Proceedings of the Post-Graduate Conference on Robotics and Development of Cognition
10-12 September 2012, Lausanne, Switzerland

A synchrony based approach for human robot interaction

Syed Khursheed Hasnain, Philippe Gaussier and Ghiles Mostafaoui

Neurocybernetic team, ETIS, ENSEA, University of Cergy-Pontoise France
syed-khursheed.hasnain@ensea.fr, gaussier@ensea.fr and ghiles.mostafaoui@ensea.fr

Abstract

As psychologists considered synchrony as an important parameter for social interaction, we hypothesize that in the case of social interaction, people focus their attention on regions of interest where the visual stimuli are synchronized with their inner dynamics. Then, we assume that a mechanism able to detect synchrony between internal dynamics of a robot and external visual stimuli can be used as a starting point for human robot interaction. Inspired by human psychological and neurobiological data, we propose a synchrony based neural network architecture capable of selecting the robot interaction partner and of locating Focus of Attention.

Index Terms: Human Robot Interaction, Synchrony, Focus of Attention, Partner Selection, Dynamical Systems.

1. Introduction

Human verbal interaction is not only speech dependent. In fact, many non-verbal behaviors such as facial expressions, pauses during discussion, hand movements etc. are also involved [1]. An important aspect of these non-verbal communications is their timing and synchrony according to the partner's behavior. Psychological Studies of dyadic interactions shows that synchrony is a necessary condition for interaction between an infant and his mother [2]. Recently, Dumas et al.[3] revealed, using hyperscanning, the emergence of inter-brain synchronization across multiple frequency bands during social interaction. Interpersonal motor coordination between people can be observed while walking along with someone [4]. Marin et al. underlined that motor resonance between robots (humanoid) and humans could optimize the social competence of human-robot interactions [5]. Qiming Shen et al. also did related experiments [6].

By the above discussion, it is clear that synchrony is an important parameter for social interaction as well as largely witnessed in natural dynamical systems. In this paper, we use immediate synchronous imitation as a communication tool. We present here a neural network architecture for socially interacting robots.

2. Materials and Methods

We used a minimal setup for our experiments as shown in figure 1. Components includes Nao robot, basic automata (1 degree of freedom), human and cameras. To avoid the frame rate limitation of the Nao's camera through the ethernet connection (limited to 10 Hz), a new camera has been added for Nao's vision. The frame rate for our experiments is 30 Hz.

To analyze synchrony, we need to investigate the dynamics of interaction between two signals. To do so, we use the Phase Locking Value (PLV) which is a practical method presented by Lachaux et al. [7]. The PLV for two signals is defined by $PLV_{n,r} = \frac{1}{T}|\sum_{t=1}^{T} exp(i(\phi_n - \phi_r))|$, where T is the number of samples and $\phi_n - \phi_r$ is the phase difference between two signals. When there is synchronization the PLV value is close to 1 otherwise the PLV value approaches to 0. Videos of our experiments can be found on: http://www.etis.ensea.fr/neurocyber/Videos/synchro/

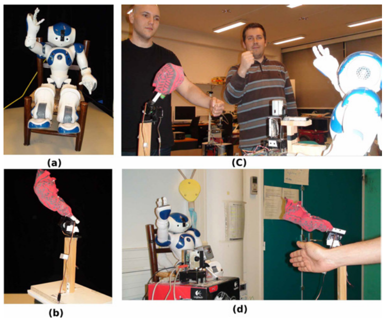

Figure 1: *Setup for our experiments. (a) Nao robot (b) Basic Automata (made in the lab) (c) and (d) Overall setup for human-robot and robot-robot interaction.*

3. Human Robot Interaction

Here, we propose a model based on dynamical interactions of two agents. Agent 1 (Nao robot) dynamically adopts or imitates the behavior of agent 2 (human / automata). Our aim is to provide to Nao limited capabilities to interact with other agents by dynamically adopting the frequency and phase of the other agents. Velocity vectors estimated by an optical flow algorithm represent the visual stimuli and inputs for our architecture.

The oscillator model is shown in figure 3(a). It is made of two neurons $N1$ and $N2$, fed by a constant signal and multiplied by the parameters $\alpha1$ and $\alpha2$ (equation 1 and 2). These two neurons inhibit each other proportionality to the parameter β.

$$N_1(n+1) = N_1(n) - \beta N_2(n) + \alpha 1 \tag{1}$$

$$N_2(n+1) = N_1(n) + \beta N_2(n) + \alpha 2 \tag{2}$$

The frequency of the oscillator depends on the parameters $\alpha1$, $\alpha2$ and β. In addition, a reservoir of oscillators (echo state network) could be used to work with a larger range of frequencies.

As shows in figure 3(a), the oscillator is connected with Nao's arm and oscillates normally at its own frequency and amplitude. Motion in the visual field of Nao is estimated by an optical flow algorithm, velocity vectors are then converted into positive and negative activities. If the perceived movements are in the upward direction, the oscillator gets the positive activity and its amplitude increases. On contrary, if the negative activity is perceived amplitude goes down. When an agent interacts with a motion frequency close to NAO's frequency, Nao's oscillator

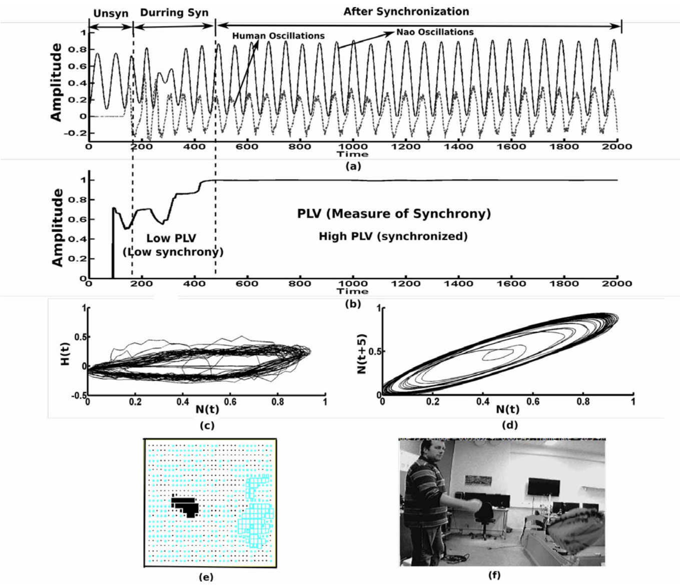

Figure 2: *(a) Shows two motion signals (human and Nao). (b) PLV measurement. (c) Lissajous curve between $N(t)$ (Nao' oscillation) and $H(t)$ (Human's movements), (d) Lissajous curve between $N(t)$ and $N(t{+}5)$. (e) Positive and Negative activities deduced from optical flow. (f) Real image seen by camera.*

can be modified within certain limits otherwise it continues to his default frequency. Mathematical equation of the oscillator can be rephrased as $N_1(n+1) = N_1(n) - \beta N_2(n) + \alpha 1 + f'$. Where f' is the induced energy.

As shown in figure 3(a) a modifiable oscillator is connected with Nao's arm. When there is no visual input it oscillates normally but if a human comes and interact with Nao by imitating him, Nao synchronizes with human by modifying his frequency and phase . Figure 2(a) shows the motion signals of both NAO's and the human arm while trying to interact by imitating games. Initially, both are unsynchronized. PLV (indicator of synchrony) has its lowest value (see figure 2(b)). As shown in Figure 2(a) and 2(b), during the interaction both Nao and human are synchronizing little by little similar to a pendulum coupling. The increasing PLV values also show the emerging synchrony. Figures 2(a) and 2(b) also clearly illustrate that, after a certain time, the agents are completely synchronized, the corresponding PLV values are at the highest possible range. Figure 2(c) shows Lissajous curve between the motion signals of Nao's ($N(t)$) and human's movements ($H(t)$). The elliptic shape of the curve indicates that both signals are almost identical. Figure 2(e) is a snapshot taken during experiment illustrating positive and negative activities in the visual field deduced using the optical flow velocity vectors. Figure 2(e) shows two moving objects in the field of view of Nao. One moves upward and induces positive activities (shown by filled black color pixels) while the other moves downward and induces negative activities (unfilled pixels). Figure 2(f) shows the real image seen by the camera.

Interesting facts are observed during experiments, some of these observations were also made by Pantaleone in his study of metronomes synchronization [11]. First, if the natural frequency of the two agents (in his case two pendulums) differs by more than a certain limit, synchronization will not occur. The range of interacting frequency (that can be synchronized with Nao) can be expanded by increasing the coupling energy f'(by scaling coupling factor) that feeds the Nao's oscillator. With low scaling factor both agents can be synchronized if their natural frequency differs by more than few percents similarly, higher scaling factor leads to higher range of frequencies. For this human/robot interaction, the default frequency of Nao's oscillator was 0.428 Hz while human's interacting frequency (measured by adding the active pixels of motion estimation) was between 0.4615 Hz to 0.476 Hz (7.8% to 11% higher than Nao's frequency) with 0.15 as a scaling factor and 15% as the corresponding Δf (difference between the natural frequencies that can be synchronization). A coupling factor of 0.3 leads to $\Delta f = 29\%$ with little variations on the amplitude, a scaling factor of 0.5 results to $\Delta f = 72\%$ but this higher coupling introduces amplitude saturation. We also observed that for the same parametric conditions, if the natural frequencies of both agents are the same no phase leg was observed but as the Δf increases to a certain limit the phase leg increases too. We experienced 0^0 to 90^0 of phase shift in our experiments.

4. Selection of Partner

We propose a neural network architecture (Figure 3(b)) that selects an interacting partner on the basis of synchrony detection among various interacting agents. Previously, the modifiable Nao's oscillator controlling the arm movement was directly connected to the visual stimuli (f'). Now, the coupling is made through an oscillator-prediction module (f''). The reason for indirect coupling is to make sure that the architecture will entertain the visual stimuli (optical flow) that is similar to its own motion (learnt by the oscillator-prediction module). Equation of modifiable oscillator can be rewritten as $N_1(n+1) = N_1(n) - \beta N_2(n) + \alpha 1 + N + f''$. Where f'' is the energy induced by the Oscillator-prediction module.

The oscillator-prediction block (represented by y') is linked to the robot's oscillator (represented by y) with a non modifiable link while the image of visual activities (represented by X) is linked with a modifiable link. The Oscillator-prediction (y') module learns the robot's oscillation as a weighted sum of active pixels. The neuron activity in the Oscillator-prediction (y') can be computed using $X \to y'$ synapses by : $y'_i(t) = \sum_{k \epsilon X} W_{X_k - y'_i} X_k$ that corresponds to the predicted future value.The learning of $X \to y'$ synaptic weights can be computed by equation 3 and is based on NLMS (Normalized Least Mean Square) algorithm (Synaptic learning modulation η is additionally added) [8].

$$W_{X_j - y'_i}(t+dt) = W_{X_j - y'_i}(t) + \alpha\eta . \frac{y_i(t) - y'_i(t)}{\sum_{k \epsilon X} X_k(t)^2 + \sigma 1} . X_j(t) \quad (3)$$

Where y' stands for the Oscillator-prediction, X for the image of visual activities and y for the NAO's arm Oscillator, α is the learning rate and $W_{X_j - y'_i}$ represents the synaptic weights from X_j to $Oscillator - prediction$ neuron i, y_i is the activity transmitted to neuron i by the oscillator, it is a target signal for the Least Mean Square (LMS) algorithm [9]. To improve the LMS convergence during the learning phase, we introduced the learning modulation η. The normalization term $\sum_{k \epsilon X} X_k(t)^2 + \sigma 1$ is specific to the NLMS and $\sigma 1$ is a small value used to avoid the divergence of the synaptic weights if the visual activities (X) values are too small.

Now we consider the complete scenario. For the selection of partner, the architecture works in two phases: learning phase and testing phase. During the learning phase, NAO oscillates according to its default frequency (no visual stimulus). NAO

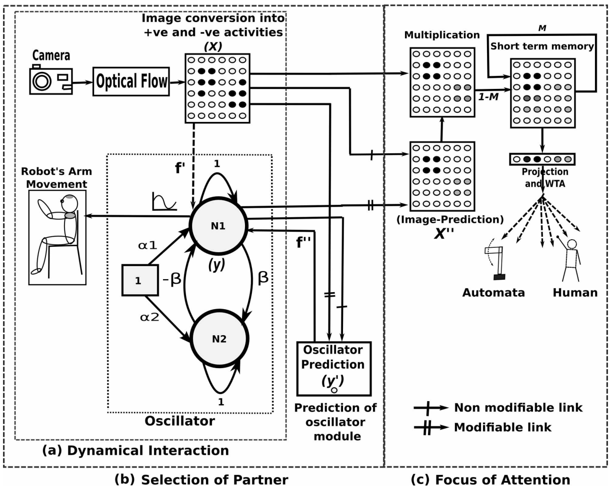

Figure 3: *(a) Dynamical Interaction model (b) Selection of Partner: select a interacting partner on the basis of synchrony detection. (c) Shows attentional mechanism architecture.*

looks at its own hand. It initiates two processes. First the oscillator prediction module which was zero due to non availability of visual stimuli starts now predicting robot's modifiable oscillator as a weighted sum of its own visual stimuli. The oscillator-prediction module learns associations between NAO's motion and the visual activities induced by NAO's arm. As a consequence, it also modifies the NAO's oscillator (as described in section 3). This process of modifying, learning and adapting continues and converge after some time. This adjustment can be assumed as a basic process by which infants gain self reflective abilities as underlined by Rochat [10]. After this phase, NAO learns to predict oscillatory movements similar to his own movement. When an agent interacts with a frequency similar to the learned one, weights (that are already learnt on modifiable links) are associated with the visual activities induced by the human movements and Nao's modifiable oscillator adopts the interacting frequency and phase. If the interacting frequency is different from the learnt one, the weights (modifiable links) could not be associated with the visual stimuli and NAO continues to move at his default frequency. Same is true for multiple agents case. Among two interactants only the agent having a similar frequency as Nao is selected. In this experiment, the coupling factor was 0.07, Nao's default frequency was 0.407 Hz, automata synchronized frequency was 0.4318 Hz (6% higher) and human synchronized frequency was 0.36 Hz (11% less). When a subject interacts with a frequency close to the learnt one, this selection of partner algorithm selects this agent as a good interacting partner and NAO modifiable oscillator synchronizes with it. Good results are obtained with this architecture, they are collectively shown in the next section.

5. Attentional Mechanism

Here, we use prediction of synchrony as a parameter to attract the attention of the robot. If two visual stimuli are presented at the same time and only one of them has the same frequency as NAO. NAO will then synchronize with the "interacting" partner corresponding to his frequency and select him as a partner (by selection of partner algorithm). However, NAO will not be able to locate the good interacting partner in its visual field, because this algorithm (partner selection) works on the perceived energy irrespective of the spatial information (agent location). To locate the correct interacting partner, the proposed FOA algorithm dynamically locate the correct interacting partner (defined by the selection of partner algorithm) using spatial predictions. Figure 3(c) shows the architecture of FOA. When a human interacts (using arm / hand), the image-prediction block (X'') learns the image of these movements as a weighted sum of Nao's synchronized frequency. This makes it possible to predict the corresponding human movements. After a short while, an other agent comes and moves with a different frequency (lower or higher than Nao), X'' which already learnt synchronized rhythmic movements predicts strongly the first synchronized agent compared to the unsynchronized one. Our algorithm modulates this predicted synchrony with the current visual stimuli and calculates the average value (acting as short term memory). As the synchronized image is well predicted its correlation values are higher than the asynchronous movements. Figure 3(c) shows that all the pixels of the memory block is projected on y axis (i.e all pixels in each column are added to find the highest correlated column). Then a Winner Takes All (WTA) selects the highest activated column. This selected column indicates the location of synchronized movement and the robot can point to the synchronized region to show the current Focus Of Attention (FOA). For this experiment the resolution of the predicted image of optical flow is 32$\times$ 24 (32 columns or location), these 32 possible locations are realized in 60^0 (-30^0 to 30^0) circular angles. The learning rule of the movement-prediction (X'') module is almost the same as the oscillator-prediction module and the weights are normalized to smooth the learning processes.

5.1. Results

we examine our selection of partner algorithm along with FOA architecture (figure 3(c)) in two situations: one Automata (1-DoF) and one human (only one of them is synchronized at a time). Results show that when the Automata moves similarly to Nao's movements while human oscillates with a different frequency, Nao synchronizes with the Automata (selection of partner) and FOA mechanisms turns towards Automata. If the human adopts his frequency close to Nao, Nao aligns himself with the human and FOA moves towards human.

These results of both algorithms are shown in figure 4 by two sets of graphs. Figure 4(a) shows the onset of the experiment, where the Automata enters in the visual field of Nao from the left side (about -20^0) and imitates him. Consequently, both become synchronized using our selection of partner algorithm. Figure 4(a1) sketches the signals of Nao modifiable oscillator and Automata illustrating how they become synchronized. Figure 4(a3) shows the PLV value (measure of synchrony) of the two agents. Initially, PLV is low but as the interaction gets longer it increases to higher value. As the Automata interacts, FOA moves towards Automata as Shown in Figure 4(a4). Figure 4(a2) shows signals of Nao and human illustrating that initially there is no interaction by human from the right side of the robot. After 700 time units (23.33 seconds) human comes with a different frequency. He does not succeeded in disturbing the selection of partner (PLV remains high for Automata) and FOA remains towards the Automata.

Now, the automata is tuned to a low frequency and human is instructed to imitate NAO (figure 4(b1) and (b2)). As a result, Nao switches the synchronized region, from left (-20^0) to right side (about 27^0). The PLV related to human increases to the highest value while the Automata PLV shifts to a lowest

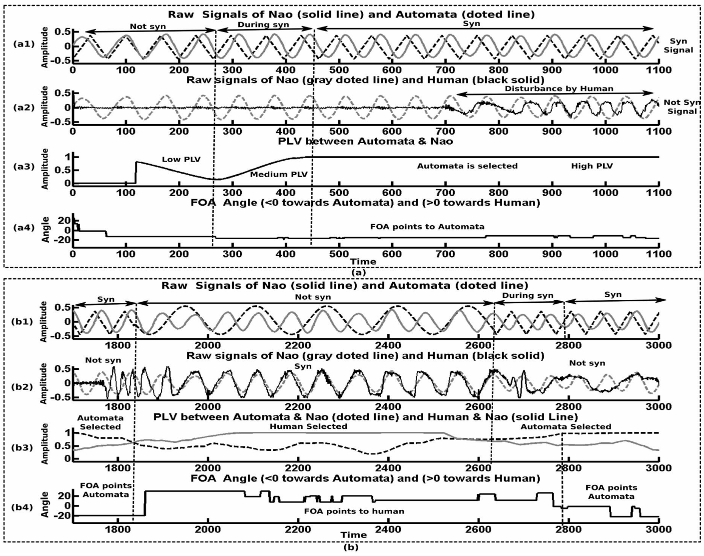

Figure 4: *Results: (a) shows start of experiment with single agent and then disturbed by the other agent. (b) Different frequency agents interact with Nao.*

one (figure 4(c3)). FOA shifts from the automata to the human (figure 4(c4)). After 2650 time units (88.3 sec), the Automata is tuned to its previous frequency again and the human is instructed to make different oscillations. Consequently, this induces a switch of the FOA and the recognized interacting partner (figure 4(b)).

6. Conclusion and discussion

We proposed a novel approach for building autonomous robots that can interact with multiple agents and select an interacting partner among several on the basis of synchrony detection. We also showed that synchrony prediction could be used as a way to establish focus of attention. From the psychological point of view, we were inspired by the unconscious communications between humans. The synchronous exchanges during social interactions are directly associated to the sensorimotor information of the two agents. These inter brain networks are ”symmetric” in low frequency band while ”asymmetric” in high frequency bands [3]. This could reflect the different processing levels of information. In our case, synchronization between two agents can be assumed as “symmetric” in low frequency band and Focus of attention can be associated with high frequency carrier.

Actually we are studying three human-robot applications for synchrony detection. The first and most obvious one is to extend the model to learn more complex interactions (complex gestures). Indeed, synchrony detection and selection of partner permit to maintain interaction with a partner moving synchronously with the robot in terms of low fundamental temporal frequency of interaction. As a result, more complex gestures (higher temporal frequencies) can be imitated and learnt autonomously by the robot while interacting with the human partner. Similarly, we aim to use our architecture for navigation tasks. A mobile robot can choose a synchronous agent to interact with and consequently learn complex navigation tasks by keeping synchrony while moving with the selected partner. Finally and in a global point of view, we question the use of synchrony detection, focus of attention and selection of partner in turn-taking games during interaction. In fact, synchrony can not only be considered as a starting point for social interaction but also as a way to re-engage the interaction with a selected partner.

7. References

[1] A. Kendon. Conducting Interaction: Patterns of behaviors in focus encounters. Cambridge University press, Cambridge, UK. 1990.

[2] J. Nadel, I. Carchon, C. Kervella, D. Marcelli, D, Reserbat. Expectancies for social contingency in 2 months old. Developmental science 2 (1999) 164-173.

[3] G. Dumas, J. Nadel, R. Soussignan, J. Martinerie, L. Garnero. Inter-Brain Synchronization during Social Interaction. PLos ONE 2010;5(8):e12166.

[4] N. R. van Ulzen, C. J. Lamoth, A. Daffertshofer, G. R. Semin and P. J. Beek. Characteristics of instructed and uninstructed interpersonal coordination while walking side-by-side. Neuroscience letters, 432, 88-93

[5] L. Marin, J. Issartel and T. Chaminade. Interpersonal motor coordination, from humanhuman to humanrobot interactions.In: Dautenhahn, Kerstin (ed.), Robots in the Wild: Exploring human-robot interaction in naturalistic environments: Special Issue of Interaction Studies 10:3 (2009). 2009. vi, 239 pp. (pp. 479504)

[6] Qiming Shen, H. Kose-Bagci, J. Saunders, K. Dautenhahn. The Impact of Participants' Beliefs on Motor Interference and Motor Coordination in HumanHumanoid Interactions. Autonomous Mental Development, IEEE Transactions on. 04/2011;

[7] J. P Lachaux, E. Rodriguez, J. Martinerie and F.J. Varela. Measuring phase synchrony in brain signals. Human Brain Mapping 8 : 194-208 (1999).

[8] J. Nagumo. A learning method for system identification. IEEE Trans. Autom. Control 12(3) (1967) 282-287 NLMS reference.

[9] B. Widrow, M. E. Hoff. Adaptive switching circuits. In IRE WESCON, pages 96-104, New York, 1960. Convention Record.

[10] P. Rochat. Ego functions of early imitiation. In A. Meltzoff & W. Prinz (Eds.), The imitative mind (pp. 85-97). Cambridge University Press, 2002.

[11] J. Pantaleone, Synchronization of metronomes. American Journal of Physics, Vol. 70, Nr. 10AAPT (2002) , p. 992–1000.