a set of 5 tests conducted with the same experimenter. Bottom: learning progress during one experiment consisting in the discovery of 10 key-sound associations among 100. Each association discovered gives 1 point, and the convergence takes about 11000 computation steps (see text for more details about convergence time).

to break the interaction. Our algorithm converges in a mean time of 445 seconds of intereaction.

The choice of this first test (under the form of a human-computer interaction using sounds, for seek of simplicity) allows the human to quickly and intuitively react to the sound emitted by the computer. With such a simple setup, all the tests where done with an expert user, that is to say a user confident with the interface, and who knows that he can use the rhythm to control the learning. The objective of this experiment was to show that the convergence of a classical reinforcement rule was possible when combined with a neural network rhythm predictor learning on-line, even with a large amount of possible sensorimotor associations (here, up to 100 associations). "Large " must be here understood as an important set of associations to be explored in a limited time of interaction. In all the experiments, the first association have always been the most difficult to discover. While none of the responses of the system are those expected by the user, it is difficult to establish a default rhythm of interaction. Conversely, once a first association is discovered and learned, the user tends to use it to support the exchanges, installing a default rate with the system before testing the new keys, and so on.

B. First Human-robot experiment

1) setup: For this experiment, we have used a Sony Aibo robot. The robot was seated on a table in front of the human (see Fig.9 for an illustration). The robot controlled four degrees of freedom (the elbow and the shoulder of each arm), and the head was still. The control architecture was implemented with four different actions: raising the left foreleg, raising the right foreleg, lowering the left foreleg or lowering the right foreleg. The task was a "mirror action" game where the robot has to learn to make the same action as the human.

Fig. 9. Setup of the a "mirror action" game with Aibo.

The robot must produce the same action, on the same side, as the actions done by the human (human and robot being face-to-face). During the game, the robot has to explore its motor repertory when stimulated by the human, and discover and reinforce the mirror gestures of the human's ones. Consequently the robot had to discover the 4 associations corresponding to the imitative actions among 16 possible ones.

2) Instructions: The subjects where told to "teach the robot to do the same action". To illustrate the instructions, a little program was run to show the 4 possible actions of the robot using the two upper legs. No additional instruction was given to the subject (if the robot is able to process speech, face, etc...). Such a setup was done to easily engage the human to lead the interaction, and also to give as less as possible cues about the functioning of the robot, especially for non-expert users. The human proposes a gesture and wait for the robot's response, but he has no access to any special interface indicating to the robot if the response was good or bad. At the end of the experiment, we questioned the subject about the quality of the interaction, the teaching procedure they used, and if they could understand how the robot was learning.

3) results: experts: For all expert subjects, the learning of the 4 *Input-Output* associations was completed in a mean time of 4min of interactions. Figure 10 shows the result of one experiment conducted with an expert according to the setup described in the previous section. The correct learning of the 4 associations took approximately 3 minutes[1]. As in the sound experiment, the experts found the first association the most difficult one to teach because they have no basis to trigger a constant rhythm and therefore no basis to control the reinforcement. Most of the experts, and especially the experts naive to robotics (but instructed to use the timing to change the robots learning) also felt the robot's action choices as inconsistent during the experiment. The interaction

[1]a video of this experiment can be seen on http://www.youtube.com/watch?v=7OQK_NjX8D8

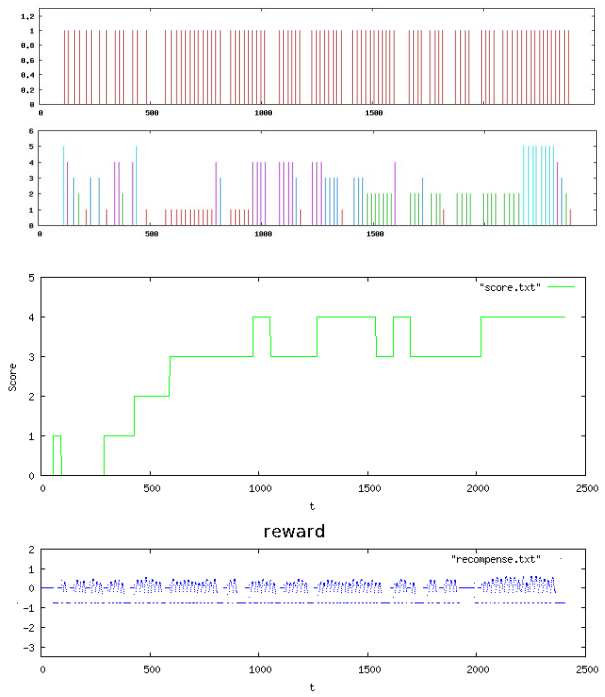

Fig. 10. Expert-robot experiment with Aibo. First Line: Activity of TD detecting the beginning of the human actions. Second Line: responses of the robot labeled from 1 to 5. Third Line: evolution of the score of the robot. Each correct association (mirror action) gives one point. Fourth Line: reward built by the robot, oscillating between positive and negative values according to the rhythm of the interaction.

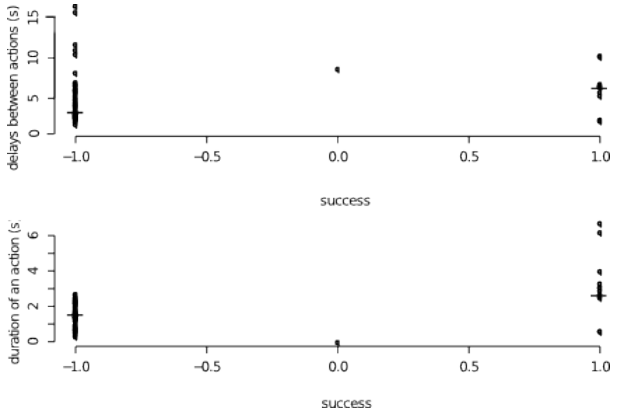

Fig. 11. Record of the delays and durations a non-expert's actions during a 3.5 min interaction with Aibo. Delays and duration are component determining the time between two actions, and therefore the rhythm of the interaction. For all experiments, we recorded the delays and durations (from the robot's perceptions) of the actions responsible for the reward attribution. On the left side ares values of delays and durations corresponding to negative rewards vs the positive ones on the right. Mean delay and duration for negative rewards are 3.2s and 1.7s vs 6.1s and 2.3s for the positive rewards. Similar distributions where found for each of the 5 non-experts.

was found to be "disturbing", or "unstable", because of the robot changing randomly the exploration of the sensorimotor associations. This behavior of the robot is directly linked to the $noise$ parameter added to the activity of O_i. If this variable is fundamental in the exploration of the search space by the algorithm, the corresponding robot's behavior is felt as inconsistent by the experimenter, changing from time to time the action, "for no reason". This situation particularly happens at the beginning of the experiment, when $|R|$ is too week to induce strong changes of the W_{ij} letting the $noise$ value decide the winning output.

4) results: non-experts: None of the 5 non-expert subjects succeeded to teach the robot the 4 $Input$-$Output$. The best score was the teaching of 2 associations (population mean score: 1.4 taught association), with a mean interaction time of 5.2 minutes. When questioned, none of the subjects detected that the timing of the exchanges played a role in the learning of the robot. When combining data record (mainly the robot's inputs) and responses of the subject to an interview about the quality of the interaction, we can formulate the following comments:

- As for the expert subjects, the non-experts felt the robot behavior as inconsistent, changing randomly the actions. A more "stable" behavior in the responses of the robot would have helped the subject to take the time to work on a given sensorimotor association, testing it iteratively and giving more chance to install an interaction rhythm.
- There is a strong overlap between the duration of actions and reaction of the non-experts whatever this duration is linked to a positive or a negative reward (Fig. 11). In this experiment, the variance and overlap for each category shows that there is no obvious distinction between the timing of successful actions and incorrect ones. We can explain the variance of positive actions as follow: cadences are not always the same, and the period of a constant satisfying interaction can slide during the experiment. This effect is taken into account by our algorithm that adapts to small successive variations of the rhythm. We can explain the variance of negative actions as follow: when observing the experiment and the inputs of the robot (reflecting the timing of the subject's actions) non-expert users often carry-on the interaction at a constant rhythm when the robot is wrong, waiting two or three wrong actions to make a strong break. This explains that the duration of negative actions and reactions overlaps with the positive ones. Such information allows us to think that if an algorithm using immediate reinforcement is suited for experts (they immediately break the rhythm after a wrong answer) it may not be the case in the frame of non-expert users, waiting a mean of 2 or 3 wrong responses before breaking the interaction. In this case, our algorithm should also take into account the delayed rewards, for example to distribute the effect of a negative reward on the 2 or 3 past associations.
- Finally, and more generally it is important to mention that experiments lasting more than 4min were found to be tiring by the non-experts. We did not conduct an intensive research on how to identify precisely the time during which a human could accept to interact with such a setup. The number of variables is too important (it would require a strong analysis of the experimental condition) and such a study is out of the scope of our competences (the human being in the loop, it would require a psychological study about the human acceptance of interaction with such artificial systems). Nevertheless, these naive but recurrent

observations of tired or bored reactions after 4 or 5 minutes of interaction let us suggest that it is a limit for studying a "one shoot" interaction with our autonomous robots (being given the entertainment level of our robots).

C. Second Human-robot experiment

To investigate the limitations of the first set of robotic experiments, we tested a second learning rule, in order to (1) take into account a delayed reward, for example if the break of the rhythm does not only concern the precedent $Input$-$Output$ association and (2) to give to the robot the possibility to maintain the associations longer in order to test more strongly if a rule must be learned or unlearned.

1) The PCR rule: The Probabilistic Conditioning Rule (PCR) [33] was proposed to allow mobiles robots to learn the correct set of perception-action associations when receiving delayed reward. A typical task solved by PCR was to allow a mobile robot to learn the right actions to do (go ahead, turn left or turn right) in order to escape from a maze. Each branch of the maze was recognized by the robot using vision (the recognition of the places being the inputs of the learning network), and the robot received negative rewards each time it bumped into a wall, and a unique positive reward when escaping from the maze. The rule was designed with the following properties:

- the robot must be able to test a given number of hypothesis during a given amount of time before deciding to learn or not (for example always turning left when sensing a wall close to the front sensors).
- the robot can manage delayed reward, thanks to a memory of the associations that where selected since the last reward (for example rewarding the robot only at the exit of the maze).

Whereas this learning rule was initially designed to solve tasks of mobile robotics with external motivations, we found interesting that its properties should fit with some of the issues of our first robotic application.

In order to be able to reward past associations, PCR uses a correlation $\overline{X}$ over a sliding window τ in order to keep a short term memory of the activated inputs ($\overline{I_i}$) , outputs ($\overline{O_j}$), and associations ($\overline{IO_{ij}}$) with:

$$\overline{X_j[t+1]} = \frac{\tau \overline{X_j[t]} + \overline{X_j[t]}}{\tau + 1} \tag{9}$$

In order to be able to test a given association longer, a *confidence* value (p_{ij}) is introduced. This value represents how the corresponding association is trusted by the architecture. A high value of p_{ij} indicate that the corresponding association between I_i and O_j should not be changed. Otherwise a weak value of p_{ij} indicate that the association must be changed because not rewarding. Consequently, R is used to change the confidence value of associations:

if $|\Delta R(t)| > \xi$ then the confidence value p_{ij} is updated:

$$\Delta p_{ij} = (\epsilon + \alpha * \Delta R) * C_{ij} * f_B(W_{ij}) - \lambda * p_{ij} \tag{10}$$

$$p_{ij}[t+1] = H(p_{ij}[t] + \Delta p_{ij}[t]) \tag{11}$$

With $R(t)$ the global reinforcement signal from PO, ϵ the learning speed, α the reinforcement factor, λ a forget factor, ξ the reward threshold. $f_B = -1$ if $W_{ij} = 0$, 1 otherwise.

Moreover, C_{ij} is the STM memory of the past associations calculated with eq. 9 and:

$$C_{ij} = \frac{\overline{IO_{ij}}}{\sqrt{\overline{I_i O_j}}} \tag{12}$$

At each iteration, the confidence of each association is tested. If a random draw $Rand$ is higher than the confidence, then the weight value and its confidence value are inverted:

if $Rand > p_{ij}$ and $\overline{I} * \overline{O} \neq 0$ then

$$\begin{array}{ll} W_{ij} & = 1 - W_{ij} \\ p_{ij} & = 1 - p_{ij} \end{array} \tag{13}$$

The less the confidence is, the higher the probability to change the rule is. The mechanisms of confidence inversion is also very important, allowing also punctual tests of new rules even when all the associations have a strong confidence. For example, there is always a very small probability to change an association that is at 0.99 of confidence. If this case happens, the benefit of the past learning (that has driven the rule to a 0.99 confidence) is not lost: the new rule will only have 0.01 value of confidence, meaning that it should come back rapidly to the previous association. Using this mechanism the network is able to test punctually new hypotheses before changing for sustainable responses.

Finally the output is selected according to:

$$Act_j = Max_i(((2 * W_{ij} - 1) * p_{ij} + 1) * I_i) + noise \tag{14}$$

with

$$O_j = \begin{array}{ll} 1 & \text{if } Act_j = Max_k(Act_k) \\ 0 & \text{otherwise} \end{array} \tag{15}$$

2) setup: Figure 12 illustrates the experimental setup. Nao is seated on a table in from of the user.

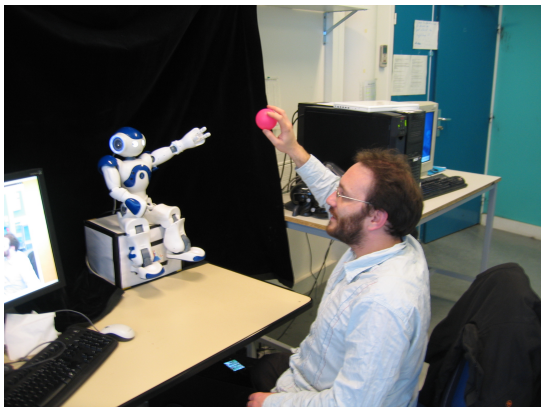

Fig. 12. Setup of the "mirror action" game with Nao. The robot is seated in front of the user. As in the Aibo experiment, a pink ball mediates the interaction, but more freedom is given to the experimenter thanks to a tracking algorithm of the ball.

The head, torso and arms are freed. We added a simple algorithm allowing to track the pink color with the head (2

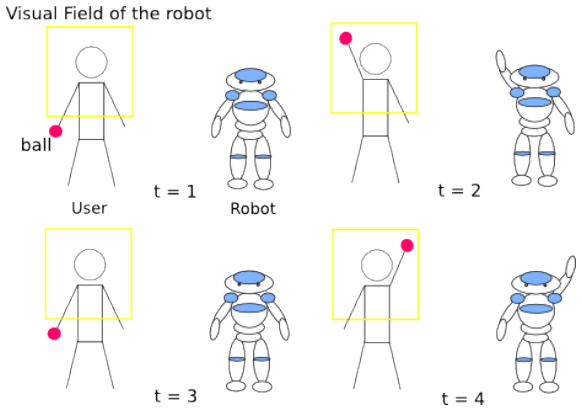

Fig. 13. Illustration of the instructions given to the experimenter: "try to teach the following actions to the robot: when you raise the left arm, the robot should respond like this, etc..."

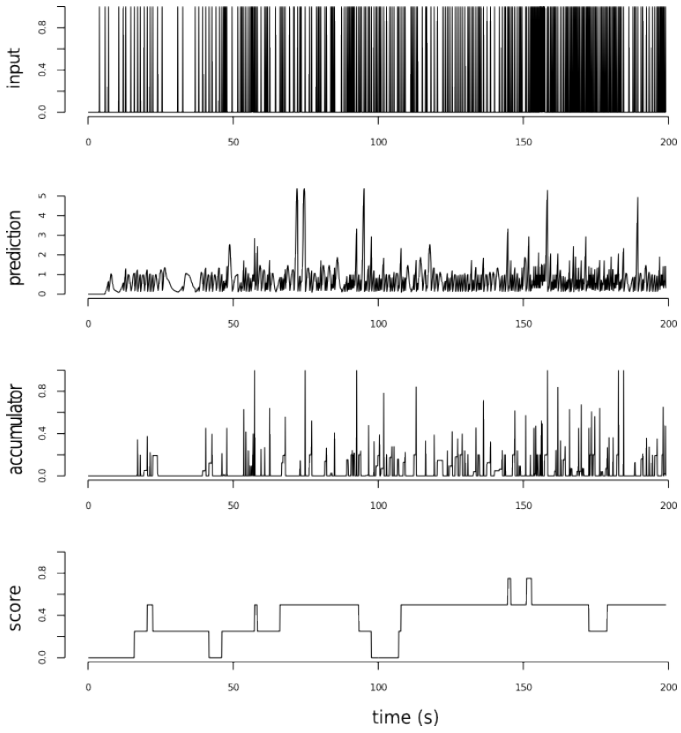

Fig. 14. Typical interaction with a non-expert. The subject alternates the strategies, using sometime a constant rhythm to strengthen the interaction with breaks when the robot does not respond correctly (from $t = 25$ to $t = 170$ seconds) and sometime a time independent behavior (just seeking the possible actions of the robot, for example from $t = 0$ to $t = 25$ seconds). After $t = 170$ the subject starts to being bored and progressively gives-up the game: he acts very rapidly, trying to "force" new responses from the robot. During this experiment the robot discovers 3 of the 4 correct associations among the 16 possible ones

degrees of freedom, pan and tilt axis). This enhancement allows Nao to follow the pink ball when the experimenter moves it. The tracking behavior gives more freedom to the experimenters moves (allowing movement with a large amplitude while being seated at approximately 1 meter from the robot) and also provides an important feedback to know if the robot is looking at the target. Nevertheless, if the experimenter decides to hide the ball from Nao's field of view, the head comes back to a central position. A side effect of the pink tracking induce sometimes that Nao looks at the head of the experimenter (depending on the skin tone) if no other pink stimulus is present in the field of view. We decided to keep this side effect, because (1) it does not change the tracking performance (the pink ball is always the winning stimuli) and (2) it adds a coherent interactive behavior to the robot (the robot is looking at the head of the experimenter if the ball is hidden, that is to say when there is nothing to do).

3) instruction: As in the previous experiment, we tested two kinds of users: experts and non-experts (as described in section IV-A and IV-B). Both types of subjects are instructed to use the pink ball to teach the robot to make the same moves, and were shown the illustration of Figure 13. All subjects are free to interact with the robot as long as they wish. Subjects were instructed that they can, if they wish, talk to the robot, but none were informed if the robot was able to process or understand what the subject was saying. No other oral or written instructions were given to the subject. The group of experts was composed of 10 subjects, and the group of non-experts was composed of 8 subjects.

4) results: experts: For the experts, the results are similar to the previous Aibo experiment. The network converges toward the imitative associations with a convergence speed similar to the ASE inspired algorithm (with a mean time of convergence of 4-5mn). The use of an algorithm able to cope with delayed rewards did not change the convergence time, since experts have a tendency to immediately break the rhythm after robot's wrong responses and therefore providing immediate feedback to the system. According to the subject's interview, the quality of the interaction had been noticeably improved thanks to the use of the confident value on the weights. When Aibo was perceived changing randomly, with no reason the actions, Nao is perceived far more stable in testing the same set of actions, before changing the weights. In the frame of expert subjects, using the rhythm of the interaction appear to be an efficient mean of teaching the robot to learn new associations. The interaction is free, without use of any programming skills or debugging information. Interestingly, when parametrizing the robotic setup, we have noticed that trying to use a debug (display of the activity of the neurons) to explicitly see when the predictions fires in order to manage our actions and test the rhythm break detection brought poorer results than naturally interacting with the robot.

5) results: non-experts: Results with non-experts were found to be in progress compared to the Aibo's experiment. None of the non-experts managed to achieve the learning of the 4 associations, all subjects obtained at least a score of 2 correct associations, and 6 subjects obtained a score of 3 correct associations (mean group score: 2.8). Experiments were lasting between 3 and 6 min (mean time of duration: 4 min 20s). The reason for not obtaining perfect scores can be explained by the fact that non-experts often use different strategies in the same experiment. As shown in Figure 14, illustrating the course of a whole experiment, the subjects can combine multiples strategies in order to teach and test the robot. Among the strategies used by the subject, we can identify 3 ones that have been often observed :

- The first strategy is consistent with our working hypotheses: the subject have a stable rate of actions when the robot responded correctly and breaks the rhythm once the robot starts to make the wrong actions, or a succession of 2 or 3 wrong actions. This strategy was observed as a part of all subject's interaction. It was more used among the subjects that did chose to speak to the robot, accompanying the correct actions with positive words during the movement, while taking the time to speak (and break the interaction) when responses were wrong. Nevertheless, if this strategy was part of the interaction, it was never the sole one, explaining why the subjects never succeed to reach a score of 4
- We observed subjects simply "scanning" the possible responses of the robot without changing the action's rate. The result of such a behavior, is that the architecture generates positive reward for all $Input$-$Output$ associations tested during this "scan", whatever they are correct or not. Such a strategy affects strongly the interaction, since it results in lowering the score and producing wrong associations that will have to be unlearned before the robot proposes new ones. Interestingly, it appear that such a strategy is also consistent with our initial hypotheses: the subject simply wanted to check if the robot responds to every possible stimulation, and the fact that the robot responds by an action -whatever the action itself- is precisely the behavior expected by the human, therefore continuing the check rhythmically (but reinforcing wrong $Input$-$Output$ associations).
- A last strategy was often observed at the end of the experiment (most often after 4 to 5 min of interaction), when the subject started to get bored by the interaction and the fact that not all the associations could be learned. In this case, the subject simply accelerate strongly the rhythm of the exchanges, in order to go as fast as possible to see if at some point, "by luck" the robot will perform the correct actions.

Table I summarizes the mean proportion of each strategy observed among the 8 non-experts. This table is based on a combination of visual observation of the experiment, subjects interview and a record of the robot' s inputs reflecting the cadency of the subject's actions. The proportion of the expected behavior represents 38% of the interaction. The subject speaking to the robot has a higher proportion of this behavior, corresponding to a mean of 56% of the interaction (observed with 3 subjects). If the proportion of the "scanning" behavior is limited, such a behavior has a strong negative effect on the course of the interaction, as explained above. The strong proportion of "other" behaviors is explained by parts of the interaction that we were not able to identify, and transitions between identified behaviors. All these data suggest that rhythmic components are present and can be exploited at the same time to improve the learning of the robot, and to enhance the quality of the interaction. Nevertheless, solutions strictly predicting the rhythm of the interaction cannot be the only ones guiding the robot's learning, as shown by the non-expert score. Of course, the performance is highly dependent

TABLE I
MEAN PROPORTION OF THE DIFFERENT STRATEGIES OBSERVED DURING THE NON-EXPERTS' EXPERIMENTS.

strategy	expected	scanning	bored	other
percent	38	12	20	30

of the instructions given to the subjects. For example, an instruction such as "you have to talk to the robot when its behavior is wrong" is a way to guide the subjects toward an appropriate behavior for such algorithms. In the present paper, we decided not to give such instruction, especially instructions that would have constrained the subject's behavior in case of wrong responses of the robot. We wanted to obtain an evaluation about the usability of rhythm's prediction in the frame of the less constrained interaction as possible. With the same idea, we have chosen to limit the robot's behavior, but improvement could lead to enhance the results. For example, some component of the behavior (such as the legs balancing, or changes in the eyes color) could indicate that the robot perceives a rhythm and provide a behavioral carrier to the caregiver, that is to say, turn the interaction in a bi-directional exchange. To illustrate the perspectives of using the rhythm of the interaction as a tool for human-robot interactions, we propose in the next section to give a brief description of an on-going research where the rhythm detection plays a different role.

V. PERSPECTIVES IN HUMAN-ROBOT INTERACTIONS

This research is about the learning and recognition of facial emotional expressions in the frame of a face-to-face interaction with a robotic head [34]. As in the previous experiments, the robot has no notion of others. The robotic head is simply expressing, alternatively, 5 different expressions (due to an arbitrary variation of its internal state). The control architecture is composed of an advanced vision system, processing the detection of feature points of the visual flow and the categorization of the position (where) and the local content (what) of the selected features points. The robot is then able to associate the result of the visual detection and categorization with the current expression or the current internal state triggering the current expression. If a human comes and starts to imitate the robot, then he closes the loop of the interaction. The human returns a visual equivalent of the robot's own "muscular" facial expression. Such an experiment shows that a categorization of the visual emotional expression is possible with an unconstrained vision system that just learn the properties of the features points of *all* the scene. Once the salient features of the scene are extracted and associated with the corresponding internal emotion, the robot is able to respond to the human with the same emotional expression. The roles are reversed and the robot is now able to recognize and imitate the facial expression of the human, without knowing what a human, a face or an agent is. But to be able to ground the visual categorization, the robot must be able to learn only when the human is interacting. And as mentioned above, the robot have no notion of what is a face or a

human. An elegant solution resides in the use of the interaction rhythm to provide the information that the robot is effectively interacting with a human face. In such an experiment, the rhythm detection and prediction mechanism is exactly the one described in section III-A. Because the robot is imposing the rhythm of the interaction, the imitating human is producing a rhythmic response that can be perceived by the robot (just a computing of the sum a visual movement detected). The rhythm prediction plays the role of a modulation, supporting the categorization of the visual field when the movement detection coincide with PO activity. Moreover, the rhythm information can then be the trigger of the learning of the difference between "face" and "non-face" visual categories, that is to say between "rhythmic" and "non-rhythmic" visual stimuli ("rhythmic" referring here to the stimuli that is locked on the rhythm of the robot actions).

VI. Conclusion

In this paper, we have presented a NN model able to learn and predict the rhythm of an interaction. Following a bottom-up perspective, the input of the rhythm detection is the sensorimotor pathway of our architecture. By monitoring its own sensorimotor dynamics, the autonomous robot has access to the rhythm of the whole interaction it is part of. This principle allows artificial systems to learn from human behavior and uses social cues without any notion of what a human is. We have shown that a combination of rhythm detection and classical learning rules allows a familiar user to drive the learning of the robot's sensorimotor associations without using any explicit setup or ad-hoc procedure to reward the robot. We have showed that the use of a more elaborated learning rule taking into account (1) a delayed reward, and (2) the sustain of the tested associations, increases the quality of the interaction. We have also tested our architecture with non-expert subjects, showing that our working hypotheses were partly covering the strategies used by the human. If rhythm variation is not the only behavior used by non-experts, it is nevertheless a mechanism that stays present in the interaction. Finally, we have suggested and illustrated another use of rhythm detection allowing a robotic head to detect that a human is interacting, thus triggering the categorization of the visual stimuli. All theses experiments show that taking into account the timing of the interaction is a necessary condition to develop adaptive skills from the social context. The timing is one of the sole *amodal* dimensions of the social interaction. Therefore, it can underly the interaction and the use of all the other modalities. If our experiments with non-experts show that such a property is not sufficient to allow an optimal learning or framing of the interaction, rhythm prediction seems to be a promising tool that needs to be combined with other learning strategies, and especially learning rules exploiting the other modalities.

Prepin et al. have recently proposed that rhythm and synchrony play the role of *phatic* signals of the pre-verbal interaction [35]. The authors highlight that the term phatic, "(...) usually used in a verbal context to define peri-verbal signals whose function is to structure and regulate the interaction, is also adapted to the non-verbal interaction where these signals are at the same time exchanged signals and regulation signals (...)". [2] In future work we wish to extend the role of the rhythm as a regulation signal. Our next working hypothesis will be to show how predicting the rhythm can play the role of an internal signal, for example generating positive (correct predictions of the timing) *vs* negative feelings (wrong predictions). These positive or negative internal values could be expressed by the robot and lead to changes of the human's behavior. Thus, changing the rhythm could modify the internal state of the robot, internal state whose expression would subsequently change the human's responses and rhythm. Testing such a model could be an interesting step toward the establishment of locked *vs* unlocked phases of interaction, allowing a progress toward the establishment of turn taking and/or role switching.

Acknowledgment

The authors wish to thanks gratefully Nicolas Garnault for the work done programming Abo and Nao, and Sofiane Boucenna for his participation to section V. This work was supported by the French Region Ile de France, the Institut Universitaire de France (IUF), the FEELIX GROWING european project FP6 IST-045169 and the French ANR Interact project.

References

[1] A.Meltzoff and W. Prinz, Eds., *The Imitative Mind: Development, Evolution and Brain Bases*. Cambridge: Cambridge University Press, 2000.

[2] J. Nadel, K. Prepin, and M. Okanda, "Experiencing contigency and agency : first step toward self-understanding ?" *Interaction Studies*, vol. 2, pp. 447–462, 2005.

[3] P. Rochat, "Five levels of self-awareness as they unfold early in life," *Conscious. Cogn.*, vol. 12, pp. 717–731, 2003.

[4] T. Kuriyama and Y. Kuniyoshi, "Acquisition of human-robot interaction rules via imitation and response observation," in *The 10th International Conference on the Simulation of Adaptive Behavior*, 2008.

[5] C. Breazeal, "Regulation and entrainment for human-robot interaction," *International Journal of Experimental Robotics*, vol. 10-11, no. 21, pp. 883–902, 2003.

[6] H. Kozima, C. Nakagawa, and H. Yano, "Can a robot empathize with people?" *Artificial Life and Robotics*, vol. 8, no. 1, pp. 83–88, September 2004.

[7] M.Masahiro, "On the uncanny valley." in *Proceedings of the Humanoids-2005 workshop: Views of the Uncanny Valley.*, Tsukuba, Japan., December 2005.

[8] M. Lungarella and L. Berthouze, "Modelling embodied cognition in humans with artifacts," *Adaptive Behavior*, vol. 10, pp. 223–241, 2003.

[9] J. Tani and S. Nolfi, "Learning to perceive the world as articulated: an approach for hierarchical learning in sensory-motor systems," in *From Animals to Animats: Simulation of Adaptive Behavior SAB'98*, R. Pfeifer, B. Blumberg, J. Meyer, and S. Wilson, Eds. MIT Press, 1998, pp. 270–279.

[10] P. Fitzpatrick., "First contact: an active vision approach to segmentation," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vagas, Nevada, 2003, pp. 27 – 31.

[11] E. Torres-Jara, L. Natale, and P. Fitzpatrick., "Tapping into touch," in *The Fifth International Workshop on Epigenetic Robotics, EPIROB'05*, Osaka, 2005.

[12] A. Barto, S. Singh, and N. Chentanez, "Intrinsically motivated learning of hierarchical collections of skills." in *In Proceedings of the 3rd International Conference on Development and Learning (ICDL 2004)*, Salk Institute, San Diego, 2004.

[13] P.-Y. Oudeyer and F. Kaplan, "Intelligent adaptive curiosity: a source of self-development," in *Fourth International Workshop on Epigenetic Robotics*, 2004, pp. 127–130.

[2] English translation of [35], the reference being available in French only.

[14] P.-Y. Oudeyer, F. Kaplan., and V. Hafner, "Intrinsic motivation systems for autonomous mental development," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 2, pp. 265–286, 2007.

[15] A. Pitti, M. Lungarella, and Y. Kuniyoshi, "Synchronization : Adaptive mechanism linking internal and external dynamics," in *Proceedings of the sixth international conference on Epigenetic Robotics, EPIROB'06*, no. 128. Lund University Cognitive Studies, 2006, pp. 127–134.

[16] K. Dautenhahn, "Getting to know each other - artificial social intelligence for autonomous robots," *Robotics and Autonomous System*, vol. 16, no. 2-4, pp. 333–356, December 1995.

[17] P. Gaussier, S. Moga, M. Quoy, and J. Banquet, "From perception-action loops to imitation processes: a bottom-up approach of learning by imitation," *Applied Artificial Intelligence*, vol. 12, no. 7-8, pp. 701–727, Oct-Dec 1998.

[18] P. Andry, P. Gaussier, S. Moga, J. Banquet, and J. Nadel, "Learning and communication in imitation: An autonomous robot perspective," *IEEE transactions on Systems, Man and Cybernetics, Part A*, vol. 31, no. 5, pp. 431–444, 2001.

[19] A. Blanchard and L. Canamero, "From imprinting to adaptation : Building a history of affective interaction," in *Proceedings of the Fifth International Workshop on Epignetic Robotics (EPIROB05)*, July 2005, pp. 42–50.

[20] K. Prepin and A. Revel, "Human-machine interaction as a model of machine-machine interaction : how to make machines interact as humans do," *Advanced Robotics. Section Focused on Imitative Robots (2)*, vol. 21, no. 15, pp. 18–31, Dec 2007.

[21] A. Blanchard and J. Nadel, "Designing a turn-taking mechanism as a balance between familiarity and novelty," in *Proceedings of the Ninth International Conference on Epigenetic Robotics (EPIROB2009)*, November 2009, pp. 199–201.

[22] A. Revel and P. Andry, "Emergence of structured interactions: From a theoretical model to pragmatic robotics," *Neural Network*, vol. 22, pp. 116–125, 2009.

[23] C. Trevarthen, *The social foundations of language and thought. Essays in Honor of Jerome Bruner*. N.Y.: Norton, 1980, ch. The foundations of intersubjectivity: Development of interpersonal and cooperative understanding in infants.

[24] J. Gewirtz, *Handbook of socialization theory and research*. Chicago : Ran McNally, 1969, ch. Mechanism of social learning : some roles of stimulation and behaviour in early development, pp. 57–212.

[25] E. Blass, J. Ganchrow, and J. Steiner, "Classical conditioning in newborn human 2-48 hours of age," *Infant behavior and development*, vol. 7, pp. 223–235, 1984.

[26] G. Gergely and J. Watson, "Early social-emotional development: Contingency perception and the social biofeedback model." *Early social cognition*, pp. 101–136, 1999.

[27] K. Hiraki, "Detecting contingency: A key to understanding development of self and social cognition," *Jpn. Psychol. Res*, vol. 48, no. 3, pp. 204–212, 2006.

[28] C. G. Prince, G. J. Hollich, N. A. Helder, E. J. Mislivec, A. Reddy, S. Salunke, and N. Memon, "Taking synchrony seriously: A perceptual-level model of infant synchrony detection," in *Proceedings of the Fifth international conference on Epigenetic Robotics, EPIROB'05*, no. 128. Lund University Cognitive Studies, 2005, pp. 89–95.

[29] M. Rolf, M. Hanheide, and K. Rohlfing, "Attention via synchrony. making use of multi-modal cues in social learning," *IEEE Transactions on Autonomous Mental Development*, vol. 1, no. 1, 2009.

[30] 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*, vol. 28, pp. 418–425, 2005.

[31] J. Banquet, P. Gaussier, J. Dreher, C. Joulain, and A. Revel, *Cognitive Science Perspectives on Personality and Emotion*. Elsevier Science BV Amsterdam, 1997, ch. Space-Time, Order and Hierarchy in Fronto-Hippocampal System: A Neural Basis of Personality, pp. 123–189.

[32] A. Barto, R. Sutton, and C. Anderson, "Neuronlike adaptive elements that can solve difficult control problems," *IEEE transactions on system, man and cybernetics*, vol. SMC-13, no. 5, pp. 834–846, Sep/Oct 1983.

[33] P. Gaussier, A. Revel, C. Joulain, and S. Zrehen, "Living in a partially structured environment: How to bypass the limitation of classical reinforcement techniques," *Robotics and Autonomous Systems*, 1997.

[34] S. Boucenna, P. Gaussier, and P. Andry, "What should be taught first: the emotional expression or the face?" in *Proceedings of the eighth International Workshop on Epigenetic Robotics, Modeling Cognitive development in robotic Systems*, . Lund University Cognitive Studies, Ed. EPIROB08, 2008, pp. 97–100.

[35] K. Prepin, "Développement et modélisation des capacités dinteractions homme-robot : L'imitation comme modle de communication," Ph.D. dissertation, Centre Emotion CNRS - Université de Cergy-Pontoise, 2008.