A neural network model for coordination of hand gesture during reach to grasp

J. Molina Vilaplana *, J. Lopez Coronado

Department of Systems Engineering and Automatics, Polytechnic University of Cartagena, Campus Muralla del Mar., C/Dr Fleming S/N. 30202, Cartagena, Murcia, Spain

Received 26 September 2003; accepted 11 July 2005

Abstract

In this paper a neural network model for spatio-temporal coordination of hand gesture during prehension is proposed. The model includes a simplified control strategy for whole hand shaping during grasping tasks, that provides a realistic coordination among fingers. This strategy uses the increasing evidence that supports the view of a synergistic control of whole fingers during prehension. In this control scheme, only two parameters are needed to define the evolution of hand shape during the task performance. The proposal involves the design and development of a Library of Hand Gestures consisting of motor primitives for finger pre-shaping of an anthropomorphic dextrous hand. Through computer simulations, we show how neural dynamics of the model leads to simulated grasping movements with human-like kinematic features. The model can provide clear-cut predictions for experimental evaluation at both the behavioural and neural levels as well as a neural control system for a dextrous robotic hand.

Keywords: Motor control; Reach to grasp; Motor and premotor cortex; Neural networks; Hand posture; Principal component analysis; Computer simulation; Dextrous robotic hands

1. Introduction

Struggling to develop robot controllers for systems working in unknown environments, roboticists address many of the same problems faced by researchers in motor behaviour: how can a computer system coordinate multiple degree of freedom limbs as well as the brain does; how is sensory information integrated with movement; how are objects perceived; and how does planning occur. In that sense, understanding human prehension has become a focus of interest to many roboticists aiming to develop control systems able to accurately perform skilled prehensile tasks with dextrous prosthetic or robotic hands. Trying to mimic human dexterity and flexibility by means of artificial robotic hands is a question that has inspired many researchers in the last 30 years. In the large spectrum of problems related to hand use, the definition of a grasping posture associated to an object has been one of the most challenging since it implies the satisfaction of a large number of constraints related not only to the hand’s and the object’s structure but also to the requirement of the task and the state of the environment. Several points make this subject complex. The first is related to the large amount of data to be considered. The second difficulty faced is that there are many ways to grasp the same object and the choice of a particular grasp is subject to object, environment and task constraints. Therefore, any mapping resolving this problem is many-to-many.

Within this work, we have assumed that control of dextrous multifingered hands comprises several partial tasks that can be classified according to hand or grasp preshaping operations and grasp synthesis. Grasp preshaping is a task oriented operation in which due to the large number of degrees of freedom of multifingered anthropomorphic hands, it is practically impossible to generate contact location from a simple description of the object to be grasped, the hand and the task. So, the hand or grasp preshaping task is aimed at reducing the number of degrees of freedom between hand and object to be grasped, and is used as pre-planner for grasp synthesis. Grasp synthesis comprises the determination of a grasp under the condition of stability, equilibrium, force closure and dexterity. Grasp synthesis is commonly approached either by optimisation-based processes or by control composition (Gorce & Fontaine,
Different human prehension classifications (Iberall, 1997; Napier, 1956) have been used in order to design solutions to the grasp shape planning problem. Engineers, mathematicians and computer and cognitive scientists have designed various forms of computational architectures that carry out the mapping between input information and the observed output prehensile behaviour. These systems have been designed using artificial intelligence programming techniques and expert systems (Cutkosky, 1989; Cutkosky & Howe, 1990; Iberall, 1987; Iberall, Jackson, Labbe, & Zamprano, 1988; Kang & Ikeuchi, 1997). Recently, another kind of approach has emerged. This new approach is based on the utilization of neural networks to define grasping configurations or to learn the mapping from an object shape to a hand configuration or a grasp choice (Kuperstein, 1991; Taha, Brown, & Wright, 1997; Uno, Fukumura, Suzuki, & Kawato, 1995). The previous studies emphasize the correspondence between an object and a hand shape. One can argue that the same grasping posture can be adopted to grasp objects of various shapes, and one important factor that affects grasping of an object may be to recognize its graspable parts (i.e. the ones that have suitable dimensions to allow a grasp) instead of considering its general shape. Following this point of view, these graspable parts of the objects were called graspable features by Moussa and Kamel (1998) or grasp affordances by Fagg and Arbib (1998). If a neural model can acquire a representation of the graspable features or affordances, it can handle more situations and thus allows an enhanced flexibility for grasp planning. Moreover, one can expect to integrate task requirements by the appropriate choice of graspable features (if we want to drink from a glass, instead of grasping it from above, its object surface has been touched. From a glass, instead of grasping it from above, its side, which is another grasp feature). Moussa and Kamel (1998) followed such an approach and proposed a computational architecture able to learn grasping rules called 'generic grasping functions'. Such grasping knowledge could be used for gross planning, while a more local grasping strategy (related with a final grasp synthesis) would be used once the object surface has been touched.

This paper focuses mainly on hand or grasp pre-shaping during prehension. Taking into account recent kinesiological data (Mason, Gomez, & Ebner, 2001; Santello, Flanders, & Soechting, 2002) characterizing the temporal kinematic coordination patterns of finger motion during the movement of grasp pre-shaping, this paper presents a new neural network model that implements a conceptual model of human prehension established by Jeannerod (1981, 1984). Computational modelling and simulations are used to study anticipatory finger pre-shaping during reach to grasp movements. The proposal involves the design and development of a Library of Hand Gestures consisting of motor primitives for hand and finger pre-shaping during the earlier phases of prehension. This model can provide clear-cut predictions for experimental evaluation at both the behavioural and neural levels as well as a neural control system for a dextrous robotic hand.

Several specific aims are treated in this paper:

(a) **Analysis and synthesis of hand gestures.** Prehension exhibits both parallel and serial combinations of tightly scheduled movement components. Among the movement components are a free movement phase involving transport and pre-shaping of the hand. This specific aim will focus in identifying the number of effective degrees of freedom involved in hand shaping during prehension. This task involves extracting the appropriate kinematic parameters for analysis and synthesis of hand pre-shaping motor programs. The principal components analysis of hand postures by Santello, Flanders, and Soechting (1998); Santello and Soechting (1997) is used as a tool to develop a suitable parametrization for these evolving motor programs.

(b) **Development of a library of hand gestures.** It is proposed that the movement pre-shaping of hand fingers can be understood in terms of a Library of Hand Gestures, based on motor primitives for various types of object oriented prehension. This specific aim focuses on the dynamic neural representations of these motor primitives based on the kinematic analysis of Mason et al. (2001); Santello et al. (2002).

(c) **Development of a cortico-subcortical neural network model of prehension.** This specific aim integrates aims (a) and (b) to develop a model of prehension based on the Library of Hand Gestures and neural networks for reaching and grasping. The hand gestural patterns emerging from this model, include the variations of temporal patterns of individual hand gestures (e.g. temporal asymmetries of wrist velocities and anticipatory intervals prior to grasping), the modulation of gestural hand components with task demands (e.g. for precision gripping of a pencil vs. a coin), and the relative timing and phasing among gestures representing a global temporal patterning of grasping actions.

The paper is organized as follows. The experimental and theoretical background of the model is presented in Section 2. Section 3 describes the computational neural model proposed in this paper. Computer simulation results carried out in order to test the model performance in different situations are presented in Section 4. In Section 5, the biological plausibility of the model, simulation results and possible implementations of the neurocontroller into real robotic or prosthetic dextrous hands are discussed.

## 2. Background

### 2.1. Jeannerod’s conceptual model of human prehension

Jeannerod (1981, 1984) has reported that prehension is characterised by two components, namely, transport and manipulation. The transport component is related to
the movement of the wrist towards the location of the object to be grasped; the manipulation component is related to the opening and closing (or pre-shaping) of the grip aperture. In particular, the movement of the hand allows for changes in the grasp orientation to match that afforded by the object. Duringprehension, typically the movement of the fingers follows a stereotyped kinematic pattern: from a closed initial position, in which the thumb and index fingers are in slight contact, the thumb and finger separate (grip opening) until maximum grip aperture is reached. As the wrist approaches the object location, the thumb and index fingers begin to close until the object is grasped. The hand pre-shaping suggests a stereotypical spatial relation between hand aperture and hand transport (Haggard & Wing, 1995).

Temporal invariances also exist in prehension in terms of a constant relative timing between some parameters of the transport and manipulation components. For example, the time to maximum grip aperture occurs between 60 and 70% of movement duration despite large variations in movement amplitude, movement speed, and different initial postures of the fingers (Jeannerod, 1984; Saling, Mescheriakov, Molokanova, Stelmach, & Berger, 1995; Wallace, Weeks, & Kelso, 1990). The time to maximum grip aperture is also well correlated with the time to maximum deceleration of the transport component (Gentilucci, Castiello, Corradini, Scarpa, Umilta and Rizzolatti, 1991).

Jeannerod also noted an important distinction between intrinsic object properties (such as size and shape) and extrinsic properties (or egocentric spatial properties such as distance and orientation). In Jeannerod’s works, it is suggested that the two types of properties are likely to be detected through different anatomical structures or channels. Specifically, he suggested that for grasping an object, separate visuomotor channels are activated in parallel by specific visual input and each channel controls a specific part of limb musculature. Extrinsic spatial properties of an object activate proximal muscles (e.g. shoulder, elbow joints) for the transport component, and intrinsic properties activate distal movements (e.g. fingers) for the grasping component.

The conceptual model of human prehension that emerges from Jeannerod’s findings specifies that the transport and hand pre-shaping components in prehension evolve independently through two segregated visuomotor channels, coordinated by a central timing mechanism. This timing mechanism ensures the temporal alignment of key moments in the evolution of the two components. In this way, Jeannerod suggests that the central timing mechanism operates such that peak hand aperture is reached at the moment of peak deceleration of the reaching component.

2.2. Hand synergies during prehension

An emerging viewpoint is that the Central Nervous System (CNS) uses synergies to simplify the control of the hand. Recently, Santello et al. (1998); Santello and Soechting (1997) have shown that static grasp posture can be described using a small number of postural synergies. These synergies could be defined as a spatial configuration or ‘primitive’ of the hand shape that is common across the various tasks. Using a principal component analysis, Santello noted that the first two components were needed to describe approximately 85% of the variance. The presence of postural synergies that contributed to the evolving hand shaping was examined in Mason et al. (2001) and Santello et al. (2002). Results in both studies agree that rather than two or more discrete grasps, hand posture may be composed of a continuum based on the temporal weighting of a few eigenpostures or synergies. In this case, eigenpostures would not be limited to or be expected to correspond to particular grasps. The temporal weightings of the eigenpostures demonstrated that hand shape evolves continuously throughout the reach-to-grasp and is unique for each object/grasp combination. Both studies concluded that temporal weightings more precisely define the shaping of the fingers for grasp than just the linear scaling of the maximum hand aperture to the object size as previously noted (Chieffti & Gentilucci, 1993; Jeannerod, 1984; Marteniuk, Leavitt, MacKenzie, & Athenes, 1990; Paulignan, MacKenzie, Materniuk, & Jeannerod, 1990). These results confirm and expand on the finding that hand shape evolved gradually throughout the movement (Santello & Soechting, 1998) and that the synergies may represent a simplifying scheme used by the CNS to control the hand.

3. Computational model

The model presented in this paper is depicted in Fig. 1. The model constitutes what Arbib (1985); Hoff and Arbib (1993) have called a coordinated control program for prehension. The proposed system captures the conceptual scheme of Jeannerod. Visual information about object’s location, size, shape and orientation is provided to the distributed motor structures implemented in the model. On the left half of Fig. 1 is depicted the computational model that implements the transport component of the movement. Visual information about the object’s location provides data to the reaching neural controller allowing it to allocate the wrist of a two joint arm (shoulder, elbow) moving along the transverse plane, in a location suitable for a correct hand prehension. The movement of the wrist/arm towards its final location has been labelled as a ‘ballistic movement’. Motor structures depicted in the right half of Fig. 1 enable suitable finger pre-shaping as well as correct palm orientation. It is important to note that we have subdivided the visuomotor channel related with grasping into two subchannels, one related with finger configuration and another related with palm orientation. The palm orientation schema uses an extrinsic property of the object such as orientation. In the model, we have included another kind of information that must be fed to the grasping channel. This information is motivational or task demands related. It appears clearly that, whatever the extrinsic orientation of an object may be, the motor program related with grasp must take into account the task we are trying to accomplish. For instance, the palm orientation and final hand posture to grasp a thin vertical cylinder with the intention of transporting it, is not going to be
the same if our intention is to make a precision grasp of the same cylinder on its upper surface aiming to make a precision insertion of that cylinder in a hole. Due to this fact, in the proposed model, motivational information is the main source in determining the palm orientation related motor program, although extrinsic orientation of the object can be used to determine the palm orientation motor program in some situations. This kind of information is also crucial in determining the hand shape in order to correctly grasp the object according to task demands.

How visual information about the object’s intrinsic properties is used by the VITE model (Bullock & Grossberg, 1988) responsible for finger pre-shaping is discussed in Section 3.1. Simultaneous activation of motor schemas through a common volitional gating signal (GO signal) initiates the ballistic movement of wrist towards the target and finger pre-shaping of the hand. The fingers are pre-shaped in advance to the size and shape of the object and according with the type of task intended, while the palm is oriented to the appropriate configuration in which the correct grasp must be achieved.

3.1. Finger pre-shaping and palm orientation

3.1.1. Model for a synergistic control of an anthropomorphic hand

Movements of the hand involve controlling a very large number of kinematic degrees of freedom and this leads to complex models that are numerically intractable. However, it is possible to approximate these models by using principal component analysis since the possible motions of the individual degrees of freedom of each of the fingers may be correlated. One of the goals of this study was to develop an analytic method that can be used to characterize succinctly the motion of a large number of degrees of freedom and to understand how properties of an object to be grasped (such as its shape, size, and its use) are mapped into hand posture. An experimental approach to this problem has been initiated by Santello and Soechting (1998) (see also Santello & Soechting, 1997). In order to model the whole finger pre-shaping during prehension and according to Mason et al. (2001) and Santello et al. (2002), we have designed a model for the synergic control of hand shape. The finding by Santello et al. (1998) that a gradual modulation of hand posture is detected along two main axes in Principal Components or eigenposture space, points to the possible existence of two main synergies through which hand shape is modulated according to different objects’ features. Based on the two first postural hand synergies found by Santello et al. (1998) in his studies about static grasps, we have designed a model in which by means of only two parameters \( w_1, w_2 \) that we have called temporal weightings of synergies, we are able to continuously generate hand postures.

The model can be described by Eq. (1)

\[
\theta = S^T w + S^2
\]  

(1)
using anthropometric data from Butchholz, Armstrong, and Goldstein (1992) (see Appendix A); \( w \) is a two-component vector containing the temporal weightings of synergies \((w_1, w_2)\), \( S^1 \) is a \( 15 \times 2 \) matrix containing a numerical model of the two real synergies found by Santello as column vectors (see Fig. 2); and \( S^2 \) is a bias vector. Allowing \( w_1 \) and \( w_2 \) to have values between \(-10\) and \(10\) generates a continuum of hand postures located within a bidimensional parametric space. According to Santello et al. (1998), the first synergy is more related with the hand aperture. At the lower extreme of \( w_1 \) \((w_1 \text{ min})\), the fingers are extended at the MCP joint and abducted, and at thumb, CMC extends and abduction increases. At the other extreme \((w_1 \text{ max})\), fingers are flexed at the MCP joint and adducted, and at the thumb, CMC abduction decreases and CMC flexion increases. The excursion at the PIP joints remains approximately constant. These angular changes can be visualized as a gradual closure of the hand. The second synergy is more related with a global measure of finger curvature. Along \( w_2 \), the changes in angular excursion are of a smaller amplitude: moving toward \( w_2 \text{ max} \), PIP and MCP joints flex. As \( w_2 \) decreases, PIP and MCP joints slightly extend.

3.1.2. Neural dynamics of finger pre-shaping

Vector integration to endpoint (VITE) (Bullock & Grossberg, 1988) dynamics is used to model the finger pre-shaping neural channel involved in prehension. In this channel, the hand posture is modelled as a 2 degree of freedom (DOF) system related with the temporal weightings of synergies \((w_1, w_2)\). VITE gradually integrates the difference (DV vector) between the desired target hand posture defined by TPV = \([w_1^d, w_2^d]\) and the actual finger configuration described by PPV = \([w_1^a, w_2^a]\) (Fig. 3). The rate of integration (i.e. the movement velocity) is controlled by the product of the DV vector and a volitional movement gating signal \((GO)\) (Bullock & Grossberg, 1988). VITE dynamics is described by

\[
\frac{d(DV)}{dt} = \varepsilon(-DV + TPV - PPV) \tag{2}
\]

\[
\frac{d(PPV)}{dt} = GO(t) \cdot DV \tag{3}
\]

\[
GO(t) = G0 \frac{t^2}{(0.06 + t^2)} \tag{4}
\]

where \( \varepsilon \) is a rate of integration and \( G0 \) is a scalar factor related with movement speed. This model is able to generate realistic joint trajectories by continuously mapping through Eq. (1),

Fig. 2. Waveform of the two synergies (SYN1, SYN2) described in text. The figure shows numerically the ‘autopostures’ conforming the matrix \( S^1 \) of Eq. (1). What is represented is the angular change in degrees for each joint due to a positive unitary variation in \( w_1 \) (SYN1) and \( w_2 \) (SYN2). Hand joints are organized as: from 1 to 4 represent MCP joints of Index, Middle, Ring and Little Finger (flexion–extension). 5–8: The same for PIP joints. 9–11 abduction between Index and Middle, abduction between Middle and Ring and abduction between Ring and Little Finger, respectively. 12–15, Thumb CMC abduction, Thumb CMC flexion–extension, Thumb MCP and Thumb IP flexion–extension, respectively. In the model DIP finger joints are stated as DIP = 2/3 PIP, from Index to Little Finger.

Fig. 3. VITE model for hand shaping formation during prehension. The finger pre-shaping motor command is coded as a two-dimensional vector formed by two temporal weightings of hand synergies. VITE gradually integrates the difference (DV vector) between the desired target hand posture defined in TPV and the actual finger configuration described by PPV. The output of VITE model at PPV stage is mapped into real hand DOFs by means of Eq. (1). Lines finishing in an arrow mean excitatory connections. Lines finishing in a circle mean inhibitory connections.
the actual hand configuration coded as $[w_1, w_2]$ in PPV, into the real hand configuration denoted by $\theta$ as seen in Fig. 3.

Our model uses a biphasic programming of the grasp component. The notion of two phased movements was introduced in the grasping literature by Jeannerod (1984). This author noted that the first phase of the movement, from initial movement to peak deceleration of the wrist, lasts for approximately 70% of the total time of the movement time. During this fast high velocity phase, the hand is opening as the fingers are extending, thereby posturing appropriately for the grasp. The peak aperture of the grip is reached at about the same time as peak deceleration. During the slow second phase, Jeannerod noted many corrective type movements in transport channel as the hand enclosed around the object. Jeannerod made an important observation, often forgotten or ignored, that these corrective type movements occurred during the deceleration phase even when only the target and not the hand was visible. It was concluded that the slow phase was not due to visual feedback processing but was a centrally generated part of the prehension pattern. According to this, the motor program for the movement of fingers has been modelled as consisting of two sequential subprograms $I = \{G_1, G_2\}$, where $G_1 = \{w_1^m, w_2^m\}$ defines a hand posture related to the maximum grip aperture and $G_2 = \{w_1^s, w_2^s\}$ defines the final static grasp posture related with object’s intrinsic properties and task demands. The biphasic patterns of temporal weightings we have designed as a dimensional reduced representation of the prehension gestures, are characterized by a $G_1$ in which $w_1^m$ has a low value, favouring an initial phase of gradual increase of hand aperture (fingers are extended and abducted at the MCP joints near the $w_1$ min) until the second phase of finger movement driven by $G_2$ takes place. In $G_2$, through the higher value of $w_1^s$ we try to match the final finger span with the width between the two opposite object’s surfaces selected to achieve the grasp. Through $w_2^s$ we adjust the final finger curvature that matches the shape of the contact surface (Santello & Soechting, 1997).

The switch between $G_1$ and $G_2$ occurs at the time of peak deceleration (tpdec) within the transport channel. This is the time at which the derivative of the unimodal velocity trace reaches its global minimum. Detection of a minimum in the acceleration of the wrist triggers the reset of $G_1$ subprogram and the read-in of $G_2$ by TPV in grasp channel (dashed line from Ballistic Movement to Finger Pre-Shaping in Fig. 1). In principle, proprioceptive information could be used by Central Nervous System to derive tpdec or a related measurement (Cordo, Schieppati, Bevan, Carlton, & Carlton, 1993). In this paper, we have used a simple algorithm to achieve this local minima detection in which we only have to determine the first minimum of wrist’s acceleration to trigger the read in of $G_2$ and reset of $G_1$. No other interactions between transport and grasp components were assumed in the model (Contreras-Vidal, Ulloa-Pérez, & López-Coronado, 2001; Molina Vilaplana, Feliu, & López Coronado, 2002).

In the model, it is also hypothesised that the motor program $I = \{G_1, G_2\}$ related with intrinsic properties of the object such as object size and shape may be computed in advance of the onset. The VITE model in the grasp channel has also been modified to account for the apparent gradual specification of hand shape from visual processing centers to the motor centers (Contreras-Vidal et al., 2001; Molina et al., 2002). The model assumes that the target aperture is not fully programmed in TPV before movement initiation; rather, it is postulated that TPV in grasp channel, sequentially and gradually specify, in a first phase the desired hand shape related with maximum grip aperture ($G_1$), and in a second phase, the hand configuration corresponding to the object size and shape ($G_2$). The proposed modification in TPV dynamics grasp channel is described by Eq. (5)

$$\frac{d(TPV)}{dt} = \mu(\text{TPV} + G_i)$$

(5)

where $\mu$ is a rate of integration.

Within this work, a Library of Hand Gestures consists in a set of biphasic motor programs $I_u (I_u = [G_1^u, G_2^u])$ such as, each motor subprogram $G h^u$ ($h = 1,2$) has a neural representation based on the hand gestural primitives described by the temporal weightings of hand synergies $[w_1, w_2]$. In the Library of Hand Gestures, each one of these motor programs is task/object oriented (e.g. $u$ can represent the motor program for the precision grip of a small rod and $u'$ can represent the motor program for the power grasp of a big cube).

3.1.3. Neural dynamics of palm orientation

As the wrist approaches its target, palm orientation must be reach a configuration that allows the actual finger configuration to achieve the correct grasp of the object. VITE dynamics is used to model the palm orientation neural channel involved in prehension. In this channel, the palm orientation is modelled as a 3 degree of freedom (DOF) system related with the flexion–extension of the wrist ($\alpha$ angle), pronation and supination of the forearm ($\beta$ angle) and ulnar and radial deviation of the wrist DOFs ($\gamma$ angle). The motor program related with palm orientation is monophasic and can be expressed as $TPV = [\alpha, \beta, \gamma]$. Neither interactions with other channels nor modifications in the classic VITE dynamics are included in palm orientation channel.

3.2. Transport component implementation

We have carried out simulations in which a two joint arm (elbow and shoulder) transports, in the transverse plane, the wrist of an anthropomorphic hand towards the location of the object to be grasped. To ensure that the model architecture would also be scalable, to enable future control of an arm with redundant degrees of freedom (e.g. a 4-joint arm that moves the wrist in 3D), a self-organizing neural network model for eye-hand coordination (DIRECT model, Bullock, Grossberg, & Guenther, 1993) has been trained in order to solve the classical inverse kinematics problem. During a motor babbling phase, the model endogenously generates movement commands that activate the correlated visual, spatial, and motor information that are used to learn its internal coordinate transformations. After learning occurs, the model is capable of controlling reaching movements of the arm to prescribed spatial targets. Training of the model has been carried out using the arm’s wrist as the end-effector system for the transport component. Degrees
of freedom related with wrist orientation and finger pre-shaping are not involved in DIRECT training. The model achieves its competence by transforming visual information about wrist’s target position and actual 3D position in space into a body centered spatial representation of the direction in 3D space that the wrist must move to locate the hand near the object to be grasped. The spatial direction vector (or difference vector Fig. 4(b)) is adaptively transformed into a motor direction vector through the direction-to-rotation transform that DIRECT neural network implements, which represents the synergistic joint rotations of shoulder and elbow that move the wrist in the desired spatial direction from the present arm configuration.

The visuomotor map (i.e. the direction-to-rotation transform) is learned through the learning rule

$$w_{ij}(t + 1) = w_{ij}(t) + \eta \left( \Delta \phi_j - \sum_{m=1}^{2} \Delta s_m w_{im} \right) \Delta s_j$$  \hspace{1cm} (6)

where \(i=[1,2]\) number of joints, in this case shoulder and elbow, \(j,m=[1,2]\) dimensionality of the workspace of arm, which moves only in the transverse plane) \(\Delta \phi_i\) are the joint angular increments; \(\Delta s_j\) are spatial increments Cartesian coordinates for each axis; \(w_{ij}\) represents the adaptive weights that implements the transformation for each joint; and \(\eta \ll 1.0\) is the learning rate.

The major obstacle in learning inverse kinematics lies in the problem that the inverse kinematics of a redundant kinematic chain has infinite solutions. Thus, the learning algorithm has to acquire a particular and a valid inverse kinematic solution. This issue was characterized by Jordan (1990) and Jordan and Rumelhart (1992) as the problem of non-convex mappings. These authors showed that in order to learn the inverse kinematics, it is necessary that all joints velocities generated during training form a convex set. Unfortunately, as shown in these works, inverse kinematics has the non-convexity property and therefore does not permit direct learning of the inverse mapping. Nevertheless, as noted by Bullock et al. (1993), it is possible to transform the non-convex problem of inverse kinematics learning into a convex problem by spatially localizing the learning task within the vicinity of each robot configuration reached during learning. With this assumption inverse kinematics problem is actually convex. Thus, inverse kinematics of a redundant system can theoretically be accomplished properly by learning an inverse mapping if a spatially localized learning algorithm is employed.

In that sense, the directional mapping learned by the DIRECT model is locally linear, even for redundant systems. This means that if one only considers a small region of joint space, the set of joint velocity vectors that produce a desired spatial velocity is convex. The linear network described below utilizes different parameters in different regions of the joint space of the manipulator (each joint configuration of the arm is related with a particular set of \(w_{ij}\) adaptive weights. These weights can be associated with a set of direction mapping neurons in the direction-to-rotation neural network, Fig. 4(b)) and smoothly interpolates between these parameter sets. In this paper, the implementation of the DIRECT model as an spatially localized and linear algorithm has been carried out through what Fiala (1995) has called ‘context fields’. A context field is a set of tonically active inhibitory neurons.
which receive broad based inputs that determine the context of a motor action. A context field neuron pauses its activity when it recognizes a particular joint configuration in its inputs. The target cells of the context field neurons are groups of direction mapping neurons, which are completely shut off when their associated context field neuron is active (Fig. 5).

As shown in Fig. 6, each context field cell projects to a set of direction mapping cells, one for each joint vector component. Each joint vector component has a set of direction mapping cells associated with it, one for each context cell. The context field neuron selects which direction mapping cells will be active in any time. In the current implementation of the model, the context field cells are assumed to tessellate the joint configuration space in a regular pattern. For a given joint configuration, the context cell which represents that region is selected and the corresponding subset of direction mapping cells is activated through the variable $c_k$.

The joint direction cell activities $\Delta \phi_i$ are driven by $V_{ik}$ during performance and by random joint rotation velocities $r_i$ from an endogenous random generator (ERG in Fig. 6) during learning.

$$\frac{dV_{ik}}{dt} = v \left( -V_{ik} + c_k \left( \sum_j w_{ijk} \Delta s_j - \Delta \phi_i \right) \right)$$

$$\frac{d(\Delta \phi_i)}{dt} = \delta \left( (e - 1) \left( \sum_k V_{ik} - \Delta \phi_i \right) + e(r_i - \Delta \phi_i) \right)$$

When $e = 1$, motor babbling is active and $\Delta \phi_i$ are driven to sensed joint velocities $r_i$. During performance, $e = 0$, and input is the sum of all $V_{ik}$, only one of which will be actively processing input. Actual joint configuration is described by VITE-like dynamic equation

$$\frac{d\phi_i}{dt} = GO(t) \Delta \phi_i$$

During motor babbling, learning is obtained by decreasing weights in proportion to the product of the presynaptic and postsynaptic activities

$$\frac{dw_{ijk}}{dt} = -\eta V_{ik} \Delta \phi_j$$

Fig. 5. Context field for inhibiting all but a subset of direction mapping neurons. Inactive neurons are shown as white disks. The context field shown in white is ‘off’ when the joint angles of arm have a certain value. The ‘off’ context cell enables a subset of direction mapping neurons (left). Adapted from Fiala (1995).

Fig. 6. Complete circuit for learning eye-hand coordination in reaching. The spatial error is computed at the top to get a spatial direction vector ($\Delta s$). $\Delta s$ is transformed by the direction-mapping network elements $V_{ik}$ to the corresponding joint direction vector ($\Delta \phi$). The context field selects the $V_{ik}$ elements based on actual joint configuration. Motor babbling is due to an endogenous random generator (ERG) circuit. Training is done by generating random movements, and using the resulting joint increments and observed spatial increments of the hand as training vectors to the direction mapping network. In the direction mapping network, $V_{ik}$ ($i = 1, 2; k = 1 \cdots 500$) receive shunting inhibition from the $k$th context field neuron. Each $V_{ik}$ receives additive inhibitory feedback from $\Delta \phi_i$ cell to which it projects.
Therefore, the learning rule can be restated in discrete form as

\[ w_{ij}(t + 1) = w_{ij}(t) + \eta \left( r_i - \sum_m w_{imk}(t) \Delta s_m \right) \Delta s_j \]  

(11)

where \( i = 1, 2; j = 1, 2; k = 1 \cdots 500. \)

We have trained this implementation of DIRECT model through a series of 10,000 babbling cycles, with a two link planar arm model whose forward kinematics in Cartesian coordinates is described by

\[ \begin{align*}
    x_1 &= l_1 \cos(\phi_1) + l_2 \cos(\phi_1 + \phi_2), \\
    x_2 &= l_1 \sin(\phi_1) + l_2 \sin(\phi_1 + \phi_2)
\end{align*} \]  

(12)

where \( \phi_1 \) and \( \phi_2 \) are the actual joint values (ranging from 0 to 135\(^\circ\)) and \( l_1, l_2 \) are the link lengths (\( l_1 = 12 = 280 \) mm).

The arm movement model of Fig. 6 reproduces key aspects of reach kinematics, such as straightline motions and unimodal, bell shaped velocity profiles. As indicated in Fig. 6, the spatial direction vector \( \Delta s \) is computed by taking the difference between the spatial target position and the spatial hand position which can be obtained from visual input or in our simulations by Eq. (12). The spatial target position is the only input information to the neural network during movement performance.

In the DIRECT model implementation described above, learning generalizes to all spatial directions at each sampled joint configuration; this is because the model learns a directional mapping that is an approximation to the Jacobian pseudoinverse at each joint configuration, and the approximate Jacobian pseudoinverse learned for one movement direction can be used for all other movement directions. It is also important to note that systems that use directional mappings as the system presented in this paper, can successfully reach targets even if the directional mapping contains a large amount of error. Therefore, any residual error that might exist; e.g. from assuming linearity over too large a region of the workspace, will not prevent the system from reaching targets, but will instead only lead to curvature in the movement trajectories.

4. Simulation and results

We have carried out systematic computer simulations of the neural model for coordination of hand gestures during prehension. In Fig. 7 are shown three key instants of two different prehension tasks. On the left column are shown the initial instant of the movement (upper left figure), the instant of maximum grip aperture at \( \sim 60\% \) of movement time (middle left figure) and final grasp configuration (lower left figure) for a power grasp of a cylinder. On the right column are shown the initial instant of the movement (upper right figure), the instant of maximum grip aperture at \( \sim 60\% \) of movement time (middle right figure) and final grasp configuration (lower right figure) for a precision grasp (thumb opposed to one finger) of the same cylinder. The simulations conditions are summarized in Table 1. In simulations, \( G_1 \) and \( G_2 \) correspond to the biphasic grasp motor program coded as weightings of synergies. \( \alpha, \beta, \gamma \) correspond to the palm orientation motor program described in Section 3.1.3. In all simulations initial hand shape is described by \( \{0, 7\} \) and initial palm orientation is \( \alpha = 0, \beta = 0, \gamma = 0. \)

In Fig. 8 are shown the kinematic curves for velocity and acceleration of wrist (left) and grip aperture and grip aperture velocity (right) for the Precision Grip task carried out under three different movement velocities. The movement velocity is controlled by varying the \( G_0 \) parameter. In our simulations the lowest movement velocity is related with a value of \( G_0 = 15 \). Medium movement velocity was simulated using \( G_0 = 20 \) (this is the \( G_0 \) value used in simulations shown in Fig. 7) and high velocity movements were simulated using \( G_0 = 25 \).

Transport velocity exhibits a bell shaped but asymmetrical velocity profile typical of point to point arm movements. The plot of hand aperture shows the opening of the hand until it gets to maximum peak aperture, then it shows a decreasing of grip aperture until it reaches object size. Due to the computational structure of the model, time of maximum grip aperture is always correlated with \( t_{dec} \), and the spatio-temporal coordination of transport and grasp channels remains unchanged through variations of speed performance (Fig. 8) or task demands (Fig. 7). Some emergent properties of the neural model match stereotyped kinematic profiles in humans. As shown in Fig. 8, as velocity of transport movement increases, the maximum grip aperture (the maximum measured distance between thumb and index finger tips along the movement) also increases. As a result of the \( G_0 \) signal shared by reaching and grasping components, this well-known kinematic feature in humans is achieved in the model without any explicit information transfer between reaching and grasping components. In Fig. 9 it is explicitly showed the temporal correlation between the kinematic variables of transport channel and finger pre-shaping channel in the task showed by Fig. 7(a).

It is important to note that the grip aperture profiles observed are driven by the evolution of temporal weightings of hand synergies. Fig. 10 shows the temporal evolution in the parametric space of the model described in Section 3.1.1, of four simulated tasks related to the grasping of four different objects (a sphere \( G_1 = \{-7, 5\}, G_2 = \{-3, 7\} \), an egg \( G_1 = \{-5, -3\}, G_2 = \{-1, 0\} \), a block \( G_1 = \{-10, -7\}, G_2 = \{-9, 2\} \) and a cylinder \( G_1 = \{-7, 4\}, G_2 = \{-3, 6\} \)). The VITE model related with finger pre-shaping generates the bidimensional continuous trajectory in the weightings of hand synergies space (Fig. 10) that through the mapping between \( w \) and \( \theta \) described by Eq. (1), automatically conforms the grip aperture realistic patterns obtained in Fig. 8 and the realistic coordination of hand joints trajectories shown in Fig. 11.

Fig. 10 shows how during the first phase of movement in all tasks, \( w_1 \) and \( w_2 \) decrease, favouring through Eq. (1) the coordinated extension/abduction of MCP, and extension of PIP joints, while thumb separates from the palm. This coordinated motion of fingers allows the observed increase in grip aperture. In the second phase (triggered by \( t_{dec} \) detection in transport

\[
\begin{align*}
    x_1 &= l_1 \cos(\phi_1) + l_2 \cos(\phi_1 + \phi_2), \\
    x_2 &= l_1 \sin(\phi_1) + l_2 \sin(\phi_1 + \phi_2)
\end{align*}
\]  

(12)
channel), $w_1$ and $w_2$ increase, favouring the coordinated flexion/adduction of MCP, and flexion of PIP joints, while thumb approaches the palm in order to match the final grasp configuration. The temporal evolution of hand joints during the grasping of the cylinder is shown in Fig. 11. Flexion–extension of MCP and PIP of index, middle, ring and little finger, are represented by MCP and PIP, respectively. Abduction/adduction of all fingers are represented by ABAD. Thumb CMC...
flexion–extension and abduction. Thumb MCP flexion–extension and Thumb IP flexion–extension are represented by THUMB.

In these simulations, the combinations of temporal weightings of hand synergies related to each of the four grasping tasks have been determined heuristically in an off-line process previous to the execution of the whole reach to grasp movement. These combinations can be stored as motor programs related to concrete grasping tasks becoming a part of what we have called a Library of Hand Gestures.

5. Discussion

In recent years, the interface between neuroscience and robotics is carried out by Neurorobotics research. Neurorobotics tries to investigate, model and validate neural dynamics underlying successful performance of complex tasks such as grasping and manipulative actions by humans, and transfer validated neural algorithms into the design of control algorithms acting on bio-inspired robotic platforms.

In this paper we have presented a new computational neural model for spatio-temporal coordination of hand gesture during prehension. The model is based on the conceptual model of human prehension proposed by Jeannerod (1984), in which two segregated visuomotor channels (one for transport component, another related with finger pre-shaping) evolve independently, both coordinated by a timing mechanism. This model presents conceptual differences and similarities with another computational model based on the same postulates, the Hoff and Arbib model (1993). According to Jeannerod, both models propose that within the different channels involved in prehension only exists a temporal interaction, while the spatial

<table>
<thead>
<tr>
<th></th>
<th>Power grasp</th>
<th>Precision grip</th>
</tr>
</thead>
<tbody>
<tr>
<td>(G_1)</td>
<td>([-10, -9])</td>
<td>([-4, -6])</td>
</tr>
<tr>
<td>(G_2)</td>
<td>([6.5, -3])</td>
<td>([2, -3])</td>
</tr>
<tr>
<td>(\alpha, \beta, \gamma)</td>
<td>(\pi/4, \pi/2, 0)</td>
<td>(0, 0, 0)</td>
</tr>
<tr>
<td>Initial wrist position (mm)</td>
<td>(X = 291; Y = 321.5; Z = 0)</td>
<td>(X = 0; Y = 200; Z = 0)</td>
</tr>
<tr>
<td>Final wrist position (mm)</td>
<td>(X = -100; Y = 400; Z = 0)</td>
<td>(X = 0; Y = 400; Z = 0)</td>
</tr>
</tbody>
</table>

Table 1

Simulation conditions for the power and precision grasp

Fig. 8. Kinematic profiles of the precision grip movement of the cylinder carried out under three different speeds. Left figures show velocity (upper figure) and acceleration (lower figure) of the wrist. Right figures show grip aperture (upper figure) and grip aperture velocity (lower figure). As speed of the transport component increase, \(t_{pde}\) instant occurs earlier in time and maximum grip aperture reached is higher.
Fig. 9. Three instants of movement evolution from left to right. Initial hand-arm configuration, instant of maximum grip aperture, final contact with the cylinder (upper figures from left to right). In the middle is shown the wrist velocity during the movement performance in (mm/sec). Lower figure shows the acceleration of arm’s wrist in (mm/s²). In abscissas is represented simulated movement time in milliseconds. Dotted lines indicate the temporal correlation of the three instants depicted in upper figures with the kinematic variables in middle and lower figures. Grasp motor program is \( G_1 = \{ -10, -9 \}, G_2 = \{ 6.5, -3 \} \).

Fig. 10. Pre-shaping trajectory in w-space related to the grasping of four different objects. A sphere (circle), an egg (triangle), a cylinder (diamond) and a block (square).
evolution of both components is carried out by independent controllers. In both models an instant is established as a temporal landmark that signals the end of the initial opening phase of the grip aperture and the beginning of the final enclose phase of the fingers. The main conceptual difference within these two models is related with the methods used in determining this referred temporal landmark instant. In our model this instant is derived from a postulated afferent signal associated with the acceleration of the hand transport. We have assumed that, in principle, CNS could derive this signal from primary proprioceptive signals although it should be noted that, there are many doubts related with the biological plausibility of this signal. In the Hoff–Arbib model, the temporal landmark instant mentioned above, is determined through a postulated temporal movement organization unit that computes this instant before movement execution. Another similarity between these two models lies in the biphasic programming of the manipulation component. In both models, manipulation component is described through a biphasic program which evolution is controlled by the instant in which the subprogram associated with reaching the maximum grip aperture is switched to the execution of the subprogram related with the final enclose phase. Of course, the explicit representation of these manipulation motor subprograms differs substantially in both models. In the Hoff–Arbib model, the motor subprograms related with manipulation are defined as desired scalars related with the appropriate distance between index and thumb fingers (grip aperture). In our model, in order to model the whole pre-shaping of fingers and due to the high number of DOFs involved in this task, we have adopted a dimensionality reduced approach based on recent findings by Mason et al. (2001), Santello et al. (1998, 2002), and Santello and Soechting (1998). The model includes a simplified and synergistic control strategy that provides realistic coordination among finger movements. This strategy is consistent with the increasing evidence that supports the view of a synergistic control of whole fingers during prehension. The low-dimensional control scheme of hand shaping presented in Section 3.1.1 allow us to flexibly plan and execute concrete grasping tasks, using only two parameters to define the evolution of hand shape during the task performance. One advantage of this simplified control scheme is that we can flexibly accommodate it to different dextrous real robotic hands depending on the physical constraints related with the real device. This fact allow us to define concrete posture primitives based on the hand synergies designed in this paper, for concrete dextrous anthropomorphic robotic hands, maintaining the simplicity of the high level control of the hand posture during reach to grasp tasks, when implementing this ideas on real robotic devices. It is also possible to expand the dimensionality of the control by adding a third synergy that could be related with the degree of opposition of the thumb in order to obtain more suitable grasping postures, especially in precision grasps. This would only affect the neural controller for finger pre-shaping
via a new temporal weighting for the third synergy, which probably should effect its major influence during the final enclosing phase.

It should also be noted that, DIRECT model, implemented as a Hebbian associative network (Bullock et al., 1993), as a linear network that uses gradient descent and context fields (Fiala (1995) and network presented in this paper) or as an hyperplane radial basis function (RBF) network with adaptive centers (Guenther & Micci Barreca, 1997; Poggio & Girosi, 1989), allows to plan reaching trajectories in 3D task coordinates (spatial coordinates) and, through the direction to rotation mapping that carries out, to implement the required joint rotations that solve the inverse kinematics problem for redundant arm manipulators moving in 3D space. So, it would be easy to scale up the DIRECT model implementation presented in this paper, to move the wrist of a redundant arm in 3D space (for instance, a 4 DoF arm with three DoF in shoulder and one DoF in elbow).

5.1. Biological plausibility of the model.

Difference vectors appearing in VITE and DIRECT models have properties that are remarkably similar to data properties reported by Georgopoulos and his colleagues. In particular, a large DV may be active without causing an overt movement. Such DV activation, called motor priming, has been reported by Georgopoulos and his colleagues. In particular, a further refining of these cells code external space during prehension. This simplifying control strategy is observed in the mentioned experiments at the output stage. Are these eigenpostures represented at the cortical level? The results of stimulation and lesion studies in the motor cortex are consistent with global control of the hand. Stimulation of one site in primary motor cortex (F1 or area 4) evokes responses in several muscles of the hand (Donoghue, Leibovic, & Sanes, 1992; Sato & Tanji, 1989) or movement around contiguous joints or fingers (Kwan, MacKay, Murphy, & Wong, 1978; Strick & Preston, 1978). Focal inactivations or focal strokes in the hand area of the motor cortex do not disrupt movements of individual fingers; rather, movement of different combinations are effected (Schieber, 1999; Schieber & Poliačkov, 1998).

Santello et al. (2002) has driven in this paper the design and modelling of a synergistic control scheme for hand posture during prehension. This simplifying control strategy is observed in the mentioned experiments at the output stage. Are these eigenpostures represented at the cortical level? The results of stimulation and lesion studies in the motor cortex are consistent with global control of the hand. Stimulation of one site in primary motor cortex (F1 or area 4) evokes responses in several muscles of the hand (Donoghue, Leibovic, & Sanes, 1992; Sato & Tanji, 1989) or movement around contiguous joints or fingers (Kwan, MacKay, Murphy, & Wong, 1978; Strick & Preston, 1978). Focal inactivations or focal strokes in the hand area of the motor cortex do not disrupt movements of individual fingers; rather, movement of different combinations are effected (Schieber, 1999; Schieber & Poliačkov, 1998).

Single-unit recordings in primary motor and premotor cortex also suggest that the hand is controlled as a unit. In monkeys a single neuron in F1 generally discharges in relation to multiple instructed finger movements (Schieber & Hibbard, 1993). Furthermore, the population of cells active with different finger movements overlaps extensively (Schieber & Hibbard, 1993).

Nevertheless, there is another cortical area which presents a more promising candidate for representing, at the cortical level, the simplified control strategy for hand grasping introduced in this paper. The hand field of monkey F1 receives a rich input from a specific sector in the ventral premotor cortex, area F5. Motor neurons in F5 have been shown to discharge selectively during a specific grasp such as precision grip, finger prehension, or whole hand prehension (Murata, Fadiga, Fogassi, Gallese, Raos, and Rizzolatti, 1997; Rizzolatti, Camarda, Fogassi, Gentilucci, Luppino and Matelli, 1988). The firing of these neurons correlated only with the specific grasps and not the individual movements made by the monkeys (Murata, Gallese, Luppino, Kaseda, & Sakata, 2000; Rizzolatti et al., 1988). These findings suggest that the hand is represented as a unit at the premotor cortex and that eigenpostures observed at the output stage may be represented in the discharge of populations of hand-related premotor cortical cells. Quite differently from what goes on in the primary motor cortex, these neurons do not simply code single movements like flexing the fingers. Area F5 neurons discharge only when these movements are performed to achieve a specific goal, namely to grasp an object. According to Jeannerod, Arbib, Rizzolatti, and Sakata (1995), the presence in F5 of these kind of neurons has some important implications. First, the number of variables to be controlled is much less than if the movements were described in terms of motoneurons or muscles. This solution for reducing the high number of degrees of freedom of hand movements comes close to that proposed theoretically with the virtual fingers (Iberall & Fagg, 1996). Second, the retrieval of the appropriate movement is simplified. Both for internally generated actions and for those that are emitted in response to an external stimulus, only a small ensemble of motor variables (such as motor programs $l = [G_1, G_2]$ encoded as temporal weightings of synergies as proposed in this paper) have to be selected or coordinated. In particular the retrieval of a movement in response to a visual object is reduced to the task of matching the intrinsic properties of the object with the appropriate motor program. If each motor program for grasping
is encoded in different F5 neurons (e.g. the motor program $I_5$ for the prehension of a sphere, which requires the opposition of all fingers, is encoded by different neurons than the prehension program $I_c$ of a cylinder, for which a palm opposition grip is used) we can argue that in F5 exists a kind of grasping ‘vocabulary’ or, in the terms expressed in this paper, we could argue that in F5 exists a Library of Hand Gestures. Under this conceptual scheme, we can define the Library of Hand Gestures as a set of predetermined biphasic motor programs encoded as a combination of temporal weightings of hand synergies. These motor programs are defined in such a way that their temporal evolution driven by the neural network model presented in this paper generates object oriented grasping movements. Third, the presence of a ‘vocabulary’ of grasping actions should facilitate greatly the learning associations, including arbitrary associations between stimuli and motor programs (for example, if red light, precision grip, if blue light, power grasp), an associative function that premotor areas such as F5 are believed to mediate (Passingham, 1993; Wise & Murray, 2000).

Area F5 is richly connected with the anterior intraparietal area (AIP) (Luppino, Murata, Govoni, & Matelli, 1999). This area is characterized by the presence of a large number of neurons that are active in association with grasping/manipulation movements (motor dominant neurons), presentation of visual stimuli (visual dominant neurons) and both hand actions and object representations (visuomotor neurons) (Murata, Gallese, Kaseda, & Sakata, 1996; 2000; Sakata, Taira, Murata, & Mine, 1995; Taira, Mine, Georgopoulos, Murata, & Sakata, 1990). On the basis of functional properties of AIP and F5, it has been proposed that the visuomotor transformations necessary for organizing grasping movements are mediated by AIP-F5 circuit (Fogassi, Gallese, Buccino, Craghiero, Fadiga and Rizzolatti, 2001). The activity of F5 neurons represents the step that transforms the object representations coded in AIP into a format suitable to activate area F1 motor neurons plus a series of subcortical centres among which are basal ganglia and cerebellum (Fagg & Arbib, 1998; Gallese, Fadiga, Fogassi, Luppino, & Murata, 1997; Jeannerod et al., 1995; Sakata et al., 1995). Gallese et al. (1997) have proposed that area AIP transforms visual information of a given 3D object into multiple descriptions, thus providing F5 with several ‘grasping possibilities’. Area F5 would then select, on the basis of contextual information, the most suitable type of prehension. This finding points to the fact that the model presented in this paper could be a first stage in the development of a biologically plausible neural model of visuomotor transformations and motor execution of grasping movements.

5.2. Future work

In the model, visuomotor transformations related with arm’s transport near the object to be grasped are driven by a self-organizing neural network (DIRECT model) that computes the inverse kinematics of a two joint arm. In this paper, the visuomotor transformations related with the grasping component of the movement have been oversimplified. Correct combinations of temporal weightings of hand synergies related with correct grasp configurations (i.e. biphasic grasp motor programs coded as a combination of two temporal weightings of hand synergies), were determined previously off-line, and later were fed to the neurocontroller. In order to achieve a complete self-organizing model, able to autonomously learn, store, plan and execute correct prehension motor plans, it is necessary to include learning capabilities in the grasp channel. This could be achieved by transforming VITE channels to Adaptive VITE (AVITE) channels (Gaudiano & Grossberg, 1991). This transformation could enable the system to autonomously find correct grasp configurations in the two-dimensional parametric space of hand synergies, adaptively transforming intrinsic properties of the objects and related task information. These correct grasp configurations can be stored as motor programs conforming the Library of Hand Gestures previously mentioned.

Finally, the high modularity imposed on the model allows it to support progressive learning of prehension tasks even on real robotic platforms. The reaching component of the proposed neurocontroller has already been implemented on a real anthropomorphic arm (Lopez Coronado & Guerrero Gonzalez, 2000). The model presented in this paper proposes a framework in which a modular self-organizing neural network progressively learn (with the mentioned inclusion of learning capabilities in the grasping channel) complex prehension tasks. This progressive adaptive learning includes an initial stage of developing reaching skills through the learning of the inverse kinematics of the arm with DIRECT model. In a second stage, the learning should consist in finding correct hand configurations in response to a concrete object’s intrinsic properties. The final stage would consist in including task demands information in order to select the more suitable grasping motor command related with the final goal of the grasp.

Results presented in this paper and partial implementations of this neurocontroller on a real anthropomorphic arm, allow the authors to be optimistic in using this new approach as an advanced and autonomous controller for learning, planning and executing complex prehension tasks on real dextrous robotic hands. Implementation of this model into a real dextrous robotic hand, can lead to the development of an autonomously generated human-like Library of Hand Gestures and can provide clear-cut predictions for experimental evaluation at both the behavioural and neural levels as well as an advanced neuromorphic control system for anthropomorphic prosthetic or robotic hands.

After this work was submitted, two highly compatible treatments of reach-grasp coordination appeared (Ulloa & Bullock, 2003; Ulloa, Bullock, & Rhodes, 2003). However, these treatments did not include a DIRECT coordinate transformation, and did not address the focal problem of this paper, namely efficient low-dimensional control of five-fingered grasp postures, including both precision and power grips. One interesting issue raised by the model of Ulloa and colleagues is the possibility of replacing biphasic control of grip aperture with velocity-modulated monophasic control.
Monophasic control has some appeal because it would simplify the processing assumed to occur in area F5. Although they achieved good data fits without biphasic control, it remains to be seen whether their proposed simplification (which could be technically easy to incorporate in our model) can produce kinematics in accord with the full range of data on reach-grasp kinematics for precision and power grips due to the fact that Ulloa and colleagues treated only precision grip data.

Acknowledgements

This work was supported in part by the National Science Foundation of Spain (CICYT TIC99-0446-C02-01) by SYNERAGH project funded by BRE-2-CT980797 BRITE EURAM-BASIC Research and by PALOMA project funded by IST 2001-33073.

Appendix A. Appendix A

A.1. Anthropomorphic hand model

As in Rezzoug et al. (2001), the adopted model for the hand is composed of five articulated rigid chains representing the fingers which are connected to a common body representing the palm (Fig. A1). Each finger possesses 4 degrees of freedom. To design the hand ‘morphologic generator’ (i.e. defining all the hand parameters trough the knowledge of hand length and breadth), we have used studies on hand anthropometry carried out by Buchholz et al. (1992). The goal of this work was to provide hand geometrical data in order to design tools adapted to the wide range of hand morphologies. Buchholz et al. (1992) defined means to know detailed representation of the hand, in particular fingers segments length and position for almost arbitrary hand morphology. Their studies were based on the dissection of six hands of cadavers (two women and four men). A method was designed to determine the rotation center of the different articulations from a measure of external hand dimensions. Regression equations were derived to determine phalanx dimensions from hand length information alone. Their structure is as follows

\[ SL_{ij} = B_{ij} \cdot HL \pm \text{error} \]  

(A1)

where \( SL_{ij} \) represents the length of segment \( i \) of finger \( j \) \((j = 1,\ldots,5, i = 1,\ldots,3)\), \( B_{ij} \) the regression coefficients and, finally, \( HL \) the total hand length. The numerical values of the \( B_{ij} \) coefficients are summarized in Table A1. Moreover, the position of the thumb CMC articulation and other fingers MCP articulations relative to the wrist reference frame can be obtained by Eqs. (A2) and (A3) from the knowledge of the hand length (HL) and breadth (HB).

\[ X_{ij} = C_{ij} \cdot HL \pm \text{error} \]  

(A2)

\[ Y_{ij} = C_{ij} \cdot HB \pm \text{error} \]  

(A3)

The \( C_{ij} \) coefficients and the associated errors for the five fingers are given in Table A2.

Fig. A1. Degrees of freedom of the anthropomorphic hand model.
Table A1
Regression equation coefficients used to define the fingers' segments length from the knowledge of hand length (HL) (from Butcholtz et al., 1992)

<table>
<thead>
<tr>
<th>Segment</th>
<th>Thumb</th>
<th>Index</th>
<th>Middle</th>
<th>Ring</th>
<th>Little finger</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proximal phal</td>
<td>0.251±0.004</td>
<td>0.245±0.001</td>
<td>0.266±0.003</td>
<td>0.244±0.003</td>
<td>0.204±0.002</td>
</tr>
<tr>
<td>Medial phal</td>
<td>0.196±0.003</td>
<td>0.143±0.003</td>
<td>0.170±0.003</td>
<td>0.165±0.002</td>
<td>0.117±0.002</td>
</tr>
<tr>
<td>Distal phal</td>
<td>0.158±0.002</td>
<td>0.097±0.002</td>
<td>0.108±0.003</td>
<td>0.107±0.004</td>
<td>0.093±0.003</td>
</tr>
</tbody>
</table>

Table A2
Regression equation coefficients used to define the fingers' root position from the knowledge of hand length (HL) and breadth (HB) (from Butcholtz et al., 1992)

<table>
<thead>
<tr>
<th>Coordinate</th>
<th>Thumb</th>
<th>Index</th>
<th>Middle</th>
<th>Ring</th>
<th>Little finger</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_i$</td>
<td>0.073±0.003</td>
<td>0.447±0.003</td>
<td>0.446±0.004</td>
<td>0.409±0.004</td>
<td>0.368±0.003</td>
</tr>
<tr>
<td>$Y_i$</td>
<td>−0.196±0.010</td>
<td>−0.251±0.007</td>
<td>0.000</td>
<td>0.206±0.005</td>
<td>0.402±0.008</td>
</tr>
</tbody>
</table>

Table A3
Segments length (in mm) of anthropomorphic hand model

<table>
<thead>
<tr>
<th>Phalanx</th>
<th>Index</th>
<th>Middle</th>
<th>Ring</th>
<th>Little finger</th>
<th>Thumb</th>
<th>Phalanx</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proximal</td>
<td>48.2895</td>
<td>52.4286</td>
<td>48.0924</td>
<td>40.2084</td>
<td>49.4721</td>
<td>MetaCarp</td>
</tr>
<tr>
<td>Medial</td>
<td>28.1853</td>
<td>33.5070</td>
<td>32.5215</td>
<td>23.0607</td>
<td>38.6316</td>
<td>Proximal</td>
</tr>
</tbody>
</table>

Table A4
Roots (in mm) of fingers of anthropomorphic hand model related to an origin of coordinates placed on the middle of the wrist

<table>
<thead>
<tr>
<th>Coordinate</th>
<th>Index</th>
<th>Middle</th>
<th>Ring</th>
<th>Little finger</th>
<th>Thumb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>X</td>
<td>88.1037</td>
<td>87.9066</td>
<td>80.6139</td>
<td>72.5328</td>
<td>14.3883</td>
</tr>
<tr>
<td>Y</td>
<td>−22.4896</td>
<td>18.4576</td>
<td>36.0192</td>
<td>−17.5616</td>
<td></td>
</tr>
</tbody>
</table>

Table A5
Degrees of freedom (DOF) of each joint of anthropomorphic hand and their associated range of motion

<table>
<thead>
<tr>
<th>Joint</th>
<th>DOF</th>
<th>Extension–flexion</th>
<th>Abduction–adduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fingers</td>
<td>DIP</td>
<td>1</td>
<td>0–80°</td>
</tr>
<tr>
<td>PIP</td>
<td>1</td>
<td>0–100°</td>
<td>-</td>
</tr>
<tr>
<td>MCP</td>
<td>2</td>
<td>−15–90°</td>
<td>Index 0–15° Middle 0° Ring 0–26° Little 0–35°</td>
</tr>
<tr>
<td>Thumb</td>
<td>MCP</td>
<td>1</td>
<td>−15–80°</td>
</tr>
<tr>
<td>IP</td>
<td>1</td>
<td>−10–55°</td>
<td>-</td>
</tr>
<tr>
<td>CMC</td>
<td>2</td>
<td>0–80°</td>
<td>0–90°</td>
</tr>
</tbody>
</table>

Table A6
Degrees of freedom (DOF) related with the wrist and their associated range of motion

<table>
<thead>
<tr>
<th>DOF</th>
<th>Range of movement</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>−90–45° (flexion–extension)</td>
</tr>
<tr>
<td>β</td>
<td>0–180° (pronation–supination)</td>
</tr>
<tr>
<td>γ</td>
<td>−30–30° (radial–ulnar)</td>
</tr>
</tbody>
</table>

Table A7
Parameters used in simulations

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Meaning</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon$</td>
<td>Integration rate, Eq. (2)</td>
<td>30</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Integration rate, Eq. (5)</td>
<td>20</td>
</tr>
<tr>
<td>$\eta$</td>
<td>Learning rate, Eq. (11)</td>
<td>0.0001</td>
</tr>
<tr>
<td>$\nu$</td>
<td>Integration rate, Eq. (7)</td>
<td>40</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Integration rate, Eq. (8)</td>
<td>80</td>
</tr>
</tbody>
</table>
Setting HB and HL to the standard values of HB = 89.6 mm and HL = 197.1 mm (Butcholtz et al., 1992), we obtain the segments lengths (in mm) and roots (related to an origin of coordinates placed on the middle of the wrist) described by Tables A3 and A4, respectively.

Table A5 describes the DOFs associated to each joint in the anthropomorphic hand model and their associated range of movement (in degrees of rotation).

Finally, Table A6 describes the range of motion associated with wrist movements: flexion–extension of the wrist (α angle), pronation and supination of the forearm (β angle) and ulnar and radial deviation of the wrist DOFs (γ angle).

A.2. Matrix $S^1$ and bias vector $S^2$

Matrix $S^1$ is described by two column vectors named $v_1$ and $v_2$, where $v_1 = [4.25, 4.25, 3.5, 3, 0, 0, 0, 0.75, 0, -1, -2, 1, 2, 0.5]$ and $v_2 = [1, 1, 0.5, 0.5, 1.5, 1.5, 1, 0.5, 0, 0, 0, 0, 0, 2]$; $S^2$ bias vector is described by $S^2 = [37.5, 37.5, 50, 55, 45, 60, 70, 72.5, -7.5, 0, 10, 15, 50, 40, 10]$.

A.3. Simulation parameters

The value of several parameters of equations used in simulations and not mentioned in main text are depicted in table A7.

References


