Cognitive Robotics: Robot Soccer Coaching using Spoken Language

Alfredo Weitzenfeld[1] and Peter Ford Dominey[2]

[1] *Instituto Tecnológico Autónomo de México*
Mexico
[2] *Institut des Sciences Cognitives – CNRS*
France

1. Introduction

The chapter describes our current work in developing cognitive robotics architectures in the context of robot soccer coaching using spoken language. The work exploits recent developments in cognitive science, particularly notions of grammatical constructions as form-meaning mappings in language, and notions of shared intentions as distributed plans for interaction and collaboration. We exploit social interaction by structuring communication around shared intentions that guide the interactions between human and robot. We demonstrate this approach in robot soccer coaching distinguishing among three levels of human-robot interaction. The first level is that of commanding or directing the behavior of the robot. The second level is that of interrogating or requesting a behavior explanation from the robot. The third and most advanced level is that of teaching the robot a new form of behavior. The chapter is organized as follows: (i) we explore language communication aspects between humans and robots; (ii) we analyze RoboCup soccer, in particular the four-legged league, as cognitive platform; (iii) we describe current experiments and results in human-robot coaching in the four-legged league; and, (iv) we provide a discussion on current and future directions for this work.

Ideally, research in Human-Robot Interaction will allow natural, ergonomic, and optimal communication and cooperation between humans and robotic systems. In order to make progress in this direction, we have identified two major requirements: First, we must work in real robotics environments in which technologists and researchers have already developed an extensive experience and set of needs with respect to HRI. Second, we must develop a domain independent language processing system that can be applied to arbitrary domains and that has psychological validity based on knowledge from social cognitive science. In response to the first requirement regarding the robotic context, we have studied two distinct robotic platforms. The first, the *Event Perceiver* is a system that can perceive human

events acted out with objects, and can thus generate descriptions of these actions. The second is the *Sony AIBO* robot having local visual processing capabilities in addition to autonomous mobility. In the latter, we explore human-robot interaction in the context of four-legged RoboCup soccer league. From the psychologically valid language context, we base the interactions on a model of language and meaning correspondence developed by Dominey et al. (2003) having described both neurological and behavioral aspects of human language, and having been deployed in robotic contexts, and second, on the notion of shared intentions or plans by Tomasello (2003) and Tomasello et al. (2006) that will be used to guide the collaborative interaction between human and robot. In section 2 we describe our spoken language approach to cognitive robotics; in section 3 we overview the RoboCup four-legged soccer league; in section 4 we describe current experimental results with the Sony AIBO platform in human-robot interaction; section 5 provides conclusions.

2. Cognitive Robotics: A Spoken Language Approach

In Dominey & Boucher (2005a, 2005b) and Dominey & Weitzenfeld (2005) we describe the **Event Perceiver System** that could adaptively acquire a limited grammar based on training with human narrated video events. An image processing algorithm extracts the meaning of the narrated events translating them into *action*(*agent, object, recipient*) descriptors. The event extraction algorithm detects physical contacts between objects, see Kotovsky & Baillargeon (1998), and then uses the temporal profile of contact sequences in order to categorize the events. The visual scene processing system is similar to related event extraction systems that rely on the characterization of complex physical events (e.g. give, take, stack) in terms of composition of physical primitives such as contact, e.g. Siskind (2001) and Steels & Bailled (2002). Together with the event extraction system, a speech to text system was used to perform translations sentence to meaning using different languages (Dominey & Inui, 2004).

2.1. Processing Sentences with Grammatical Constructions

Each narrated event generates a well formed <*sentence*, *meaning*> pair that is used as input to a model that learns the sentence-to-meaning mappings as a form of template in which nouns and verbs can be replaced by new arguments in order to generate the corresponding new meanings. These templates or grammatical constructions, see Goldberg (1995) are identified by the configuration of grammatical markers or function words within the sentences (Bates et al., 1982).

Each grammatical construction corresponds to a mapping from sentence to meaning. This information is also used to perform the inverse transformation from meaning to sentence. For the initial sentence generation studies we concentrated on the 5 grammatical constructions shown in Table 1. These correspond to constructions with one verb and two or three arguments in which each of the different arguments can take the focus position at the head of the sentence. On the left example sentences are presented, and on the right, the corresponding generic construction is shown. In the

representation of the construction, the element that will be at the pragmatic focus is underlined.

	Sentence	Construction <sentence, meaning>
1	The robot kicked the ball	<*Agent event object, event(agent, object*>
2	The ball was kicked by the robot	<*Object* was *event* by *agent, event(agent, object*>
3	The red robot gave the ball to the blue robot	<*Agent event object* to *recipient, event(agent, object, recipient)*>
4	The ball was given to the blue robot by the red robot	<*Object* was *event* to *recipient* by *agent, event(agent, object, recipient)*>
5	The blue robot was given the ball by the red robot	<*Recipient* was *event object* by *agent, event(agent, object, recipient)*>

Table 1. Sentences and corresponding constructions.

This construction set provides sufficient linguistic flexibility, for example, when the system is interrogated about the red robot, the blue robot or the ball. After describing the event *give(red robot, blue robot, ball)*, the system can respond appropriately with sentences of type 3, 4 or 5, respectively. Note that sentences 1-5 are specific sentences that exemplify the 5 constructions in question, and that these constructions each generalize to an open set of corresponding sentences.

We have used the CSLU Speech Tools Rapid application Development (RAD) (CSLU, 2006) to integrate these pieces, including (a) scene processing for event recognition, (b) sentence generation from scene description and response to questions, (c) speech recognition for posing questions, and (d) speech synthesis for responding.

2.2. Shared Intentions for Learning

Perhaps the most interesting aspect of the three part "command, interrogate, teach" scenario involves learning. Our goal is to provide a generalized platform independent learning capability that acquires new <*percept*, *response*> constructions. That is, we will use existing perceptual capabilities, and existing behavioral capabilities of the given system in order to bind these together into new, learned <*percept*, *response*> behaviors.

The idea is to create new <*percept*, *response*> pairs that can be permanently archived and used in future interactions. Ad-hoc analysis of human-human interaction during teaching-learning reveals the existence of a general intentional plan that is shared between teachers and learners, which consists of three components. The first component involves specifying the percept that will be involved in the <*percept*, *response*> construction. This percept can be either a verbal command, or an internal state of the system that can originate from vision or from another sensor. The second component involves specifying what should be done in response to this percept. Again, the response can be either a verbal response or a motor response from the existing behavioral repertoire. The third component involves the binding together of

the <*percept*, *response*> construction, and validation that it was learned correctly. This requires the storage of this new construction in a construction database so that it can be accessed in the future. This will permit an open-ended capability for a variety of new types of communicative behavior.

In the following section this capability is used to teach a robot to respond with physical actions or other behavioral responses to perceived objects or changes in internal states. The user enters into a dialog context, and tells the robot that we are going to learn a new behavior. The robot asks *what is the perceptual trigger of the behavior* and the human responds. The robot then asks *what is the response behavior*, and the human responds again. The robot links the <*percept*, *response*> pair together so that it can be used in the future.

Having human users control and interrogate robots using spoken language results in the ability to ergonomically teach robots. Additionally, it is also useful to execute components of these action sequences conditional on perceptual values. For example the user might want to tell the robot to walk forward until it comes close to an obstacle, using a "command *X* until *Y*" construction, where *X* corresponds to a continuous action (e.g. walk, turn left) and *Y* corresponds to a perceptual condition (e.g. collision detected, ball seen, etc.).

3. RoboCup Soccer: Four-Legged League

In order to demonstrate the generalization of the spoken language human-robot interaction approach we have begun a series of experiments in the domain of RoboCup Soccer (Kitano, 1995), a well documented and standardized robot environment thus provides a quantitative domain for evaluation of success. For this project we have chosen as testing platform the Four-Legged league where ITAM's Eagle Knights team regularly competes (Martínez-Gómez et al. 2005; Martínez-Gómez & Weitzenfeld, 2005). In this league two teams of four robots play soccer on a 6m by 4m carpeted soccer field using Sony's Four-Legged AIBO robots (RoboCup, 2004), as shown in Figure 1. Robots in this league have fully autonomous processing capabilities including a local color-based camera, motors to control leg and head joints, and a removable memory stick where programs can be loaded. The robots also include wireless communication capabilities to interact with other robots in the field as well as computers outside. In addition to the two colored goals, four colored cylinders are used in helping the robots localize in the field. To win, robots need to score as many goals as possible in the opposite goal. Ball is orange and robots use either a red or blue colored uniform. As in human soccer, good teams need to perform better that opponents in order to win, this includes being able to walk faster, process images, localize and kick the ball in a more efficient way, and have more advanced individual behaviors and more evolved team strategies.

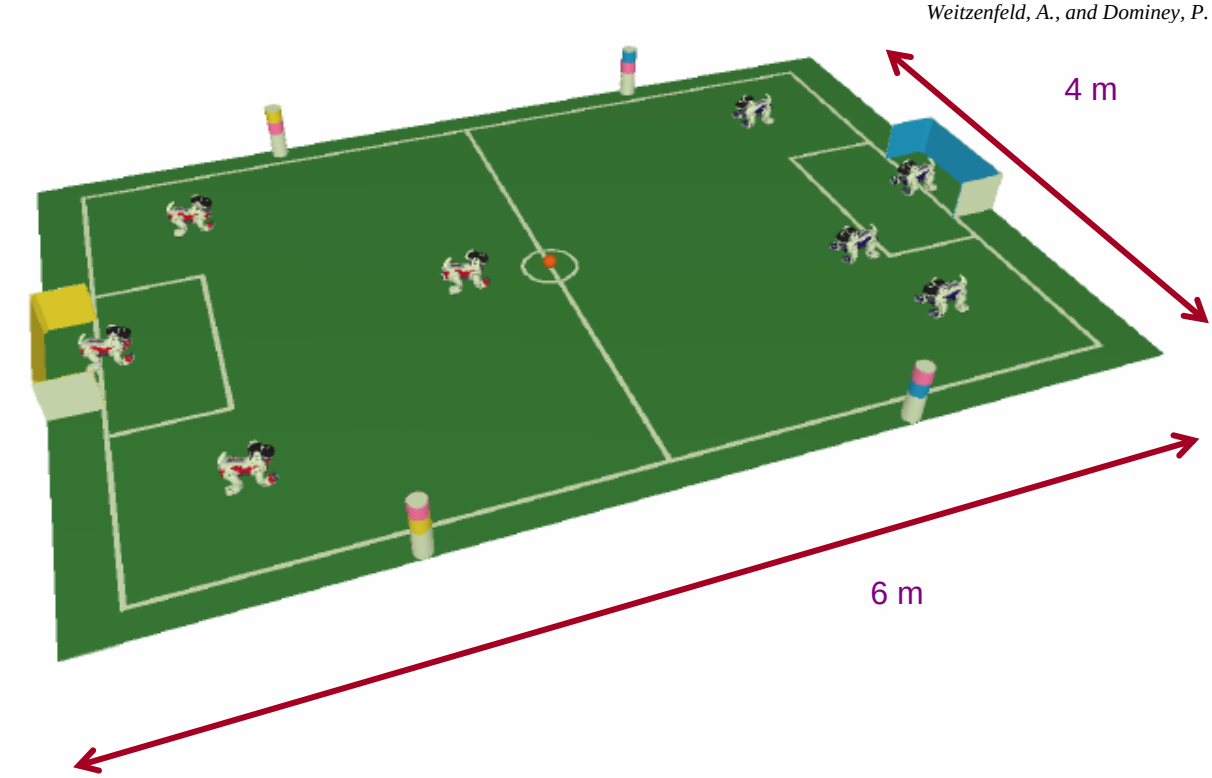

Fig. 1. **Four-Legged Soccer League**. Two teams of four robots play against each other in a 6m by 4m carpet. AIBOs from Sony are used having fully autonomous control. Sensors include a local color-based camera having images processed by a local CPU sending output commands to motors controlling joints in the four legs and heads. The AIBO also includes wireless communication capabilities to interact with other robots in the field and computers outside.

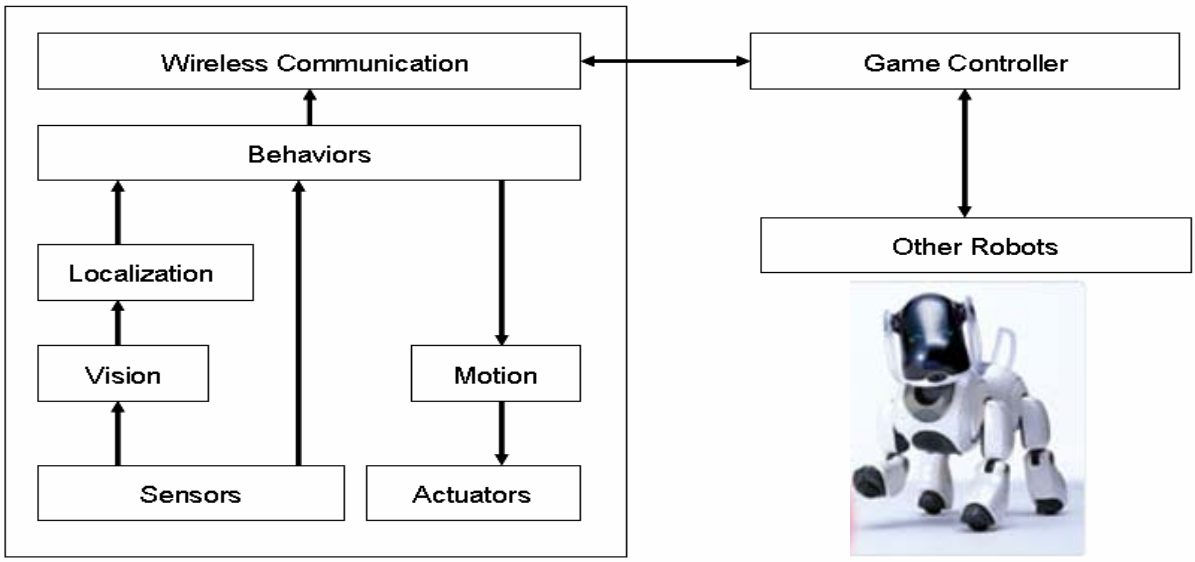

Fig. 2. **Four-Legged System Architecture**. The system includes the following processing components: sensors, actuators, motion, vision, localization, behaviors and wireless communication. The latter is used to share information with other robots through the game controller responsible of informing all robots of the state of the game.

A typical four-legged system architecture shown in Figure 2 consists of the following modules:

- **Sensors**. Primary sensors include a color camera and feedback from motors. The raw camera image is passed to a vision module segmenting objects of interest. Other sensors are used to obtain complementary information on the robot such as joints.
- **Actuators**. The main robot actuators are head and leg motor controlling joints for walking and head turns.
- **Motion**. The motion module is responsible for robot movements, such as walk, run, kick, pass, turn, move the head, etc. It receives commands from the behavior module with output sent to the corresponding actuators representing individual leg and head joint motor control. Robot motions are adapted according to team roles, for example, the goalie has different defensive poses in contrast to other team players. This also applies to different head and ball kicks.
- **Vision**. The vision module receives a raw image from the camera segmenting objects according to color and shape. Objects recognized include ball, robots, cylinder landmarks and goals. More details are given in Section 3.1.
- **Localization**. The localization module uses visual information to provide a reliable localization of the robot in the field. Colored cylinders, goals and white lines are used for this task. More details are given in Section 3.2.
- **Behaviors**. The behaviors module makes decisions affecting individual robots and team strategies. It takes input from sensors and localization system to generate commands sent to motion and actuators modules. Further details are given in Section 3.3.
- **Wireless Communication**. Robots include wireless communication to share information and commands with the external Game Controller or among robots. Data transmitted includes information such as player id, location of ball if seen, distance to the ball, robot position and ball position.

3.1. Vision

The vision module segments incoming camera images to recognize objects of interest from color and shape information. The vision architecture consists of the following stages: image capture, color calibration, segmentation, recognition and identification, as shown in Figure 3. The vision architecture is common to many leagues in RoboCup, including the Mid-size league havign a global viewing camera and the Small-size league having an aereal camera (Martínez-Gómez & Weitzenfeld, 2004).

In order to recognize and identify objects of interest in the image appropriate calibration needs to be performed to adapt to existing lighting conditions. We take initial photos of objects of interest and then select colors that we want to distinguish during segmentation. Figure 4 shows sample output of the segmentation calibration process. Images on the first and third columns are segmented to images in the second and fourth columns, respectively. Colors of interest are orange, green, yellow, pink and blue. All other colors are left as black in the images.

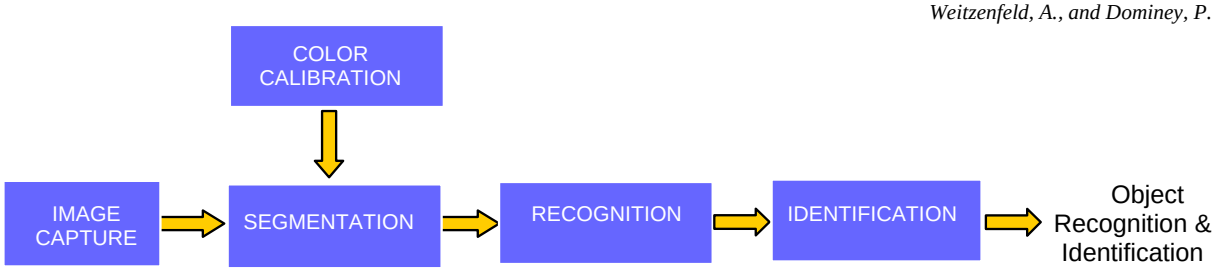

Fig. 3. **Vision Architecture**. The vision module includes the following stages: image capture, color calibration, segmentation, recognition and identification. Output is in the form of recognized and identified objects.

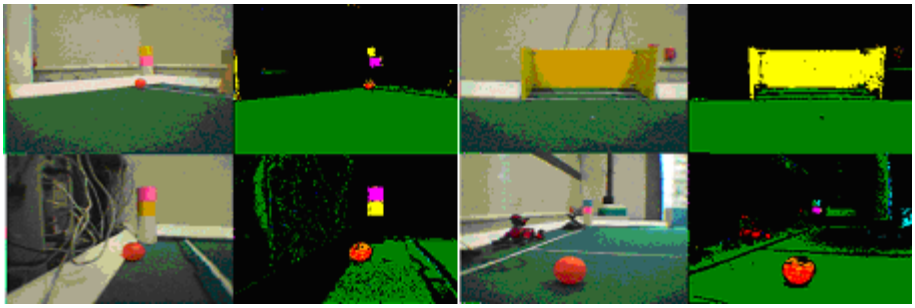

Fig. 4. Color segmentation. Images on the first and third columns are segmented to images in the second and fourth columns, respectively. Colors of interest are orange, green, yellow, pink and blue. All other colors are left as black in these images.

After color regions are obtained, objects are recognized according to certain requirements to allow some confidence that the region being analyzed corresponds to an object of interest. For example, the ball must have green in some adjacent area with a similar criteria used to identify goals. The recognition of landmarks is a little more complex, nevertheless after more elaborated comparison landmarks are identified. Finally, recognized objects are identified depending on color combinations, for example, preestablished landmark "1" and landmark "2".

3.2. Localization

To be successful robots need to localize in the field in an efficient and reliable way. Localization includes computing distances to known objects or landmarks, use of a triangulation algorithm to compute exact positioning, calculate robot orientation angles, and correct any resulting positioning errors, as described in Figure 5.

Distances to Objects. The first step in localizing is to obtain distances from the robot to identified objects in the field. The more objects the robot can distinguish in the field the more reliable the computation. After making several experiments, we developed a simple algorithm that computes distances to objects by using a cubic mathematical relationship that takes as parameter the viewed object area and returns the distance to that object. The distance range tested was from 15 centimeters to 4 meters. Closer than 15 cm or beyond 4 m it becomes harder to compute distances or distinguish between objects and noise, respectively.

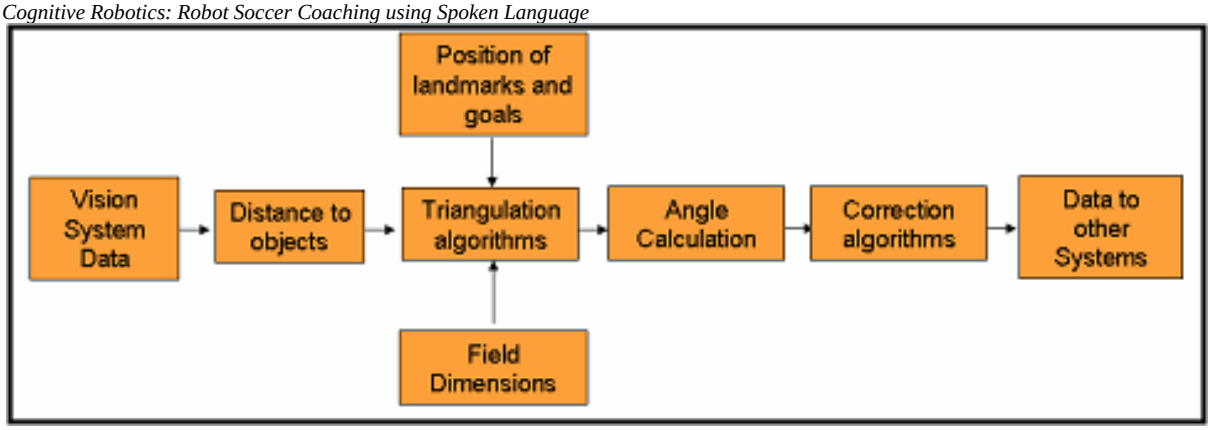

Fig. 5. Localization system block diagram. Localization involves computing distances to known objects in the field, triangulations based on landmarks and goals, angles to landmarks and goals, and computing error corrections to obtain reliable positioning.

Triangulation Algorithm. Following distance computation we apply a triangulation method from two marks to obtain the position of the robot in the field. Triangulation results in a very precise positioning of the robot in a two dimensions plane. If a robot sees one landmark and can calculate the distance to this landmark, the robot could be anywhere in a circumference with origin in the landmark, and radio equal to the distance calculated. By recognizing two landmarks the robot can compute its location from the intersection of two circumferences, as shown in Figure 6. Note that the robot could be in one of two intersection points in the circumferences, although one of these two points will fall outside the field of play.

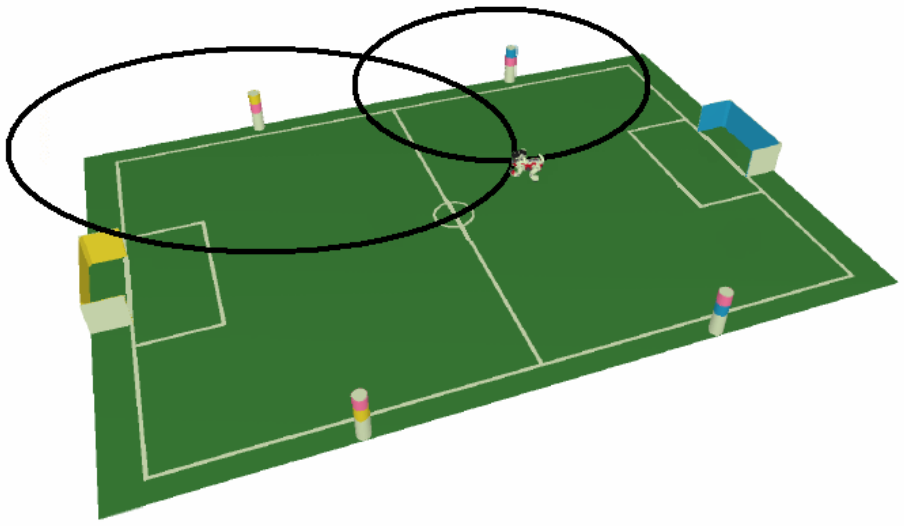

Fig. 6. **Triangulation from two landmarks**. By calculating distance to two different landmarks the robot can compute its position in the field.

Angle Calculation. Once robot position is computed orientation is calculated to complete localization. Two vectors are calculated with origin at the robot pointing to the marks used as references for triangulation, as shown in Figure 7. Orientation calculation is usually more precise than positioning while also being more important to the game. Kicking the ball in the right direction is quite critical to winning.

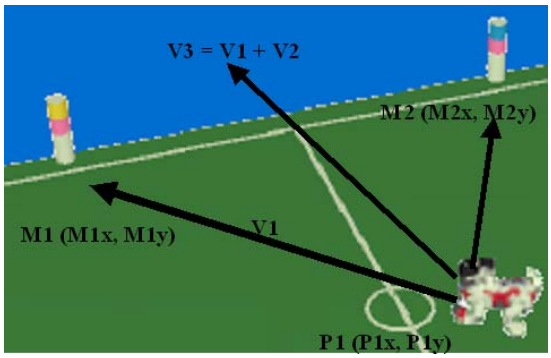

Fig. 7. Robot orientation. By computing vectors to marks (or goals) the robot may calculate its orientation in the field.

Correction Algorithm. Robot positioning resulting from the localization algorithm usually results in inconsistencies between contiguous frames. To stabilize localization computation correction algorithms are necessary starting by smoothing historic measurements. To reduce variation of the output signal for the triangulation algorithm the following filter function shown in equation 1 can be used where *s(x)* represents the updated position value as a function of previously computed *x* position values taking an average over *n* historic samples.

$$s(x) = \frac{\sum_{i=1}^{n} x(i)}{n} \tag{1}$$

Figure 8 shows sample output from this filter correction. Original signals can produce variations of 10% in contiguous positions. By applying the filter this variation can be reduced to less than 3%, see Martínez-Gómez & Weitzenfeld (2005) for more details.

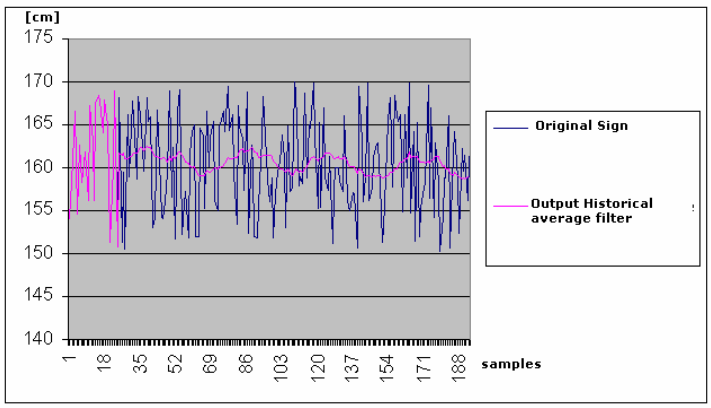

Fig. 8. Correction Algorithm. Abrupt variations in positioning are caused by resolution related errors. To reduce this variability data can be smoothed by averaging with a historic filter.

In addition to variances in positioning, the large size of the robots, around 20cm in length and 10cm in width, requires a positioning precision of at least half the size of the robot. Furthermore, positioning in the field like in human soccer does not require exact knowledge of location as opposed to orientation. For this purpose localization

by field regions can be more effective than knowing exact positioning. In Figure 9 the complete field is divided into twelve similarly sized areas to provide rough localization in the field. During experimentation, localization in some regions resulted in larger errors due to changes in illumination.

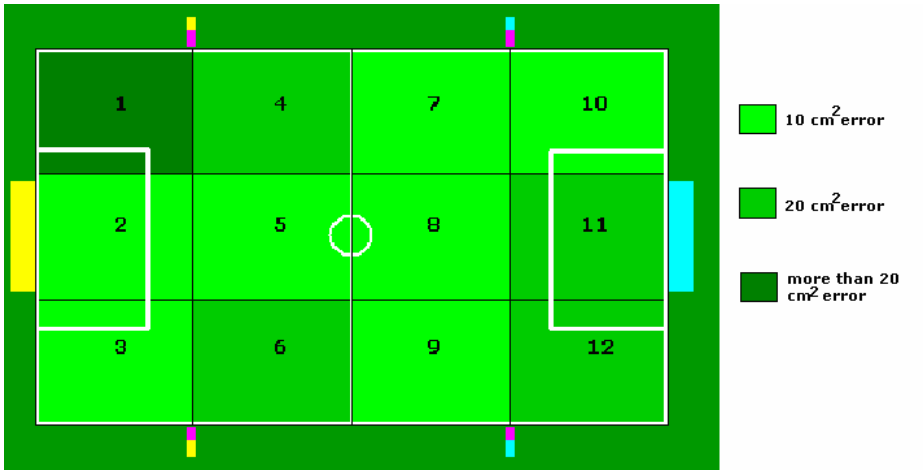

Fig. 9. Localization by regions. While orientation in the field is critical in moving and kicking the ball in the right direction, positioning does not require very high precision as in human soccer. Knowing positioning in relation to certain regions in the field provides enough information for play as shown in the diagram with a 3 by 4 field subdivision.

Regions can be defined in a heterogeneous way as well, e.g. having a specific local goal area while dividing middle field areas with coarser granularity than areas closer to the goals. Also, probabilistic methods not described in the chapter are usually used to localize when occlusions occur during a game.

3.3. Behaviors

The behavior module receives input information from sensors, vision and localization in order to compute individual and team behavior. Output from behavior decisions are sent to motion and actuators. In defining team robot behaviors, we specify different player roles, e.g. Goalkeeper, Attacker, and Defender. Each role behavior depends on ball position, state of game and overall team strategy.

Goalkeeper basic behaviors are described by a state machine as shown in Figure 10:

- **Initial Position**. Initial posture that the robot takes when it is turned on. Depending on its ability to localize robot may autonomously move to its initial position.
- **Search Ball**. An important aspect of the game is searching for the ball around the field. If communication is enabled among robots, searching may be made more efficient by doing this task as a team with individual robots informing others where the ball is when found.
- **Search Goal**. The ultimate objective of the goalkeeper is to defend its own goal. To achieve this it must always know its relative position. Sometimes the

goalkeeper gets away from its area for a special defensive move and needs to return to the goal by searching around.

- **Reach Ball**. After searching and finding the ball, the robot can walk towards it in order to take possession or simply kick the ball. Additional plays include reaching the ball up to certain distance in order to defend the goal from a possible ball kick by an opponent.
- **Reach Goal**. After searching and finding the goal, the goalkeeper moves towards it to relocate on the goal line in the middle of the goal.
- **Kick ball**. The goalkeeper can kick the ball in different ways to get it out of its own goal.

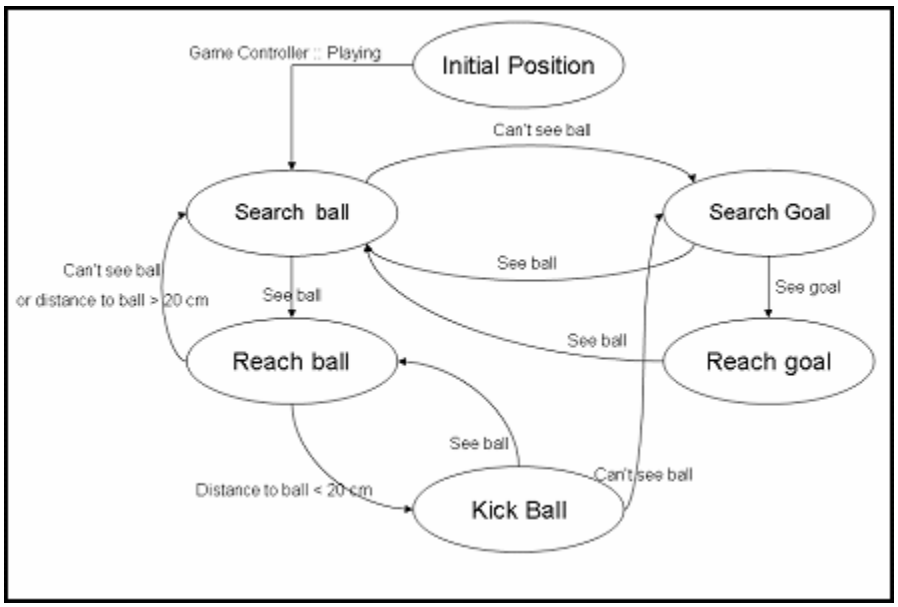

Fig. 10. **Goalkeeper Basic Behavior**. The basic individual goalkeeper state machine includes activities starting from an initial position followed by search ball, search goal, reach ball, reach goal and kick ball.

Attacker and **Defender** basic common individual behaviors are described by a state machine as shown in Figure 11:

- **Initial Position**. Initial posture that the robot takes when it is turned on. Depending on the ability to localize robot may autonomously move to their initial positions.
- **Search Ball**. An important aspect of the game is searching for the ball around the field. If communication is enabled among robots, searching may be made more efficient by doing this task as a team with individual robots informing others where the ball is when found.
- **Reach Ball**. After searching and finding the ball, the robot can walk towards it in order to move with it or kick the ball. Additional plays include turning with the ball and reaching the ball with different orientations.

- **Explore Field**. Exploring the field is a more extended search than the search ball behavior in that it can walk throughout the arena not only looking for the ball but also searching for goals and landmarks.
- **Kick Ball**. The robot can kick the ball in different ways such as to score a goal in the case of an attacker or to simply kick it forwards in the case of a defender.

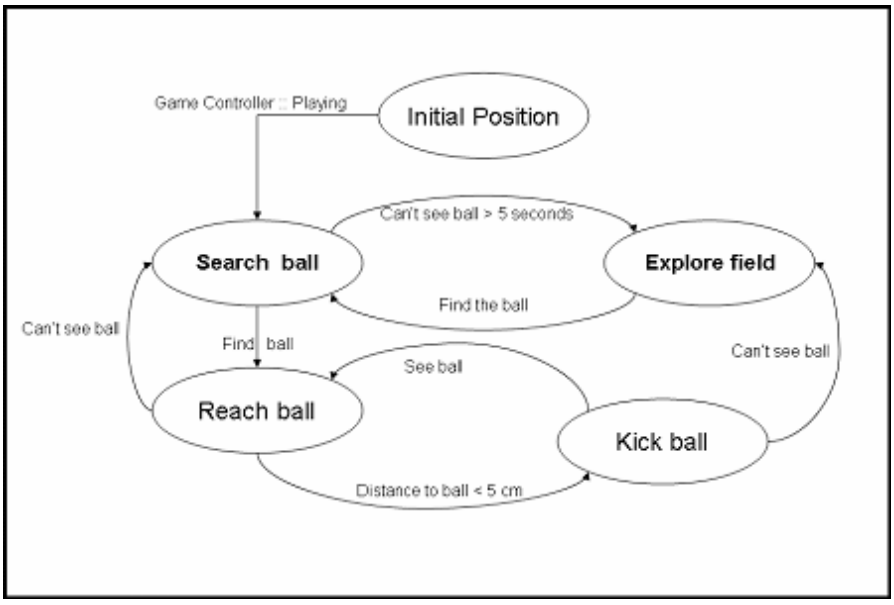

Fig. 11. Attacker and Defender Basic Behavior. The basic individual attacker and defender state machine includes common activities starting from an initial position followed by search ball, explore field, reach ball, and kick ball.

4. Human-Robot Coaching in RoboCup Soccer

While no human intervention is allowed during a RoboCup Four-Legged soccer league game, in the future humans could play a decisive role analogous to real soccer coaches. Coaches could be able to adjust in real-time their team playing strategies according to the state of the game. RoboCup already incorporates a simulated coaching league where coaching agents can learn during a game and then advice virtual soccer agents on how to optimize their behavior accordingly, see (Riley et al., 2002; Kaminka et al. 2002). In this section we describe our most recent work in human-robot interaction with Sony AIBOs.

4.1. Human-Robot Architecture

The human-robot interaction architecture is illustrated in Figure 12. The spoken language interface is provided by the CSLU-RAD framework while communication to the Sony AIBO robots is done in a wireless fashion via the URBI platform (URBI, 2006). The URBI system provides a high level interface for remotely controlling the AIBO. Via this interface, the AIBO can be commanded to perform different actions as well as be interrogated with respect to various internal state variables. Additionally, URBI provides a vision and motion library where higher level perceptions and movements can be specified. (The AIBO architecture shown at the right hand side of Figure 12 describes the robot processing modules previously shown in Figure 2.)

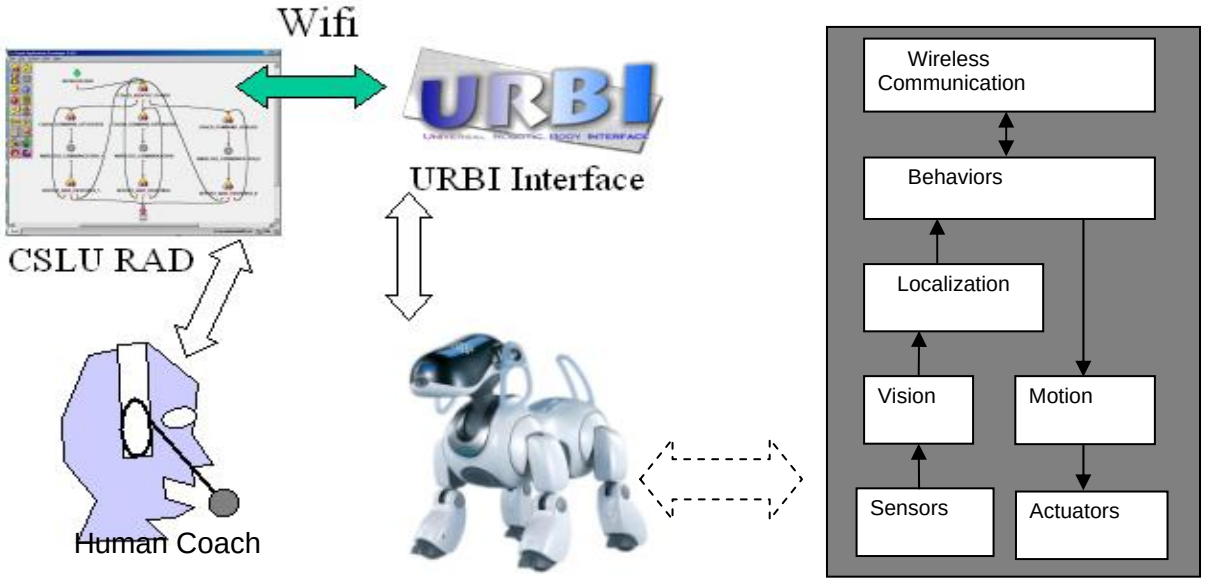

Fig. 12. CSLU-URBI-AIBO system architecture. The left portion of the diagram shows a Human Coach interacting with the CSLU RAD spoken language system that in turns interacts via wireless communication with the URBI interface at the AIBO. The diagram to the right shows the internal AIBO processing modules: Sensors, Actuators, Vision, Motion, Localization, Behaviors and Wireless Communication.

4.2. Command, Interrogate and Teach Dialogs

In order to demonstrate the human coaching model we have developed and experimented with simple dialogs that let the user: (1) *command* the robot to perform certain actions including perception related actions; (2) *interrogate* the robot with specific questions about its state and corresponding perceptions; and (3) *teach* the robot to link a sequence of lower level behaviors into higher level ones.

Command. We define a set of action-only and action perception commands. Action-only commands i.e. no perception, include: *Stop, Move, Turn*, *Turn Head*, and *Kick Ball*. Depending on the commands, these may include arguments such as magnitude of rotation, and movement in degrees or steps, etc. For example a rotation command would be *Turn 180 degrees* and a movement command would be *Move 4 steps*. It should be noted that at this level commands such as *Kick Ball* would not use any perceptual information, i.e. the resulting kick will depend on the current robot orientation. We also define a set of action-perception commands requiring the full perception-action cycle, i.e. the action to be performed depends on the current robot perceptions. These commands include: *Kick Ball* with a specified direction; *Reach Ball* moving to a position behind the ball pointing towards the goal; *Initial Position* during game initialization requiring localization in the field; *Pass the Ball* to gently kick the ball to another team robot; *Move to Location* specifying a position in the field where to move; *Search Ball* resulting in robot looking for a ball nearby; *Explore Field* resulting in a more extensive search for the ball; *Defend Goal* resulting in all robots moving close to the goal requiring knowledge of the robot location in the field; *Defend Kick* in

trying to block a kick from the other team, requiring knowledge of ball location, and *Attack Goal* similar although opposite in behavior to defending goal.

Interrogation. We define state and perception interrogation commands returning information on current actions or behaviors. State interrogations include for example: *What was your last action,* e.g. kicked the ball; *Why did you take the last action,* e.g., I saw the ball, so I moved towards it; *What is your current behavior*, e.g. I'm searching for the ball; *What is your current role in the game*, e.g. I am the goalie. Perception interrogations include for example: *Do you see the ball* returning e.g. *I do*, *I don't*; *What is your distance to the ball*, returning e.g. *30 centimeters*; *What is your current orientation,* returning e.g. *45 degrees* (in relation to field coordinate system); *What is your current position,* returning e.g. *I am in region 9*; *What is the position of object X* returning an estimate of its position.

We have already implemented in the AIBO a number of these commands and interrogations. A sample command and interrogation dialog is shown in Table 2 corresponding to "*Go get the ball and walk it into the goal*".

Robot: *Welcome to the LAFMI Human-Robot Interaction Demonstration. I am the AIBO…*
Human: *Do you see the ball?*
Robot: *Looking for the ball…I don't see anything.*
Human: *Turn right.*
Robot: *Turning right. (Turns right)*
Human: *Do you see the ball?*
Robot: *Looking for the ball… I see the ball.*
Human: *Get the ball.*
Robot: *Going for the ball. (Starts walking to the ball).*
Human: *Stop. Turn right.*
Robot: *Turning right. (Turns right but misses the ball)*
Human: *Back up.*
Robot: *Back up. (Backs up)*
Human: *Turn left.*
Robot: *Turning left. (Turns left but misses the ball)*
Human: *Turn right.*
Robot: *Turning right. (Turns right and catches the ball)*
Human: *Walk.*
Robot: *Walking. (Starts walking, pushing the ball towards the goal)*
Human: *Walk.*
Robot: *Walking. (Keeps walking, pushing the ball and "scoring")*

Table 2. **Sample human-robot dialog**. The dialog shows sample spoken interaction with the human for commanding and interrogating the robot.

Teach. The ultimate goal in human-robot coaching in the context of soccer is being able to positively affect the team performance during a game. While part of this interaction can eventually be carried out by agent coaches inside the robot, it is our

goal to define the basic capabilities and communication interactions that human coaches should have. For example, being able to transmit strategy knowledge in the form "*if blocked pass the ball to player behind*". Such a command will modify an internal robot database with "*if possess(ball) and goal(blocked) then pass(ball)*".

Previous systems allowed the user to use spoken language to teach the AIBO robot the association between a name and a single behavior in the robot's repertoire (Dominey et al. 2005). More recently, we have extended this so that the system can associate a sequence of commands with a new name in a macro-like capability. The limitations of this approach result from the fact that all of the motor events in the sequence are self contained events whose terminations are not directly linked to perceptual states of the system. We can thus teach the robot to walk to the ball and stop, but if we then test the system with different initial conditions the system will mechanically reproduced the exact motor sequence, and thus fail to generalize to the new conditions.

Nicolescu & Mataric (2001, 2003) developed a method for accommodating these problems with a formalized representation of the relations between pre-conditions and post-conditions of different behaviors. In this manner, after the robot has performed a human guided action, such as following the human through an obstacle course and then picking up an object, the system will represent the time ordered list of intervals during which each of the component behaviors is active. From this list, the pre- and post-condition relations between the successive behaviors can be extracted, generalized over multiple training trials, and finally used by the robot to autonomously execute the acquired behavior.

Boucher and Dominey (2006) builds upon these approaches in several important ways. First they enrich the set of sensory and motor primitives that are available to be used in defining new behaviors. Second, they enrich the human-robot interaction domain via spoken language and thus allow for guiding the training demonstrations with spoken language commands, as well as naming multiple newly acquired behaviors in an ever increasing repertoire. Third, they ensure real-time processing for both the parsing of the continuous valued sensor readings into discrete parameterized form, as well as the generalization of the most recent history record with the previously generalized sequence. This ensures that the demonstration, test, correction cycle takes places in a smooth manner with no off-line processing required.

Here is a simple example scenario with the AIBO. In this case the user will teach the robot a form of collision avoidance through demonstration. The user initiates the learning by commanding the robot with a spoken command "turn around" that does not correspond to a primitive command nor to a previously learned command. The robot thus has no knowledge of what to do, and awaits further instructions. The user commands the robot to "march forward" and the robot starts walking. The user sees that the robot is approaching a wall, and tells the robot to stop. He then tells the robot to turn right. Behind and to the right of the robot is the red ball. When the

robot has turned away from the wall and is facing the ball the user tells it to stop turning, and then tells it that the learned behavior demonstration is over.

Now let us consider the demonstration in terms of the commands that were issued by the user, and executed by the robot, and the preconditions that could subsequently be used to trigger these commands. The robot was commanded to "turn around." Because it had no representation for this action, it awaited further commands. The robot was then commanded to walk. Before it collided with the wall the robot was commanded to stop walking. It was then commanded to turn right, and to stop when it was in front of the red ball. Now consider the perceptual conditions that preceded each of these commands, which could be used in a future automatic execution phase to sequentially trigger the successive commands. The pertinent precondition to start walking was that the command to "turn around" had been issued. The pertinent precondition to stop walking was the detection of an obstacle in the "near" range by the distance sensor in the robots face. The pertinent preconditions for subsequently turning right are that the robot is near something, and that it has stopped walking.

The goal then is for the system to encode the temporal sequence of all relations (which include user commands and perception values) in a demonstration run, and then to determine what are the pertinent preconditions for each commanded action relation. Likewise, it may be the case that perceptual relations were observed during the demonstration that were not pertinent to the behavior that the human intended to teach the robot. The system must thus also be able to identify such "distractor" perceptions that occurred in a demonstration, and to eliminate these relations from the generalized representation of the behavioral sequence.

5. Conclusions

The stated objective of the current research is to develop a generalized approach for human-machine interaction via spoken language that exploits recent developments in cognitive science - particularly notions of grammatical constructions as form-meaning mappings in language, and notions of shared intentions as distributed plans for interaction and collaboration. In order to do this, we tested human-robot interaction initially with the Event Perceiver system and later on with the Sony AIBOs under soccer related behaviors. We have presented the system architecture for the Eagle Knights Four-Legged team as a testbed for this work.

With respect to social cognition, shared intentions represent distributed plans in which two or more collaborators have a common representation of an action plan in which each plays specific roles with specific responsibilities with the aim of achieving some common goal. In the current study, the common goals were well defined in advance (e.g. teaching the robots new relations or new behaviors), and so the shared intentions could be built into the dialog management system. We plan to continue this work by experimenting with more evolved behaviors in testing full

coaching capabilities in the soccer scenario. Videos for several human-robot dialogs, including the previous one, can be found in Dominey & Weitzenfeld (2006).

Acknowledgements

This work has been supported in part by the French-Mexican LAFMI, the ACI TTT Projects in France and grants from UC-MEXUS CONACYT, CONACYT (grant #42440) and "Asociación Mexicana de Cultura A.C." in Mexico.

6. References

Bates E, McNew S, MacWhinney B, Devescovi A, and Smith S, (1982) Functional constraints on sentence processing: A cross linguistic study, Cognition (11): 245-299.

Boucher J-D, Dominey PF (2006) Perceptual-Motor Sequence Learning via Human-Robot Interaction, Proceedings of the The Ninth International Conference on the Simulation of Adaptive Behavior

CSLU, (2006) Speech Tools Rapid Application Development (RAD), http://cslu.cse.ogi.edu/toolkit/index.html.

Dominey PF and Boucher JD, (2005a) Developmental stages of perception and language acquisition in a perceptually grounded robot, Cognitive Systems Research, Volume 6, Issue 3, Pages 243-259, September.

Dominey PF and Boucher JD, (2005b) Learning to talk about events from narrated video in a construction grammar framework, Artificial Intelligence, Volume 167, Issues 1-2, Pages 31-61, September.

Dominey PF, Hoen M, Lelekov T and Blanc JM, (2003) Neurological basis of language in sequential cognition: Evidence from simulation, aphasia and ERP studies, Brain and Language, 86(2):207-25.

Dominey PF and Inui T, (2004) A Developmental Model of Syntax Acquisition in the Construction Grammar Framework with Cross-Linguistic Validation in English and Japanese, Proceedings of the CoLing Workshop on Psycho-Computational Models of Language Acquisition, Geneva, 33-40.

Dominey PF and Weitzenfeld A, (2005) Robot Command, Interrogation and Teaching via Social Interaction, IEEE-RAS International Conference on Humanoid Robots, Dec 6-7, Tsukuba, Japan.

Dominey PF and Weitzenfeld A, (2006) Videos for command, interrogate and teach AIBO robots, ftp://ftp.itam.mx/pub/alfredo/COACHING/

Goldberg A, (1995) Constructions. U Chicago Press, Chicago and London.

Kaminka G, Fidanboylu M, Veloso M, (2002) Learning the Sequential Coordinated Behavior of Teams from Observations. In: RoboCup-2002 Symposium, Fukuoka, Japan, June.

Kitano H, Asada M, Kuniyoshi Y, Noda I, and Osawa, E., (1995) Robocup: The robot world cup initiative. In Proceedings of the IJCAI-95 Workshop on Entertainment and AI/ALife.

Kotovsky L and Baillargeon R, (1998) The development of calibration-based reasoning about collision events in young infants , Cognition, 67, 311-351.

Martínez-Gómez, L.A., and Weitzenfeld, A., (2004) Real Time Vision System for a Small Size League Team, Proc. 1st IEEE-RAS Latin American Robotics Symposium, ITAM, Mexico City, October 28-2.

Martínez-Gómez JA, Medrano A, Chavez A, Muciño B and Weitzenfeld, A., (2005) Eagle Knights AIBO Team, Team Description Paper, VII World Robocup 2005, Osaka, Japan, July 13-17.

Martínez-Gómez J.A and Weitzenfeld A, (2005) Real Time Localization in Four Legged RoboCup Soccer, Proc. 2nd IEEE-RAS Latin American Robotics Symposium, Sao Luis Maranhao Brasil, Sept 24-25.

Nicolescu M.N., Matari M.J. (2003) Natural Methods for Robot Task Learning, Proc. AAMAS '03.

Nicolescu M.N., Mataric M.J. (2003) Learning and Interacting in Human-Robot Domains, IEEE Trans. Sys. Man Cybernetics B, 31(5) 419-430.

Riley P, Veloso M, and Kaminka G, (2002) An empirical study of coaching. In: Distributed Autonomous Robotic Systems 6, Spring-Verlag.

RoboCup, (2004) "Sony Four Legged Robot Football League Rule Book.", RoboCup Official Web Site URL:http://www.robocup.org, May.

Siskind JM, (2001) Grounding the lexical semantics of verbs in visual perception using force dynamics and event logic. Journal of AI Research (15) 31-90.

Steels L and Baillie JC, (2002) Shared Grounding of Event Descriptions by Autonomous Robots. Robotics and Autonomous Systems, 43(2-3):163—173.

Tomasello M, (2003) Constructing a language: A usage-based theory of language acquisition. Harvard University Press, Cambridge.

Tomasello M, Carpenter M, Call J, Behne T, Moll H, (2006) Understanding and sharing intentions: The origins of cultural cognition, Behavioral and Brain Sciences.

URBI (2006) http://www.urbiforge.com/

Weitzenfeld A and Dominey PF, (2006) Cognitive Robotics: Command, Interrogation and Teaching in Robot Coaching, RoboCup Symposium 2006, June 19-20, Bremen, Germany.