E. Bicho, L. Louro, N. Hipólito, W. Erlhagen, "*A dynamic field approach to goal inference and error monitoring for human-robot interaction*", In: Dautenhahn, E. (Ed.), Proceedings of the 2009 **International Symposium on New Frontiers in Human-Robot Interaction.** AISB (Adaptive & Emergent Behaviour & Complex Systems) Covention, pp. 31-37, Heriot-Watt University, Edinburgh, Scotland, 8-9 April, 2009.

A dynamic field approach to goal inference and error monitoring for human-robot interaction[1]

Estela Bicho[2] and **Luís Louro**[3] and **Nzoji Hipólito**[4] and **Wolfram Erlhagen**[5]

Abstract. In this paper we present results of our ongoing research on non-verbal human-robot interaction that is heavily inspired by recent experimental findings about the neuro-cognitive mechanisms supporting joint action in humans. The robot control architecture implements the joint coordination of actions and goals as a dynamic process that integrates contextual cues, shared task knowledge and the predicted outcome of the user's motor behavior. The architecture is formalized by a coupled system of dynamic neural fields representing a distributed network of local but connected neural populations with specific functionalities. We validate the approach in a task in which a robot and a human user jointly construct a toy 'vehicle'. We show that the context-dependent mapping from action observation onto appropriate complementary actions allows the robot to cope with dynamically changing joint action situations. This includes a basic form of error monitoring and compensation.

1 INTRODUCTION

As robot systems are moving as assistants into human everyday life, the question how to design robots capable of acting as sociable partners in collaborative joint activity becomes increasingly important ([4], [12]). Useful and efficient human-robot interaction requires that both teammates coordinate and synchronize their actions and decisions in a shared task. In order to decrease the workload of the human and to increase user satisfaction, the robot should equally contribute to this coordination effort. This necessarily means that the robot should be endowed with cognitive capacities such as action understanding and goal inference. Humans achieve their remarkable fluent organization of joint action by anticipating the intentions of others [21]. In our everyday social interactions we continuously monitor the actions of our partners, interpret them effortlessly in terms of their outcomes and use these predictions to select adequate complementary behaviours. Very often this happens without the need for explicit verbal communication. Imagine for the instance the joint action task of preparing a dinner table. The way how a partner grasps a certain object, e.g., a coffee cup, transmits to the observer important information about the ultimate goal of the action. Depending on the grip type, the partner may want to place the cup on the table or, alternatively, has the intention to hand it over. Being able to predict the goal of the whole action sequence at the time of the grasping allows the observer to timely prepare for receiving the cup, or to initiate the selection of another object for the dinner table.

This paper presents our ongoing research towards creating socially intelligent robots that are able to flexibly adjust their goal-directed behaviours in dependence of the predicted outcomes of actions of their human partners [3]. Our approach is heavily inspire by recent experimental and theoretical findings about the neuro-cognitive mechanisms underlying joint action in humans and other primates ([18], [25]). We believe that designing cognitive control architectures on the basis of these mechanisms defines a very promising research direction to reduce the significant imbalance in social and cognitive skills between human and robot that still exists today. Ultimately, implementing a human-like joint action model in the robot will contribute to more natural HRI since the teammates will become more predictable for each other. This in turn will increase the acceptance by humans. A recent HRI user study with a simulated robotic teammate revealed that anticipatory action selection seems to be a natural expectation of a robotic assistant in known joint action tasks [17]. The robot is perceived as a full partner that contributes to the team's fluency and success only if it acts in anticipation of the needs of the human user.

Several neuro-cognitive mechanisms that are believed to underlie successful human joint action define fundamental components of the robot control architecture. An impressive body of experimental evidence from studies investigating action and perception in a social context suggests that motor simulation routines in the brain support the understanding of other's actions and facilitate overt imitation [25]. The fundamental idea is that perceived actions are automatically mapped onto corresponding motor representation of the observer to predict or replicate the action effect. Over the last couple of years, the suggested close perception-action link has inspired robotics work mainly in the domain of learning by imitation and social development (e.g., [5], [1], [10], [15], [19]). For implementing a high-level goal inference capacity in the context of HRI, it is important that the matching takes place on a sufficiently abstract level related to the goal or desired end state of an action sequence. This allows the robot to predict actions of the teammate despite the obvious differences in embodiment and motor skills between human and robot. However, for action selection in cooperative tasks an automatic and direct resonance of matching motor structures is in general not beneficial. Normally, action observation should facilitate the selection of a non-imitative, complementary behaviour. Moreover, to cope with dynamically changing joint action conditions, the decision about what defines the most adequate complementary action

[1] The present research was conducted in the context of the fp6-IST2 EU-project JAST (proj.nr. 003747) and partly financed by the FCT grants POCI/V.5/A0119/2005 and CONC-REEQ/17/2001.

[2] Dept of Industrial Electronics, University of Minho, Portugal. Email: estela.bicho@dei.uminho.pt.

[3] Dept of Industrial Electronics, University of Minho, Portugal. Email: llouro@dei.uminho.pt.

[4] Dept of Industrial Electronics, University of Minho, Portugal. Email: nhipolito@dei.uminho.pt.

[5] Dept of Mathematics for Sciences and Technology, University of Minho, Portugal. Email: wolfram.erlhagen@mct.uminho.pt.

should depend on additional contextual cues. Recent evidence from neurophysiological and behavioural studies shows that the automatic mapping from action observation onto action execution is indeed more complex and flexible as previously thought ([18], [23]). The robot control architecture reflects these findings by implementing a context-dependent mapping that is biased by the inferred goal of the human user.

As a theoretical framework for the high-level control of the robot we have used the Dynamic Neural Field (DNF) approach to robotics [8]. Originally introduced as a simplified mathematical model for pattern formation in neural populations [2], DNFs have been later generalized and applied to the cognitive domain (for a recent review see [20]). The architecture of DNFs reflects the hypothesis that strong recurrent interactions in local populations of neurons form a basic mechanism for cortical information processing. These interactions support the existence of self-stabilized inner states that allow the cognitive agent for instance to compensate for temporally missing sensory input, or to anticipate future environmental inputs that may inform the decision about a specific goal-directed behaviour. The DNF-based model for joint action consists of a distributed network of reciprocally connected neural populations that represent in their activation patterns specific task-relevant information. It implements the idea that the coordination of actions and decisions among the teammates is a dynamic process that builds on the continuous integration of input from representations of the inferred goal of observed actions, contextual cues and shared task knowledge. The representation of the complementary action that gets the strongest support from all connected populations will win the dynamic competition process among all possible actions.

The dynamic field architecture has been validated in a joint construction task in which the human-robot team assembles a toy 'vehicle' from its components knowing the construction plan. The study differs from conceptually related HRI work [17] in the sense that the robot is not only serving the user (e.g., holding out pieces for the user) but is able to perform itself the assembly task. This symmetric situation challenges the joint coordination of decisions and actions. The focus of the results reported here is on successful trials in which the robot shows anticipatory action selection. Since coordination and other errors may occur even in tasks that are well known to the teammates [22], performance monitoring and error detection is another topic that we have addressed. In the present implementation, complementary action selection in error trials may range from simple head nodding to pointing. In addition, we have integrated and tested a speech production system that allows the robot not only to explain the error in some more detail but to send in general feedback about its reasoning to the user.

The paper is organized as follows: Section 2 introduces the joint construction task and the robotic platform. Section 3 gives an overview about the cognitive control architecture. Section 4 presents the basic concepts of the dynamic field framework. The results of the human-robot interactions are described in section 5. The paper ends with a discussion of concepts, results and a short outlook.

2 JOINT CONSTRUCTION TASK

To validate the dynamic field architecture for human-robot interaction we have chosen the joint construction of a toy 'vehicle'from components that are initially distributed on a table. (Figure 1). The task requires only a limited number of different motor actions to be performed by the team but is complex enough to show the impact of action monitoring and evaluation on action selection. The components that have to be manipulated by the robot were designed to limit the workload for the vision and the motor system of the robot. The toy object consists of a round platform with an axle on which two wheels have to be attached and fixed with a bolt. Subsequently, 4 columns have to be plugged into holes in the platform. The placing of another round object on top of the columns finishes the task. It is assumed that each teammate is responsible to assemble one side of the toy. Since the working areas of the human and the robot do not overlap, the spatial distribution of components on the table obliges the team to coordinate in addition handing-over sequences. It is further assumed that both partners know the construction plan and keep track of the subtasks which have been already completed by the team. Since the desired end state does not uniquely define the logical order of the construction, at each stage of the construction the execution of several subtasks may be simultaneously possible. The main challenge for the team is thus to efficiently coordinate in space and time the decision about actions to be performed by each of the teammates.

Figure 1. Human-robot team for the joint construction of a toy 'vehicle'. The vehicle consists of a (red) round platform with an axle where two (green) wheels have to be mounted and fixed with (magenta) bolts.

For the HRI experiments we used a robot built in our lab. It consists of a stationary torus on which a 7 DOFs AMTEC arm (Schunk GmbH) with a 3-fingered BARRET hand (Barrett Technology Inc.) and a stereo camera head are mounted. A speech synthesizer (Microsof Speech SDK 5.1) allows the robot to communicate the result of its reasoning to the human user. For the control of the arm-hand system we applied a global planning method in posture space that allows us to integrate optimization principles derived from experiments with humans [10]. The information about object type, position and pose is provided by the camera system. The object recognition combines color-based segmentation with template matching derived from earlier learning examples [24]. The same technique is also used for the classification of object-directed, static hand postures such as grasping and communicative gestures such as pointing or demanding an object.

3 COGNITIVE CONTROL ARCHITECTURE

Figure 2 presents a sketch of the multi-layered robot control architecture for dynamic decision making and performance monitoring in joint action that is based on known neuro-cognitive mechanisms. Ultimately, the architecture implements a context-dependent mapping between observed action and executed action. The fundamental idea is that the mapping takes place on the level of abstract motor primitives defined as whole goal-directed motor acts

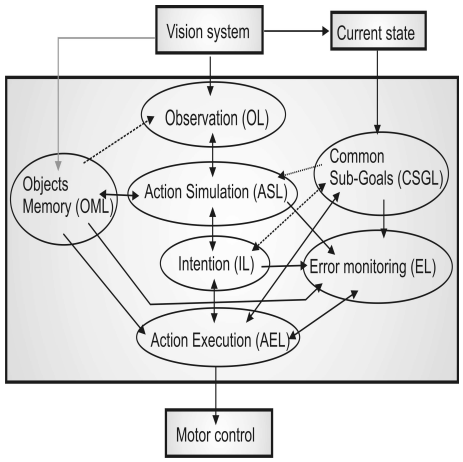

Figure 2. Cognitive control architecture for joint action. It implements a mapping from observed actions (layer OL) onto complementary actions (layer AEL) taking into account the inferred action goal of the partner (layer IL), detected errors (layer EL), contextual cues (layer OML) and shared task knowledge (layer CSGL). The goal inference capacity is based on motor simulation (layer ASL).

like reaching, grasping, placing, attaching or plugging an object [19]. An observed hand movement that is recognized by the vision system as a particular primitive is represented in the action observation layer (OL). The action simulation layer (ASL) encodes entire chains of action primitives that are in the motor repertoire of the robot (e.g., reaching-grasping-placing/plugging a particular object). These chains are linked to representations of specific goals or end states (e.g., attach right wheel to base) in the intention layer (IL). The basis of the goal inference capacity is the activation of a particular chain and its associated goal during action observation. It is important to stress that due to the self-stabilizing properties of the chain representations, goal inference is possible even if the action sequence performed by the human is only partially observable [11]. The object memory layer (OML) encodes the memory about the position of objects in each of the working areas. The common subgoal layer (CSGL) contains the information about currently active and future subgoals as well as memorized information about subtasks which have been already completed by the team. The construction plan is encoded in the connections between neural populations in 3 different layers representing past, current and future subtasks, respectively. The representations are updated in accordance with the construction plan and real or anticipated feedback from the vision system and/or layer IL. The error monitoring layer (EL) represents a detected discrepancy between the inferred goal of the human partner and the subgoals that are currently available. This error-related activity is functionally relevant since it is linked to representations of compensatory behaviour in the action execution layer (AEL). This layer integrates input from IL, OML, CSGL and EL to select among all possible action sequences the most appropriate complementary sequence. It is worth noting that this layer contains also the representation of an 'action' linked to the speech synthesizer that allows the robot to verbally inform the user about its reasoning.

4 BASIC CONCEPTS OF THE DYNAMIC NEURAL FIELD FRAMEWORK

Each layer of the distributed control architecture is formalized by one or more Dynamic Neural Fields (DNFs). DNFs implement the idea that task-relevant information about action goals, motor primitives or context is encoded by means of activation patterns of local pools of neurons. These patterns are initially triggered by transient input from connected populations and sources external to the network. They may become self-sustained in the absence of any external input due to the recurrent interactions within the population. Functionally, these patterns may thus serve a working memory function. We employed a particular form of a DNF first analyzed by Amari (1977) [2]. In each model layer i, the activity $u_i(x,t)$ at time t of a neuron at field location x is described by the following integro-differential equation (for an overview about analytical results see [8]):

$$\tau_{\mathrm{i}}\frac{\delta u_{\mathrm{i}}(x,t)}{\delta t} = -u_{\mathrm{i}}(x,t) + S_{\mathrm{i}}(x,t) + \int w_{\mathrm{i}}(x-x')f_{\mathrm{i}}(u_{\mathrm{i}}(x',t))dx' + h_{\mathrm{i}} \quad (1)$$

where the constants $\tau_i > 0$ and $h_i < 0$ define the time scale and the resting level of the field dynamics, respectively. The integral term describes the intra-field interactions. It is assumed 1) that the interaction strength, $w(x,x')$, between any two neurons x and x' depends only on the distance between locations, and 2) that nearby cells excite each other, whereas separated pairs of cells have a mutually inhibitory influence. For the present implementation we used the following integral kernel of lateral-inhibition type:

$$w(x) = A\exp(-x^2/2\sigma^2) - w_{inhib} \quad (2)$$

where $w_{inhib} > 0$ is a constant and $A > 0$ and $\sigma > 0$ describe the amplitude and the standard deviation of a Gaussian, respectively. Only sufficiently activated neurons contribute to interaction. The threshold function $f(u)$ is chosen of sigmoidal shape with slope parameter β and threshold u_0:

$$f_{\mathrm{i}}(u_{\mathrm{i}}) = \frac{1}{1+exp\left(-\beta(u_{\mathrm{i}}-u_0)\right)}. \quad (3)$$

The model parameters are adjusted to guarantee that the field dynamics is bi-stable, that is, the attractor state of a localized activation pattern coexists with a stable homogenous activation distribution that represents the absence of specific information. If the summed input to a local population is sufficiently strong, the homogeneous state loses stability and a localized pattern evolves. Weaker external signals lead to a subthreshold, input-driven activation patten in which the contribution of the interactions is negligible. Normally, a constant input from a single population does not drive directly connected populations. It may play nevertheless an important role for the processing in the joint action circuit. The preshaping by weak input brings populations closer to the threshold for triggering the self-sustaining interactions and thus biases the decision processes linked to behavior. Much like prior distributions in the Bayesian sense, multi-modal patterns of subthreshold activation in for instance the action execution layer (AEL) may represent the probability of different complementary actions [6].

The summed input from connected fields $\mathrm{u_j}$ is given as $S_{\mathrm{i}}(x,t) = k\sum_{\mathrm{j}} f_{\mathrm{j}}(u_{\mathrm{j}}(x,t))$. The parameter k scales the total input relative to the threshold for triggering a self-sustained pattern.

This guarantees that the inter-field coupling is weak and the field dynamics is dominated by the recurrent interactions. The external inputs from the vision system to layers OL and MOL that initiate the dynamic interplay of the different populations in the network are modeled as Gaussian functions.

The existence of a single self-stabilized pattern of activation in a dynamic field is closely linked to decision making. In layers ASL, IL and AEL subpopulations encoding different chains (ASL), goals (IL) and complementary actions (AEL), respectively, interact through lateral inhibition. This inhibitory interaction leads to the suppression of activity below threshold in competing neural pools whenever a certain subpopulation becomes activated above threshold. To represent and memorize simultaneously 1) the location of several objects of a certain type, and 2) multiple common subgoals, the interaction kernels in layers OML and CSGL were adapted to allow for the existence of multiple patterns of activation. [8]. OML contains individual fields for each of the object classes. They are labeled by the workspace to which the object belongs, that is, each class is represented by two separate fields.

5 RESULTS

In the following we validate the dynamic field architecture by presenting snapshots of the human-robot interaction in the construction task. The examples illustrate the impact of action observation on action selection in varying context from the perspective of the robot. The videos of the human-robot interaction and the associated dynamics of the fields can be found at
http://dei-s1.dei.uminho.pt/pessoas/estela/JASTvideos.htm.
In each case, the robot is supposed to know what common subgoals are currently active and can be selected by the team. In the present implementation real or anticipated feedback about accomplished subtasks trigger directly through hand-coded connections the population representing subsequent assembly steps.

5.1 Anticipatory action selection

The capacity to simulate the consequences of observed actions allows the robot to act in anticipation of the partner's motor intentions. Depending on the predicted outcome of the ongoing action, a sociable robot may for instance decide to already prepare for an action that best serves future needs of the user. Within the dynamic field architecture this is possible since current subgoals of the team are updated based on the inferred goal. The anticipatory action selection is illustrated in Fig. 3. All components are distributed on the table and the user starts the joint construction by grasping a wheel (green object) from above (full grip). The robot has sequences in its motor repertoire that associate the type of grasping with specific goals. A grasping from above is used to attach a wheel to the platform whereas using a side grip is the most comfortable and secure way to hand the wheel over to the teammate. The observation of the full grip (represented in OML) triggers an activation peak in ASL that represents the simulation of the respective reaching-grasping-plugging chain. Since attaching a wheel on the side of the user is a current subgoal for the team, the inputs from layers ASL and CSGL automatically activate the representation of that goal in the intention layer (IL). The existence of this activation pattern initiate a dynamic updating process in layer CSGL (not shown here). The peak representing the subgoal "attach wheel" disappears and an activation pattern representing the new subgoal "fix wheel with bolt" evolves. Since all bolts (magenta objects) are in the workspace of the robot, the inputs from layers OML and CSGL converge on a population in the action execution layer (AEL) that represents a decision for a "hand over bolt" sequence as a complementary behaviour of the robot. As can be seen in the activation pattern of layer AEL, the possible alternative to select a wheel in it's working area with the goal to attach it is also represented by a weaker, subthreshold peak. The decision to serve the user first is the result of small biases in the connection strengths to the populations in CSGL that favor the subtasks and intentions of the user over the subtasks to be realized by the robot. For HRI this offers interesting perspectives since a simple adjustment of these weights will affect how social the robot companion behaves.

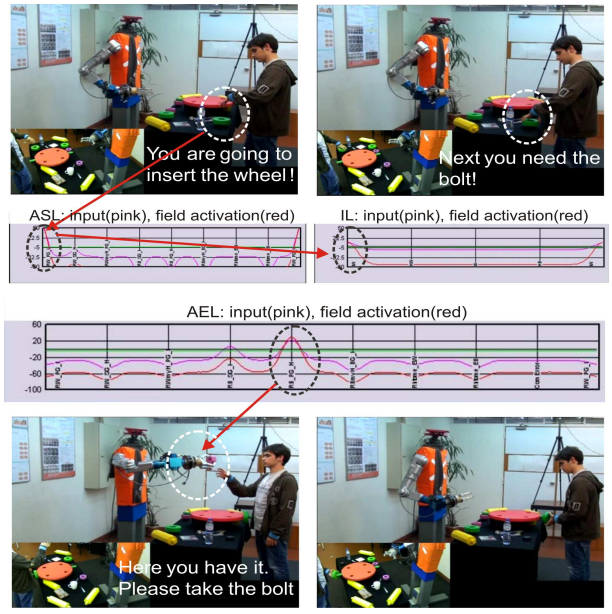

Figure 3. Anticipatory action selection. The human reaches and grasps the wheel from above. The robot infers that the human is going to attach it to the platform. The robot decides to grasp the wheel for handing it over since the wheel is the next component the human will need. The green line in the plots indicates the resting level of the field dynamics.

5.2 Impact of shared task knowledge and context

Very often motor simulation alone is not sufficient to read the motor intentions of the human user. The integration of shared task knowledge is equally important for the decision process [21]. This is illustrated in panel A and B of Fig. 4 In both situations the user reaches his open hand towards the robot. The robot has this gesture which is associated with the goal "request object" in its motor repertoire, but needs additional information from the common subgoal layer (CSGL) to disambiguate what object the human user is requesting. In panel A the self-stabilized bi-modal activation pattern in CSGL indicates that the two wheels have still to be attached. Since both wheels are located in it's workspace, the robot is able to infer that the user is asking for wheel to attach it (compare the peak in IL). The inputs from layers OML and IL activate a population representation in AEL representing the handing-over sequence. In panel B, the user shows again the requesting gesture. However, the state of the construction

and therefor the current subgoals for the team have changed. The bi-modal activation pattern in CSGL now represents the information that the two wheels have to be fixed with bolts. The robot is thus able to infer what object the human wants. A possible complementary action is again to serve the user by handing over a bolt. However, the robot decides instead to attract the attention of the user to the fact that he has a bolt in his working area. The robot performs a pointing gesture in the direction of the bolt. This action selection, which overrides the prepotent tendency to satisfy a user request, is possible because of additional input from the error layer (EL). Population activity in the object memory layer representing the bolt in the workspace of the user together with the input from the population encoding the inferred goal in IL automatically activate a self-stabilized peak in EL. This pattern is associated with the pointing gesture and generates the strongest input to AEL.

5.3 Error detection

The last example shows that even in well known joint action tasks the user can easily make errors that should be compensated by the teammate if possible. Errors may occur for different reasons. The user may have overlooked an object or may be confused about the state of the construction and its temporal order. Different error categories [22] affect joint action on different levels (e.g., error in intention versus error in the selection of action means). The following example illustrates a case in which the robot detect a mismatch between the inferred intention of the user and the state of the construction, that is, between the intention and possible subgoals. In panel A of Fig. 5, the robot observes the human user grasping a wheel from the side which it interprets via action simulation as belonging to a handing over sequence. However, on the side of the robot the wheel is already attached. The information about the already accomplished subtask is memorized by a self-stabilized activation peak in CSGL (compare the snapshot of "past" field). Input from this field together with input from the intention layer (IL) trigger the emergence of a suprathreshold activation pattern in the error layer (EL). In this case, the error related activity is linked with a population in AEL that initiates speech to explain the nature of the error to the user. The content of the speech combines the information represented in the activation patterns that have initially triggered the error-related activity ("I do not need a wheel since a wheel is already attached on my side"). The example in panel B shows that the information represented in the activation patterns of the various populations of the distributed network can be used to give an even more detailed explanation of a detected error. In this case, the user holds out a bolt to hand it over so that the robot may fix the wheel on its side. Since both wheels have been attached but not yet fixed, the inferred goal is valid. However, the activation pattern in the object memory layer indicates that the robot does not need a bolt since it has one in its working area. Moreover, the user still has to use the bolt himself to continue the assembly of the toy on his side. The activation peak that evolves in EL in response to the converging inputs from layers OML and CSGL controls the speech output. The robot refuses the offered object and informs the user in addition about the missing bolt on his side.

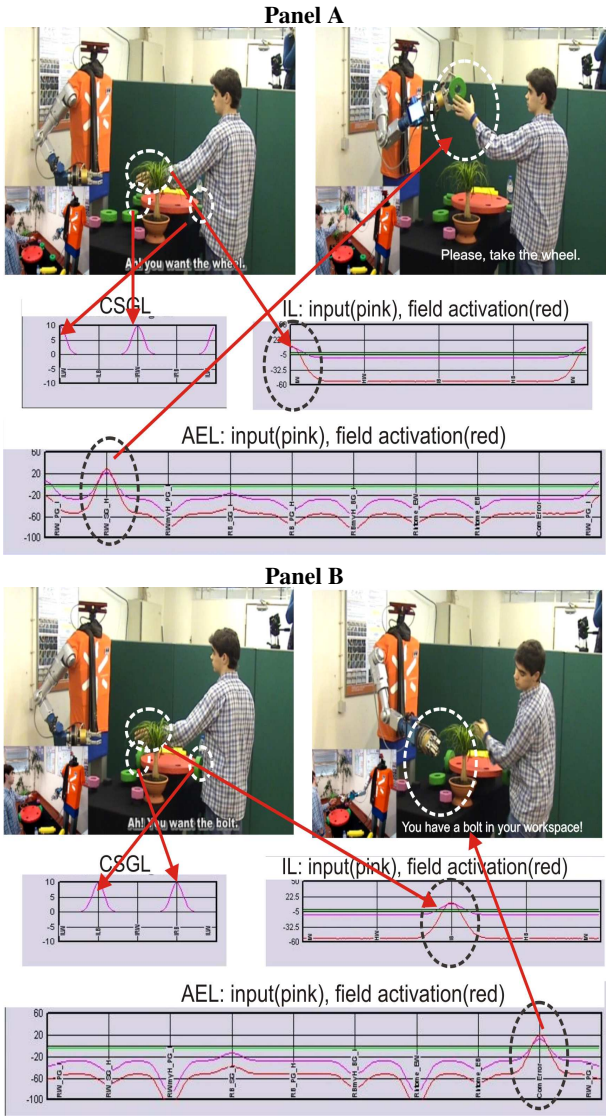

Figure 4. The same observed action may have different meanings. In two different contexts, the human reaches his empty hand toward the robot. Panel A: The robot infers that the human is requesting a wheel with the intention to attach it. The robot decides to grasp it for handing it over. Panel B: The robot infers that the human is asking for a bolt but interprets this request as an error since the human has a bolt in his workspace. The robot decides to communicate this error to the human by pointing to the bolt and speaking to the user.

6 DISCUSSION

The capacity to anticipate and take into account action goals of a partner is considered a fundamental cognitive capacity for successful joint action [21, 18]. We have presented a robot control architecture for human-robot interaction that is based on theories about how humans perceive and act in a social context. The ease with which humans coordinate in routine joint tasks their actions and decisions in space and time is impressive. The capacity to quickly register the intention of the teammate before the action sequence is completed is essential for a fluent team performance. The dynamic field architecture implements the idea that in known tasks dynamic decision making, goal-directed action selection and performance monitoring occur rather effortlessly and do not require a fully developed human capacity for conscious control [16]. As the representation of context, goals and shared task knowledge are interconnected, the observation of a motor act together with situational cues may directly activate the self-sustained population representations of the related goal and the

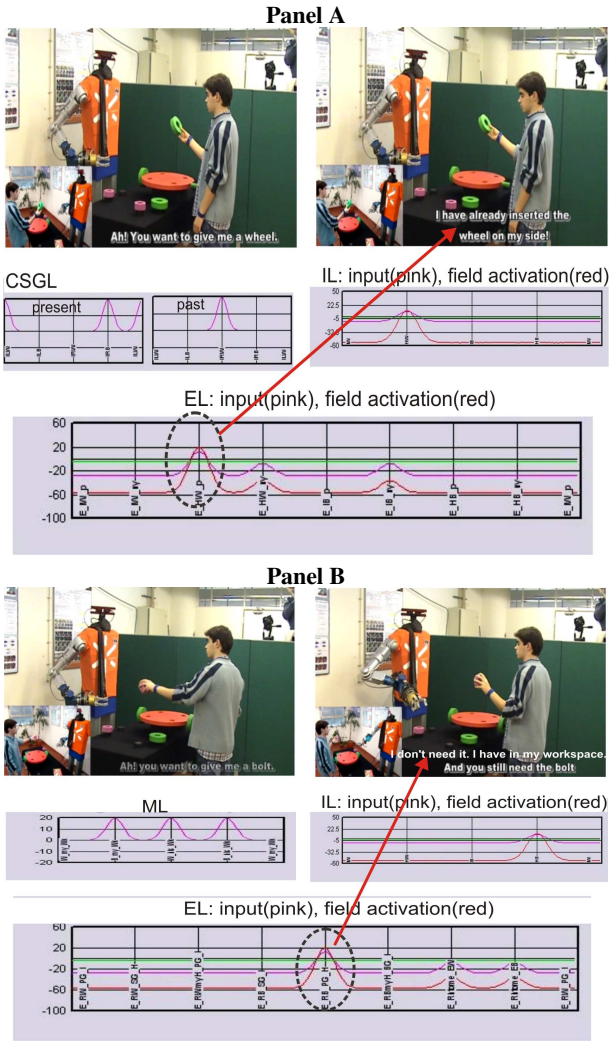

Figure 5. : Error monitoring. Panel A: The robot infers that the human's intention is to hand-over a wheel. The robot interprets this is an error, because the wheel on its side of the platform has been already attached, and decides to communicate this to the human by speaking and nodding its head. Panel B: The robot infers that the human intents to hand over the bolt. The robot interprets the action as an error because the robot has a bolt in its workspace and there is still the need to attach the bolt at the side of the human.

most appropriate complementary action. This automatic process includes basic forms of error monitoring and error compensation

More traditional probabilistic approaches have been applied in the past as well to model and implement cognitive skills like goal inference and decision making for joint action ([7],[17]). Hoffman and Breazeal for instance modeled anticipatory decision making in a Baysian framework to study team fluency in a simulated construction task. In general, Baysian statistics offers a powerful tool for describing human behaviour under circumstances of uncertainty [14]. In our view, a major advantage of the dynamic field approach is that it represents explicitly the important temporal dimension of goal coordination in joint action [21]. Importantly, Dynamic Neural Fields can be used in the Baysian sense by exploiting that multi-modal, subthreshold activation patterns may encode the probability of choices [6]. We are currently testing the joint action model in more complex construction tasks in which the robot has first to infer from observed actions and contextual cues which of several possible toy objects the user is going to built. The accumulated evidence for each of the possible choices is represented by the level of pre-activation of neural populations encoding the different objects.

The decision process linked to complementary actions unfolds over time under multiple influences which are themselves modelled as dynamic representations with proper time scales. This is the basis of flexible behaviour in dynamic joint action conditions. The absence or delay of information about the intention of the user for instance will automatically lead to a decision about an action that does not take into account the other [3]. A challenge for the future will be to endow the robot with the capacity to self-adapt the time window for the integration of input to the dynamics of the different users.

Learning is in general an important research topic of our group. For the present experiments, all inter-field connections were hand-coded. It is certainly not realistic to assume for the next future that for a complex joint action model these connections will self-organize with only modest intervention by the human designer. However, using correlation based learning rules, we have shown in previous work for instance how the goal-directed mappings of the action understanding model may develop during learning and practice [10, 9]. Interestingly, the development process includes the emergence of new task-specific populations which have not been introduced to the architecture by the human designer [11].

The focus of the presented work on action understanding does not mean of course that other information channels of human-human cooperation are not equally important. Our robot is equipped with a speech synthesizer to communicate the state of its reasoning to the user. An obvious extension of this work is to close the loop and advance toward natural-language dialogue. We are currently starting to test a hybrid control architecture that allows us to combine non-verbal and verbal communication skills [13].

ACKNOWLEDGEMENTS

We would like to thank Rui Silva, Eliana Costa e Silva, Toni Machado and Emanuel Sousa for their contribution in many ways.

REFERENCES

[1] A Alissandrakis, C L Nehaniv, and K Dautenhahn, 'Imitation with alice: learning to imitate corresponding actions across dissimiliar embodiments.', *IEEE Transactions on Systems, Man and Cybernetics-PartA*, **32**, 482–496, (2002).

[2] S Amari, 'Dynamics of pattern formation in lateral-inhibitory type neural fields.', *Biological Cybernetics*, **27**, 77–87, (1977).

[3] E Bicho, L Louro, N Hiplito, and W Erlhagen, 'A dynamic neural field architecture for flexible and fluent human-robot interaction', in *Proceedings of the 2008 International Conference on Cognitive Systems*, pp. 179–185. University of Karlsruhe, Germany, (2008).

[4] C Breazeal, 'Social intercations in HRI: The robot view', *IEEE Transactions on Systems, Man and Cybernetics-PartC: Applications and Reviews*, **34**, 181–186, (2004).

[5] S Calinon, F Guenter, and A Billard, 'On learning, representing and generalizing a task in a humanoid robot.', *IEEE Transactions onSystemsm man and Cybernetics, Part B*, **37**, 286–298, (2007).

[6] R Cuijpers and W Erlhagen, 'Implementing Bayes' rules with neural fields', in *ICANN 2008, Part II, LNCS 5164*, ed., V Kurkov, pp. 228–237. Springer Verlag, (2008).

[7] R H Cuijpers, H T van Schie, M Koppen, W Erlhagen, and H Bekkering, 'Goals and means in action observation: A computational appraoch', *Neural Networks*, **19**, 311–322, (2006).

[8] W Erlhagen and E Bicho, 'The dynamic neural field approach to cognitive robotics', *Journal of Neural Enginering*, **3**, R36–R54, (2006).

[9] W Erlhagen, A Mukovskiy, and E Bicho, 'A dynamic model for action understanding and goal-directed imitation.', *Brain Research*, **1083**, 174–188, (2006).
[10] W Erlhagen, A Mukovskiy, E Bicho, G Panin, C Kiss, A Knoll, H van Schie, and H Bekkering, 'Goal-directed imitation for robots: a bio-inspired approach to action understanding and skill learning', *Robotics and Autonomous Systems*, **54**, 353–360, (2006).
[11] W. Erlhagen, A. Mukovskiy, F. Chersi, and E. Bicho, 'On the development of intention understanding for joint action tasks.', in *6th IEEE Int. Conf. on Development and Learning*, pp. 140–145. Imperial College London, (11-13 July, 2007).
[12] T Fong, I Nourbakhsh, and K Dautenhahn, 'A survey of socially interactive robots', *Robotics and Autonomus Systems*, **42**, 143–166, (2003).
[13] M E Foster, M Giuliani, T Mller, M Rickert, A Knoll, W Erlhagen, E Bicho, N Hiplito, and L Louro, 'Combining Goal Inference and Natural-Language Dialogue for Human-Robot Joint Action', in *Proceedings of the 1st International Workshop on Combinations of Intelligent Methods and Applications (CIMA 2008)*. Patras, Greece, (July, 2008).
[14] J I Gold and M N Shadlen, 'Banburismus and the brain: the relationship between sensory stimuli, decisions, and reward.', *Neuron*, **36**, 299–308, (2002).
[15] S Gray, C Breazeal, M Berlin, A Brooks, and J Lieberman, 'Action parsing and goal inference using self as simulator', *IEEE International Workshop on Robots and Human Interacive Communication*, 202–209, (2005).
[16] R R Hassin, H Aarts, and M J Ferguson, 'Automatic goal inferences', *Journal of Experimental Social Psychology*, **41**, 129–140, (2005).
[17] G Hoffman and C Breazeal, 'Cost-based anticipatory action selection for human-robot fluency', *IEEE Transactions on Robotics*, **23**, 952–961, (2007).
[18] R D Newman-Norlund, M L Noordzij, R G J Meulenbroek, and H Bekkering, 'Exploring the basis of joint action: Coordination of actions, goals and intentions', *Social Neuroscience*, **2**, 48–65, (2007).
[19] S Schaal, 'Is imitation learning the route to humanoid robots?', *Trends in Cognitive Science*, **3**, 233–242, (1999).
[20] G Schöner, 'Dynamical systems approaches to cognition', in *The Cambridge Handbook of Computational Psychology*, ed., R Sun, pp. 101–125. Cambridge University Press, (2008).
[21] N Sebanz, H Bekkering, and G Knoblich, 'Joint action: bodies and minds moving together', *Trends in Cognitive Sciences*, **10**, 70–76, (2006).
[22] T P Spexard, M Hanheide, S Li, and Wrede B, 'Oops, something is wrong: Error detection and recovery for advanced human-robot intercatios', in *Proceedimgs of the ICRA 2008 Workshop on Social Intercation with Intelligent Indoor Robots*, eds., G J Kruijff, H Zender, M Hanheide, and B Wrede, (2008).
[23] H T van Schie, B M van Waterschoot, and H Bekkering, 'Understanding action beyond imitation: Reversed compatibilty effects of action observation in imitation and joint action', *Jurnal fo Experimental Psychology: Human Perception and Performance*, (in press).
[24] G. Westphal, C. von der Malsburg, and R. P. Würtz, *Applied Pattern Recognition*, chapter Feature-Driven Emergence of Model Graphs for Object Recognition and Categorization, 155–199, Springer Verlag, 2008.
[25] M Wilson and G Knoblich, 'The case for motor involvment in perceiving conspecifics', *Psychological Bulletin*, **131**, 460–473, (2005).