Grant Agreement no. 214668

ITALK
Integration and Transfer of Action and Language Knowledge in robots

Seventh Framework Programme (FP7)
Cognitive Systems, Interaction, Robotics
(ICT-2007.2.1)

Deliverable 6.3 (Y2)
Peer-review scientific publications in international journals and conferences

Start Date: 2008-03-01  Duration: 48 months

Project coordinator name: Angelo Cangelosi
Project coordinator organization name: University of Plymouth
### Table of contents

**Journals (8) .................................................................................................................................. 5**


**Conferences papers (20).................................................................................................................. ........................................................................ 7**


Conference abstracts and posters (10).............................................................................................................................. ............................................................. 13

Book Chapters (8)......................................................................................................................................................... 15

Submitted Papers (10).............................................................................................................................................. 15

This paper presents a cognitive robotics model for the study of the embodied representation of action words. The present research will present how a iCub humanoid robot can learn the meaning of action words (i.e. words that represent dynamical events that happen in time) by physically acting on the environment and linking the effects of its own actions with the behaviour observed on the objects before and after the action. The control system of the robot is an artificial neural network trained to manipulate an object through a Back-Propagation Through Time algorithm. We will show that in the presented model the grounding of action words relies directly to the way in which an agent interacts with the environment and manipulates it.


This study shows that a mixture of RNN experts model can acquire the ability to generate sequences that are combination of multiple primitive patterns by means of self-organizing chaos. By training of the model, each expert learns a primitive sequence pattern, and a gating network learns to imitate stochastic switching of the multiple primitives via a chaotic dynamics, utilizing a sensitive dependence on initial conditions. As a demonstration, we present a numerical simulation in which the model learns Markov chain switching among some Lissajous curves by a chaotic dynamics. Our analysis shows that by using a sufficient amount of training data, balanced with the network memory capacity, it is possible to satisfy the conditions for embedding the target stochastic sequences into a chaotic dynamical system. It is also shown that reconstruction of a stochastic time series by a chaotic model can be stabilized by adding a negligible amount of noise to the dynamics of the model.


A difficulty in robot action learning is that robots do not know where to attend when observing action demonstration. Inspired by human parent-infant interaction, we suggest that parental action demonstration to infants, called motionese, can scaffold robot learning as well as infants'. Since infants' knowledge about the context is limited, which is comparable to robots, parents are supposed to properly guide their attention by emphasizing the important aspects of the action. Our analysis employing a bottom-up attention model revealed that motionese has the effects of highlighting the initial and final states of the action, indicating significant state changes in it, and underlining the properties of objects used in the action. Suppression and addition of parents' body movement and their frequent social signals to infants produced these effects. Our findings are discussed toward designing robots that can take advantage of parental teaching.


Social robots are designed to interact with humans. That is why they need interaction models that take social behaviors into account. These usually influence many of a robot's abilities simultaneously. Hence, when designing robots that users will want to interact with, all components need to be tested in the system context, with real users and real tasks in real interactions. This requires methods that link the analysis of the robot's internal computations within and between components (system level) with the interplay between robot and user (interaction level). This article presents Systemic Interaction Analysis (SiNIA) as an integrated method to (a) derive prototypical courses of interaction based on system and interaction level, (b) identify deviations from these, (c) infer the causes of deviations by analyzing the system's operational sequences, and (d) improve the robot iteratively by adjusting models and implementations.

The issue of how children learn the meaning of words is fundamental to developmental psychology. The recent attempts to develop or evolve efficient communication protocols among interacting robots or virtual agents have brought that issue to a central place in more applied research fields, such as computational linguistics and neural networks, as well. An attractive approach to learning an object-word mapping is the so-called cross-situational learning. This learning scenario is based on the intuitive notion that a learner can determine the meaning of a word by finding something in common across all observed uses of that word. Here we show how the deterministic Neural Modeling Fields (NMF) categorization mechanism can be used by the learner as an efficient algorithm to infer the correct object-word mapping. To achieve that we first reduce the original on-line learning problem to a batch learning problem where the inputs to the NMF mechanism are all possible object-word associations that could be inferred from the cross-situational learning scenario. Since many of those associations are incorrect, they are considered as clutter or noise and discarded automatically by a clutter detector model included in our NMF implementation. With these two key ingredients - batch learning and clutter detection - the NMF mechanism was capable to infer perfectly the correct object-word mapping.


A brain-inspired computational system is presented that allows sequential selection and processing of objects from a visual scene. The system is comprised of three modules. The selective attention module is designed as a network of spiking neurons of the Hodgkin-Huxley type with star-like connections between the central unit and peripheral elements. The attention focus is represented by those peripheral neurons that generate spikes synchronously with the central neuron while the activity of other peripheral neurons is suppressed. Such dynamics corresponds to the partial synchronization mode. It is shown that peripheral neurons with higher firing rates are preferentially drawn into partial synchronization. We show that local excitatory connections facilitate synchronization, while local inhibitory connections help distinguishing between two groups of peripheral neurons with similar intrinsic frequencies.

The module automatically scans a visual scene and sequentially selects regions of interest for detailed processing and object segmentation. The contour extraction module implements standard image processing algorithms for contour extraction. The module computes raw contours of objects accompanied by noise and some spurious inclusions. At the next stage, the object segmentation module designed as a network of phase oscillators is used for precise determination of object boundaries and noise suppression. This module has a star-like architecture of connections. The segmented object is represented by a group of peripheral oscillators working in the regime of partial synchronization with the central oscillator. The functioning of each module is illustrated by an example of processing of the visual scene taken from a visual stream of a robot camera.


This paper addresses the problem of planning the movement of highly redundant humanoid robots based on non-linear attractor dynamics, where the attractor landscape is obtained by combining multiple force fields in different reference systems. The computational process of relaxation in the attractor landscape is similar to coordinating the movements of a puppet by means of attached strings, the strings in our case being the virtual force fields generated by the intended/attended goal and the other task dependent combinations of constraints involved in the execution of the task. Hence the name PMP (Passive Motion Paradigm) was given to the computational model. The method does not require explicit kinematic inversion and the computational mechanism does not crash near kinematic singularities or when the robot is asked to achieve a final pose that is outside its intrinsic workspace: what happens, in this case, is the gentle degradation of performance that characterizes humans in the same situations. Further, the measure of inconsistency in the relaxation in such cases can be directly used to trigger higher level reasoning in terms of breaking the goal into a sequence of subgoals directed towards searching and perhaps using tools to realize the otherwise unrealizable goal. The basic PMP model has been further expanded in the present paper by means of (1) a non-linear dynamical timing mechanism that provides terminal attractor properties to the relaxation process and (2) branching units that allow to ‘compose’ complex PMP-networks to coordinate multiple kinematic chains in a complex structure, including manipulated tools. A preliminary evaluation of the approach has been carried out with the 53 degrees of freedom humanoid robot iCub, with particular reference to trajectory formation and bimanual/whole upper body coordination under the presence of different structural and task specific constraints.


A robot that can communicate with humans using natural language will have to acquire a grammatical framework. This paper analyses some crucial underlying mechanisms that are needed in the construction of such a framework. The work is inspired by language acquisition in infants, but it also draws on the emergence of language in evolutionary time and in ontogenic (developmental) time. It focuses on issues arising from the use of real language with all its evolutionary baggage, in contrast to an artificial communication system, and describes approaches to addressing these issues. We can deconstruct grammar to derive underlying primitive mechanisms, including serial processing, segmentation, categorization, compositionality, forward planning. Implementing these mechanisms are necessary preparatory steps to reconstruct a working syntactic/semantic/pragmatic processor which can handle real language. An overview is given of our own initial experiments in which a robot acquires some basic linguistic capacity via interacting with a human.

The concept of semantic bootstrapping of grammar, though tracing back at least to [?], has not been explored fully in applied computational linguistics as a procedure for learning grammars. This has largely been due to the lack of an adequate workbench on which extralinguistic contexts, the supposed trigger of grammar learning, can be precisely extrapolated for the use of machine learning. Cognitive robotics, based on experimental methods with embodied robots, offers such a facility. A humanoid robot with fine sensori-motor capabilities in various modalities can ‘perceive’ its environment while ‘hearing’ a speech stream, in a way that an experimenter can measure the robot’s perceptive fields quantitatively in real time. Furthermore, more recent cognitively oriented robots (e.g. [?]) are designed to have an attentional state at a given time, which is regarded as vital for language learning by ‘constructivists’ of grammar learning [?]. Given the feasibility of extracting the measurements that constitute the robot’s attention, we can hope that the plausible semantic representation relevant to a given context may be constructed adequately. We describe below the outline of the proposed grammar induction method for an embodied robot as well as our initial experiments and their results.


This paper describes a HRI case study which demonstrates how a humanoid robot can use simple heuristics to acquire and use vocabulary in the context of being shown a series of shapes presented to it by a human and how the interaction style of the human changes as the robot learns and expresses its learning through speech. The case study is based on findings on how adults use childdirected speech when socially interacting with infants. The results indicate that humans are generally willing to engage with a robot in a similar manner to their engagement with a human infant and use similar styles of interaction varying as the shared understanding between them becomes more apparent. The case study also demonstrates that a rudimentary form of shared intentional reference can sufficiently bias the learning procedure. As a result, the robot associates humantaught lexical items for a series of presented shapes with its own sensorimotor experience, and is able to utter these words, acquired from the particular tutor, appropriately in an interactive, embodied context exhibiting apparent reference and discrimination.


We summarise the experimental design issues related to timing in several human-robot interaction scenarios investigating turn-taking or synchronization between child-sized humanoid robots and human participants. Our aim is not to have the humanoid robots just replicate the human’s behaviours (e.g. waving, peek-a-boo, or drumming), but to engage in interactions in a socially appropriate manner. From these various studies, we have identified several ways in which time has an impact on interaction. We have also identified practical concerns about data collection for time-dependent interactions and ways to address them. The conclusions drawn from this work is likely to be useful in informing the design of systems which engage in synchronized or turn-taking interactions with people.


This paper gives a short overview of time representations in current symbol grounding architectures. Furthermore we report on a recently developed embodied language acquisition system that acquires object words from a linguistically unconstrained human-robot dialogue. Conceptual issues in future development of the system towards the acquisition of action words will be discussed briefly.


This paper reports on foundational considerations for experiments into the acquisition of human-like use and understanding of negation in linguistic utterances via a developmental robotics approach. For this purpose different taxonomies of negation in early child language are analysed in order to show the large variety of communicative functions that these different types of negation have. Requirements for robotic systems that aim at acquiring these utterances in a linguistically unconstrained human-robot dialog are derived from this analysis.
This paper introduces a novel neuro-dynamical model that accounts for possible mechanisms of action imitation and learning. It is considered that imitation learning require at least two classes of generalization. One is generalization over sensory-motor trajectory variances, and the other class is on cognitive level which concerns on more qualitative understanding of goal-directed actions by own and others which do not necessarily depend on exact trajectories. This paper describes one of the possible model dealing with these classes of generalization by focusing on the combination of behavior primitives. The model was evaluated in the experiments using a small humanoid robot. The robot was trained with a set of different goal-directed actions concurring manipulation of an object which can be decomposed into sequences of behavior primitives. Then the robot was asked to imitate a novel goal-directed action demonstrated by human subject which are composed from prior-learned behavior primitives. The results showed that novel goal-directed action was successfully generated by means of combining prior aquired behavior primitives even if the appearances of the same goal-directed actions by the human and the robot were significantly different.


In order to learn and interact with humans, robots need understand actions and make use of language in social interactions. The use of language for the learning of actions has been emphasized by Hirsh-Pasek & Golinkoff introducing the idea of Acoustic Packaging [1]. Accordingly, it has been suggested that acoustic information, typically in the form of narration, overlaps with action sequences and provides infants with a bottom-up guide to attend to relevant events and to find structure within them. Following the promising results achieved by Brand & Tapscott for infants who packaged sequences together when acoustic narration was provided, in this paper, we make the first approach towards a computational model of the multimodal interplay of action and language in tutoring situations. For our purpose, we understand events as temporal intervals, which have to be segmented in both the visual and the acoustic signal in order to perform Acoustic Packaging. For the visual modality, we looked at the amount of motion over time via a motion history image based approach. The visual signal is segmented by detecting local minima in the amount of motion. For the acoustic modality, we used a phoneme recognizer, which currently segments the acoustic signal into speech and non-speech intervals.

Our Acoustic Packaging algorithm merges the segments from both modalities based on temporal synchrony. First evaluation results show that our computational model is receptive to parental cues during child-directed tutoring. The findings discussed in this paper are consistent with recent results from developmental psychology but for the first time are obtained employing an objective, computational model. The presence of “multimodal motherese” is verified directly on the audio-visual signal. Lastly, we hypothesize how our computational model facilitates tutoring interaction and discuss its application in interactive learning scenarios, enabling social robots to benefit from adult-like tutoring.


In developmental research, tutoring behavior has been identified as scaffolding infants’ learning processes. It has been defined in terms of child-directed speech (Motherese), childdirected motion (Motionese), and contingency. In the field of developmental robotics, research often assumes that in humanrobot interaction (HRI), robots are treated similar to infants, because their immature cognitive capabilities benefit from this behavior. However, according to our knowledge, it has barely been studied whether this is true and how exactly humans alter their behavior towards a robotic interaction partner. In this paper, we present results concerning the acceptance of a robotic agent in a social learning scenario obtained via comparison to adults and 8-11 months old infants in equal conditions. These results constitute an important empirical basis for making use of tutoring behavior in social robotics. In our study, we performed a detailed multimodal analysis of HRI in a tutoring situation using the example of a robot simulation equipped with a bottom-up saliency-based attention model [1]. Our results reveal significant differences in hand movement velocity, motion pauses, range of motion, and eye gaze suggesting that for example adults decrease their hand movement velocity in an Adult-Child Interaction (ACI), opposed to an Adult-Adult Interaction (AAI) and this decrease is even higher in the Adult-Robot Interaction (ARI). We also found important differences between ACI and ARI in how the behavior is modified over time as the interaction unfolds. These findings indicate the necessity of integrating top-down feedback structures into a bottom-up system for robots to be fully accepted as interaction partners.
Learning is a social endeavor, in which the learner generally receives support from his/her social partner(s). In developmental research – even though tutors/adults behavior modifications in their speech, gestures and motions have been extensively studied – studies barely consider the recipient’s (i.e. the child’s) perspective in the analysis of the adult’s presentation. In addition, the variability in parental behavior, i.e. the fact that not every parent modifies her/his behavior in the same way, found less fine-grained analysis. In contrast, in this paper, we assume an interactional perspective investigating the loop between the tutor’s and the learner’s actions. With this approach, we aim both at discovering the levels and features of variability and at achieving a better understanding of how they come about within the course of the interaction. For our analysis, we used a combination of (1) qualitative investigation derived from ethnomethodological Conversation Analysis (CA), (2) semi-automatic computational 2D hand tracking and (3) a mathematically based visualization of the data. Our analysis reveals that tutors not only shape their demonstrations differently with regard to the intended recipient per se (adult-directed vs. child-directed), but most importantly that the learner’s feedback during the presentation is consequential for the concrete ways in which the presentation is carried out.


If robot systems are being deployed in real world settings with untrained users who happen to accidentally pass by or could leave at any moment in time, then this places specific demands on the robot system: it needs to secure and maintain the user’s engagement. In this, a common and critical problem consists of entering into a focused encounter. It requires each interactional partner to closely react upon the other’s actions on a very fine-grained level engaging in a stepwise and dynamic process of mutual adjustments. We report initial findings from a study in which we have developed a preliminary, simple solution to this problem inspired by work from interaction analysis [7]. Using this as an instrument to explore the impact of a ‘contingent’ (CE) vs. ‘non-contingent entry’ (NCE), we find that users who enter into the interaction in a dynamic and contingent manner show a significantly different way of interacting with the robot than the NCE group.


As it is often assumed that an interaction with a robot is similar to an interaction with an infant, one would expect similar characteristics in tutoring behavior in human-robot interaction from which the robot’s learning processes can benefit. Especially learning actions can profit from the use of language. As it has been shown by Hirsh-Pasek and Golinkoff introducing the idea of Acoustic Packaging, language not only serves as a social cue but also provides functional structure. In order to better understand how robot-directed interaction is structured and how the modifications may help to extract cues relevant for learning actions from a tutoring situation, in the present paper we analyzed the structure of Acoustic Packages in the adult-robot situation. Our results indicate that in human-robot interaction, the phenomenon of Acoustic Packaging is similar in adult-child interaction: In both situations, more Acoustic Packages with less content can be found, compared to adult-adult interaction. This can be interpreted as an encouraging result towards the goal of building robots that can learn from a tutoring situation.


In developmental research, tutoring behavior has been identified as scaffolding infants' learning processes. Infants seem sensitive to tutoring situations and they detect these by ostensive cues. Some social signals such as eyegaze, child-directed speech (Motherese), child-directed motion (Motionese), and contingency have been shown to serve as ostensive cues. The concept of contingency describes exchanges in which two agents interact with each other reciprocally. Csibra and Gergely argued that contingency is a characteristic ostensive stimulus of a tutoring situation. In order for a robot to be treated similar to an infant, it has to both, be sensitive to the ostensive stimuli on the one hand and induce tutoring behavior by its feedback about its capabilities on the other hand. In this paper, we raise the question whether a robot can be treated similar to an infant in an interaction. We present results concerning the acceptance of a robotic agent in a social learning scenario, which we obtained via comparison to interactions with 8-11 months old infants and adults in equal conditions. We applied measurements for motion modifications (Motionese) and eye-gaze behavior. Our results reveal significant differences between Adult-Child Interaction (ACI), Adult-Adult Interaction (AAI) and Adult-Robot Interaction (ARI) suggesting that in ARI, robot-directed tutoring behavior is even more accentuated in terms of Motionese, but contingent responsivity is impaired. Our results confirm previous findings concerning the differences between ACI, AAI, and ARI and constitute an important empirical basis for making use of ostensive stimuli as social signals for tutoring behavior in social robotics.
Cross-situational learning is based on the idea that a learner can determine the meaning of a word by finding something in common across all observed uses of that word. Although cross-situational learning is usually modeled through stochastic guessing games in which the input data vary erratically with time (or rounds of the game), here we investigate the possibility of applying the deterministic Neural Modeling Fields (NMF) categorization mechanism to infer the correct object-word mapping. Two different representations of the input data were considered. The first is termed object-word representation because it takes as inputs all possible object-word pairs and weighs them by their frequencies of occurrence in the stochastic guessing game. A re-interpretation of the problem within the perspective of learning with noise indicates that the cross-situational scenario produces a too low signal-to-noise ratio, explaining thus the failure of NMF to infer the correct object-word mapping. The second representation, termed context-word, takes as inputs all the objects that are in the pupil’s visual field (context) when a word is uttered by the teacher. In this case we show that use of two levels of hierarchy of NMF allows the inference of the correct object-word mapping.

The subjective ease with which we move gracefully in constraint filled uncertain environments often masks the enormously complex representations of the input data that contribute to any act's planning and execution. In this paper, we apply the computational framework of passive motion paradigm for task specific composition and coordination of the movements of a limb, network of limbs (e.g. left arm–waist–right arm) or networks of external objects coupled to the body of the 53 degrees of freedom humanoid robot ‘Cub’. The basic PMP model is further extended by formulation of a pair of branching nodes that allow compositionality and transfer of force fields from one relaxation network to another. The generality of the proposed approach is further illustrated using simulations of whole body reaching (WBR) tasks from a quiet standing posture that recruits virtually all the joints of the upper limbs, lower limbs, and trunk, binding together a large number of degrees of freedom into a functional unit that combines a focal task (reaching a target with the hand) and a postural task (keeping the projection of the center of mass within the bipedal support area). Preliminary comparisons of the solutions generated by the computational model with the movements of human subjects performing similar WBR tasks are presented.

The first reaching movements of human infants lack limb coordination leading to ataxic-like hand trajectories. Kinematically, these early trajectories are characterized by multiple peaks in the hand velocity profile which gradually decrease in frequency during development. In this paper we explore the hypothesis that the jerky hand trajectories seen in early infancy can be the result of imprecise internal motor models. Results from our simulation suggest that imprecise estimations of multi-joint inter-segmental torques (e.g., Coriolis forces) by the controller may induce multi-peak hand velocity profiles. When the system was allowed to use delayed peripheral feedback (300 ms after reaching onset), the resulting kinematics began to resemble those seen in early infancy. This suggests that the output of an imprecise internal model of limb dynamics coupled with delayed feedback may be sufficient to explain early human hand trajectories. Our data provide an alternative to previous hypotheses theorising jerky trajectories as the result of concatenated mini ballistic movements.

Active perception refers to a theoretical approach grounded on the idea that perception is an active process in which the actions performed by the agent play a constitutive role. In this paper we present two different scenarios in which we test active perception principles using an evolutionary robotics approach. In the first experiment, a robotic arm equipped with coarse-grained tactile sensors is required to perceptually categorize spherical and ellipsoid objects. In the second experiment, an active vision system has to distinguish between five different kinds of images of different sizes. In both situations the best individuals develop a close to optimal ability to discriminate different objects/images as well as an excellent ability to generalize their skills in new circumstances. Analyses of evolved behaviours show that agents are able to solve their tasks by actively selecting relevant information and by integrating these information over time.

Active perception refers to a theoretical approach to the study of perception grounded on the idea that perceiving is a way of acting, rather than a process whereby the brain constructs an internal representation of the world. In this paper, we complement previous studies by illustrating the operational principles of an active categorization process in which a neuro-controlled anthropomorphic robotic arm, equipped with coarse-grained tactile sensors, is required to perceptually categorize spherical and ellipsoid objects.
Active perception refers to a theoretical approach to the study of perception grounded on the idea that perceiving is a way of acting, rather than a cognitive process whereby the brain constructs an internal representation of the world. The operational principles of active perception can be effectively tested by building robot-based models in which the relationship between perceptual categories and the body-environment interactions can be experimentally manipulated. In this paper, we study the mechanisms of tactile perception in a task in which a neuro-controlled anthropomorphic robotic arm, equipped with coarse-grained tactile sensors, is required to perceptually discriminate between spherical and ellipsoid objects. The results of this work demonstrate that evolved continuous time nonlinear neural controllers can bring forth strategies to allow the arm to effectively solve the discrimination task.


We investigate preliminary stages in enabling robots to talk with humans in a natural manner, and outline experiments. The process is inspired by language acquisition in infants, and by recent empirical evidence of neuronal organisation. In this paper we describe preliminary stages in enabling robots to communicate with humans, using natural language. This starts with babbling, analogous to the pre-linguistic infant in a proto-conversation with its carer, progressing to learning the meaning of utterances through mediated physical interaction (Saunders et al., 2009). This work is inspired by the acquisition of language by human infants, and by recent empirical evidence of neuronal organisation. Participants in our experiments talk to the robot in natural, unrestricted language, about a blocks world with objects of various shapes. The robot must learn to “understand” this highly redundant natural language, in which the same concept can be expressed in a number of alternative ways (e.g. “push the red box”, “give the red box a push”). Its own productions may be more limited: asymmetrical development is typical of human infant language acquisition.

Conference abstracts and posters (10)


This paper presents a model of sensorimotor learning grounded in the sensory streams of a real humanoid robot (the iCub robot). The robot participates in a replication of two developmental psychology experiments, in which it is shown how spatial cues are sufficient for associating linguistic labels with objects. The robot, using auto-associated self-organizing maps connecting is perceptual input and motor control, produces similar performance and results to human participants. This model confirms the validity of a body centric account of the linking of words to objects as sufficient to account for the spatial biases in learning that these experiments expose.


Human language is a formidable communication system. It allows to describe the world around us and exchange our thoughts. Among humans language is used in many different ways, such as describing what we perceive, asking others to perform certain actions or simply engage in conversation (Siskind, 2001). At the core of this description there is our ability to understand and use correctly the meaning that words represent. Especially the first two cases above require to ground language in perception and action processes. Moreover, by focusing the attention on our abilities to easily describe dynamics that happen in time and entails specific relations between objects and object properties, the process of grounding language in perception and actions means that, when we describe a given scene or we ask someone to perform a certain action, the words used must be linked with physical entities in the scene or in actions that can be either observed or desired.


In the presentation we will discuss ongoing results on the ITALK project. Highlights of year 1 results includes: (i) Roadmap. The consortium has produced a RoadMap for research in developmental robotics, with respect to future challenges on the integration of action and language knowledge in robots. See milestones in table below. (ii) iCub Simulator. A new Open Source iCub simulator software was produced. This is available at: http://eris.liralab.it/italk Social learning experiments. We extended the ROSSUM learning architecture to a humanoid robot platform, and work has commenced on the experimental issues for various aspects of negation and grammar induction. (iii) Cognitive linguistics grammar learning scenarios. Empirical analyses of child-directed and robot-directed speech interactions led to the definition of incremental cognitive linguistic scenarios for language learning experiments. (iv) Acoustic packaging. Acoustic packaging has been observed as a means of communication that is used towards infants. We applied the concept of contingency to our studies on human-robot-interaction and analysed it qualitatively. We found that in comparison to adult-adult and child-adult interaction, people show little eye gaze towards the robot suggesting that contingency is
impaired in this situation. (iv) Cognitive biases. Initial experiments on category learning and naming in human-robot interactions have demonstrated the presence of a spatial/location bias, previously observed in developmental psychology.


In our approach, we aim at an objective measurement of synchrony in multimodal behavior. The use of signal correlation provides a well formalized method that yields gradual information about the degree of synchrony. For our analysis, we used and extended an algorithm proposed by Hershey & Movellan (2000) that correlates single pixel values of a video signal with the loudness of the corresponding audio track over time. The results of all pixels are integrated over the video to achieve a scalar estimate of synchrony.


Acoustic packaging makes use of the synchrony between the visual and audio modality in order to detect temporal structure actions that are demonstrated to children and robots [1, 2]. Support for action learning in robots Acoustic packages form early units for further learning processes. Feedback generation during tutoring.


In developmental research, tutoring behavior has been identified as scaffolding infants’ learning processes. It has been defined in terms of child-directed speech (Motherese), child-directed motion (Motionese), and Contingency. Contingency describes situations in which two agents socially interact with each other and Csibra and Gergely showed that contingency is a characteristic aspect of social interaction [3]. In the field of developmental robotics, research often assumes that in human-robot interaction (HRI), robots are treated similar to infants, because their immature cognitive capabilities benefit from this behavior. Here we present results concerning the acceptance of a robotic agent in a social learning scenario obtained via comparison to adults and 8-11 months old infants in equal conditions. These results constitute an important empirical basis for making use of tutoring behavior in social robotics. Our results reveal significant differences between Adult-Child Interaction (ACI), Adult-Adult Interaction (AAI) and Adult-Robot Interaction (ARI) in eye-gaze behavior suggesting that contingency is impaired in the analyzed ARI situation.


While it is already known that parents modify their demonstrations towards children [Brand et al., 2002; Brand et al., 2007] and that young infants aged 6 to 8 months prefer “motionese” [Brand & Shallcross, 2007], little is known about whether the modified behavior can also be found in interaction with older children. Here, we therefore seek to investigate the effects of children’s age on motionese, defined as modified action demonstration [Brand et al., 2002; Rohlfing et al., 2006]. In our study, parents demonstrated a function of an object (stacking cups) towards their infant and towards another adult. We analyzed parental behavior in three different age groups: parents of prelexical (8 – 11), early lexical (12 –24) and advanced lexical (25 – 30 months olds) children. In our analysis, we use objective measurements of hand trajectories providing data about their shape and time structure. Results suggest that actions chosen to attract attention by providing more range can primarily be found in interaction with younger infants, whose attention needs more guidance. Interactions with older children seem to benefit either from the increase of children's attention abilities or that parents use other means (such as language) to attract their attention. In contrast, parameters that appear to be more in charge of structuring the action by organizing it in motion pauses seem to persist over the age and verbal capabilities.


Prediction is a central skill in human life. As our environment is constantly changing, both as a consequence of our actions and independently of us, it’s necessary to anticipate when and where future events will happen, in order to be able to synchronize our actions with them and to proficiently interact with the world (e.g. Gredebäck et al. 2002). The study of interception abilities represents a good option to investigate this topic, as interception is a quite common task and, at the same time, it strongly requires anticipation skills. In fact it wouldn’t be possible to catch any target acting in a purely reactive manner, due to the visuo-motor transmission delays of the human body (Zago & Lacquaniti, 2005). In an interception study we evaluated which aspects of target behavior are important to improve prediction, focusing in particular on acceleration and motion direction. We also compared interception results with the ones of a similar purely visual task to evaluate whether different mechanisms are involved in prediction in dependence of the goal of the predictive effort.
Experiments using natural, spontaneous speech, speaking to the robot as if it were a small child. Inspired by the acquisition of language in human infants and by evidence of neuronal organization.

In previous studies we observed that intercepting a target may require modeling the force field which drives its motion. The question we address in this study is what force field parameters are to be modeled. Therefore, we performed an experiment to understand whether force field direction and modulus are both involved in the modeling process. We investigated how the stability of these force field features affects the prediction and in particular the motion strategy adopted in the interception. Moreover we compared interception results with the ones of a similar purely visual task to evaluate whether different mechanisms are involved in prediction in dependence of the goal of the predictive effort.


Experiments using natural, spontaneous speech, speaking to the robot as if it were a small child. Inspired by the acquisition of language in human infants and by evidence of neuronal organization.

**Book Chapters (8)**


In this chapter we summarize the progresses that have recently been made in the study of the emergence of communication in artificial embodied agents along different dimensions, including the understanding of the adaptive roles of communication, the expressive power and organization complexity of signalling systems, the stability, robustness, and evolvability of communication, and the knowledge gain obtained with such models. Finally, we briefly discuss which we think are the most important open challenges for future research in this area.


In this chapter we introduce the area of research that attempts to study the evolution of communication in embodied agents through adaptive techniques, such us artificial evolution. More specifically, we illustrate the theoretical assumptions behind this type of research, we present the methods that can be used to realize embodied and communicating artificial agents, and we discuss the main research challenges and the criteria for evaluating progresses in this field.


When testing linguistic knowledge and behavior in experimental conditions, we try to systematically vary factors constituting the external situation. We therefore make effort to create novel, neutral or abstract situations and attempt to control for the extraneous variables in the environment that might have an impact on the dependent variable accessed in the experiment. However, when we adapt this method to studies on children’s language development, we have to calculate the risk of eliminating important components for children’s reasoning: Children are good learners because they are biased towards certain solutions (Dabrowska, 2005). For this reason, it seems to be problematic to create situations, in which children cannot draw from their nonlinguistic experiences. In this paper, I am arguing that objects being present in the external situation ground the linguistic meaning. Referring to previous literature and the results of a pilot study, I will show that the nature of objects can change not only linguistic but also gestural behavior. Instead of simply excluding materialistic factors, I therefore suggest a careful inclusion of object knowledge into experimental conditions.


We present an evaluation of different techniques for the estimation of forces and torques measured by a single six-axis force/torque sensor placed along the kinematic chain of a humanoid robot arm. In order to retrieve the external forces and detect possible contact situations, the internal forces must be estimated. The prediction performance of an analytically derived dynamic model as well as two supervised machine learning techniques, namely Least Squares Support Vector Machines and Neural Networks, are investigated on this problem. The performance are evaluated on the normalized mean square error (NMSE) and the comparison is made with respect to the dimension of the training set, the information contained in the input space and, finally, using a Euclidean subsampling strategy. Key words: Force sensing, machine learning, humanoid robotics.
Recent advances in different disciplines, ranging from cognitive sciences and robotics, biology and neurosciences, to social sciences and philosophy are clarifying that intelligence re-sides in the circular relationship between the brain of an individual organism, its body, and the environment (including the social environment). In this chapter we will focus our attention on the evidence collected in robotic research with particular reference to results obtained in experiments in which the robots develop their skill autonomously while they interact with the external environment through an adaptive process. In particular we will demonstrate how the behavioural and cognitive skills developed by the robots can be properly characterized as complex adaptive systems which: (a) arise from the interactions between the brain of the robots, their body, and the environment and eventually between the dynamical process occurring within the robot and within the environment, (b) display a multi-level and a multi-scale organization in which the interaction between behavioural and cognitive properties at a certain level of organization lead to higher-level properties and in which higher-level properties later affect the lower-level properties.


Natural/Artificial systems that are capable of utilizing thoughts at the service of their actions are gifted with the profound opportunity to mentally manipulate the causal structure of their physical interactions with the environment. A cognitive robot can in this way virtually reason about how an unstructured world should ‘change’, such that it becomes a little bit more conducive towards realization of its internal goals. In this article, we describe the various internal models for real/mental action generation developed in the GNOSYS Cognitive architecture and demonstrate how their coupled interactions can endow the GNOSYS robot with a preliminary ability to virtually manipulate neural activity in its mental space in order to initiate flexible goal directed behavior in its physical space. Making things more interesting (and computationally challenging) is the fact that the environment in which the robot seeks to achieve its goals consists of specially crafted ‘stick and ball’ versions of real experimental scenarios from animal reasoning (like tool use in chimps, novel tool construction in Caledonian crows, the classic trap tube paradigm and their possible combinations).

We specifically focus on the progressive creation of the following internal models in the behavioral repertoire of the robot a) a passive motion paradigm based forward inverse model for mental simulation / real execution of goal directed arm (and arm+tool) movements; b) a spatial mental map of the playground; c) an internal model representing the causality of pushing objects and further learning to push intelligently in order to avoid randomly placed traps in the trapping groove. After presenting the computational architecture for the internal models, we demonstrate how the robot can use them to mentally compose a sequence of ‘Push-Move-Reach’ in order to Grasp (an otherwise unreachable) ball in its playground.


This chapter reports to the development of the tools and methodologies that are in development within the EU, with an emphasis on the Open Source approaches with a view to performance analysis and comparison, and to provide an overview of cooperative research and especially on the use of Open platforms.


We present an evaluation of different techniques for the estimation of forces and torques measured by a single six-axis force/torque sensor placed along the kinematic chain of a humanoid robot arm. In order to retrieve the external forces and detect possible contact situations, the internal forces must be estimated. The prediction performance of an analytically derived dynamic model as well as two supervised machine learning techniques, namely Least Squares Support Vector Machines and Neural Networks, are investigated on this problem. The performance are evaluated on the normalized mean square error (NMSE) and the comparison is made with respect to the dimension of the training set, the information contained in the input space and, finally, using a Euclidean subsampling strategy. Key words: Force sensing, machine learning, humanoid robotics.


