Vocabulary Learning and Sound Localization in Autonomous Robots

***Abstract*— The simulation of vocabulary acquisition in embodied artificial agents (robots) has been studied by various methods and protocols. Here we simulate the emergence of a common repertoire of signs applying unsupervised learning and phonotaxis to enable a hearer robot to locate a speaker robot and the object referred during communication. Results show that the proposed scenario and control mechanisms allow the robots to find the sound source and lead to robust learning and establishment of a shared repertoire of sign by robots, even in the presence of noise and sensor data identification errors.**

I. INTRODUCTION

SIMULATION of vocabulary acquisition in embodied artificial agents (robots) has been studied through various experimental methods and computational protocols (for a review of works, see [2],[3],[4]). It usually involves a community of agents that initially do not have a repertoire of signs[1] to refer to objects and, based on interaction events and learning/evolution processes, ends up with a shared vocabulary of such entities.

Such kind of research contributes to the scientific agenda of understanding the evolutionary conditions involved in the emergence of such complex traits and behaviors. Besides, there is also a complementary perspective related to the development of 'new technology for communication between humans and robots, or among robots' [5]. Allowing robots to adapt and acquire the necessary means for communicating and use language is a better solution than pre-programming them, especially when dealing with real world and open-ended applications involving humans or collective robotics, and makes such systems more flexible and robust, able to continuously cope with new demands.

Here we present initial results from an experiment with robots that are able to learn a shared repertoire of signs in an unsupervised manner based on communication and relying on phonotaxis (sound localization) to find the speaker and the referring object. Phonotaxis is a crucial cognitive ability to approach sound sources and can be considered a fundamental competence in the emergence of intra- and inter-specific communication [11].

Manuscript received May 31, 2010.

First, we review some of the related works, and then we present our experimental set up. Next, we exhibit our initial results and discuss them, and, finally, we outline our conclusions and point to future perspectives in this experiment.

II. RELATED WORK

Studies on the emergence of a communication system for artificial agents apply a common methodology: they start with a community of agents that do not have a previously established communication system, but are able to interact with each other and acquire a shared communication system based on learning or evolutionary processes.

Floreano and coleagues [6] simulated the evolutionary conditions that might allow the emergence of a reliable communication system in a community of simulated robots, based on biological motivations on animal communication. The robots could use a visual signal, turning on or off a light ring, to communicate with other robots about the position of food source. They found that if selection acts on group level instead of individual level, or if members of a community are genetically similar, a reliable communication system could emerge. The robots simulated in this experiment were controlled by artificial neural networks, with a direct connection between the input layer and the output layer, so a light signal received was directly mapped to a displacement speed. The robots use only one signal to communicate and the robots do not relate this signal with any object in the environment, they only reactively respond to them.

Marroco and Nolfi [7] developed an experiment for the emergence of communication in robots to solve a collective navigation problem. The task was for a group of four robots to find and remain in two different areas, with two robots in each of them. The robots were also controlled by neural networks that are evolved to allow their adaptation to the task. They have an actuator to send signals of varying frequencies and a sensor to detect signals in four different directions up to a certain distance. At the end of the evolutionary process, they noticed that robots produce different types of signal depending on the situation they are at (close or near to another robot, inside or outside the target areas), and these signals modify the behavior of other robots to help solve the task. There is therefore a repertoire of

[1] We use the term 'sign' here as something that refers to something else for someone, on whom it produces an effect, following the definition of C.S.Peirce [1].

signals to communicate, but they are not related with any object in the environment and robots reactively respond to them.

Language games were used by Steels [5][8] as the interaction protocol for language acquisition between artificial robotic agents. In his experiments, he applied the guessing game to allow for vocabulary learning between fixed robots, computer-controlled pan-tilt cameras facing a white board with geometric figures, the topics of the game. In the guessing game, two robots interact at each time, one as a speaker and the other as a hearer. The speaker chooses a topic in the board and vocalizes an utterance to the hearer, who has to guess the topic chosen. If the hearer fails in his guess, the speaker indicates the expected topic. At the end of each game round, the robots adjusts their memory associations based on the outcome of the game in a supervised manner.

Vogt [9,10] also experimented with language games between robots, but he used mobile lego robots, instead of fixed camera robots like Steels. Besides, the topics of the language games in Vogt's experiments were light towers sensed located through light sensors when the robots makes a complete turn. Vogt used the guessing game of Steels, but proposed an observational game where the hearer did not receive a feedback from the speaker, so the robots had to learn in an unsupervised manner, but he points out that results were not good in this new language game.

Besides these works, others have studied the emergence of a communication system among artificial agents (see [2][3][4]). But, as far as we know, this is one of the few studies to apply unsupervised learning for the emergence between robots of a common repertoire of signs, which refer to objects on the environment, and additionally it is the first work to apply phonotaxis to enable a hearer robot to locate a speaker robot and the object referred by it during communication.

III. The Experiment

Our experiment involves robotic agents that are able to learn a shared repertoire of signs in an unsupervised manner based on autonomous communication and relying on phonotaxis (sound localization) to find the speaker and the referring object. The robots interact in a manner similar to the guessing language game, but there is no feedback from the learner and the robots do not have far vision sensors (such as cameras or long range light sensors).

At the start of a guessing game round, the speaker senses the object of communication, a colored paper, then vocalizes repeatedly a sign for this topic. The hearer searches for the speaker and the object of communication by trying to locate the sound source. When the hearer approaches the speakers and finds a color, the game ends and the hearer adjusts its associations between the sign and the color. These rounds are repeated for a fixed number of times.

The scenario is a white ground arena, where the colored paper are placed. The agents were assembled as mobile Lego Mindstorms NXT robots. Each robot has a control unit, the NXT brick, which has a loudspeaker, a sound actuator. The speaker robot also has a light sensor direct towards the ground, and the hearer has the same light sensor and two microphones as sound sensors. To expand control programming capabilities, the original NXT firmware was changed to leJOS NXJ, a Java based replacement firmware for the Lego Mindstorms NXT microcontroller.

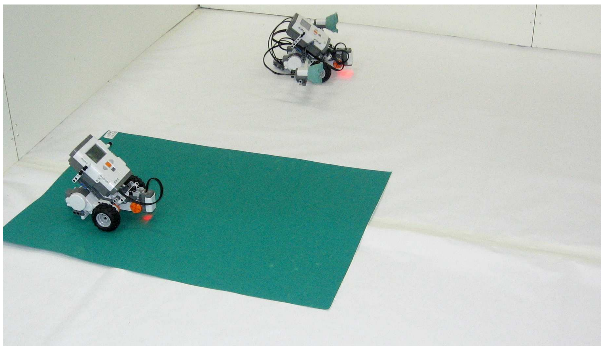

Fig. 1. The scenario for the experiment. The speaker stays on top of the topic (a colored paper). The hearer searches for the speaker based on sound localization.

The first challenge in this experiment was to deal with phonotaxis. Phonotaxis is the sound localization ability of the hearer robot that will allow it to find the speaker along with the referent of the speaker vocalized sign (for more on phonotaxis modeling in robots, see [11]). To implement this ability in the hearer, we used two microfones placed far away from each other, each one inside a cone to give better directional response. The difference of signal amplitude between these two sound sensors can be used to determine the direction of the sound source. If there is no significant difference in amplitude, the sound source should be ahead of the robot. We determined empirically that an appropriate sound frequency for speaker robot to use was 950 Hz for a good reception by the hearer robot. And the distance between the microphones should be 35 cm, to allow an appropriate difference in signal amplitude.

Besides being used as a mean for phonotaxis, the sound signals are also a medium for the speaker robot to encode up to eight different signs for the hearer robot. The speaker robot maintains associations between colors and signs in its associative memory. As the speaker robot senses a color through its light sensor, it looks up for a sign for this color. If it finds an associated sign, it uses the sound actuator to vocalize this sign. If no sign is found associated with this color, the speaker robots randomly chooses a new sign among the ones not used yet.

The speaker robot repeats the selected sign until the end of each round. The sign is modulated in a carrier, a continual and constant signal with fixed frequency, by a simple amplitude shift. Each sign is composed by three binary

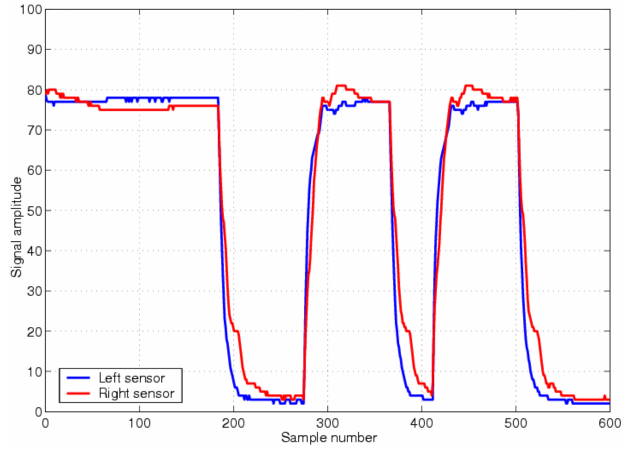

Fig. 2. Reception of the sign with a binary encoding of 101, with both microphones equidistant from the sound source. The sensor was sampled every 10 ms.

TABLE I
COLORS AND LIGHT SENSOR VALUES FOR THE SPEAKER AGENT

Color	Sensor Value Range
White	110-94
Yellow	87-81
Blue	64-57
Green	15-11
Black	3-(-2)

TABLE II
COLORS AND LIGHT SENSOR VALUES FOR THE HEARER AGENT

Color	Sensor Value Range
White	115-94
Yellow	93-83
Blue	67-58
Green	18-12
Black	3-(-3)

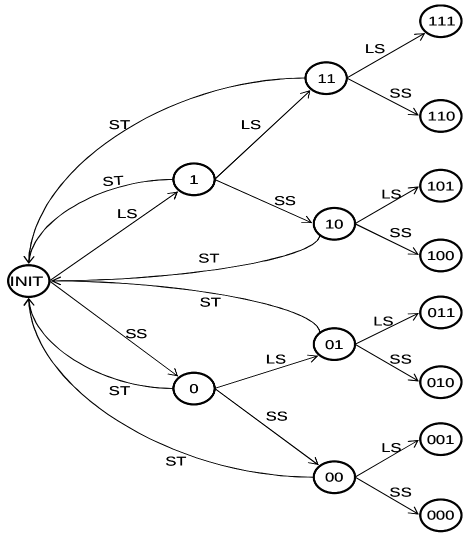

Fig. 3. State diagram of the parser. ST is the marker for the start of a sign (continual signal for 5000 ms). LS is the long silence period of 1000 ms. SS is the short silence period of 500 ms.

characters, thus ranging from the decimal numbers from 0 to 7. Each binary character corresponds to a silence period, with either a 500 ms duration (character 0) or a 1000 ms duration (character 1). Characters are separated by a period of 2000 ms of continual signal, and each sign repetition is preceded by a period of 5000 ms of continual signal. Figure 2 exhibits the reception of a sign by the hearer when both microphones are equidistant from the sound source.

The signs refer to colors captured by the light sensor directed to the ground, which is able to distinguish only brightness level. The sensor readings are encoded as a 10 bit value, with smaller values corresponding to darker colors. The light sensor was calibrated for a scale between 0% and 100%, with the white ground color being 100% and the black color paper, 0%. The other colors – yellow, blue, and green – have intermediary values. Tables 1 and 2 show values of the light sensor for each robot for each color. Notice that the values correspond to value ranges since different light and shade conditions, as well as proximity of the sensor, produce variations in the light sensor reading. To deal with these variations, both robots categorize the light sensor reading based on a prototype-based categorization. Each category has a prototypical value, and, if the value read from the light sensor is ± 9 from a prototypical value, this new value is classified as the corresponding category, but if it does not fall within any category then a new category is created and this new value is set as its prototypical value.

To engage in the guessing game, the robots have to perform a series of actions, and the hearer robot has a greater set of them. It has to perform phonotaxis, find the speaker and the colored paper, sense the color, recognize the sign, and learn from this communicative interaction. The speaker robot only senses the colored paper, selects the sign to communicate and vocalizes it repeatedly.

Phonotaxis behavior corresponds to a closed loop control sequence which starts with the robot staying still and reading its sound sensors. If there is a significant amplitude difference between sound sensors readings, the robot takes a small rotation in this direction. If signals are about the same amplitude, the agent moves forward for a few seconds. After rotating or moving forward the robot reads its light sensor to determine if it reached the colored paper. If not, it reads its sound sensors again and repeats the sequence.

When a colored paper is sensed by the light sensor, the hearer robot starts a sign identification procedure, using the parser presented in figure 3. The parser decodes the sound sensor readings by measuring the silence periods and its intervals and, using a tolerance range, it can identify the start of a sign and each binary character.

With a color value and an identified sign, the hearer robot can update its associative memory to this perceived relation. Each association in the memory involves a color, a sign and an associative strength value (between 0 and 10), and associations can either be created or adjusted according to the perceived color-sign pair. If there is no association between the given color and sign, a new association is created for them with a strength value of 1. If there is already an association, then its strength is increased by 1. Besides, all competing associations have their strength values decreased by 1. Competing associations are other associations that have the same color associate with a different sign or the same sign with a different color. These strength adjustments correspond to an associative learning mechanism with lateral inhibition and allow the hearer robot to robustly acquire a repertoire of signs based on communicative events.

IV. RESULTS

To run our experiment, we execute several rounds (or iterations). Each round starts with the upload of the robots associative memories saved in a PC for persistence and data analysis, followed by calibration of the light. The speaker robot is placed on top of a colored paper and the hearer robot is placed at another part of the arena away from the colored paper. At the end of the round, the associative memories are downloaded back to the PC.

We executed a total of 30 rounds with the robots to verify if the hearer robot could learn from the speaker robot to name the colors available, and how and what was learned, based on communicative interactions. Ideally, we expected the hearer robot to always relate one color category only with one sign, so the associative strength values would always increase and there would be no competing associations. This was not the overall result obtained, however.

According to the results, there were competing associations and thus decrements happened in association strengths. Due to variations in sensor readings and also in sign identification, the hear robot and also the speaker robot could behave erroneously but results show that the associative learning mechanism can robustly deal with these false data, by applying reinforcement and weakening cycles together with lateral inhibition of competing associations. To observe the outcome of the learning process, we show in figure 4 the strength values of the hearer robot associations at every round.

In the graphs of figure 4, each increase in association strength value corresponds to a reinforcement due to co-occurrence of a color category and a sign, according to sensors readings. On the other hand, decrements in strength value occur for associations competing with the reinforced association, either because they share a common color category or a common sign.

The speaker robot used the signs consistently. It created and always used sign 010 for color category 1, sign 001 for color category 2, sign 110 for the color category 3, and sign 101 for color category 4. So the speaker robot did not have incorrect readings from the color sensor that could mislead it, but the hearer robot did have problems with sensorial data in quite a few rounds.

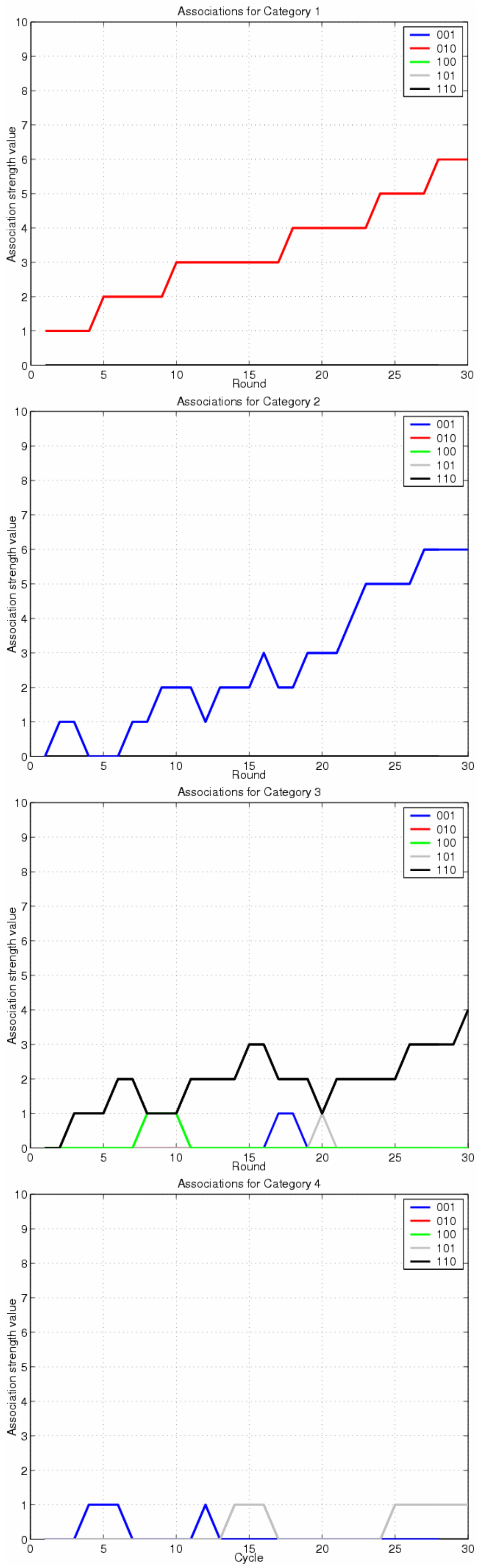

Fig. 4. The strength value of the hearer robot associations for each sign associated with the color categories 1 (blue), 2 (yellow), 3 (black), 4 (green).

For the hearer robot, the was also only one sign (010) associated with color category 1, and its strength value followed a monotonically increasing behavior, with only reinforcements of this single association (figure 4). This means that the hearer robot correctly identified the color and the sign in every round involving the colored paper described by category 1. But that was not the case with color category 2, as we can observe, once in some rounds the association with sign 001 did have decrements, although no other sign was associated with color category 2. Since every decrease in association strength corresponds to an increase in other association, we can identify that the decreases in the association between sign 001 and color category 2 were due to the sign 001 been associated, at some rounds, with color category 3 and color category 4. Since the speaker robot vocalized sign consistently, what happened is that the hearer robot misidentified the sign heard as 001 instead of 101 vocalized for the color category 4, for example. Similar association errors due to sign identification problems can also be observed in the graphs for color category 3 and 4.

The sign identification problem was probably due to noise or signal degradation. Even though the sign is correctly outspoken by speaker robot, the laboratory is not free of other sound sources that can be captured by that hear robot along with the sign emitted. Besides, depending on the sound trajectory and facing direction the sound signal of the sign can be degraded, lowering signal amplitude and mislead the hear robot sound readings.

Even though it was not the case in this experimental sequence, another possible source of incorrect association between colors and signs is the light sensor, depending on the colors involved. In this experiment round, the light sensor did not read a misleading value, since the colors chosen could be well distinguished. Nevertheless, if the colors chosen were more similar in brightness value, then this type of error could also lead to the appearance of more competing associations.

In spite of errors in sensor readings, that caused instability on the learning process, at the end of the experiment, we can observe that the strongest associations established by the hearer robot were the ones that the speaker robot applied. This means that the speaker robot created its sign-color associations and, based on communicative interactions, the hearer robot was able to learn from the speaker robot, although the latter did not provide any explicit feedback to the former during the learning process.

V. CONCLUSION

According to our experiment, communicative agents are able to create and establish a shared repertoire of signs in an unsupervised manner, with the hearer relying on phonotaxis to find the speaker and the referring object. The robots are able to sense objects and established its own categories, with no transmission from the speaker of the exact categories used to interpret sensory data. On the other hand, there was no a priori definition on what signs should be used for each object, and the speaker selected them randomly and the hearer was able to robustly learn how to associate them with the objects.

The robots had limited sensory and motor capabilities, there was no camera for acquiring images from the environment, neither laser nor ultrasound sensor for spatial localization. Nevertheless, the phonotaxis proved to be a viable mechanism for speaker localization along with the object referred during communication, and this is probably the first work to simulate this sound localization mechanism for the emergence of a communication system.

In an ideal scenario, robots only receive correct sensory data, but this is clearly not the case in real application, as our experiment shows. Our associative learning mechanism applying of reinforcement and weakening cycles, establishing a lateral inhibition of competing associations demonstrated to be robust in coping with erroneous sensory information.

The autonomous development by robots of its own sign system, in an unsupervised way and applying phonotaxis as the mechanism for sound source localization, permits new ways of dealing with communication and with representation in artificial systems. Visual information is not always available as visual obstacles might be present and explicit feedback on the learning process might neither be always available. This contributes for a better understanding of how natural systems can cope with these limitations, as much as it expands the technological perspectives in building new system with more restricted conditions.

REFERENCES

[1] C.S. Peirce, *Collected Papers of Charles Sanders Peirce*. Cambridge, Mass. : Harvard University Press, 1958.

[2] S. Nolfi, and M. Mirolli, *Evolution of Communication and Language in Embodied and Situated Agents*, Berlin: Springer Verlag, 2010.

[3] M.H. Christiansen, and S. Kirby, “Language evolution: consensus and controversies,” *Trends in Cognitive Sciences*, vol., no.7, pp.300-307, 2003.

[4] K. Wagner, J.A. Reggia, J. Uriagereka, and G.S. Wilkinson, “Progress in the simulation of emergent communication and language,” *Adaptive Behavior*, vol. 11, no.1, pp.37-69, 2003.

[5] L. Steels, “Evolving grounded communication for robots,” *TRENDS in Cognitive Sciences*, vol.7, no.7, July 2003.

[6] D. Floreano, S. Mitri, S. Magnenat, and L. Keller, “Evolutionary Conditions for the Emergence of Communication in Robots,” *Current Biology*, vol.17, pp. 514-519, 2007.

[7] D. Marocco, and S. Nolfi, “Emergence of communication in embodied agents evolved for the ability to solve a collective navigation problem”, *Connection Science*, vol.19 , no. 1, pp. 53-74 , March 2007.

[8] L. Steels, “Language games for autonomous robots,” *IEEE Intelligent Systems*, vol. 16, no. 5, pp. 16–22, 2001.

[9] P. Vogt, “Language evolution and robotics: Issues in symbol grounding and language acquisition,” in *Artificial Cognition Systems*, A. Loula, R. Gudwin, J. Queiroz, Eds., Hershey, PA.: Idea Group. pp. 176-209, 2006.

[10] P. Vogt, “ Bootstrapping grounded symbols by minimal autonomous robots,” *Evolution of Communication*, vol.4, no.1, pp.89-118, 2000.

[11] R. Reeve, and B. Webb, “New neural networks for robot phonotaxis,” Philosophical Transactions of the Royal Society A, vol.361, pp.2245-2266, 2002.