

Chapter 6

Visuomotor Transformations for Grasp Planning and Execution

The previous chapter was dedicated to the processing of basic visual information aimed at extracting object properties, both spatial and cognitive, relevant for grasping purposes. This chapter deals with the tasks of transforming such properties into suitable hand configurations, and executing an appropriate grasping action on the target object. As a first step, the data extracted so far have to be expressed in a format especially dedicated to transformation into hand shapes. In Sect. 6.1, new computational descriptions are offered of the tasks performed by CIP as a fundamental relay station between the visual cortex and the visuomotor areas downstream. Analytical expressions of the transfer functions realized by surface and axis orientation selective neurons (SOS and AOS) of CIP are derived and discussed. Section 6.2 has a more practical stance, and describes how the obtained representations are used in grasp planning and execution. The different projections to AIP and its job as the fundamental hub in programming and monitoring grasping actions are discussed. Practical solutions are proposed for a working model of its connections with ventral stream, premotor cortex, somatosensory areas, basal ganglia and cerebellum (Fig. 6.1). Robotic grasping experiments based on such connections and exploiting tactile feedback for increased reliability are described.

6.1 Neural Coding in the Caudal Intraparietal Sulcus

The caudal intraparietal area, CIP, constitutes a central node in the spatial analysis processing of the dorsal stream, which endows the subject with the ability of interacting with her/his surrounding peripersonal environment. Neuroscience studies both on monkeys and humans have depicted a rather clear image of the sort of processing performed by CIP (see Sect. 2.3.1). At the computational level, though, this area has been rather neglected compared to its downstream neighbor AIP, more directly related to grasping actions.

As explained in Sect. 2.3.1, two main neuronal populations have been distinguished in CIP: surface orientation selective (SOS) and axis orientation selective

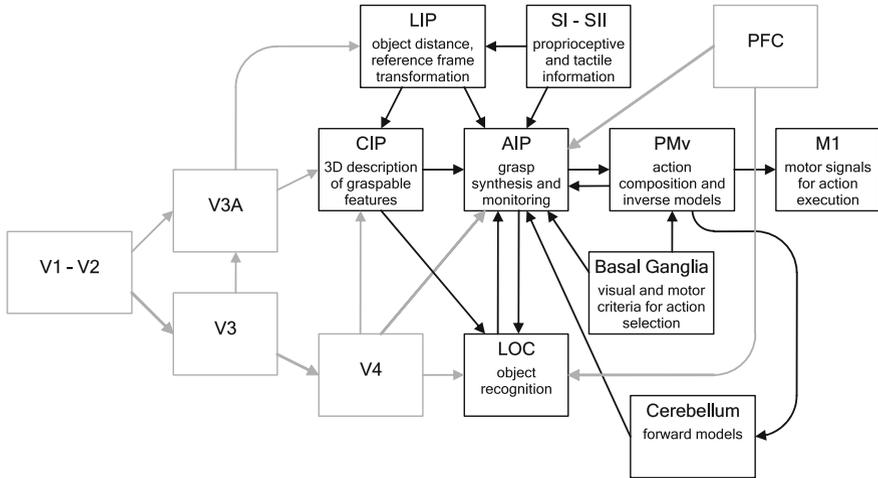


Fig. 6.1 Areas of the model framework involved in the planning and execution of grasping actions. The function of all highlighted areas is discussed in the text, but the implementation is especially focused on the job of AIP and its connections to the other areas

(AOS) neurons. SOS neurons code for the orientation of rather flat objects. Square shapes are preferred, and elongation in either width or length inhibits the neuronal response. The thickness of the object strongly inhibits the response only above a certain threshold. It can be hypothesized that such threshold represents the graspability of the feature, as it appears close to the size of the hand.

AOS neurons represent the 3D orientation of elongated objects, preferring thin and long features. It is not clear from the provided data if the reduced responsiveness to thicker objects is only due to the relative proportion between the object dimensions or also by a comparison with the hand size. The proposed model promotes this last possibility, for consistency with the role of CIP in providing AIP information regarding graspable features. Indeed, at least an approximate absolute object size estimation is available to CIP. Curvature coding, very likely also maintained in CIP, is not modeled at this point.

Overall, a population of mixed CIP neurons, including differently tuned SOS and AOS neurons, is able to provide full information about 3D proportion and orientation of a target shape. This information is forwarded to AIP, where 3D orientation, shape and size are jointly coded, and possible grip configurations generated.

Let us consider the situation in which a simple object (possibly box-like, or cylindrical) lies on a table, slanted about a vertical axis, as the ideal objects in Fig. 6.2. The goal is to generate, using only binocular visual information, possible grips on the object, emulating as much as possible the data flow connecting V3/V3A—CIP—AIP. In particular, the focus is on the tasks performed by the caudal intraparietal area, which can be schematized as in Fig. 6.3. The module on the left of the schema is the integration of proprioception with stereoscopic and perspective visual information

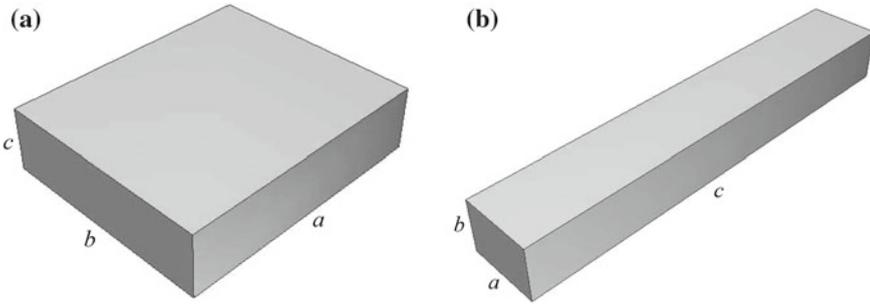


Fig. 6.2 Examples of SOS and AOS dominant objects and size naming convention. **a** Flat (SOS dominant) object. **b** Long (AOS dominant) object

in order to estimate position, orientation and size of simple 3D objects, introduced in the previous chapter.

The following step (right module of Fig. 6.3) requires an action-based point of view, to assess the intermediate level object features with the purpose of evaluating their suitability for grasping. Orientation, relative and absolute size of the major axes of the object are thus compared and the response is synthesized in the SOS and AOS neurons output. The activation of these two kinds of neurons depends on the relation between object dimensions. Considering the three main inertia axes, if two dimensions are similar and the third clearly smaller there is a high SOS activation; if two dimensions are similar and the third bigger AOS activation will prevail. In the case of three different dimensions, SOS and AOS responsiveness is modulated by the actual proportion between sizes. As a convention, from now on the three dimensions are called a , b and c , where a and b are close in size, whereas c is the smaller dimension for SOS activation (Fig. 6.2a) and the bigger dimension for the AOS case (Fig. 6.2b).

6.1.1 Understanding and Interpreting the Available Data

Despite the recent efforts and encouraging advancements (Naganuma et al. 2005), the most important insights regarding the nature of 3D object representation by CIP neurons date back to the second half of the last decade (Shikata et al. 1996; Sakata et al. 1998). The basic concepts were clear from the beginning, such as the distinction between the two classes of orientation responsive neurons, SOS and AOS, and their

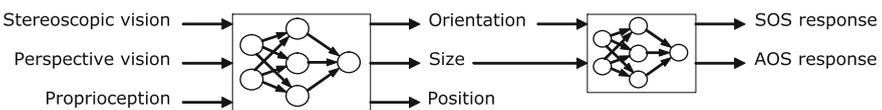


Fig. 6.3 Elaboration of visual data in the posterior intraparietal sulcus CIP

responsiveness trend as a function of an object relative dimensions. The number and variety of different experiments is nevertheless reduced, and their characterization remains at most qualitative. The current goal is to analyze such experiments with modeling purposes, and possibly advance new interpretation hypotheses deriving from a pragmatic point of view. One such hypothesis concerns the quality of absolute size representation in CIP, which will be more thoroughly discussed in Sect. 6.2.1.

Figure 6.4a reproduces the response of an AOS neuron to the view of a slanted elongated object as a function of object width (Sakata et al. 1998). The authors of the original study briefly comment on it suggesting that neuronal response and object width are inversely proportional. A sigmoidal, or logistic response function constitutes an alternative explanation. This solution fits very well with the observed data, as can be observed in Fig. 6.4b, where two differently parameterized sigmoids are superposed to the data of Fig. 6.4a. The sigmoidal is a transfer function very commonly found in brain mechanisms (see e.g. Hu et al. 2005), especially when some threshold effects have to be taken into account. Indeed, in this case there is a very important threshold to consider, that is, the size of the grasping hand. The assumption is that the cut-off value for the sigmoid is the dimension of the open hand or even better the extension of a comfortable grip. For the monkey performing the experiment of Fig. 6.4a this value is reasonably around 12–15 cm. Indeed, CIP neurons seem to be sensitive not only to relative object dimensions (and thus shape) but also to its absolute size (Sakata et al. 1997). Available experimental data are not conclusive to this regard though. In any case, if the size of a potentially graspable object has to be represented in the brain, hand size is a very useful and convenient unit of measure to use. In the next section, this principle is further developed and exploited for defining the analytic expressions which model the function of SOS and AOS neurons.

Overall, CIP is responsive for all the following features of an object: relative size of main axes, absolute size, orientation in 3D, local curvature. Studies reported

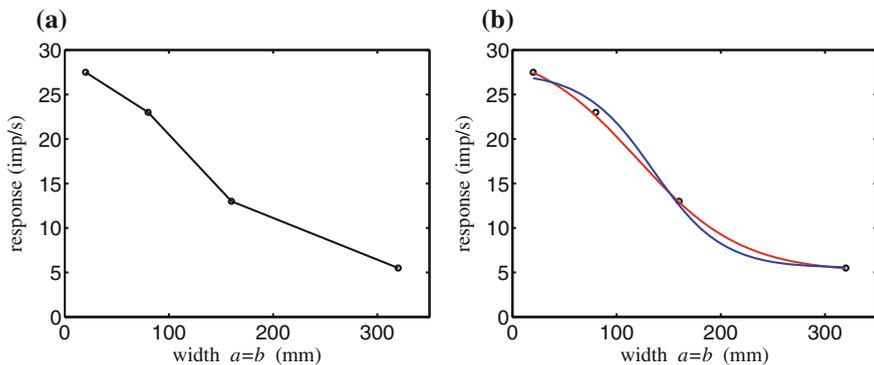


Fig. 6.4 Response of an AOS neuron as a function of object width (length $c = 300$ mm). Experimental data (adapted from Sakata et al. 1998) and interpolation with sigmoidal functions. **a** Experimental data. **b** Sigmoidal interpolation

in the literature describe SOS and AOS neurons that are selective only for width and not for thickness, or only for relative and not for absolute size. Indeed, just a minority of CIP neurons are selective for all the features at the same time, but globally, at a population level, all relevant information regarding object shape in relation to potential grasping actions is processed by the posterior intraparietal area (Sakata et al. 2005). The proposed transfer functions take into account dimensional aspects at a neural population level, leaving aside orientation extraction (modeled in Chap. 5) and curvature.

6.1.2 SOS Neurons Transfer Function

As a general principle, SOS neurons are preferentially activated when two dimensions of the object are similar, while the third is sensibly smaller: $a \geq b \gg c$. Experiments performed varying the width and the thickness of the object gave the results reproduced in Fig. 6.5 (Shikata et al. 1996). These graphs and the comments of the authors, together with the principles previously introduced, are the bases for defining a transfer function which models the behavior of a population of SOS neurons.

The proposed transfer function depends on three main factors represented by three penalty, or inhibition terms, that take into account different aspects of SOS neurons responsiveness. In a hypothetical ideal situation, all inhibition terms would be zero and activation maximal.

The first component of the transfer function is I_s , the *symmetry inhibition term*. This term takes into account the difference between the two major dimensions of the object a and b : responsiveness is maximal, and inhibition minimal, for equal major axes. Asymmetrical situations are given higher penalties. The value of I_s is 0 when the major dimensions are equal, and increases with their difference:

$$I_s = \left(\frac{a - b}{a + b} \right)^{k_s} \quad (6.1)$$

Constant k_s modulates the effect of the difference between a and b on I_s . The exact value of k_s can be deduced only experimentally, and is not necessarily stable across conditions.

The second term considers the relation between the minor, most easily graspable dimension, c and the major ones a and b . It is called I_f , *flatness inhibition term*, and it increases with dimension c , representing the thickness of the object:

$$I_f = \frac{c}{a + b} \quad (6.2)$$

The two previous terms are independent from the absolute size of the object. As discussed in the previous section, it is though likely that the hand size is playing an important role in determining the global responsiveness of CIP to a given target

object. The *graspability inhibition term* I_g was thus introduced. As anticipated, it is expressed as a sigmoidal function. I_g decreases when increasing the graspable dimension c , and its symmetry point is the limit of a comfortable hand opening, called H :

$$I_g = \sigma(c, H) = \frac{1}{1 + e^{-k_g(c-H)}} \quad (6.3)$$

Constant k_g affects in this case the non-linearity of the equation: the larger k_g , the steepest the slope of the sigmoid function, and thus the influence of hand size H on SOS activation.

The global response R_{SOS} of a population of SOS neurons is thus estimated detracting the inhibitory quantities, appropriately weighted, from the theoretical 100% activation:

$$R_{SOS} = 1 - w_s \cdot I_s - w_f \cdot I_f - w_g \cdot I_g \quad (6.4)$$

The given expression is still undetermined, as the two parameters k_s and k_g and the three weights w have not been assigned any value yet. Starting with the symmetry term alone, least squares fitting can be used to compute the value of k_s and w_s that best fits (6.1) to the data corresponding to Fig. 6.5a. This gives $k_s = 1.948$, and $w_s = 1.059$. It looks reasonable to simplify setting $k_s = 2$ and $w_s = 1$. In this way, I_s is the square of the fraction $(a - b)/(a + b)$ and its weight can be omitted. Similarly, (6.3) can be fitted to the data of Fig. 6.5b. A value of 0.042 is obtained for the estimation of k_g , and 0.458 for w_g . With a little approximation, $k_g = 0.04$ and $w_g = 0.5$. Finally, the only remaining coefficient w_f is estimated through least squares fitting of (6.2) to the data of Fig. 6.5b (taking into account the contribution of (6.3)). The final result is $w_f = 0.030$. After substituting all these values in the corresponding formulas, the response of (6.4) remains more explicitly defined as:

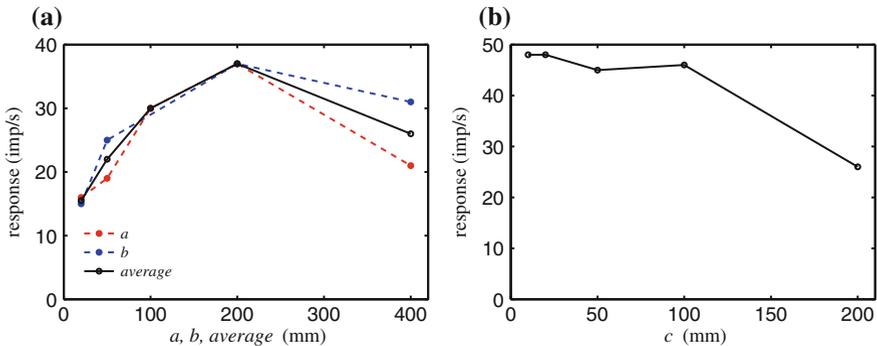


Fig. 6.5 Response of an SOS neuron as a function of object width and thickness. For width, the two major size responses and their average are plotted; the constant major size is 200 mm. Experimental data adapted from Shikata et al. (1996). **a** Width response (thickness $c = 20$ mm). **b** Thickness response (width $a = b = 200$ mm)

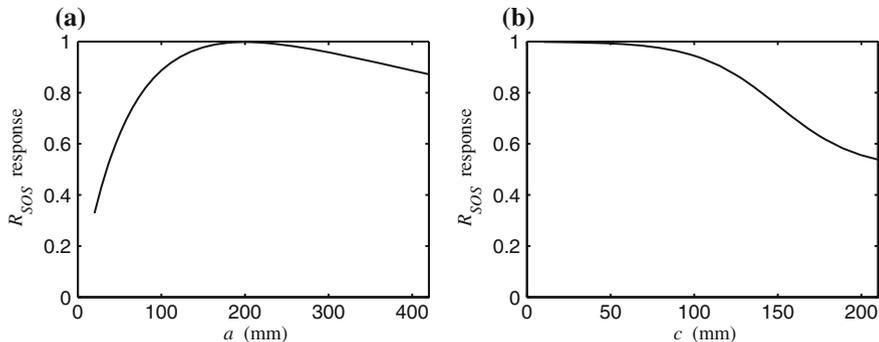


Fig. 6.6 Response of an SOS neuron as a function of object width and thickness. Simulated data obtained with (6.5), for the same width and thickness values of Fig. 6.5. **a** Width response ($b = 200$ mm, $c = 20$ mm). **b** Thickness response ($a = b = 200$ mm)

$$R_{SOS} = 1 - \left(\frac{a-b}{a+b} \right)^2 - 0.03 \frac{c}{a+b} - 0.5 \frac{1}{1 + e^{-0.04(c-H)}} \quad (6.5)$$

The global SOS response according to 6.5 was calculated as a function of object width and thickness. The results depicted in Fig. 6.6 show how the proposed model, properly parameterized, nicely fits the experimental data of Fig. 6.5 ($H = 150$ mm).

6.1.3 AOS Neurons Transfer Function

Axis orientation selective (AOS) neurons activate when one of the three dimensions of the object is quite larger than the other two, which are closer in size: $c \gg a \geq b$. Compared to SOS, less numerical results are available in the literature, and the main source of information is Fig. 6.4a, with the description of the corresponding experiments (Sakata et al. 1998). SOS and AOS neurons are intermixed in CIP, and it is thus plausible to assume that their response functions are similar. The hypothetical transfer function of AOS neurons was thus composed starting from the same three inhibition terms introduced in the previous section.

AOS *symmetry inhibition term* is equal to 0 when dimensions a and b are equal, and increases proportionally with their difference, exactly as in (6.1):

$$I_s = \left(\frac{a-b}{a+b} \right)^{k_s} \quad (6.6)$$

No experiments explicitly designed to verify the effect of differences between the two minor dimensions have been carried out for AOS neurons. This effect is probably not very strong, but it can be reasonably assumed that a large asymmetry would indeed

affect the perception of the elongated object. Such reduced influence of the fraction $(a - b)/(a + b)$ on the total response can be obtained changing the constant k_s .

Similarly to (6.2), the next term compares the major and minor dimensions of the object. This time, it is called I_l , *length inhibition term*, as it decreases with increasing the major dimension c of the object:

$$I_l = \frac{a}{c} \quad (6.7)$$

The graspable dimension, a in this case, is again the numerator of the fraction, as was c in (6.2). In this case the numerator could also be $(a + b)/2$, but if a and b are very similar this would likely be a pointless calculation.

The *graspability inhibition term* is again a sigmoidal function decreasing with the increasing of the minor dimension a , having as symmetry point the limit of a comfortable hand opening H .

$$I_g = \sigma(a, H) = \frac{1}{1 + e^{-k_g(a-H)}} \quad (6.8)$$

Again, the activation of a population of AOS neurons is estimated detracting the inhibition quantities from the theoretical 100% activation:

$$R_{AOS} = 1 - w_s \cdot I_s - w_l \cdot I_l - w_g \cdot I_g \quad (6.9)$$

Due to the limited availability of data, a bigger extrapolation effort is needed in the AOS case to estimate appropriate values for parameters and coefficients. The case of the symmetry term is the most critical, as there is no published numerical data which can help in determining the values of k_s and w_s . This second coefficient can be set to the same value as for SOS neurons, $w_s = 1$, whilst k_s should be assigned a value such that the influence of the term on the overall response is reduced with respect to the SOS case. The easiest solution, but certainly no the only possible one, is to set $k_s = 1$, and leave only the fraction component. Response would thus linearly increase when reducing the difference between a and b . Regarding graspability, there are no reasons to believe that parameter k_g and weight w_g should be much different from the SOS case. Least squares fitting of (6.8) to the data of Fig. 6.4a gives values included in $[0.02, 0.05]$ for k_g and in $[0.5, 0.8]$ for w_g , depending on the initial conditions. It seems thus reasonable, for symmetry and ecological reasons, to set $k_g = 0.04$ and $w_g = 0.5$, as in (6.5).

Sakata et al. (1998) state that: “discharge rate of the AOS neurons increased monotonically with increasing length of the stimulus”. The authors did not provide further information on this issue, but this comment describes how to generate additional data which could help in fitting the functions. A small additional dataset of 6 points in which response linearly increases with c was thus prepared. The newly generated dataset was used to fit (6.7) and thus set the value of w_l . Values between 0.2 and 1 were obtained using different graspable sizes of a . There is no reason why

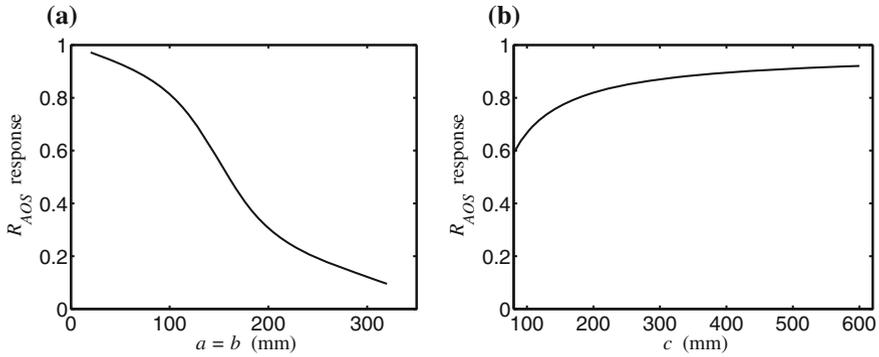


Fig. 6.7 Response of an AOS neurons as a function of object width and length. Simulated data obtained with (6.10). **a** Width response ($c = 300$ mm). **b** Length response ($a = b = 80$ mm)

the value of w_l should not change dynamically, but for the moment an intermediate value of $w_l = 0.374$, obtained for $a = 80$, is chosen. The overall formula for AOS response is thus defined as:

$$R_{AOS} = 1 - \frac{a - b}{a + b} - 0.37 \frac{a}{c} - 0.5 \frac{1}{1 + e^{-0.04(a-H)}} \quad (6.10)$$

The behavior of (6.10) with changing thickness and length of the object is shown in Fig. 6.7 ($H = 150$ mm). Figure 6.7a tries to reproduce the effect depicted in Fig. 6.4a, whilst Fig. 6.7b shows how the response grows when increasing c . Again, the effects described in the neuroscience literature are well reproduced.

6.1.4 Robotic SOS and AOS

After definition of the transfer functions and comparison with the available neuroscience data, the CIP neuron model can be tested on the robotic setup with images of real objects. Using the object pose estimation procedure described in Chap. 5, the dimensions of twelve shapes, depicted in Fig. 6.8, were estimated and used to compute simulated AOS and SOS activations for each shape. The shapes used for this purpose were eight boxes and four cylinders of different size and proportions. In principle, the modeled activations do not take into account curvature and do not distinguish between cylinders and parallelepipeds. Nevertheless, all cylinders have the same a and b dimensions (their diameter), and this increases their AOS responsiveness, because the length inhibition term (6.7) is always 0.

The constants of the final activation functions were employed, with the exception of k_s in (6.5), which was set to 1 as in (6.10), in order to improve the equilibrium between SOS and AOS activations. Average activation across 10 trials for all the

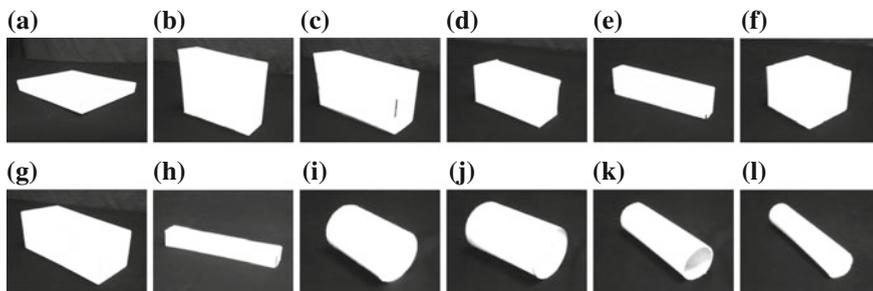
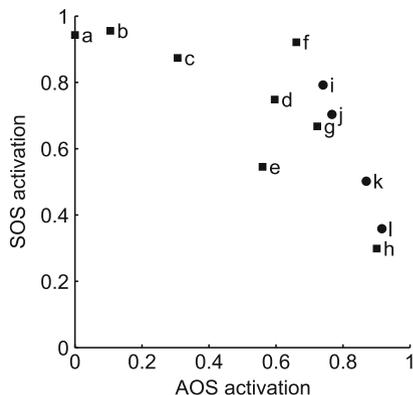


Fig. 6.8 Shapes for which experimental SOS and AOS activations are computed

Fig. 6.9 Experimental SOS/AOS activation for the shapes of Fig. 6.8



shapes in Fig. 6.8 is mapped on an SOS/AOS graph displayed in Fig. 6.9. Standard deviations are very low and hence not plotted. The comfortable grasp size H was maintained equal to 150 mm: although the robot can grasp larger objects, 150 mm is about the limit that allows for full contact of the tactile sensors placed on the robot hand fingertips.

Even though there is no direct comparison available for validating the obtained results, a visual assessment of the activations plotted in Fig. 6.9 reveals that activations look appropriate for the objects' characteristics. The only clearly elongated box, (h), shows a clear dominance of AOS over SOS response. On the other extreme, boxes (a) and (b) are undoubtedly assessed as completely flat. For box (c) the SOS activation is still clearly superior to the AOS activation, whilst for (d) and (f) the difference is much reduced. Boxes (d), (e) and (g) demonstrate a substantial equilibrium between activations, with a light bias toward SOS for (d) and toward AOS for (g). It is interesting to observe that nearly all boxes are disposed along an arc from (a) to (h), with only (e) and (f) deviating from the main path. Such deviations can be simply a side-effect of the model approximation, but could also reflect an increased suitability for grasping actions of (f) with respect to (e). For what concerns cylinders, (i) is the only one having a larger SOS activation, while for (j) it is

the AOS responsiveness that is slightly prevailing. Cylinders (k) and (l) are clearly elongated, and their AOS activation is dominant. Qualitatively, these results seem to properly represent the range of possible object proportions. From a robotic point of view, they show that the system is able to properly detect and code absolute and relative dimensions of target objects. For the model, these results suggest that it goes in the right direction, but more neuroscience experiments of different kinds would be needed for refinement and further validation.

Anticipating the possible use of SOS and AOS activations for the generation of hand configurations, the analyzed objects can be easily clustered in three groups. Objects on the top left of the graph, (a), (b) and (c), are definitely flat and will likely be grasped with a pad opposition between thumb and fingers. Elongated objects (k), (l) and (h) form the bottom right cluster, denoting AOS dominance, and can be grasped with either a precision grip or an involving grip. More complicated is the situation for the six objects in the central cluster. In fact, for hand shaping, they seem to be more different one another than represented in the graph. Size and curvature are probably the factors that would further distinguish between them to drive the selection of suitable grips.

6.1.5 Discussion and Future Developments

The above model offers some solutions to the problem of identifying the transfer functions of the different areas of the dorsal stream, but opens at least as many questions. More experiments are needed to validate the proposal. The actual importance of hand size on SOS and AOS activation should be explicitly analyzed, through experimental protocols designed to distinguish the effect of relative and absolute size of features. For example, no experiments are reported in the CIP literature regarding non graspable (or strangely shaped) objects, and these are definitely required at this point. Similarly, there is the need to disambiguate the influence of shape and size on neuronal response. This can be done by gradually changing the proportion and size of objects, and analyzing the response as a function of only one driving variable at a time. The responsiveness to object curvature should also be further explored. As the robotic simulation pointed out, it is very likely that the proposed functions will need to be updated and suited to new findings and requirements, but they constitute a helpful tool for orienting the future studies on the subject. The next step in the grasp planning process, performed by AIP, is to join SOS and AOS activations with data coming from other brain areas for deciding how to grasp possible target objects.

6.2 Planning and Executing the Grasping Action

The coding of SOS and AOS neurons for the visual characteristics of objects relevant for grasping purposes is the ideal input for AIP so that it can process the visual data and transform it to suitable hand configurations. A critical question is how much

of the information that AIP needs is provided by CIP and how much needs to be complemented by other areas.

6.2.1 Characteristics of the Visual Input to AIP

Fukuda et al. (2000) accurately measured, with a data glove, human grasping configurations on fifteen different objects of three classes (spheres, cylinders and parallelepipeds) and five different sizes for each class. They registered the values of 18 joint angles of the hand at the time of contact in real and *pantomimed grasping* (see related text box), extracting the two principal components of the joint space for each condition. They found statistically significant differences between real and pantomimed grasping. To a minor extent, they also found a difference between real and 2D object stimuli in pantomimed grasping. Both findings are consistent with the two streams literature. They also demonstrated with a neural network implementation that the visual information provided to the subject could account for 99 % of the variability observed in joint configurations, suggesting that indeed grasping was purely based on the available visual data. These results would have probably been very different if task and object identity were taken into account.

A comparison of the modeled SOS/AOS representation with the work of Fukuda et al. (2000) can help in drawing some conclusions on the completeness of the proposed model. In Fig. 6.10 an adaptation of the results provided in the cited work is compared with a representation of the same objects in a simulated SOS/AOS activation space. Figure 6.10a shows an average, across the three experimental subjects, of the first and second principal components of the joint space when grasping spheres of different sizes and cylinders and parallelepipeds of different thicknesses. Objects

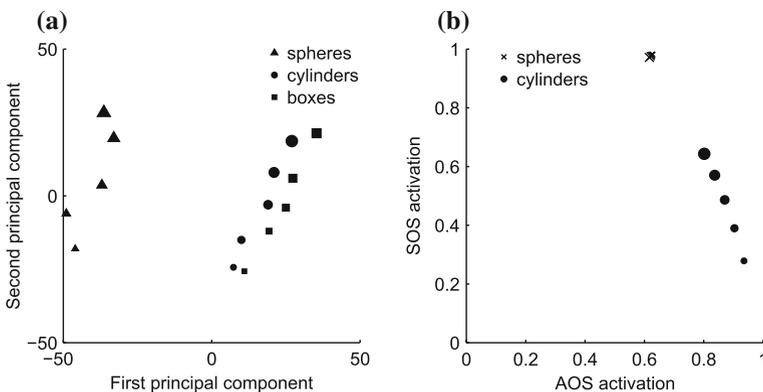


Fig. 6.10 Comparison between principal components of joint space during grasping (adapted from Fukuda et al. 2000) and AOS/SOS coding of similar objects. **a** Grasping joint space. **b** AOS/SOS representation

of the same dimensions were used to calculate the AOS and SOS activations of Fig. 6.10b.

The model does not recognize curvature, so parallelepipeds and cylinders with similar proportions have similar representations, and the same happens for cubes and spheres. The comparison between joint spaces for grasping cylinders and parallelepipeds (Fig. 6.10a), which are very similar, partially justifies this simplification. Whereas Fig. 6.10a represents an output of AIP processing, Fig. 6.10b constitutes a possible, probably partial, input. A qualitative comparison suggests that, while for cylinders the visual information provided by AOS and SOS neurons seems to be enough for generating an appropriate joint space for grasping, spheres of different sizes show nearly the same representation, suggesting that there is not enough data for deciding on how to grasp them. The reason for this discrepancy is probably twofold. First, AOS and SOS activation models were built on the existing data, which do not take into account all aspects of shape estimation. Size effects in the model are clearly observable only for the biggest shapes, close to the hand opening threshold. Indeed, activations of CIP neurons for equal objects of different sizes are not provided in the literature, and such aspect is not fully taken into account.

The second reason is related to the tasks and connectivity of AIP. Distance estimation, performed in LIP, is critical for the reliability of objects' size estimation. Ventral stream information regarding recognized objects also carries very good size estimates derived from experience. Some CIP neurons also code for distance (Shikata et al. 1996; Sakata et al. 1998), but it is likely that projections from LIP and from the ventral stream provide AIP with more exact estimates of the object size. It is thus likely that CIP uses a distance, and hence a size estimate, less precise than the one available to AIP. A possible hypothesis is that object size representation in CIP is exploited only with the purpose of filtering graspable from non graspable features, leaving exact estimation to LIP and AIP. The coding of potentially graspable object features, conveyed by the firing of SOS and AOS neurons, needs hence to be completed by accurate data on object size and location.

In spite of its critical importance, visual information is only one ingredient of the complex grasping recipe. In the next section, additional important factors, such as the criteria to follow in order to obtain reliable grasping actions, are described.

6.2.2 The Search for Grasp Quality

After AIP has gathered the available visual data regarding a target object, a number of issues have to be taken into account in order to produce an appropriate grasp. A very important aspect that strongly affects grasp planning is the search for quality in a grasping action. The same definition of quality is controversial, as strictly related to the task to execute. If the task is to handle a pen for writing, quality is measured in terms of manipulability. If the goal is to lift a heavy object, stability in the grip has to be pursued. For the limited scope of robotic grasping, often quality has been interpreted as a synonym of stability, or reliability of the action. The reduced tactile

skills of robots compared to primates makes of reliability a critical factor for the selection of a grip. Especially if no or little experience is available, deciding among candidate affordances represents a key task (Borst et al. 2004; Morales et al. 2004). Human beings take into account a number of aspects which help in ensuring at least a minimum level of reliability to their grasping actions, as described in Sect. 4.3.3.1.

Multiple criteria and measures have been employed for assessing and predicting robot grasp quality (Roa and Surez 2015). In Chinellato et al. (2005), a set of visual criteria for the reliability assessment of planar grips was defined, taking as reference physiological studies of human grasping, and robotic research on grasp stability. The criteria were used to predict the outcome of future robot grasping actions (Chinellato et al. 2003; Morales et al. 2004). Some of those criteria can be extended, maintaining their plausibility and usefulness, to the three-dimensional case, and are presented below. An important conceptual difference is that in Chinellato et al. (2005) criteria were computed for a number of pre-defined candidate grips, in order to globally evaluate them and select one of them for execution, whilst here the task is to generate a grasp plan that implicitly achieves good quality values. The criteria are consistent with the findings described in Sect. 4.3.3.1, and it can be hypothesized that there is an important contribution of the basal ganglia in providing AIP with the signals required to select between alternative grasping patterns (Clower et al. 2005). In fact, it has been suggested that the basal ganglia is a key area in the development of action selection tasks through reinforcement learning (Doya 1999).

The quality criteria can be subdivided into two classes: visual criteria, that mostly affect the selection of the contact points on the object; and motor criteria, that mostly affect hand shaping.

6.2.2.1 Visual Criteria

At least three visual criteria important in human grasping are also useful for robotic implementation.

Center of mass. The opposition axis of the grip should always pass close to the object center of mass, in order to minimize the effect of gravitational and inertial torques, especially if the object is heavy. If possible, grasping along the main inertia axes is preferred for the same reasons. Moreover, heavy objects are often grasped above the center of mass for increased stability (Bingham and Muchisky 1993). In many cases, this criterion is predominant in human grasping.

Grasping margin. This criterion aims at minimizing the risk of placing the fingers on unsuitable object features which could result in unstable contacts. It builds on the assumption that fingers should be placed far from edges, and that large grasping surfaces, at least above a given threshold, should be chosen if available.

Curvature. Grips on slightly concave surfaces are normally considered more reliable than grips on convex ones, because contact surface and thus friction is higher in the first case (Jenmalm et al. 2000). To implement this criterion computationally, the curvature of graspable features could be calculated at different frequencies, as

a slowly changing curvature is normally preferred. On the other hand, very high frequency curvature changes may indicate the presence of a rough surface, which is good to grasp because it offers high contact friction.

6.2.2.2 Motor Criteria

Finger extension This criterion aims at maximizing the contact surface between fingertips and object. The goal is to have a substantial equilibrium between the opening of all fingers. Moreover, an average finger aperture is preferred as it allows for bigger contact surfaces. Even though for humans this aspect is less important, because of the number of degrees of freedom of the hand and the high compliance of the fingertips, a grasping action in which some fingers are extended and other flexed is usually clumsy.

Force distribution In many cases, the optimization of the previous criterion ends in an optimal distribution of forces as well. Nevertheless, in case of complex objects and grasps with abducted fingers, a homogeneous distribution of the contact forces can be a critical issue. Usually, badly equilibrated grips can be improved through tactile feedback, but if the force asymmetry is too high, the object could slip or rotate due to unwanted torques.

An estimation of the movement cost should be added to these criteria to take into account the reaching comfort (see Sect. 4.3.3.1). This can be done by computing the expected joint rotations required to achieve a given goal position.

6.2.2.3 Modulation of the Effect of Quality Criteria

The ventral stream contribution provides the use of the quality criteria with an important flexibility. Knowledge regarding object characteristics, such as weight or compliance, and the outcome of previous grasping experiences can be used as modulation factors which assign different importance to the above criteria in different conditions. Default, prudential solutions are adopted in the case of failed recognition or low classification confidence, to respect the uncertainty of the situation. If the object properties are known, the biasing toward one criteria or another can be much stronger. To give a simple example, if the object is big and heavy, the center of mass criterion and the force distribution are very important, whilst for a small light object the grasping margin is probably the critical criterion. Recent psychophysiological findings support this hypothesis (Eastough and Edwards 2007). Computationally, criteria weighting can be initially hard-wired, but when the system increases its knowledge of the graspable world, this aspect should acquire a more dynamical behavior, especially if feedback is available regarding the appropriateness of grasping decisions.

6.2.3 Grasp Planning

All elements necessary to generate a grasping action suitable to a given condition have been described, and they have now to be joined in a grasp plan. Area AIP is in charge of transforming the visual data provided by CIP and other areas into an appropriate hand configuration for grasping the target feature. The goal is to translate information about size, location and AOS/SOS representation into hand joint space. The specification of exact contact locations is not necessary, the finger placing and the movement trajectory being dependent on contextual optimization of the quality criteria. Grasp planning can be performed following a short sequence of logical steps.

The task-driven decision on the type of grip to perform (precision or power) is taken in advance. If the goal is to perform a power grip, visual analysis is usually very simplified (as suggested by the reduced activity in AIP), and only the object center of mass has to be approximately calculated. In this case, the hand has to move toward the object center, and the opposition axis of the fingers on the object is determined by the motor cost, which is minimized avoiding unnecessary rotations. The reaching action that requires minimum joint movements is thus executed, and once the palm gets in contact with the objects, the fingers close around it.

For precision grips, the requirements are different according to the AOS/SOS coding. If the object has a prevailing AOS activation, and the long axis is free for grasping, like for a standing cylinder, there is no preferential approaching direction apart from the one provided by the hand pose. The grasp action will be performed so that the opposition axis between thumb and fingers passes close to the center of mass, maximizing the correspondent criterion, and in a way that minimizes the cost of the movement, maintaining the trajectory as straight as possible and avoiding unnecessary rotations (Fig. 6.11a). If the object with prevailing AOS activation is laying on its long axis then wrist rotation is required, as only one approaching direction carries to the correct grasping position, from above and toward the center of mass (Fig. 6.11b). In the former case, an involving grip which includes contact between object and hand palm can be executed if required, in the latter case only fingertip grips are possible.

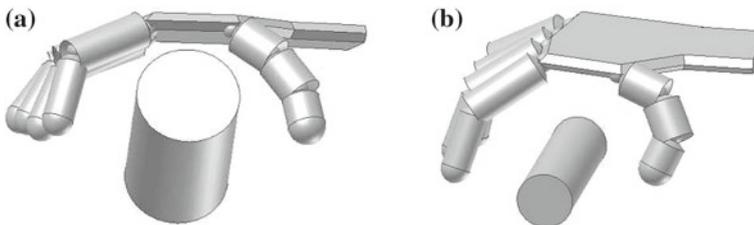


Fig. 6.11 Grasp approaching direction for standing and lying AOS dominant objects. **a** Standing AOS object. **b** Lying AOS object

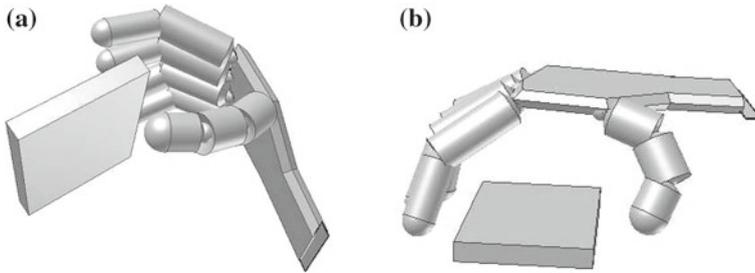


Fig. 6.12 Grasp approaching direction for standing and lying SOS dominant objects. **a** Standing SOS object. **b** Lying SOS object

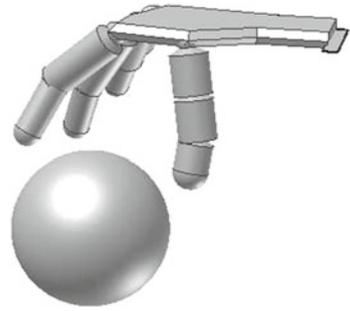
If the object has a prevailing SOS activation, and the thin, grasping dimension is free, the direction of grasping is the one which makes the fingers oppose on the minor dimension of the object. The final part of the reaching movement is constrained to a plane, and there is still one degree of freedom for optimizing movement cost and center of mass approaching. Both visual and motor criteria have thus to be taken into account. If there are no other constraints, it is safer in this case to grasp an object from above and not from the side, in order to minimize the effect of gravitational torques. For light objects destabilizing torques are unlikely, and reaching comfort prevails (Fig. 6.12a). Object identification would help in this case. If the object is laying on its preferred graspable feature, the grasping action will have to be performed on a different dimension, along the main inertia axis (Fig. 6.12b), or not performed at all.

As can be observed in Fig. 6.9, some shapes show no clear dominance of either SOS or AOS activation. Objects with this characteristic can be grasped with both strategies, and again movement economy can be the determinant factor for establishing the final grasping position. Activation thresholds could be employed to distinguish the cases of SOS dominant, AOS dominant and neutral objects.

The described procedure leaves out the cases of objects that are approximately spherical, or that simply do not offer clearly graspable axes or surfaces. In such cases, a grasp in which fingers are abducted is preferable, to distribute the grasping force around the object surface. Most commonly in these situations, the ring and small fingers are used just for providing support and additional stability, whilst the index and medium fingers create a triangular force distribution with the thumb (see Fig. 6.13). For this reason the resulting grasp is called the *tripod grasp*.

Human grasping experimental results are consistent with a two virtual-fingers, hierarchical model of the control of a tripod grasp. First, an opposition space is selected between the thumb and the index and medium fingers, which form together the second virtual finger. Index-medium finger abduction depends on the object size (Gentilucci et al. 2003), and the direction of the forces exerted by the two fingers are symmetric with respect to the opposition axis (Baud-Bovy and Soechting 2001). If the object is irregular, tactile feedback is required in order to adjust finger position and force distribution in order to find a stable configuration before lifting the object. fMRI research support a substantial identity in the processing of two finger precision

Fig. 6.13 Tripod grasping on a spherical object



grasps, tripod grasps and extended tripod grasps in which all five fingers contact the object (Cavina-Pratesi et al. 2007).

The above description, although valid for robotic implementation, is just a qualitative description of the results of AIP processing. Suggestions for an implementation closer to the cortical mechanisms are provided in Sect. 7.1.

6.2.4 Grasp Execution

The final grasping action is executed following the above guidelines, making use of visual information regarding object pose and location, and taking into account relevant grasp quality criteria. The grasping system introduced in Sect. 5.4.1 allows for the execution of differently shaped precision grips, including the tripod grasp. Power grips, although possible, are not controllable and thus avoided, as the hand palm is not endowed with tactile sensors necessary to detect the contact with the object. A wrong positioning could hence result in an excessive solicitation of the hand, with risk of damaging it.

6.2.4.1 The Reach and Grasp Movement

Before the movement onset, the goal position and direction of the opposition axis are defined as described in the previous section, and computed using the estimated location, pose and size of the object. The initial position of the arm, corresponding to the fixation period, before the movement starts, is shown in Fig. 6.14a. The first part of the reaching movement is just aimed at reducing the distance between object and effector. The final stage of the reaching action is more precise, and has to be performed moving perpendicularly to the opposition axis, from a short distance to the goal position. A *via posture*, i.e. an intermediate position and orientation goal (Meulenbroek et al. 2001), is defined in order to allow the correct execution of such stage, ensuring at the same time that no collisions with the target object are possible. The intermediate goal position has the hand in the correct grasping

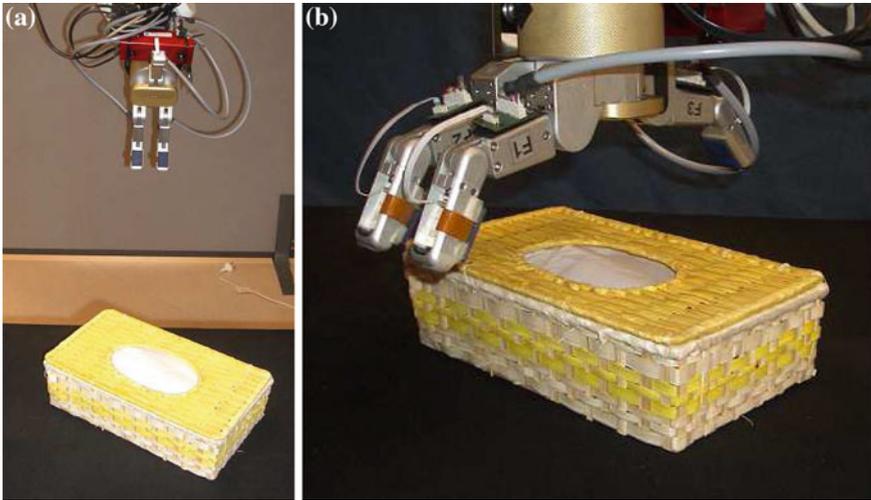


Fig. 6.14 First stages of grasping action execution, during object fixation before movement onset, and during reaching, before the hand closes on the object. **a** Fixation position. **b** Via-posture

direction, and the distance from the object is such that movements different from the approaching one would not result in unwanted contacts. Figure 6.14b shows an example of appropriate via-posture. A safety margin is also added to the expected object size to compensate for possible estimation errors. This is consistent with the findings of Hu et al. (1999), suggesting that hand preshape is performed taking into account also the object dimensions not directly involved in grasping. Such dimensions affect the security of the transport movement, which could be at risk of collision if during hand approaching the fingers pass too close to the object.

During the first stage of the transport movement the hand rotates toward the correct orientation while the arm moves to the intermediate position. Once the via-posture is reached, the hand is in the correct direction and, without stopping the movement, the robot arm reaches further toward the object until the fingertips are at level with the estimated object center of mass (Fig. 6.15a). Until this point the process

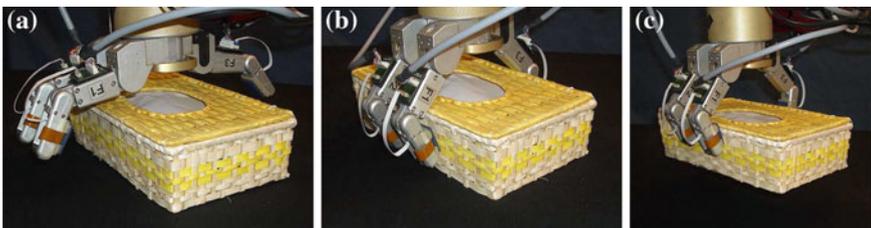


Fig. 6.15 Last stages of grasping action execution, during finger closing and object lifting. **a** Closing position. **b** Grasping. **c** Lifting

is fully open-loop, and only driven by the initial grasp plan. Once the estimated final position is reached, the fingers close, and tactile sensors are used to determine the moment of contact between fingertips and object. As soon as a contact is detected, the corresponding finger stops moving. The grasping movement is completed when all fingertips have contacted the object, as depicted in Fig. 6.15b. The grasp configuration is then checked, as described in the next section, and if it is considered correct, the object is lifted (Fig. 6.15c). If a finger misses the object, and thus no contact is ever detected, it stops as a security measure when it reaches a minimum extension threshold, currently set to 54 mm.

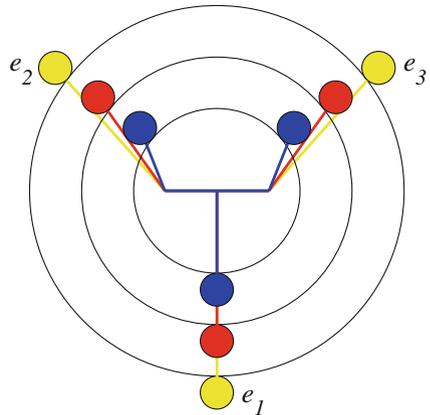
For what concerns tripod grasping, it is at the moment executed for objects classified as spheres (Sect. 5.4.2), which are approached from above, and grasped so that the opposition axis passes ideally through their center of mass. The only addition to the usual procedure is the separation between the fingers. This is done setting the finger opening angle θ (see Fig. 5.11) proportional to the object size: the bigger the object, the higher the finger separation. The opening angle, in radians, is given by:

$$\theta = \frac{D - i}{D} \quad (6.11)$$

where D is the object diameter and i the inter-finger distance (see Fig. 6.19). The outcome of using (6.11) can be observed in Fig. 6.16, in which finger positions and force directions for spheres of different sizes are shown. This solution allows to homogeneously distribute the contact points around the shape while maintaining the force directions as orthogonal as possible to the object side. In all other cases of normal opposition grips, in which fingers e_2 and e_3 are parallel, $\theta = 0$.

Examples of via-posture and grasp execution for a vertically placed cylindrical shape and a spherical shape are shown in Fig. 6.17.

Fig. 6.16 Finger positions and force directions obtained with expression (6.11) for grasping spheres of different sizes (80, 130, 180 mm)



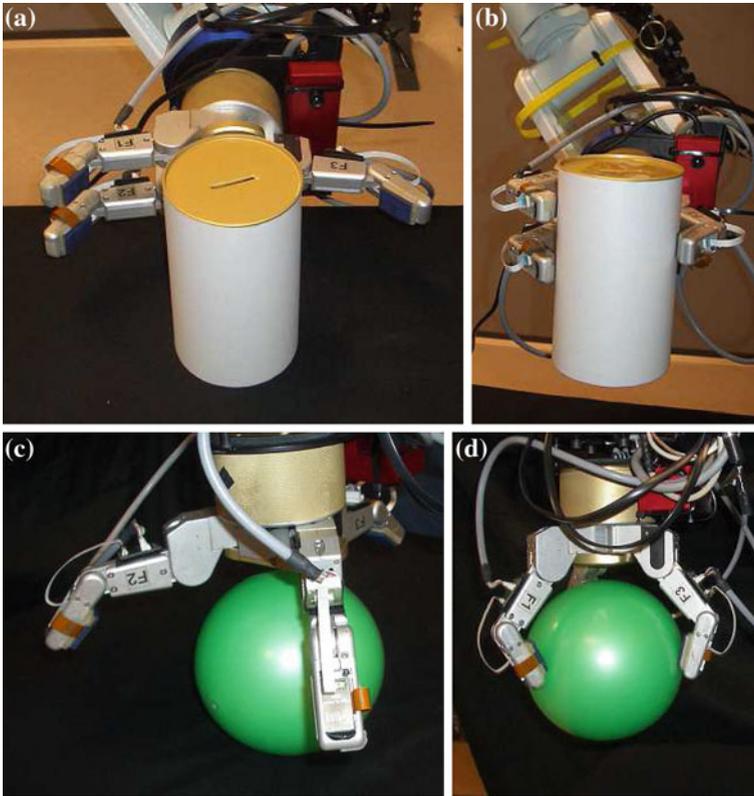


Fig. 6.17 Via-postures and grasp execution for a vertically placed cylindrical shape and a spherical shape. **a** Via-posture for cylinder. **b** Cylinder grasping. **c** Via-posture for sphere. **d** Sphere grasping

6.2.4.2 A Helping Tactile Hand

When the fingers close on the object, they stop at the moment of contacting the object surface. If the pose estimation was correct, the movement performed as required, and the object did not move, the fingers should present a substantial equilibrium in their final extensions. The expected proprioceptive state of the hand, corresponding to a symmetric grasp with respect to the object grasping axis, is of equal extensions for the three fingers. This constitutes a basic forward model of the expected action outcome. If, for a divergence between the estimated and the real values of distance, pose and size of the object, or for any other unexpected factor, the fingers touch the object with different extensions and orientations, the grasp could be unstable (see Fig. 6.18). In these cases, differences between finger extensions are detected, and proprioceptive hand feedback is used to adjust the grasping action to the real conditions, and thus achieve the necessary grip stability. An adaptation of the finger extension criterion

Fig. 6.18 Example of unstable grasp requiring a correction movement



provides the feedback on the actual conditions and suggests a correction movement if necessary. A proper action for adapting the hand pose to the new situation can be computed from the difference between finger extensions. As represented in Fig. 6.19, any correction movement is made of two components: z for translation and α for rotation.

Orientation correction is necessary when the orientation of the object is different from expected. This situation is identified by comparing the extensions of the two parallel fingers e_2 and e_3 . The required rotation correction α is given by:

$$\alpha = \arctan \frac{e_3 - e_2}{i} \quad (6.12)$$

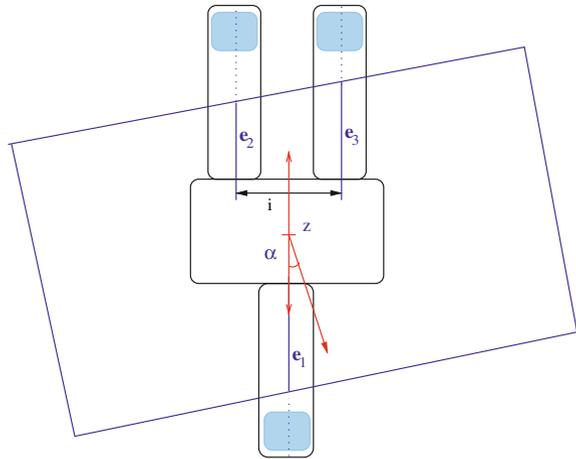
where i is the inter-finger distance. If the hand is rotated by α in the direction of the finger with shorter extension, it will get in the situation in which the parallel fingers should contact the object with the same extension. A threshold is used to allow for minimum unavoidable extension differences which do not affect grasp stability. With the current settings, the correction movement is executed only if $|\alpha| > 2^\circ$.

Translation correction is performed when the position of the object is different from the estimation. This case corresponds to a difference in extension between the thumb and the opposing fingers. The thumb extension is compared to the average extension of the opposing fingers, taking into account also possible extension differences between them. The displacement required for position correction is computed with the following expression (see again Fig. 6.19):

$$z = \frac{1}{2} \left(\frac{e_2 + e_3}{2} - e_1 \right) \quad (6.13)$$

Again, this is the displacement that would carry the hand to the planned grasping axis. The threshold for moving is set at $|z| > 5$ mm. This threshold value and the one for rotation correction could be more appropriately set through a learning framework driven by the results of experiments performed in different conditions. Different thresholds could also be used for different objects.

Fig. 6.19 Representation of the correction movements in rotation, α , and translation, z



During action execution, once all fingers have contacted the object and stopped moving, (6.12) and (6.13) are computed. If at least one of them is above threshold, the fingers open again, and the movement that compensates for the required orientation and translation correction is calculated and executed. If needed, the process repeats until no other corrections are required, then the fingers close firmly on the object and lift it.

The described grasping technique has been tested in two different conditions, i.e., without or with object displacement. The first condition, corresponding to a normal working situation, usually ends with a successful grasping action without performing any correction movement. In fact, in almost all cases the input provided by the visual system is good enough to allow the execution of the grasping action without the need of correcting hand position or orientation.

During the second type of tests, in perturbed conditions, changes in the object position and/or orientation were introduced on purpose, to check if the system was able to deal with unexpected and suddenly changing situations. The changes were applied after the visual analysis had been finished so that the real pose of the object was different from the estimated one, like in the example of Fig. 6.18. In this situations the robot might not be able to grasp the object without the support of the tactile feedback. Using information about finger extensions and hand contacts with the object surface, hand orientation and position are corrected as described above. When the difference between the real and estimated object pose is big, more than one correction movement might be required.

This framework presents two major limits. The first is that only displacement errors parallel to the grasping axis can be corrected. Any deviations from the object center of mass along the other two directions will not be detected, unless one of the fingers misses the contact. The second problem is with objects which edges are not parallel. In such cases, a rotation correction movement will be performed although not required, as α is always above threshold. A possible solution is to increase the

threshold to a trade-off value which does not affect the reliability of normal grasping actions and, at the same time, is suitable for many unusual conditions.

This correction method models, in a simple way and at high level, the comparison between expected and real somatosensory input, described in Sect. 4.3.4, and determines a correction movement that aims at reducing such difference. In this way, grasp stability is implicitly achieved, through minimization of the difference between detected proprioceptive state and expected goal state given by a basic forward model. Following a simplified version of the schema of Demiris and Hayes (2002), very simple inverse models (Eqs. 6.12 and 6.13) compare the goal state of the hand with its actual condition and generate a motor command suitable to approach the goal. The forward model evaluates the outcome of the current motor command and guides the following step in order to keep improving the quality of the ongoing situation estimated by the finger extension criterion.

6.3 Conclusions

Compared to its neighbor grasping area AIP, the knowledge, and especially the modeling regarding one of the most fundamental areas of the dorsal stream, the posterior intraparietal area CIP, remains relatively undeveloped. In the first part of this chapter, a detailed analytical interpretation of CIP tasks is provided which takes into account both the computational and the neurophysiological points of view. The coding of visual features as it is thought to be performed by CIP neurons is employed in the second part of the chapter for generating appropriate grasp configurations. The integration of different kinds of grasp-related information and constraints, as performed by area AIP, is modeled and adapted to the requirements of the robotic system. Grasping experiments, performed with the aid of tactile feedback, confirm the suitability of the model to real robotic setups.

Neuroscientific plausibility and practical usefulness of the proposed vision-based grasping model have been justified, but several directions for further development in both regards can be devised. The next chapter will present a number of optional developments and required improvements for the robotic application. Issues regarding necessary refinements and possible alternatives for the model are also discussed.

References

- Baud-Bovy G, Soechting JF (2001) Two virtual fingers in the control of the tripod grasp. *J Neurophysiol* 86(2):604–615
- Bingham GP, Muchisky MM (1993b) Center of mass perception: perturbation of symmetry. *Percept Psychophys* 54(5):633–639
- Borst C, Fischer M, Hirzinger G (2004) Grasp planning: how to choose a suitable task space. In: IEEE international conference on robotics and automation, New Orleans, USA, pp 319–325

- Cavina-Pratesi C, Monaco S, McAdam T, Milner D, Schenk T, Culham JC (2007) Which aspects of hand-preshaping does human AIP compute during visually guided actions? Evidence from event-related fMRI. In: Annual meeting of the Society for Neuroscience
- Chinellato E, Morales A, Fisher RB, del Pobil AP (2005) Visual quality measures for characterizing planar robot grasps. *IEEE Trans Syst Man Cybern Part C* 35(1):30–41
- Chinellato E, Morales A, Sanz PJ, del Pobil AP (2003) Validation of features for characterizing robot grasps. In: Mira J, Alvarez J (eds) *Artificial neural nets problem solving methods*, LNCS 2687, Springer, pp 193–200
- Clower DM, Dum RP, Strick PL (2005) Basal ganglia and cerebellar inputs to ‘AIP’. *Cereb Cortex* 15(7):913–920. doi:[10.1093/cercor/bhh190](https://doi.org/10.1093/cercor/bhh190)
- Demiris Y, Hayes G (2002) Imitation as a dual-route process featuring predictive and learning components: a biologically-plausible computational model. In: Dautenhahn K, Nehaniv C (eds) *Imitation in animals and artifacts*, Chap 13. MIT Press, Cambridge, pp 327–361
- Doya K (1999) What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Netw* 12(7–8):961–974
- Eastough D, Edwards MG (2007) Movement kinematics in prehension are affected by grasping objects of different mass. *Exp Brain Res* 176(1):193–198. doi:[10.1007/s00221-006-0749-3](https://doi.org/10.1007/s00221-006-0749-3)
- Fukuda H, Fukumura N, Katayama M, Uno Y (2000) Relation between object recognition and formation of hand shape: a computational approach to human grasping movements. *Syst Comput Jpn* 31(12):11–22
- Gentilucci M, Caselli L, Secchi C (2003) Finger control in the tripod grasp. *Exp Brain Res* 149(3):351–360. doi:[10.1007/s00221-002-1359-3](https://doi.org/10.1007/s00221-002-1359-3)
- Hu Y, Eagleson R, Goodale MA (1999) The effects of delay on the kinematics of grasping. *Exp Brain Res* 126(1):109–116
- Hu Y, Osu R, Okada M, Goodale MA, Kawato M (2005) A model of the coupling between grip aperture and hand transport during human prehension. *Exp Brain Res* 167(2):301–304. doi:[10.1007/s00221-005-0111-1](https://doi.org/10.1007/s00221-005-0111-1)
- Jenmalm P, Dahlstedt S, Johansson RS (2000) Visual and tactile information about object-curvature control fingertip forces and grasp kinematics in human dexterous manipulation. *J Neurophysiol* 84(6):2984–2997
- Meulenbroek RG, Rosenbaum DA, Jansen C, Vaughan J, Vogt S (2001) Multijoint grasping movements. Simulated and observed effects of object location, object size, and initial aperture. *Exp Brain Res* 138(2):219–234
- Morales A, Chinellato E, Fagg AH, del Pobil AP (2004) Using experience for assessing grasp reliability. *Int J Humanoid Rob* 1(4):671–691
- Naganuma T, Nose I, Inoue K, Takemoto A, Katsuyama N, Taira M (2005) Information processing of geometrical features of a surface based on binocular disparity cues: an fMRI study. *Neurosci Res* 51(2):147–155. doi:[10.1016/j.neures.2004.10.009](https://doi.org/10.1016/j.neures.2004.10.009)
- Roa M, Surez R (2015) Grasp quality measures: review and performance. *Auton Robots* 38(1):65–88. doi:[10.1007/s10514-014-9402-3](https://doi.org/10.1007/s10514-014-9402-3)
- Sakata H, Taira M, Kusunoki M, Murata A, Tanaka Y (1997) The TINS lecture. The parietal association cortex in depth perception and visual control of hand action. *Trends Neurosci* 20(8):350–357
- Sakata H, Taira M, Kusunoki M, Murata A, Tanaka Y, Tsutsui K (1998) Neural coding of 3D features of objects for hand action in the parietal cortex of the monkey. *Philos Trans R Soc B: Biol Sci* 353(1373):1363–1373
- Sakata H, Tsutsui KI, Taira M (2005) Toward an understanding of the neural processing for 3D shape perception. *Neuropsychologia* 43(2):151–161. doi:[10.1016/j.neuropsychologia.2004.11.003](https://doi.org/10.1016/j.neuropsychologia.2004.11.003)
- Shikata E, Tanaka Y, Nakamura H, Taira M, Sakata H (1996) Selectivity of the parietal visual neurones in 3D orientation of surface of stereoscopic stimuli. *Neuroreport* 7(14):2389–2394