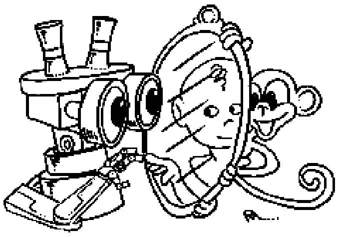

MIRROR

IST–2000-28159

Mirror Neurons based Object Recognition

Deliverable Item 1.7

Periodic Progress Report N°: 2

Covering period 1.9.2002-31.8.2003

Delivery Date: November 15th, 2003
Classification: Internal
Responsible Person: Prof. Giulio Sandini – University of Genova
Partners Contributed: ALL

Contract Start Date: September 1st, 2001 Duration: 30 Months

Project Coordinator and Partners:
DIST - University of Genova (Prof. Giulio Sandini and Dr. Giorgio Metta)
Department of Biomedical Sciences – University of Ferrara (Prof. Luciano Fadiga)
Department of Psychology – University of Uppsala (Prof. Claes von Hofsten)
Instituto Superior Técnico – Computer Vision Lab – Lisbon (Prof. Jose Santos-Victor)

Project funded by the European Community under the "Information Society Technologies" Programme (1998-2002)

Content list

1. Scientific background

[1]Vision and manipulation are inextricably intertwined in the primate brain. Neuroscientists are doing a very good job in elucidating the mixed structure of action and perception. We now know a great deal about this structure. By providing a plausible model of these same functions we can delve deeper into the whys: i.e. is this integration functionally important? If the answer is yes, how much is it important? A physical implementation, in the form of a robotic system, can shed new light into the linkage between acting and perceiving.

In fact, according to a prevalent view, actions do not constitute a semantic category such as animals, objects, people or buildings. Actions are defined as 'actions' because they are external, physical expressions of our intentions. It is true that very often actions are the response to external contingencies and/or stimuli but it is also certainly true that (at least in the case of human beings) actions can be generated on the basis of internal aims and goals, they are possibly symbolic and not related to immediate needs. Typical examples are communicative actions.

Actions are represented in the brain as words in a vocabulary (Rizzolatti and Fadiga, 1998) and experimental evidence (see Jeannerod, 1995) shows that action representation can be voluntarily evoked (as in the case of an individual imagining performing an action) without any detectable motor activation. However, what is somewhat surprising is the fact that during *motor imagery*, premotor and motor areas are selectively activated (Porro et al. 1996, Fadiga et al. 1999).

Another piece of evidence favors the unique semantic value of actions: the best way to describe a complex act to someone else is to show it directly. This is not true for objects such as animals or buildings (that we describe by using size, weight, color, texture, etc.). In other terms we describe 'things' by using visual categories and 'actions' by using motor categories.

Actions are all characterized by the presence of goals, and possibly an object: e.g. an apple, immediately evoking the action of biting. The target object represents a powerful cue in activating the brain's motor representations. In favor of this idea are electrophysiological studies of monkey premotor area F5 (see also Figure 1), showing that a subset of neurons (the so called *canonical neurons*) are activated not only when the monkey executes a goal directed hand action (such as grasping, manipulating, tearing, etc.) but also when it observes an object, pragmatically congruent with that action. It is remarkable that this specific visual activation is present also when the animal refrains to act on the basis of a previous no-go command (Murata et al. 1997).

More recently, "motor resonant" neurons have been observed both in premotor (area F5) and parietal (area PF) cortices of the macaque monkey (Gallese et al. 2002). These neurons are visuomotor neurons that are active when the monkey acts on an object and when it observes another monkey, or the experimenter, making similar goal-directed actions. Due to this property, these neurons have been called *mirror neurons*.

Typically, F5's mirror neurons, in order to respond, require the interaction between hand and an object. The sight of the object alone or of the agent pretending an action is not effective in getting their activation. The specific significance of the object for the monkey has no direct influence on the mirror neuron measured responses. Grasping a piece of food or a geometric solid produces responses of the same intensity.

---

[1] A longer and somewhat more thorough development of the same argument is included in last year's progress report. The discussion here borrows much of it.

An important functional aspect of mirror neurons is the correspondence between their visual and motor properties. Virtually all mirror neurons show congruence between the visual actions they respond to and the motor response they code. According to the type of congruence they exhibit, mirror neurons have been subdivided into “strictly congruent” and “broadly congruent” neurons (Gallese et al. 1996). “Strictly congruent” mirror neurons are those in which the effective observed and effective executed actions correspond both in terms of goal (e.g. grasping) and means, that is how the action is executed (e.g. precision grip). They represent about 30% of F5 mirror neurons. “Broadly congruent” are those that do not require the observation of exactly the same action they code motorically. Some of them discharge during the execution of a particular type of action (e.g. grasping) when executed using a particular grasp type (e.g. precision grip). However, they respond to the observation of grasping made by another individual, regardless of the type of grip used. Other broadly congruent neurons discharge in association with a single motor action (e.g. holding), but respond to the observation of two actions (e.g. grasping and holding). Broadly congruent neurons are the class of mirror neurons mostly represented (about 60%).

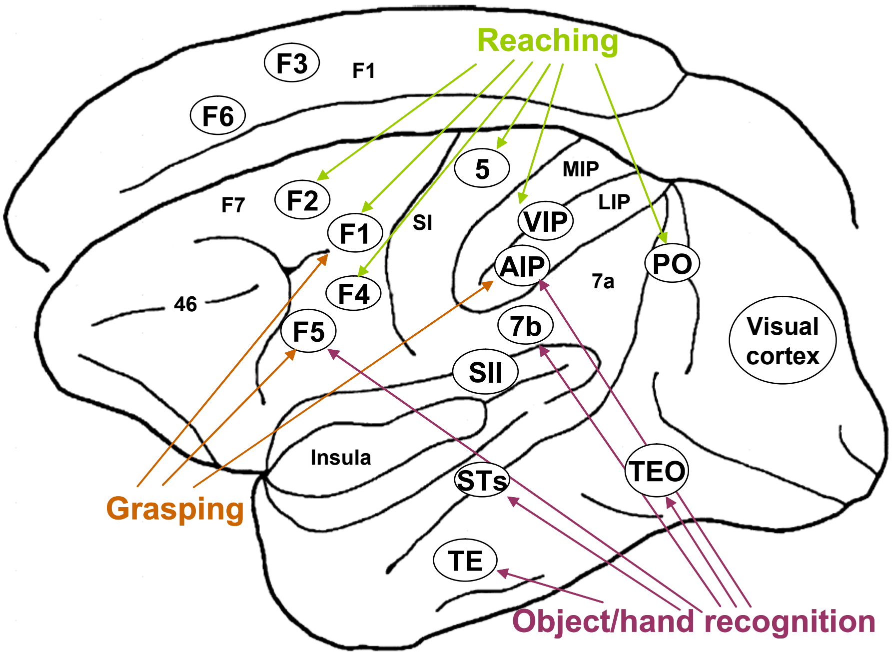

Figure 1 Monkey brain with indication of the main areas participating in object oriented actions (adapted from (Fagg and Arbib 1998)). Three main functions can be identified: object recognition, reaching, and grasping. These form three parallel yet connected streams of processing. The circuit connecting the visual cortex to the inferior parietal lobule VIP, F4 and F1 is thought to compute the visuomotor transformations required to control reaching. Some evidence also suggests a possible role in the organization of reaching played by the posterior parietal cortex PO and dorsal premotor area F2, reciprocally connected. AIP and F5 are responsible for grasping. Temporal areas (TE, TEO) and STS are correlated to the semantic of object recognition.

From this short review of basic properties of F5 neurons, it appears that this area is storage of potential actions or as we previously described it of a “vocabulary of actions” (Rizzolatti et al 1988). The activation of F5 neurons does not necessarily imply an actual action. It seems to only evoke the action’s representation. If other contingencies are met, this potential action becomes an actual motor action (see Rizzolatti and Luppino 2001). F5 potential actions can be activated endogenously or exogenously. Exogenous (visual) activation is caused by the observation of objects (canonical neurons) or by the observation of actions made by others (mirror neurons).

Another cortical area where there are mirror neurons is area PF (Fogassi et al. 1998; Gallese et al. 2002). This area forms the rostral part of the inferior parietal lobule. PF receives input from STS, where there are many neurons that become active during the observation of action (Perrett at al. 1989), and sends output to area F5. Neurons in area PF are functionally heterogeneous. Most of them (about 90%) respond to sensory stimuli (Hyvarinen 1982; Leinonen and Nyman 1979; Fogassi et al. 1998; Gallese et al. 2002). About 50% of them discharge also in association with monkey active movements. Neurons responding to sensory stimuli have been subdivided into three categories: “somatosensory” neurons (33%), “visual” neurons (11%), and “bimodal” somatosensory and visual neurons (56%). Among the neurons with visual responses (“visual neurons” and “bimodal neurons”), 41% respond to the observations of actions made by another individual. One third of them, however, similarly to STS neurons, do not appear to have motor-related activity. The other two-third discharge also during the monkey movement and, in most cases, showed the visuo-motor congruence typical of mirror neurons (Gallese et al. 2002).

With respect to development action plays a role too. In fact, animals are actors in their environment, not simply passive bystanders. They have the opportunity to examine the world using causality, by performing probing actions and learning from the response. In other words animals can act and consequently observe the effects of their actions. Effects can be more or less direct, e.g. I feel my hand moving as the direct effect of sending a motor command, or they can be eventually ascribed to complicate chains of causally related events producing what we simply call “a chain of causality”. For example, I see the object rolling as a result of my hand pushing it as a result of a motor command. Tracing chains of causality from motor action to perception (and back again) is important both to understand how the brain deals with sensorimotor coordination and to implement those same functions in an artificial system (such as Mirror’s humanoid robot). We proposed that such causal probing could be arranged in a developmental sequence leading along the way to a manipulation-driven representation of objects, to the perception/interpretation of manipulative actions, and to perceiving our own body. The same analysis could be used to explain why we observe certain developmental patterns or behaviors. Vice versa, by understanding development we can probe deeper the structure of a particular brain’s function.

Table 1 shows three levels of causal complexity. The simplest causal chain that an actor – whether *robotic or biological* – may experience is the perception of its own actions. The temporal aspect is immediate: visual information is tightly synchronized to motor commands. Once this causal connection is established, it/he/she can go further and use this knowledge about its own body to actively explore the boundaries of the environment (specifically objects). In this case, there is one additional step in the causal chain, and the temporal nature of the response may be delayed since initiating a reaching movement does not immediately elicit consequences in the environment. Finally, we argue that extending this causal chain further will allow the actor to make a connection between its own actions and the actions of another. This is clearly reminiscent of what has been observed in the response of the monkey's premotor cortex (area F5).

Type of activity	*Nature of causation*	*Time profile*
Sensorimotor coordination	Direct causal chain	Strict synchrony
Object probing	One level of indirection	Fast onset upon contact, potential for delayed effects

Constructing mirror representation	Complex causation involving multiple causal chains	Arbitrary delayed onset and effects
Object recognition	Complex causation involving multiple observations	Arbitrary delayed onset and effects

Table 1 Degrees of causal indirection. There is a natural trend from simpler to more complicated tasks. The more time-delayed an effect the more difficult it is to model.

An important aspect of the analysis of causal chains is the link with objects. Many actions are directed towards objects, they act on objects, and the goal eventually involves to some extent an object. For example, Woodward (Woodward, 1998), and Wohlschlager and colleagues (Wohlschlager and Bekkering, 2002) have shown that the presence of the object and its identity change the perception and the execution of an action.

1.1. Do F5 mirror neurons play a role in action understanding?

From the very first moment of the discovery of mirror neurons it was suggested that they could play a role in action understanding. The core of this proposal is the following:

> *When an individual acts he selects an action whose motor consequences are known to him. The mirror neurons allow this knowledge to be extended to actions performed by others.*

Each time an individual observes an action executed by another individual, neurons that represent that action are activated in his or her premotor cortex. Because the evoked motor representation corresponds to that internally generated during active action, the observer understands the observed action (see Rizzolatti et al 2001).

This action recognition hypothesis was recently tested by studying mirror neuron responses in conditions in which the monkey was able to understand the meaning of the occurring action, but without the visual stimuli that typically activate mirror neurons. The rationale of the experiments was the following: if mirror neurons are involved in action understanding, their activity should reflect the meaning of the action and not the specific sensory contingencies. In a series of experiments the hypothesis was tested by presenting auditory stimuli capable of evoking the ‘idea’ of an action (Kholer et al. 2002).

F5 mirror neuron activity was recorded while the monkey was either observing an action producing a sound (e.g. ripping a piece of paper), or hearing the same noise without visual information. The results showed that most mirror neurons that discharge to presentation of actions accompanied by sounds, discharge also in response to the sound alone (“audio-visual” mirror neurons). The mere observation of the same “noisy” action without sound was also effective. Further experiments showed that a large number of audiovisual mirror neurons respond selectively to specific sounds (linked to specific actions). These results strongly support the notion that the discharge of F5 neurons correlates with action understanding and not simply with the stimuli that determine it. The effective stimulus might be visual or acoustic. The neuron fires as soon as the stimulus has specified the meaning.

Another series of experiments studied mirror neurons’ responses in conditions where the monkey was prevented from seeing the final part of the action (and listening to its sound), but were provided with clues on what the action might be. If mirror neurons are involved in action understanding they should discharge also in this condition. This experiment was recently carried out by Umiltà et al. (2001). The experimental paradigm consisted of two basic conditions. In the first condition, the monkey was shown a fully visible action directed toward an object (“full vision” condition). In the second condition, the monkey watched at the same action but with its final critical part hidden (“hidden” condition). Before each trial the experimenter placed a piece of food behind the screen so that the monkey knew that there was an object behind it. The main result of the experiment was that more than half of the tested neurons discharged in hidden condition. Some did not show any difference between hidden and full vision condition, others responded stronger in full vision. In conclusion, both these experiments showed that F5 mirror neuron activation correlates with action representation rather than with the stimulus properties leading to it. This finding strongly supports the notion that F5 activity plays a fundamental role in the understanding of the meaning of action.

1.2. Modeling at large

The first attempt of modeling perception and action altogether was started several decades ago by Alvin Liberman, aiming at the construction of a ‘speech understanding’ machine (Liberman et al. 1967, Liberman and Mattingly, 1985, Liberman and Whalen, 2000). As one can easily imagine, the first effort of Liberman’s team was directed at analyzing the acoustic characteristics of spoken words, to investigate whether the same word, spoken by different subjects, possessed any common phonetic invariant. Soon Liberman and his colleagues understood that speech recognition on the basis of acoustic properties could not be achieved with the limited computational power available at that time. Somewhat stimulated by this negative result, they put forward the hypothesis that the ultimate constituents of speech are not sounds but rather articulatory gestures that have evolved exclusively at the service of language. Accordingly, a cognitive translation into phonology is not necessary because the articulatory gestures are phonologic in nature. This elegant idea was however strongly debated, mainly because its implementation into a real system was impossible and it only now supported by experimental evidence (Kerzel and Bekkering 2001, Fadiga et al. 2002).

Why is it that, normally, humans can visually recognize actions (or, acoustically, speech) with an approximation of about 99-100%? Why the inter-subject variability typical of motor behavior does not represent a problem for the brain while it is troublesome for machines? One possibility is that Liberman was right in saying that speech perception and speech production use a common repertoire of motor primitives that during speech production are at the basis of articulatory gestures generation, while during speech perception are activated in the listener as the result of an acoustically-evoked motor “resonance”. With the only difference of the sensory modality, this sentence might be true also for other, visually perceived, actions. What, in both cases, the brain needs is a “resonant” system that matches the observed/listened actions on the observer/listener motor repertoire. Note that, an additional advantage of such an empathic system would be the capability to automatically “predict”, at some extent, the future development of somebody else’s actions on the basis of the observer implicit knowledge (on the same actions).

1.3. A working hypothesis for modeling mirror neurons

Taken together the results from neuroscience suggest a critical role for motor action in perception. Certainly vision and action are intertwined at a very basic level. While an experienced adult can interpret visual scenes perfectly well without acting upon them, linking action and perception seems crucial to the developmental process that leads to that competence. We can construct a working hypothesis: that action is required whenever the animal (or our artifact in the following) has to develop autonomously. Further, as we argued above, the ability to act is also fundamental in interpreting actions performed by a conspecific. Of course if we were in standard supervised learning setting action would not be required since the trainer would do the job of pre-segmenting the data by hand and providing the training set to the machine. In an ecological context, some other mechanism has to be provided. Ultimately this mechanism is the body itself and the ability of being the initiator of actions that by means of interaction (and under some suitable developmental rule) generate percepts informative to the purpose of learning.

Grossly speaking, a possible developmental explanation of the acquisition of these functions can be framed in terms of tracing/interpreting chains of causally related events. Although it is still speculative, this analysis predicts that i) development of functions roughly follows a dorsal to ventral temporal gradient (i.e. see reaching, grasping, recognition in Figure 1); ii) the ability to probe longer chains triggers the emergence of new functionality and/or a new set of behaviors.

We can distinguish four main conceptual functions (similar to the schema of Arbib et al. (Arbib, 1981)): reaching, grasping (manipulation), and object recognition. These functions correspond to the three levels of causal understanding introduced in Table 2. They form also an elegant progression of abilities which emerge out of very few initial assumptions. All that is required is the interaction between the actor and the environment, and a set of appropriate developmental rules specifying what information is retained during the interaction, the nature of the sensory processing, the range of motor primitives, etc.

The results outlined in the previous sections can be streamlined into a developmental sequence roughly following a dorsal to ventral gradient. Unfortunately this is a question which has not yet been investigated in detail by neuroscientists, and there is very little empirical support for this claim (apart from (Bertenthal and von Hofsten 1998) and (Kovacs, 2000)).

What is certainly true is that the three modules/functions can be clearly identified. If our hypothesis is correct then the first developmental step has to be that of transporting the hand close to the object. In humans, this function is accomplished mostly by the circuit VIP-7b-F4-F1 and by PO-F2-area 5. Reaching requires at least the detection of the object and hand, and the transformation of their positions into appropriate motor commands. Parietal neurons seem to be coding for the spatial position of the object in non-retinotopic coordinates by taking into account the position of the eyes with respect to the head. According to (Pouget, Ducom, Torri, & Bavelier, 2002) and to (Flanders, Daghestani, & Berthoz, 1999) the gaze direction seems to be the privileged reference system used to code reaching. Relating to the description of causality, the link between an executed motor action and its visual consequences can be easily formed by a subsystem that can detect causality in a short time frame (the immediate aspect). A system reminiscent of the response of F4 can be developed by the same causal mechanism.

Once reaching is reliable enough, we can start to move our attention outwards onto objects. Area AIP and F5 are involved in the control of grasping and manipulation. F5 talks to the primary motor cortex for the fine control of movement. The AIP-F5 system responds to the “affordances” of the observed object with respect to the current motor abilities. Arbib and coworkers (Fagg & Arbib, 1998) proposed the FARS model as a possible description of the computation in AIP/F5. They did not however consider how affordances can be actually learned during interaction with the environment. Learning and understanding affordances requires a slightly longer time frame since the initiation of an action (motor command) does not immediately elicit a sensory consequence. In this example, the initiation of reaching requires a mechanism to detect when an object is actually touched, manipulated, and whether the collision/touch is causal to the initiation of the movement.

The next step along this hypothetical developmental route is to acquire the F5 mirror representation. We might think of canonical neurons as an association table of grasp/manipulation (action) types with object (vision) types. Mirror neurons can then be thought of as a second-level associative map which links together the observation of a manipulative action performed by somebody else with the neural representation of one's own action. Mirror neurons bring us to an even higher level of causal understanding. In this case the action execution has to be associated with a similar action executed by somebody else. The two events do not need to be temporally close to each other. Arbitrary time delays might occur.

The conditions for when this is feasible are a consequence of active manipulation. During a manipulative act there are a number of additional constraints that can be factored in to simplify perception/computation. For example, detection of useful events is simplified by information from touch, by timing information about when reaching started, and from knowledge of the location of the object.

Subsequently object recognition can develop. Object recognition can build on manipulation in finding the physical boundaries of objects and segmenting them from the background. More importantly, once the same object is manipulated many times the brain can start learning about the criteria to identify the object if it happens to see it again. These functions are carried out by the infero-temporal cortex (IT). The same considerations apply to the recognition of the manipulator (either one's own, or another's). In fact, the STS region is specialized for this task. Information about object identity is also sent to the parietal cortex and contributes to the formation of the affordances. However object recognition is performed, at a minimum all information (visual in this case) pertaining to a certain object needs to be grouped during development so that a model of the object can be constructed.

Nature of causation	*Main path*	*Function and/or behavior*
Direct causal chain	VC-VIP/LIP/7b-F4-F1	Reaching
One level of indirection	VC-AIP-F5-F1	Grasping
Complex causation involving multiple causal chains	VC-AIP-F5-F1+STs+IT	Mirror neurons, mimicry
Complex causation involving multiple instances of manipulative acts	STs+TE-TEO+F5-AIP(?)	Object recognition

Table 2 Degrees of causal indirection, localization and function in the brain.

1.4. Core model

From the above discussion, two core elements of the prospective mirror neurons model emerge: they are the use of motor information (or coding) also during the recognition of somebody else's actions and the use of object affordances (we provided support for the relevance of the target object during action execution).

In practice, many objects are grasped in very precise ways, since they allow the object to be used for some specific purpose or goal. A pen is usually grasped in a way that affords writing and a glass is hold in such a way that we can use it to drink. Hence, if we *recognize* the object being manipulated, then recognition immediately provides some information about the most likely grasping possibilities (expectations) and hand appearance, simplifying the task of gesture recognition.

The affordances of the object possess an attentional-like[2] property because the number of possible (or likely) events is reduced. Affordances provide expectancies that can be used to

---

[2] *Attention* in the sense of selecting relevant information out of a possibly much larger space.

single out possible ambiguities. This has clearly to be a module of our overall system architecture.

The common approach to recognition involves comparing acquired visual features to data from a training set. Differently, our approach is based on the use a *Visual-Motor Map* (VMM) to convert such measurements to a motor space and then perform the comparison/recognition in terms of motor representations. The advantage of doing this inference in motor space is two-fold. Firstly, while visual features can be ambiguous, we were able to show that converting these features to the motor space might reduce ambiguity. Secondly, since motor information is directly exploited during this process, imitative behaviors could be trivially implemented given that all the information/signals are already available.

To use motor representations for grasp recognition, we need to define *Visuo-Motor maps* to transform visual data onto motor information. The VMM can be learnt during an initial phase of self-observation, while the robot performs different gestures and learns their visual effects. The question that remains to be addressed is that of choosing what **visual features** to use. As we will focus on the classification and imitation of coarse gestures (e.g. power grasp and precision grip), we will rely on global appearance-based image methods. Together with the prior information provided by the “canonical neurons” (or their artificial implementation), appearance based methods offer an easier, fast and more robust representation than point tracking methods. In the next section we will present a Bayesian approach for a gesture recognition that includes models of the *canonical* and *mirror* neurons, using visual appearance methods.

1.5. A Bayesian model for canonical and mirror neurons

Gesture recognition can be modeled within a Bayesian framework, which allows naturally combining *prior* information and knowledge derived from observations (likelihood). The role played by canonical and mirror neurons will be interpreted within this setting.

Let us assume that we want to recognize (or imitate) a set of gestures, G_i, using a set of *observed* features, $\boldsymbol{F}$. For the time being, these features can either be represented in the motor space (as mirror neurons appear to do) or in the visual space (directly extracted from images). Let us also define a set of objects, O_k, that can happen to be observed in the scene (not simultaneously) and which are the goals of a certain grasp actions.

Prior information is modeled as a probability density function, $p(G_i|O_k)$, describing the probability of each gesture given a certain object. The observation model is captured in the *likelihood function*, $p(F|G_i,O_k)$, describing the probability of observing a set of (motor or visual) features, conditioned to an instance of the pair gesture and object. The *posterior* density can be directly obtained through Bayesian inference:

$$\begin{aligned} p(G_i|F,O_k) &= p(F|G_i,O_k)p(G_i|O_k)/p(F|O_k), \\ \hat{G}_{MAP} &= arg\max_{G_i} p(G_i|F,O_k) \end{aligned} \quad (1)$$

where $p(F|O_k)$ is just a scaling factor that will not influence the classification. The *MAP* estimate, G_{MAP}, is the gesture that maximizes the posterior density in Equation (1). In order to introduce some temporal filtering (since the information over time is available), features of several images can be considered:

$$p(G_i|F, O_k) = p(G_i|F_t, F_{t-1}, ..., F_{t-N}, O_k)$$

where F_j are the features corresponding to the image at time instant *j*. The posterior probability distribution can be estimated using a naive approach, assuming independence between the observations at different time instants. The justification for this assumption is that recognition does not necessarily require the accurate modeling of the density functions. We then have:

$$p(G_i|F_t, ..., F_{t-N}, O_k) = \prod_{j=0}^{N} \frac{p(F_{t-j}|G_i, O_k)p(G_i|O_k)}{p(F_{t-j}|O_k)}$$

The role of canonical neurons in the overall classification system lies essentially in providing affordances, modeled here as the *prior* density function, $p(G_i|O_k)$ that, together with evidence from the observations, will shape the final decision. This density can be estimated by the relative frequency of gestures in the training set. In practice, if we were to work with a complete system estimation and learning of affordances would require a much more complex learning procedure. The ultimate goal would still be the estimation of prior probabilities (that could still be done by estimating the relative frequencies of actions) but acquiring the visuo-motor information autonomously is perhaps a feat in itself.

Canonical neurons are also somewhat involved in the computation of the likelihood function since they respond both on the *gesture* and *object* (and in the model $p(G_i|O_k)$ shows this relationship), thus implicitly defining another level of association. Computing the likelihood function, $p(F| G_i, O_k)$, might be difficult since the shape of the data clusters might be quite complicate. We modeled these clusters as mixtures of Gaussian and the Expectation-Maximization algorithm was used to determine both the number of the Gaussian terms and their coefficients.

Mirror neurons are clearly represented by the responses of the maximization procedure since both motor and visual information determine the activation of a particular unit (for real neurons) and the corresponding probability (for artificial neurons).

1.6. A caveat

Although on a superficial reading it might seem that the Bayesian model encompasses all what it has to be said about mirror neurons, in fact it is substantially a supervised learning model. To relax the hypothesis of having to “supervise” the machine during training by indicating which action is which we need to remind what the evidence on mirror neurons tells us. First of all, it is plausible that the ‘canonical’ representation is acquired by self exploration and manipulation of a large set of different objects. F5 canonical neurons represent an association between objects’ physical properties and the action they afford: e.g. a small object affords a precision grip, or a coffee mug affords being grasped by the handle. This understanding of object properties and the goal of actions is what can be subsequently factored in while disambiguating visual information. There are at least two level of reasoning: i) certain actions are more likely to be applied to a particular object – that is, probabilities can be estimated linking each action to every object, and ii) objects are used to perform action – e.g. the coffee mug is used to drink coffee. Clearly, we tend to use actions that proved to lead to certain results or, in other words, we trace backward the link between action and effects: to obtain the effects apply the same action that earlier led to those effects.

Bearing this is mind, when observing some other individual's actions; our understanding can be framed in terms of what we already know about actions. In short, if I see someone drinking from a mug I can hypothesize a particular action (that I know already in motor terms) is used to obtain that particular effect (of drinking). This link between mirror neurons and the goal of the motor act is clearly present in the neurons' response. It is also the only possible way of autonomously learning a mirror representation. Technically speaking, the learning problem is still a supervised one but the information can now be collected autonomously. The association between the canonical response (object-action) and the mirror one (including vision of course) is made when the observed consequences (or goal) are recognized as similar in the two cases – self or others acting. Similarity can be evaluated following different criteria ranging from kinematic (e.g. the object moving along a certain trajectory) to very abstract (e.g. social consequences such as in speech).

2. Experiments and status of the project

The organization of the experiments conducted so far is sketched in Figure 2. The formalization of the model is what we have described above (sections 1 to 1.5). The outcome of this conceptual modeling activity is both a general outline of the development of motor skills in humans and a specific probabilistic model of the functioning of the mirror system.

The investigation within Mirror has two specific and somewhat interrelated goals: i) clarification of some of the questions or gaps within the model or understanding of the functioning of the mirror system, and ii) implementation of the model in an autonomous learning/developing artifact. Questions asked include for instance "What is the contribution of the vision of the hand in the response of mirror neurons?" or "What is the time course of the development of manipulative skills?" Experiments cover aspects ranging from the development of reaching/hand localization towards understanding of the contribution/role of mirror neurons to communicative behaviors. In the following we describe in some detail each experiment and their contribution to the final picture. Experiments were conducted on one side on animals or human subjects and, on the other side on various robotic prototypes. Part of the activity of Mirror (especially during the first year) consisted in preparing the experimental setups – for example, the robot has been now equipped with a five-finger hand, or new equipment for recording children's behavior has been acquired. Results are encouraging. The next phase of the project will concentrate on integrating the different parts and experimental components in the final demo setup and in comparing the implementation with predictions derived from the model or biological considerations.

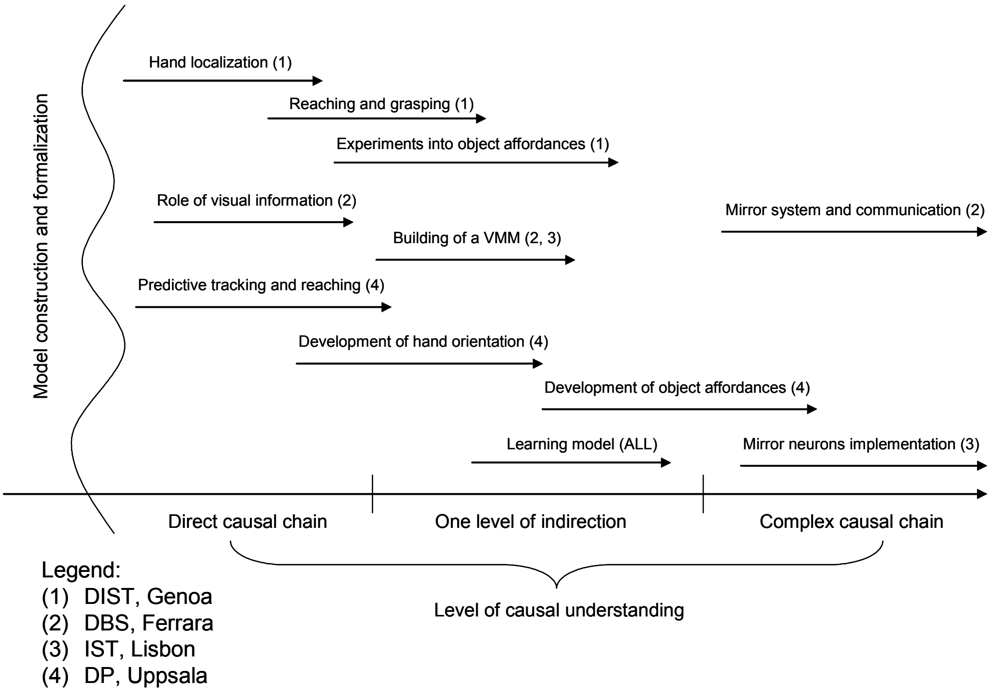

Figure 2 Experiments conducted within Mirror. Indication of the placement of the each experiment within the model and the partner that contributed most to it is indicated.

The following table contains a short description of each experiment as reported also in Figure 2 together when available with the reference to a paper or manuscript describing the experiments in full details. It is worth stressing that this is the complete collection of experiments performed so far (from project start date on September 2001). For details on the specific advancement during Y2 please refer to the following sections.

Name	Description
Hand localization DIST	Learning how to reach for visually identified object requires the identification and the recognition of where the arm/hand is and how to place it with respect to the environment. Similarly multi-sensory responsive neurons of F4 or VIP seem to be coding for the position in space of the body with respect to the environment. We implemented a procedure where the robot starting with minimal assumptions learns an egocentric representation of the body as it moves. Feedforward neural networks arranged within a self-supervised schema were used. Paper: NO PAPERS YET.
Reaching and grasping DIST, DP	Reaching requires the appropriate timing of motor commands to move the hand near the object of interest. Two aspects have been investigated: i) how a robot can control movements so to reach for visually identified objects and ii) how children learn to predict the position in space where contact with a moving object will occur. Paper: See below.
Predictive tracking and reaching	An experimental paradigm has been established to investigate the development of infants' predictive reaching for moving

DP	objects. We are running an experiment where we are asking 7/8-month-old infants to catch objects moving on an ellipsoid trajectory with either constant angular velocity (sinusoidal modulation) or constant tangential velocity (constant speed). Both types of motion are found in nature. Animals decelerate when they go around a bend and the pendulum motions that underlie our body movements also adhere to these principles. On the other hand, objects that move passively continue with the same speed if they are forced to turn. The reason why adults perceive object motion in this way could be because they have much experience with biological motion or it could be an inherent constraint on the perception of motion. Also, infants' ability to smoothly track objects of different size, track them along different trajectories, and over occlusions has been studied. The focus has been on the emergence of predictive tracking of temporarily occluded objects. We have pursued this effort with two kinds of eye movement recordings: EOG and cornea reflection. Paper: von Hofsten, C. (2003) *The development of prospective control in looking*. To be published in J. Lockman J. Riese (Eds.) Action as an Organizer of Perception and Cognition during Learning and Development. Minnesota symposium on Child Psychology.
Experiments on object affordances DIST	In a set of experiments we explored the acquisition of object affordances in a simplified situation where the robotic artifact could only poke objects. Objects where chosen carefully so to behave differently if poked from different directions. It was possible to show that the robot could autonomously acquire the information to characterize this simple affordance. It was also possible to show that this information could be reused later on if the robot encounters the same object again. A simple form of mimicry was demonstrated. Simple lookup-table based learning was used throughout this set of experiments. Paper: Paul Fitzpatrick and Giorgio Metta (2003) *Grounding vision through experimental manipulation*. Philosophical Transactions of the Royal Society: Mathematical, Physical, and Engineering Sciences, 361:1811, pp. 2165-2185.
Role of visual information UNIFE	We recorded more than 100 neurons both in F5 and F1, and we submitted to formal testing more than 80% of them. Briefly, the goal of our exploration was to establish if F5 premotor neurons are sensitive to the vision of one's own acting hand. The experiment aimed at verifying one of the possible explanations of the origin of mirror neurons: i.e. they might have developed from the visual feedback system that controls own execution of grasping. Thus, first a visuomotor association is established between a given motor command and the visual cues arising during its execution; subsequently the system might have been capable of generalizing the association to other people's hand, guided by the invariance of the motor information. In short the experiment consists in testing the same set of mirror neurons in darkness and in various lighting conditions.

	Paper: Fadiga L., Craighero L. *Electrophysiology of action representation*. Journal of Clinical Neurophysiology, In press.
Building a VMM IST, UNIFE, DIST	One of the questions in building a VMM is whether stereoscopic vision is really necessary to create the visuomotor map and how much finger occlusion during grasping influences action recognition. To answer to some of these questions we are now investigating the capability of predicting the outcome of observed actions while experimentally manipulating the various viewing conditions. To this purpose we are mainly focused on a particularly relevant phase of the grasping actions: the instant at which the surface of the fingers touches the target object. The capability of predicting the action's outcome is indicated by the difference in time between the actual time of contact with the object and the occurrence of subject's response. Preliminary results are encouraging. Peculiarly, it seems that subjects' performance is not significantly influenced by monocular-binocular vision. Another investigation into the building of a VMM consisted in building a map from visual space to motor space. We collected an extensive data set of a person grasping a set of objects using our data-glove setup (data-glove and cameras). Motor (the position of the hand's joints) and visual information was recorded. We were able to train a feedforward neural network to correctly approximate the mapping from a certain set of global visual features (PCA components in practice) into the hand's joint angles. Paper: M. Cabido Lopes and J. Santos-Victor (2003). *Visual Transformations in Gesture Imitation: what you see is what you do*. ICRA – IEEE International Conference on Robotics and Automation, Taiwan, September 2003.
Mirror system and communication UNIFE	We are investigating in humans how the mirror system could be involved in communication. By using transcranial magnetic stimulation (TMS) we made some preliminary observations showing that a motor resonance, similar to that observed in monkey mirror neurons, can be evoked not only by action viewing but also when a subject is passively listening verbal stimuli. It is clear that, in this case, the "mirror" effect involves at the cortical level not the hand but the tongue motor representation. On one side we are investigating whether the motor resonance induced by speech listening represents a mere epiphenomenon or rather it reflects an involvement of motor centers in speech perception (as suggested by the Liberman's motor theory of speech perception). We are currently designing an fMRI experiment aiming to investigate whether cortical speech centers, and particularly the frontal ones, are specifically tuned for "verbal" communication or provide a neural substrate in which sequences of movements, individually meaningless, are translated into meaningful representations. We are therefore investigating the effect of gestural, non-symbolic, non-verbal communication on inferior frontal gyrus and particularly on area 44, which is considered

	the human homologue of the monkey premotor area (F5) where mirror neurons have been located. Paper: Fadiga L., Craighero L., Buccino G., Rizzolatti G. (2002). *Speech listening specifically modulates the excitability of tongue muscles: a TMS study*, European Journal of Neuroscience, Oxford, 15: 399-402.
Development of hand orientation DP	The aim of this experiment is to investigate when and how infants start to control hand posture in relation to the shape of the object to be grasped. This ability is shown very clearly in adults by the "pre-shaping" of the hand during the transport phase of grasps. For practical reasons it is very hard to precisely measure the posture of the hand of infants during grasping (e.g. no "data-glove" is available for infant-hand size) and therefore, it is experimentally difficult to investigate the onset of pre-shaping abilities. It was decided to simplify the measure by assuming that the orientation of the hand with respect to a rod-like object can be studied as an example of pre-shaping ability. In several ways, the results indicated that approaching and grasping an object are independent actions. First, the analysis of movement units showed that the rotation of the rod affected the rotational adjustments of the hand but not the approach to the rod. Second, the maximum approach velocity was not dependent on the rotational velocity of the rod but the maximum rotational velocity of the hand was. Finally, the small correlations between the rotational velocity and approach velocity support the conclusion that these two actions are relatively independent. Also, the results showed that the grasping of the rod is prospectively controlled. Paper: von Hofsten, C. and Johansson, K. (2003). *Planning to reach for a rotating rod: developmental aspects*. Manuscript.
Development of object affordances DP	Children's ability to adjust the orientation of objects with various shapes in order to fit them into holes is studied. This project began already during year one. The experiments utilize the natural interest of young children in fitting objects into holes. By varying the form of the objects and the holes, the difficulty of the task can be manipulated. Pre-adjustments of the orientation of the various objects before trying to push them through the holes, give information about the subjects spatial cognition as well as their ability to plan these actions. Some experiments have been completed and others are planned. The fitting task reflects basic abilities of spatial cognition. The subject has to realize how the 3-dimensional object has to be turned to fit into the 2-dimensional aperture. The main interests of this study are to investigate at what age infants can perceive properties of objects such as their shape and their orientation in relation to the goal of inserting it into the aperture, and whether they can already elaborate an action plan. At the moment we are completing a study that included infants from 12 to 24 months of age. 71 infants participated.

	Paper: NO PAPERS YET.
Learning model ALL partners	The model is the Bayesian one described in the previous sections. Investigation proceeded along various directions. As we have seen above part of learning consists in acquiring a representation of the artifacts own body. This allows localizing the hand in space by vision and proprioception and enables a whole new set of exploratory behaviors. More focused on the final goal of acquiring a mirror-like representation additional sub-problems have been tackled: • Learning of VMM. In the following model the VMM is a mapping from the visual space to motor space that allows running the MAP classifier in motor space rather than directly in visual space. Experiments using feedforward neural networks were executed. • Learning classification. The estimation of the various probabilities was carried out simply by estimating the frequency of occurrences in the simple cases or by the EM algorithm in some of the most complicated cases. • Learning affordances. In some of the experiments lookup table methods have been used. Given the focus of the specific experiments and the difficulty of the learning problem we found them to be adequate. Papers: M. Cabido-Lopes, J. Santos-Victor (2003). *Motor representations for hand gesture recognition and imitation*. IROS workshop on robot programming by demonstration. Las Vegas, USA, October 2003.
Mirror neurons implementation IST	Learning of mirror neurons in our model is equivalent to building the probabilistic association between observed visual and motor features, the grasp type, and the object towards which the action is directed to. As we have already mentioned the EM algorithm was employed in this case. Also it is important to stress the role of the VMM in transforming information into motor terms. We proved that this transformation improves classification performances at least in a simple case. The goal was here to show how real world image information and motor data (acquired with the data-glove setup) could be used in training a mirror-like representation. Papers: M. Cabido-Lopes, J. Santos-Victor (2003). *Motor representations for hand gesture recognition and imitation*. IROS workshop on robot programming by demonstration. Las Vegas, USA, October 2003.

2.1. Some specific experiments towards integration

We have already discussed to some extent the role of canonical neurons and their combination with mirror neurons especially for what relevant to our model. Another fundamental component of the model is a procedure for mapping visual features into motor ones. This is called a Visuo-Motor Map (VMM) and requires i) the definition of a suitable set of visual (image) features, and ii) the design of a machine learning procedure. Within the formalism developed in our model the **Visuo-Motor Map** transforms the features **F** from the visual space to the motor space.

$$VMM : \mathbf{F}^V \to \mathbf{F}^M$$

In short, visual features were chosen to be a set of the 15 first PCA components of the input images. PCA proved to be robust, being a global method, and efficient. Learning of the VMM was carried out by using a multi-layer perceptron neural network trained with the backpropagation algorithm. For the results presented here, several subjects were asked to perform different types of grasp on different objects. The experiment begins with the subject sitting in a chair, with the hand on the table. Then, the subject is told to grasp the object that is in front of him. The data set was collected using the data-glove setup developed as part of Mirror.

The experiments included two types of grasp: **power grasp and precision grip**. Power grasp is defined when all the hand fingers and palm are in contact with the object. Instead, in precision grip, only the fingertips touch the object. The three objects considered were a small sphere, a large sphere and a box. The size of the small sphere is such that only precision grip is possible. The big sphere affords only power grasps. The box is ambiguous, as it could be grasped with all grasps with different orientations. Each grasp type was repeated several times under varying conditions by many subjects. Figure 3 shows sample images of the data set acquired according to process just described. Notice the multiplicity of grasps and view points.

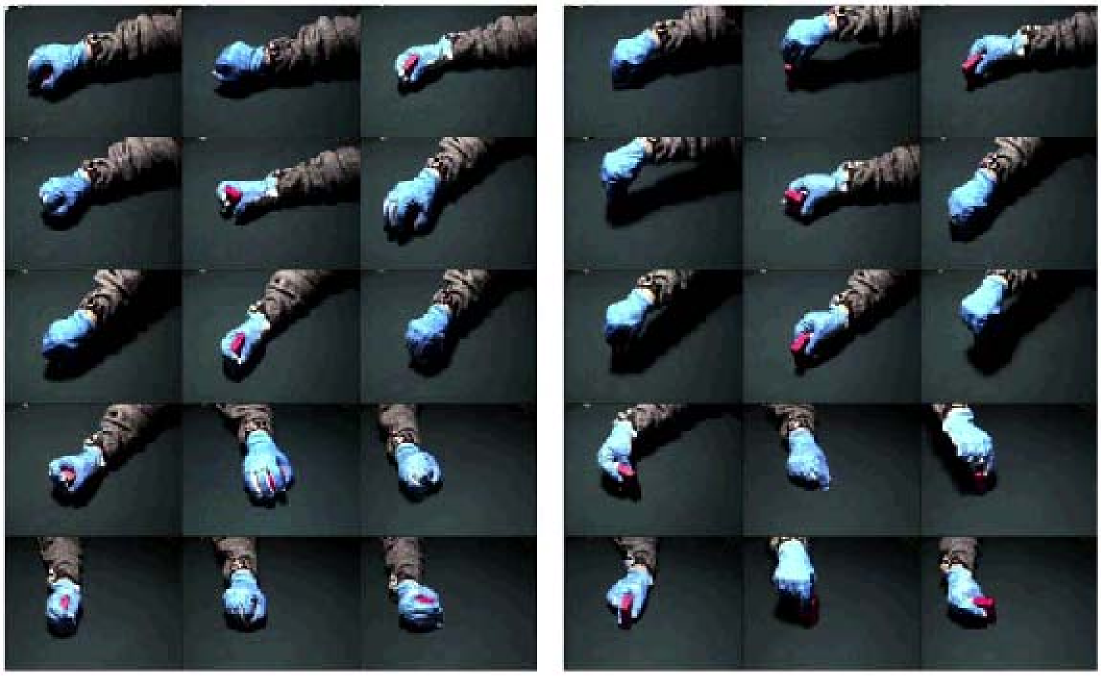

Figure 3: Data set illustrating some of the grasp types used: power (left) and precision (right). Altogether the tests were conducted using 60 sequences, from which a total of about 900 images were processed.

Table 3 shows the obtained classification rates. It allows us to compare the benefits of using motor representations for recognition as opposed to visual information only. The results shown correspond to the use of the ambiguous objects only, when the recognition is more challenging. We varied the number of viewpoints included in both the training and test sets, so as to assess the degree of view invariance attained by the different methods.

In a first set of three experiments the performance of the Bayesian classifier was evaluated in visual space: i.e. **F** is simply the set of PCA components. Results are reported in the first three columns of Table 1. Different conditions with different training sets were analyzed. Experiments 1 refers to the single viewpoint case (i.e. images are taken from the same constant camera orientation): clearly performance is good (100%) since no much generalization is required. Experiment 2 shows that the same training as in experiment 1 does not generalize. Testing of the classifier in 1 in the general case gives less than chance scores (30%). Experiment 3 contains a fairer training set from many possible viewpoints. Performances are back to 80%. The **fourth experiment** corresponds to the modeling of mirror neurons. The system uses the learned VMM to transform the (segmented) hand images into motor space (joints) where classification is performed. Very high scores were achieved (97%). Interestingly, the number of modes (we are talking about the mixture of Gaussian mentioned above) needed for the learning is only 1 to 2 in this case as opposed to 5 to 7 when recognition takes place in the visual domain. This also shows that mapping visual data to motor representations helps clustering the data which is now viewpoint invariant. Note also that viewpoint invariance is achieved when the training set only contains sequences from one single view point (this is a non trivial achievement).

	Exp. I (visual)	Exp. II (visual)	Exp. III (visual)	Exp. IV (motor)
	Training			
# Sequences	16	24	64	24
View Points	1	1	4	1
Classif. Rate	100%	100%	97%	98%
# Features	5	5	5	15
# Modes	5-7	5-7	5-7	1-2
	Test			
# Sequences	8	96	32	96
View Points	1	4	4	4
Classif. Rate	100%	30%	80%	97%

Table 3. Recognition results. To note: the improvement obtained in classification rates and viewpoint invariance due to the use of motor features.

We believe that this first set of experiments provides automatically an outline of the integration plan. Since the final goal of the project is that of demonstrating a complete robotic system that learns autonomously the complete mirror model, the next step is clearly that of integrating the learning algorithm and the classifier into the robot. We already started experimenting with hand/arm localization and object visual feature collection within a more dynamic context (e.g. moving hand/arm, moving cameras, etc.).

In summary, one of the major issues we foresee for the last year of the project is the integration of the results into a coherent whole: that is the developing artifact. Additional experiments are certainly planned but they are less demanding in terms of overall effort. The plan for integration is to start as soon as possible integrating a good part of the robotic/machine learning experiment into the humanoid robot in Genoa. The setup has been lately completed with a five-finger hand that should allow proper experimentation with different action/grasp types in a reasonably natural environment. It is worth noting that we are not starting from a "blank slate" since many sub-components both visual (sensorial) and behavioral are already functional on the robot (processing of motion, color, and binocular disparity for example). Simple primitive hand motions have already been implemented as part of the testing of the hand mechanics. The control of the head and in particular of gaze has been already implemented in the past.

We are well aware of the difficulty of combining machine learning and real-time control in a complex robotic system. A bit of uncertainty comes also from the mechanics of the interaction of the hand with the environment. Although this has been already considered at design stage by including elastic elements into the actuation structure, it might still be possible to encounter unexpected difficulties. The consortium nonetheless has a proven and long-lasting experience in many different aspects of robotics and biologically as well as traditional control theory.

3. References

Bertenthal, B., & Von Hofsten, C. (1998). Eye, Head and Trunk Control: the Foundation for Manual Development. *Neuroscience and Behavioral Reviews*, 22(4), 515-520.

Fadiga L, Buccino G, Craighero L, Fogassi L, Gallese V, Pavesi G. Corticospinal excitability is specifically modulated by motor imagery: a magnetic stimulation study. *Neuropsychologia* 1999;37:147-58.

Fadiga L, Craighero L, Buccino G, Rizzolatti G. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur J Neurosci* 2002;15:399-402.

Fagg AH, Arbib MA. Modeling parietal-premotor interactions in primate control of grasping. *Neural Netw* 1998;11:1277-303.

Fogassi L, Gallese V, Fadiga L, Rizzolatti G. Neurons responding to the sight of goal directed hand/arm actions in the parietal area PF (7b) of the macaque monkey. *Society of Neuroscience Abstracts* 1998;24:257.5.

Gallese V, Fadiga L, Fogassi L, Rizzolatti G. Action recognition in the premotor cortex. *Brain* 1996;119 (Pt 2):593-609.

Gallese V, Fogassi L, Fadiga L, Rizzolatti G. Action representation and the inferior parietal lobule. In *Attention & Performance XIX. Common mechanisms in perception and action*, ed. W Prinz, B Hommel. Oxford: Oxford University Press 2002.

Hyvarinen J. Posterior parietal lobe of the primate brain. *Physiol Rev* 1982;62: 1060-129.

Jeannerod M. Mental imagery in the motor context. *Neuropsychologia* 1995;33:1419-32.

Kerzel D, Bekkering H. Motor activation from visible speech: evidence from stimulus response compatibility. *J Exp Psychol Hum Percept Perform* 2000;26:634-47.

Kohler E, Keysers C, Umilta MA, Fogassi L, Gallese V, Rizzolatti G. Hearing sounds, understanding actions: action representation in mirror neurons. *Science* 2002;297:846-8.

Leinonen L, Nyman G. II. Functional properties of cells in anterolateral part of area 7 associative face area of awake monkeys. *Exp Brain Res* 1979;34:321-33.

Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. *Psychol Rev* 1967;74:431-61.

Liberman AM, Mattingly IG. The motor theory of speech perception revised. *Cognition* 1985;21:1-36.

Liberman AM, Whalen DH. On the relation of speech to language. *Trends Cogn Sci* 2000;4:187-96.

Murata A, Fadiga L, Fogassi L, Gallese V, Raos V, Rizzolatti G. Object representation in the ventral premotor cortex (area F5) of the monkey. *J Neurophysiol* 1997;78:2226-30.

Perrett DI, Harries MH, Bevan R, Thomas S, Benson PJ, Mistlin AJ, Chitty AJ, Hietanen JK, Ortega JE. Frameworks of analysis for the neural representation of animate objects and actions. *J Exp Biol* 1989;146:87-113.

Porro CA, Francescato MP, Cettolo V, Diamond ME, Baraldi P, Zuiani C, Bazzocchi M, di Prampero PE. Primary motor and sensory cortex activation during motor performance and motor imagery: a functional magnetic resonance imaging study. *J Neurosci* 1996;16:7688-98.

Rizzolatti G, Camarda R, Fogassi L, Gentilucci M, Luppino G, Matelli M. Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. *Exp Brain Res* 1988;71:491-507.

Rizzolatti G, Fadiga L. Grasping objects and grasping action meanings: the dual role of monkey rostroventral premotor cortex (area F5). *Novartis Found Symp* 1998;218:81-95; discussion 95-103.

Rizzolatti G, Fogassi L, Gallese V. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat Rev Neurosci* 2001;2:661-70.

Rizzolatti G, Luppino G. The cortical motor system. *Neuron* 2001;31:889-901.

Umilta MA, Kohler E, Gallese V, Fogassi L, Fadiga L, Keysers C, Rizzolatti G. I know what you are doing. a neurophysiological study. *Neuron* 2001;31:155-65.

Wohlschlager A, Bekkering H. Is human imitation based on a mirror-neurone system? Some behavioural evidence. *Exp Brain Res* 2002;143:335-41.

Woodward AL. Infants selectively encode the goal object of an actor's reach. *Cognition* 1998;69:1-34.

4. Executive Summary

In the second year of the project, our main objective was to investigate: i) how visual and motor information can be used to learn to discriminate grasping actions; ii) the role of visual feedback in the ontogenesis of mirror neurons in monkeys; iii) the temporal sequence of the emergence of manipulative skills in human infants.

Relative to point (i), these are the main activities:

- i) We proposed a methodology for gesture recognition based on motor representations. The approach takes object affordances into account and experiments were carried out using the data-glove setup.
- ii) Approaches for acquiring models of both objects and the robot's hand were developed. The object model is based on a sequence of foveations at distinctive object points where the object appearance is stored. The hand modeling/segmentation is based on the exploitation of visuo-motor associations while the system generates repetitive hand movements.
- iii) A methodology that allows the robot to learn how to adapt to gravity was also proposed and developed.
- iv) Initial grasp strategies were implemented using the Mirror robotic hand, equipped with touch sensors. The strategies were proposed in the form of grasping reflexes.

Relative to point (ii), the following main activities can be emphasized:

- i) Electrophysiological recording of single neurons in monkey ventral premotor (area F5) and primary motor (area F1) cortices.
- ii) Psychophysical assessment of critical factors for biological data acquisition system.
- iii) Role of the mirror system in interindividual communication.

Finally, in point (iii), the following experiments and studies:

- i) The ability of children to adjust the orientation of objects with various shapes in order to fit them into holes.
- ii) Study the development of infants' predictive reaching to moving objects.
- iii) Study of the development of predictive visual tracking.

Cooperation among the partners is well established and led to a conspicuous exchange of information and know-how also outside the specific goals of the project. Effort and funding are being used as planned apart from minor changes.

According to the initial plans, there is a small delay regarding both the implementation on the robotic artifact and some of the biological experiments. For this reason an unpaid 6-month extension period was requested, which will allow us to meet the proposed objectives in the final year.

5. Second year activities

> *The goals of MIRROR are: 1) to realize an artificial system that learns to communicate with humans by means of body gestures and 2) to study the mechanisms used by the brain to learn and represent gestures. The biological base is the existence in primates' premotor cortex of a motor resonant system, called mirror neurons, activated both during execution of goal directed actions and during observation of similar actions performed by others. This unified representation may subserve the learning of goal directed actions during development and the recognition of motor acts, when visually perceived. In MIRROR we investigate this ontogenetic pathway in two ways: 1) by realizing a system that learns to move AND to understand movements on the basis of the visually perceived motion and the associated motor commands and 2) by correlated electrophysiological experiments.* `(From MIRROR's Technical Annex)`

The second year activity of Mirror has been formally reported in the deliverables listed in the following table:

DELIVERABLES TABLE

Project Number: IST-2000-28159

Project Acronym: MIRROR

Title: Mirror Neurons Based Object Recognition

Del. No.	Title	Leader	Type	Classifi-cation	Due (new date)
1.6	Management Report 3	DIST	Report	Public	18
1.7	Progress Report 2	DIST	Report	Public	24
1.8	Management Report 4	DIST	Report	Public	24
2.5	Architecture of the learning artifact	DIST	Report	Public	18
2.6	Robot testing and technology assessment	DIST	demo	Public	24(30)
3.4	Modeling of mirror neurons representation	DIST	demo	Public	18
4.5	Final results of the biological experiments	UNIFE	Report	Public	24(30)

Due to delays both in the biological experiments and on the implementation of the developed methodologies in the artificial setup, a 6-month unpaid extension was formally requested. As a consequence of this delay and extension, deliverables D2.6 and D4.5 are now due at month 30 (shown in red in the table).

5.1. Workpackage 1 – Management and Coordination

During the second year of the project, three meetings were held. All meetings were organized as a two day events. The first (full) day was devoted to technical presentations where the various partners present progress on their scientific work in the context of Mirror. The second day was generally devoted to the evaluation of the progress in the different work-packages and deciding further directions of work and research.

The **fourth project meeting** took place in Genoa on February 10-11th 2003 and it was devoted to aspects of the implementation on the robotic artifact. As such it involved mainly the technological partners in Mirror (DIST and IST) as well as UNIFE. One of the main focuses of this meeting was to define a set of experiments of data acquisition to be conducted with the data-glove setup. The data set collected following what planned at the meeting eventually was used in testing the first implementation of the probabilistic model already described.

The **fifth meeting** took place in Uppsala on the 9th and 10th May, 2003. All partners attended the meeting. The main objective consisted in reporting the progress attained in each partner site, with a special emphasis on the research conducted at the Department of Psychology at UU.

The **sixth meeting** took place in Lisbon, on September 27-28th, 2003, with the participation of all partners. In addition to presentations describing new results obtained by each group, there was a discussion as to the organization of the second year review and the formal reports to be finalized. Future directions for Mirror were discussed. Among these some planning for integration and the realization of the final demo artifact was carried out.

In addition to these formal meetings, cooperation was also achieved by direct contact between the different partners allowing identifying data to share and experiments to conduct.

Globally, the research activity is proceeding as planned. However, in order to achieve the final results of Mirror, the consortium asked for a 6-month extension of the project that will allow for a better integration in the robotic setup, as well as for a consolidation of the biological experiments.

5.1.1. Activity at DIST – University of Genoa

Besides the management activity the work at DIST during Y2 consisted mainly in:

1) Realization of the robotic hand. The five finger hand is a mechatronic device with 15 joints and 6 controlled degrees of freedom. Joints were coupled through elastic elements in order to offer intrinsic safety and compliance. Compliance can be used when grasping to get an approximate estimation of the shape and size of the object. The work on the hand involved the debugging of the mechanics, preparation of the interfacing electronics, harnessing, and software layer. We have shown now examples of grasping and simple control synergies.

2) Realization of multiprocessor control architecture. Drawing on our past experience we re-implemented a good part of the software control architecture of the robotic setup in Genoa. The supporting software allows now a much better and seamless experimentation with different control strategies, better modularization, and parallelism. The hardware structure consists of a set of Pentium based machines connected using standard 100Mbit Ethernet. The new architecture support a range of features that allows also experimenting with learning at a larger scale than what was allowed with our old system.

3) Hand localization and improved arm control. To support the range and precision of the new essential behaviors supporting object manipulation, we started improving the robot arm and hand control. A procedure to locate the hand in the images by employing a combination of vision and motor/proprioceptive information was developed. The control of the arm now benefits from the estimation (learned) of the gravitational load on the joints and can consequently position much better. Improved control is essential during reaching transport phase and for preshaping of the hand. Also some effort was devoted to image processing with the goal of exploring what visual information might be retained while exploring object and into a procedure for segmenting and tracking the hand. The latter was implemented in the data-glove setup.

4) Recording and experimentation. All recording and data collection with the data-glove setup were also part of the activity in Genoa.

References:

P. Fitzpatrick and G. Metta. *Grounding vision through experimental manipulation.* In the Philosophical Transactions of the Royal Society: Mathematical, Physical, and Engineering Sciences, 361:1811, pp. 2165-2185.

G. Metta and P. Fitzpatrick. *Early Integration of Vision and Manipulation.* Adaptive Behavior special issue on Epigenetic Robotics. Vol 11 Issue 2 (2003) pp. 109-128.

L. Natale, S. Rao, G. Sandini. *Learning to act on objects. 2nd Workshop on Biologically Motivated Computer Vision (BMCV).* Tübingen (Germany), November 22-24, 2002

G.Metta, L.Natale, S.Rao, G.Sandini. Development of the "mirror system": a computational model. In Conference on Brain Development and Cognition in Human Infants. Emergence of Social Communication: Hands, Eyes, Ears, Mouths. Acquafredda di Maratea - Napoli. June 7-12, 2002.

L. Natale, G. Metta, and G. Sandini. Development of Auditory-evoked Reflexes: Visuo-acoustic Cues Integration in a Binocular Head. Robotics and Autonomous Systems, vol. 39/2 pp. 87-106, 2002.

Paul Fitzpatrick, Giorgio Metta, Lorenzo Natale, Sajit Rao, Giulio Sandini. What am I doing? Initial steps toward artificial cognition. (Submitted to IEEE Conference on Robotics and Automation)

5.1.2. Activity at UNIFE – University of Ferrara

During the second year of the MIRROR project, UNIFE activity was mainly focused onto three topics: (i) Electrophysiological recording of single neurons in monkey ventral premotor (area F5) and primary motor (area F1) cortices; (ii) Psychophysical assessment of critical factors for biological data acquisition system and (iii) Role of the mirror system in inter-individual communication. A more detailed description of these activities is given below:

1) Monkey experiments: The goal of these experiments is to establish if F5 premotor neurons are sensitive to the vision of monkey's own acting hand. This because one possible explanation of the origin of mirror neurons is that they might have developed from the visual feedback system that visually controls grasping execution. After electrophysiological mapping (recording and microstimulation) of the frontal cortex in order to delimitate the regions of interest (areas F1 and F5), we recorded more than 100 neurons in both areas and we submitted to formal testing more than 80% of them. There were four experimental

conditions: grasping in light, grasping in dark, grasping in dark with a brief visual feedback of the grasping hand touching the object, grasping in dark with a brief visual feedback of the grasping hand before touching the object. Data are currently under analysis and we are collecting additional data on another hemisphere. We are also planning to introduce visual feedback perturbation (e.g. by prisms).

2) What visual information is necessary to build the visuomotor map? UNIFE is actively involved in setting up of the biological data acquisition setup. This implementation requires the optimization of several factors that might strongly influence the performance of the action-recognition system. Some of the questions posed by robotics partners are: is a stereoscopic vision really necessary to create the visuomotor map, how much does finger occlusion during grasping influence action recognition, is the embodiment characterizing the mirror system based on movements recognition or on actions recognition, etc. To answer these questions we set up a paradigm in which subjects are looking at the experimenter grasping objects in different ways. Subjects are requested to indicate the exact instant of object touching by tapping with their index finger. Both subjects and demonstrator fingers are provided with resistive touch sensor which permits to acquire the instant at which the object is touched and the instant at which the subject taps. These measures allow calculating the subject's error in touch evaluation. Observation of different grasping movements (with different degrees of finger occlusion) is performed binocularly and monocularly to assess the relevance of stereoscopic vision in evaluating others' action outcome. Preliminary results are encouraging and show that to accomplish the task subjects use an internal model of the observed action.

3) Mirror system and communication: As we have already described in the previous year report, still in the framework of the scientific problem of action recognition on which the MIRROR project is based upon, we are investigating in humans how the mirror system could be involved in communication. On the basis of preliminary observations (Fadiga et al, European Journal of Neuroscience, 2002;15, 399-402), we are now investigating the role of premotor cortex in speech perception by applying transcranial magnetic stimulation to inferior frontal cortex during phonological tasks. In addition we have set up an fMRI experiment aiming to clarify the role of Broca's region in action understanding and interindividual communication.

References:

Craighero L. Bello A., Fadiga L., Rizzolatti G., *Hand action preparation influences the responses to hand pictures*, Neuropsychologia, Amsterdam, 2002, 40: 492-502.

Gallese V., Fadiga L., Fogassi L., Rizzolatti G. *Action representation and the inferior parietal lobule. In: Common Mechanisms in Perception and Action* - Attention and Performance - Volume XIX. Eds. Prinz W. e Hommel B. (New York: Oxford University Press) 2002.

Rizzolatti G., Fadiga L., Fogassi L., Gallese V. *From mirror neurons to imitation: facts and speculations.* In: The Imitative Mind Development, Evolution and Brain Bases. Eds. Meltzoff A.N., Prinz W. (Cambridge: CUP (Cambridge studies in cognitive perceptual development), 2002

Rizzolatti G., Fadiga L. *The mirror-neuron system and action recognition.* In Higher-order motor disorders: from Neuroanatomy and Neurobiology to Clinical Neurology. Eds. Freund H.J., Jeannerod M., Hallett M. (New York: Oxford University Press), (2004, in press).

Fadiga L., Craighero L. *New insights on sensorimotor integration: From hand action to speech perception*, Brain and Cognition (2003, in press).

Fadiga L., Craighero L. *Electrophysiology of action representation.* Journal of Clinical Neurophysiology, (2004, in press)

Rizzolatti G., Craighero L. *The mirror-neuron system.* Annual Reviews of Neuroscience, (2004, in press)

5.1.3. Activity at ISR – Instituto Superior Técnico

In addition to the regular activities of the project (meetings communication, etc) during the second year of MIRROR, IST has worked primarily on *WP2 – Artifact Realization* and in WP3 – *Biological Setup*.

The work developed in WP3 consisted in developing a methodology for gesture recognition exploiting motor information as well as visual data. Inspired after findings and possible models for Mirror neurons, one key the approach consists in a Visuo-Motor Map (VMM) that establishes an association between the appearance of images of a hand and the corresponding motor information. Training is performed based on the data-glove experiments and data set. The classification strategy encompasses also the affordances of each manipulated object in the sense of coding the likelihood of different grasp types for different objects. It is assumed that affordances are learnt *a priori* even if the model allows for updating the affordance representation.

Based on this mapping a classifier is proposed that allows recognizing grasp gestures irrespective of the view point based on the motor information. It is further shown that by using visual data alone the same level of performance is not attained, thus demonstrating the advantages of relying on motor information for gesture recognition or (eventually) imitation. This work is also related to WP2 in the aspect of providing a possible model for the mirror neurons representation to be implemented into the robotic setup.

Future work will focus on extending the methodology and using more varied experimental data and conditions as well as implementing this approach in the real artifact.

This work is extensively described in deliverables D3.4. In addition to technical reports, this research was published in:

References:

Raquel Vassallo, José Santos-Victor, Hans-Jorg Schnebeli, “Using Motor Representations for Topological Mapping and Navigation,” Intl. Conference on Intelligent Robots and Systems, IROS 2002, Lausanne, Switzerland, October 2002,

Manuel Cabido Lopes, José Santos-Victor, “Visual Transformations in Gesture Imitation: what you see is what you do,” ICRA - IEEE International Conference on Robotics and Automation, Taiwan, September 2003.

Manuel Cabido Lopes, José Santos-Victor, “Motor Representations for Hand Gesture Recognition and Imitation,” IROS Workshop on Robot Programming by Demonstration, Las Vegas, SA, October 31st, 2003.

5.1.4. Activity at DP – University of Uppsala

During the second year of the project, UU has worked on two kinds of problems related to the development of manual control. In addition, UU has also started to investigate action control when visual information is temporarily absent due to occlusion of the external object.

1. Children’s ability to adjust the orientation of objects with various shapes in order to fit them into holes is studied. This project was begun during Y1. The experiments utilize the natural interest of young children in fitting objects into holes. By varying the form

of the objects and the holes, the difficulty of the task can be manipulated. Pre-adjustments of the orientation of the various objects before trying to push them through the holes, give information about the subjects spatial cognition as well as their ability to plan these actions. Some experiments have been completed an others are planned. Right now we are completing a study that evaluates these abilities in children from 1 to 3 years of age.

2. During Y2, an experimental paradigm has been established to the development of infants' predictive reaching for moving objects. Two orthogonal servomotors drive an object on a 1 x 1 m planar surface. The motors are placed behind the surface and transmit the motion to the object magnetically. Software for this device has been developed during Y2 and now we are able to construct any almost arbitrary motion with any velocity profile. The device is going to be used to explore predictive reaching and the development of extrapolation rules in infant catching. Right now we are running an experiment with this device where we are asking 7-8-month-old infants to catch objects moving on an ellipsoid trajectory with either constant angular velocity (sinusoidal modulation) or constant tangential velocity (constant speed). Both kinds of motion are found in nature. Animals decelerate when they go around a bend and the pendulum motions that underlie our body movements also adhere to these principles. On the other hand, objects that move passively continue with the same speed if they are forced to turn. The reason why adults perceive object motion in this way could be because they have much experience with biological motion or it could be an inherent constraint on the perception of motion.
3. During Y2, UU has proceeded with the work on predictive visual tracking. Infants' ability to smoothly track objects of different size, track them along different trajectories, and over occlusion has been studied. The focus during this year has been on the emergence of predictive tracking of temporarily occluded objects. We have pursued this effort with two kinds of eye movement recordings: EOG and cornea reflection.

References:

Rosander, R. and von Hofsten, C. (2003) *Infants' emerging ability to represent object motion.* Cognition, (in press).

Gredebäck, G. and von Hofsten, C. (2003) *Infants' evolving representation of moving objects between 6 and 12 months of age.* Infancy, (in press).

von Hofsten, C. (2003) *The development of prospective control in looking.* To be published in J. Lockman and J. Rieser (Eds.) Action as an Organizer of Perception and Cognition during Learning and Development. Minnesota symposium on Child Psychology.

von Hofsten, C. Johansson, K. (2003) *Planning to reach for a rotating rod: Developmental aspects.* Unpublished manuscript.

5.2. Workpackage 2 – Artifact

5.2.1. Deliverable 2.5 – Architecture of the learning artifact

Deliverable 2.5 describes mostly the architecture of the robot in part overlapping with Deliverable 3.4. In this deliverable though we report on the final hardware and software architecture where the mirror neurons model will be running.

In particular, the description of the robotic hand testing and expected mechanical behavior is included together with a general description of the whole robotic setup. During Y2 a newer version of the software control architecture has been produced and fully implemented on the existing setup in Genoa. This is described to some extent in its main features; among them the ability to dynamically start, stop and connect different sub-parts of the system at run-time.

Finally, a part of the lowest level set of behaviors already implemented on the robot is briefly commented since they represent the layer where the learning architecture plugs in. As mentioned in Deliverable 3.4, hand localization, arm control and some visual routines are described.

5.2.2. Deliverable 2.6 – Robot testing and technology assessment

This deliverable has been postponed to month 30.

5.3. Workpackage 3 – Biological Setups development and test

This Workpackage is devoted to the definition, realization and test of the experimental setups to be used to investigate the biological bases of the project. For the purpose of the project it will be necessary to acquire information about the trajectory and posture of a human arm as well a synchronized sequence of images of the arm performing the action. This information will be used to test the correlation between motor and visual data in the discrimination of different grasping actions. Therefore it is important that both the visual as well as the kinematic data is, as much as possible, analogue/similar to what perceived by the person executing the grasping.

Which visual information is necessary to build the visuomotor map?

UNIFE is actively involved in setting up of the biological data acquisition setup. This implementation requires the optimization of several factors that might strongly influence the performance of the action-recognition system. Some of the questions posed by robotics partners are: if a stereoscopic vision is really necessary to create the visuomotor map, how much finger occlusion during grasping influences action recognition, does the embodiment characterizing the mirror system is based on recognition of movements or actions, etc.

To answer to some of these questions we are now investigating the capability to predict the outcome of observed actions while experimentally manipulating the various viewing conditions. To this purpose we are mainly focused on a particularly relevant phase of the grasping actions: the instant at which the pulpar surface of the fingers touches the to-be-grasped object. It is plausible that in order to generate smooth and continuous actions, the brain should predict this instant and that grip force cannot be correctly provided only on the basis of proprioceptive/tactile feed-back. Subjects are therefore instructed to detect (by tapping once with their index finger on the table) the precise instant at which a demonstrator’s hand grasps a target object. The capability to predict this action outcome is indicated by the difference in time between the actual object touch instant and the occurrence of subject's response. Temporal data are acquired from tactile sensors attached to pulpar surfaces of both the demonstrator's index finger and thumb and subjects' index finger. During the touch detection task, participants are presented with: (1) different types of grasping movements directed to different objects; (2) different degrees of finger visual occlusion (e.g. a precision grip on a small cube seen laterally or frontally); (3) touch detection with binocular and monocular (both dominant and non-dominant eye) vision; (4) touch detection during observation of actions performed by a mechanical hand. Preliminary results are quite encouraging. Peculiarly, it seems that subjects' performance is not significantly influenced by monocular-binocular vision and that when the performing hand is a mechanical device, an anticipation of the touch instant is measured.

5.3.1. Deliverables 3.4 – Modeling of the mirror neurons representation

This deliverable describes the work done during the second year of Mirror that directly addresses some of the fundamental scientific issues set at the beginning of the project. A formal probabilistic model of the emergence of mirror neurons has been formulated and the first implementation produced. This deliverable also includes results from such implementation. The main problems addressed were:

> How can "visual-only" information of a motor act be used to index the self-centered visuo-motor representation, coding the action (this indexing ability is the core "response" of a mirror neuron)? We propose a methodology that allows an artificial system to perform gesture recognition relying on motor representations.
>
> How can the system learn how to represent (and segment) the objects of interest in the scene and its own hand in a cluttered environment? We started experimenting with a method whereby the representation of objects is acquired through a sequence of fixations at interesting points of the object. The problem of representing/segmenting the robot's hand is addressed by exploiting periodic movements and by correlating visual information to motor regularities.
>
> How can the robot-arm controller learn to adapt to gravity and how to develop grasp strategies? The problem of learning how to compensate the gravity force acting upon the arm is addressed through a learning/adaptation mechanism. Experiments on grasping using the anthropomorphic hand developed within Mirror are presented.

5.4. Workpackage 4 – Experiments

Besides the robotic experiments described earlier, additional experimental setups and related pilot/preliminary experiments were realized with monkeys and young children.

5.4.1. Deliverable 4.5 – Final results of the biological experiments

The deliverable has been postponed to month 30 following the extension of the project. The status of the activity at month 24 is sketched below.

Monkey experiments

After electrophysiological mapping (recording and microstimulation) of the frontal cortex in order to delimitate the regions of interest (areas F1 and F5), we recorded more than 100 neurons in both areas and we submitted to formal testing more than 80% of them. Briefly, the goal of our exploration was to establish if F5 premotor neurons are sensitive to the vision of one's own acting hand. This because one possible explanation of the origin of mirror neurons is that they might have developed from the visual feedback system that controls own execution of grasping. Thus, first a visuomotor association is established between a given motor command and the visual signals arising during its execution; then the system might have been capable to generalize the association to others' hand, guided by the invariance of the motor schemata (see Annex 1 of the project). The experimental paradigm was the following: during each trial, the monkey grasps a small handle to open a plastic door covering some reward (pieces of apple). The handle (translucent Plexiglas) is dimly illuminated by a LED positioned behind it. After single neuron isolation, the grasping of the handle is repeatedly executed under four different conditions: in full vision; in darkness (the handle is however visible because of the LED), in darkness but with a brief flash of light when the fingers are approaching the handle, in darkness but with a brief flash of light when the fingers touch the handle. In order to be sure that differences in neural discharge are not due to different movement patterns in light and in dark, we recorded the reaching-grasping kinematics through a catadioptric stereo system realized in our laboratory. In addition, F1 neurons served as an additional control, the prediction being that their discharge should reflect also very small changes in kinematics, whereas they should be very little influenced by visual inputs. The analysis of preliminary results shows that while about 30% of F5 motor neurons are sensitive to the different experimental conditions, only a negligible percentage of F1 motor neurons is influenced by the task. We are currently replicating these observations on a second hemisphere and we are also aiming to test F5 neurons by modulating the behavioral relevance of the visual feedback relative to the grasping hand. We are therefore planning to manipulate monkey's visual perception by applying prisms with different degrees of refractivity in order to challenge the recorded neurons in condition where the visual control becomes necessary to accomplish the task.

Mirror system and communication

As we have already described in the previous year report, still in the framework of the scientific problem of action recognition on which the MIRROR project is based upon, UNIFE is investigating in humans how the mirror system could be involved in communication. By using transcranial magnetic stimulation (TMS) we made some preliminary observations showing that a motor resonance, similar to that observed in monkey mirror neurons, can be evoked not only by action viewing but also when a subject is passively listening verbal stimuli acoustically presented (Fadiga et al, European Journal of Neuroscience, 2002;15, 399-402). It is obvious that, in this case, the "mirror" effect involves at the cortical level not hand but tongue motor representation.

The experiments we are currently carrying out are two-folded: On one side we are investigating whether the motor resonance induced by speech listening represents a mere epiphenomenon or if it reflects an involvement of motor centers in speech perception (as suggested by the Liberman's motor theory of speech perception). To this purpose we are applying repetitive TMS on several sites of premotor cortex and Broca's region, to interfere with subjects' performance during phonologically and/or semantically related perceptual tasks. On the other side, we are investigating the role of area 44 during communication. We are currently designing an fMRI experiment aiming to investigate whether cortical speech centers, and particularly the frontal ones, are specifically tuned for "verbal" communication or provide a neural substrate in which sequences of movements, individually meaningless, are translated into meaningful representations. We are therefore investigating the effect of gestural, non-symbolic, non-verbal communication on inferior frontal gyrus and particularly on area 44, which is considered the human homologue of the monkey premotor area (F5) where mirror neurons have been located.

Behavioral Development

The understanding of relationships between objects and holes was examined in 14- to 26-month-old toddlers. The task was to insert short rods with various cross-sections (circular, square, rectangular, and triangular) into apertures into which they fitted snugly. Task difficulty was varied from a circular rod to a triangular one with a cross section of unequal sides. The cylinder fits into the aperture as long as its long axis is perpendicular to it, while the triangular rod, in addition, has to be turned in a specific way. Results show that toddlers can fit the cylinder into the circular aperture by 14 months of age while the most difficult triangular rod was only mastered by 26 months of age. The youngest children took more time to transporting the objects to the aperture lids, and tended to make more adjustments and changes. 14-month-old children moved the object from one hand to the other, transported it to the mouth, and inspected it closely before transporting it to the lid. Such transactions were less common in the 26-month-olds. The actions of the 14-month-olds were more explorative than those of the 26-month-olds. The success rate of the younger infants was also more influenced by the mode of presentation. If the rod was lying down, they often failed to raise it up before trying to insert it into the aperture.

6. Deviations from planned activities

As indicated previously, the consortium has requested a 6-month unpaid extension of the project. Even though we cannot explicitly refer to a delay in the different workpackages of the project, this extra time, will allow for a better integration of the developed methodologies as well as a better consolidation of the analysis of the biological experiments. Also, it should be mentioned that some of the experiments currently conducted, exceed what was originally planned in Mirror's technical annex.

7. Plans for next period

The overall goal of the next period will be to integrate the developed methods in the robotic artifact and eventually to compare experiments conducted with the artificial system to results obtained in experiments with monkeys or children.

With specific reference to the scientific workpackages of the project, the planned activity is briefly described.

7.1. WP2 – Robot

In the final year of the project, we will further refine the control of grasping and transfer to the robotic system the methodologies developed for hand/object segmentation and representation as well as the proposed approach for (motor based) gesture recognition and learning.

7.2. WP3 – Biological Setup and Test

The data acquired with the biological setup was extensively used during the second year of the project (refer to D3.4). In the final year we expect to proceed with the effort of planning data acquisition sessions that will allow the (off-line) testing of the segmentation and learning processes, as well as a comparative analysis for benchmarking, before transferring some methods to the robotic artifact.

7.3. WP4 – Experiments

The “behavioral development” experiments will continue during this period with essentially the same agenda as during Y2. We will continue to study how infants learn to fit objects into holes and how they develop their ability to catch objects moving on complex trajectories with complex velocity profiles. We will also continue with the experiments exploring infants developing ability to track moving objects that are temporarily occluded.

As to the **monkey experiments** the final year of the project will be devoted to acquire data, to validate data from the first monkey on other animals and, possibly, to explore manipulative neurons in the parietal cortex. We will continue the investigation the role of visual feedback in the ontogenesis of the mirror system.

Also the results obtained in these experiments will be transferred to the robot setup where it will further be used to validate the implementation. UNIFE will continue the investigation on the possible relationships between motor resonance and speech perception with transcranial magnetic stimulation.

8. Effort in person hours in the period 1.9.2002 – 31.8.2003

Contractors	Person hours	
	HOURS Y1	HOURS Y2
Coordinator DIST	1887	3340
Contractor UNIFE	2258	3173
Contractor UU	2880.5	3993
Contractor IST	2023.41	2045
TOTAL	9,048.91	12,551.00

9. Cost breakdown for the Reporting period

Contractors	Costs											
	Costs	Personnel	Durable equipment	Subcontrac-ting	Travel and subsistence	Consumables	Computing	Protection of *knowledge*	Other specific costs	Administrative and financial coordination costs	Overheads	TOTAL
Coordinator DIST	D[5]	128,605.42	7,687.50	61,019.91	28,450.75	18,203.95	0.00	0.00	0.00	14,541.10	91,222.68	349,731.31
	A[6]	45,721.63	3,277.18	54,150.70	11,059.78	10,326.25	0.00	0.00	0.00	2,967.50	33,874.28	161,377.32
Contractor UNIFE	D	54,310.00	9,352.84	5,289.00	7,902.48	37,179.75	0.00	0.00	8,839.00	0.00	23,516.81	146,389.88
	A	22,580.00	4,018.76	5,289.00	3,684.00	36,808.00	0.00	0.00	0.00	0.00	13,418.15	85,797.91
Contractor UU	D	92,296.96	1,287.56	248.16	9,195.32	3,317.56	455.64	0.00	716.27	0.00	21,503.49	129,020.96
	A	43,430.41	99.51	0.00	5,602.77	716.27	0.00	0.00	0.00	0.00	9,969.79	59,818.75
Contractor IST	D	96,041.45	0.00	3,280.00	16,550.22	4,372.15	0.00	0.00	0.00	0.00	156,448.00	276,691.82
	A	43,253.95	0.00	2,400.00	4,326.81	151.82	0.00	0.00	0.00	0.00	74,648.00	124,780.58
TOTAL		154,985.99	7,395.45	61,839.70	24,673.36	48,002.34	0.00	0.00	0.00	2,967.50	131,910.22	431,774.56

1 - To be filled in by the *coordinator* /administrative and financial *coordinator* (in case of split between administrative and financial coordination and scientific coordination) starting from the second period.

2 - Insert the *project commencement date* .

3 - Insert the end date of the last period covered by the integrated cost statement.

4 - The administrative and financial *coordinator* , in case of split between administrative and financial coordination and scientific coordination.

5 - Costs declared and subject to acceptance of the Commission for the current and previous periods.

6 - Costs accepted by the Commission for previous period(s).

10. Tentative Agenda of Review Meeting

Venue: Avenue de Beaulieu 24, room 0/61, Brussels

Date: Monday December 1, 2003

Time: 13:30-15:30.

Attendees:

DIST: Giulio Sandini, Giorgio Metta

Universty of Uppsala: Claes von Hofsten

University of Ferrara: Luciano Fadiga

IST-Lisbon: Josè Santos-Victor

TENTATIVE AGENDA

Introduction

13:30	Overview of Mirror, status of the project, and overview of the results	Giorgio Metta or Giulio Sandini

Highlights of the second year results

	Architecture of Learning Artifact	José Santos-Victor
	Biological experiments	Luciano Fadiga
	Development	Claes von Hofsten

Discussion

14:30	Questions/Answers
15:30	End of meeting