

# Hierarchical and multiple action representation using temporal postural synergies

G. Tessitore · C. Sinigaglia · R. Prevede

the date of receipt and acceptance should be inserted later

**Abstract** The notion of synergy enables one to provide simplified descriptions of hand actions. It has been used in a number of different meanings ranging from kinematic and dynamic synergies to postural and temporal postural synergies. However, relatively little is known about how representing action by synergies might take in account the possibility to have a hierarchical and multiple action representation. This is a key aspect for action representation as it has been characterised by action theorists and cognitive neuroscientists. Thus, the aim of the present paper is to investigate whether and to what extent a hierarchical and multiple action representation can be obtained by a synergy approach. To this purpose, we took advantage of representing hand action as a linear combination of Temporal Postural Synergies (TPSs), but provided that TPSs have a tree-structured organisation. In a tree-structured organisation an hand action representation can involve a TPS only if the ancestors of the synergy in the tree are themselves involved in the action representation. The results showed that this organisation is enough for forcing a multiple representation of hand actions in terms of synergies which are hierarchically organised.

## 1 Introduction

The human hand is a very complex system with more than twenty degrees of freedom (DOF) allowing to dex-

terously perform actions. This complexity has provided a key challenge for action representation research. For it seems unlikely that all DOF are individually represented and controlled during the execution of hand actions such as grasping, tearing, holding. Several studies have highlighted the need for a simplified way of computing hand actions (Iberall et al 1986; Santello et al 1998; Mason et al 2001). The notion of *synergies* is crucial to this regard. Originally coined by Bernstein (Bernstein 1967) in the motor control domain, this notion refers to specific patterns in muscle activities or movement kinematics/dynamics as building blocks for representing and controlling actions. It has been used in a number of different meanings ranging from kinematic and dynamic synergies (Grinyagin et al 2005; d'Avella et al 2006; Santello and Soechting 2000) to postural (Tessitore et al 2010; Mason et al 2001; Santello et al 1998) and temporal postural synergies (Vinjamuri et al 2010a,b; Santello et al 2002).

Thus, for instance, it has been shown that hand actions can be represented by linear combinations of postural synergies, where the coefficients consist of temporal weightings of the synergies (Mason et al 2001). In a similar vein, it has been proposed (Santello et al 2002; Vinjamuri et al 2010b) that hand actions, expressed as temporal sequences of hand-joint configurations, should be represented by linear combinations of a small number of temporal postural synergies, that is, of specific patterns in the space of hand-joint configurations varying over time (see Figure 1).

In spite of these studies, relatively little is known about how representing action by synergies might take into account a characteristic aspect of action representation, that is, that an action can be represented in a variety of ways differing from each other in generality (Goldman 1970; Mele 1997). Consider, for in-

---

G. Tessitore, R. Prevede  
University of Naples Federico II, Department of Physical Sciences, via Cinzia I-81100 Napoli, Italy.  
E-mail: tessitore,prevete@na.infn.it

C. Sinigaglia  
University of Milan, Department of Philosophy, via Festa del Perdono 7, I-20122 Milano, Italy  
E-mail: corrado.sinigaglia@unimi.it

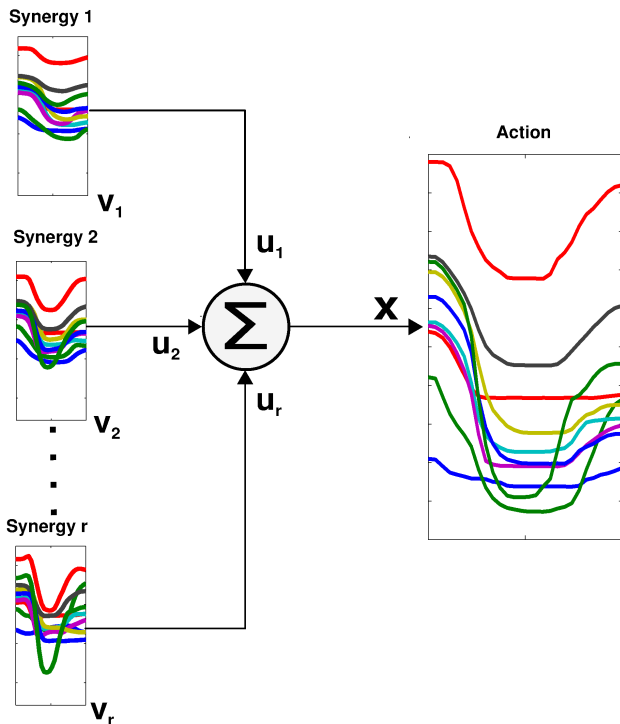


Fig. 1: *Temporal Postural Synergies*. Hand actions can be expressed by linear combinations of a small number of temporal postural synergies. This kind of synergies are specific patterns in the space of hand-joint configurations varying over the time. In figure  $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_r$  refer to the temporal postural synergies,  $u_1, u_2, \dots, u_r$  to the coefficients of the linear combination.

stance, a hand action like a pinch precision grip: it can be represented as a mere hand action, but it can be also represented as a grasping action, as a grasping action performed with a precision grip, and finally as a grasping action performed with a pinch precision grip. Accordingly a given action can be represented in multiple ways which are hierarchically organized (Goldman 1970). Strikingly, similar multiple action representations have been also found at the brain level. Indeed, neurophysiological and brain imaging studies have shown that the cortical motor system may encode actions such as grasping hand actions with different degrees of generality (Jeannerod 1988; Jeannerod et al 1995; Rizzolatti et al 2001; Nelissen et al 2005; Grafton and Hamilton 2007; Rizzolatti et al 2008), being this encoding also crucial for cognitive functions such as action recognition (Pellegrino et al 1992; Gallese et al 1996; Rizzolatti et al 1996); for a review see (Rizzolatti and Sinigaglia 2010). Thus, a natural question arises as to whether

and to what extent multiple levels of action generality can be represented by a synergy approach.

Hence, the aim of the present paper is to tackle this question. The question is not completely new, of course, and some answers have been already provided. Notably, several models of basic hand actions such as grasping actions rely on a two-hierarchical synergy organisation (Gorniak et al 2009; Latash et al 2007; Shim et al 2005; Zatsiorsky et al 2003). The first level of the hierarchy consists in the synergies between thumb and virtual finger, being the latter an imaginary finger which produces a mechanic effect (force and moment) equal to the sum of the mechanic effects produced by all the fingers (excluding the thumb) (Iberall and Fagg 1996; Arbib et al 1985; Shim et al 2003). The second level is composed by the synergies of the individual fingers forming the virtual finger.

No doubt a hierarchical synergy organisation of this kind might provide a useful way of representing action. However, we investigated here whether and to what extent a more general hierarchical representation without any virtual finger assumption would be able to fulfil the need of a multiple action representation as it has been characterised by action theorists (Goldman 1970; Mele 1997) and by cognitive neuroscientists (Jeannerod 1988; Jeannerod et al 1995; Rizzolatti et al 2001; Nelissen et al 2005; Grafton and Hamilton 2007; Rizzolatti et al 2008). To this purpose, we took advantage of representing hand action as a linear combination of temporal postural synergies (TPSs) involving all the hand's digits, but provided that TPSs assume a specific organisation. More precisely, we proposed that the hand action representation has to fulfil a tree constraint such as the following:

*Tree-constraint* The set of TPSs  $S = \{\mathbf{V}^i\}_{i=1}^r$  is organized as a rooted-tree  $T$  composed of  $r$  nodes, one for each TPS  $\mathbf{V}^i$ , such that each action representation can involve a synergy  $\mathbf{V}^j$  (i.e., the corresponding coefficient in the linear combination is different from zero) only if the ancestors of  $\mathbf{V}^j$  in  $T$  belong themselves to the action representation, as an example see Figure (2).

Accordingly, an hand action can involve a TPS only if the ancestors of the synergy in the tree  $T$  are themselves involved in the action representation. In the present paper we show that this constrain is enough for forcing a multiple representation of hand actions in terms of synergies which are hierarchically organised. More specifically we expected that: 1) any tree level is related to a certain degree of generality of the action representation (**multiple representation**); 2) the nearer a tree level is to the root, the higher is the degree of generality of its action representation; synergies at higher tree levels will

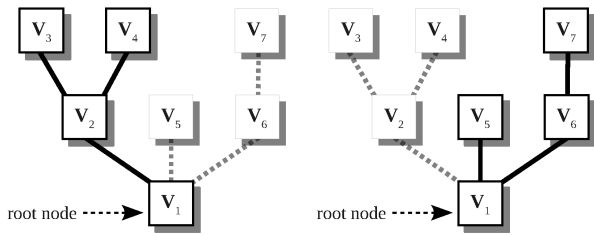


Fig. 2: *Tree-structured synergies*. The synergies  $S = \{\mathbf{V}^i\}_{i=1}^r$  are organized in accordance with a rooted-tree  $T$  composed of  $r$  nodes, one for each synergy  $\mathbf{V}^i$ . Different actions can be represented using different  $T$ 's subtrees. In (a) and (b) an example is reported of how two different actions can be represented using two different sub-trees of the same rooted-tree composed of seven nodes. In (a) the action representation uses the synergies  $\mathbf{V}^1$ ,  $\mathbf{V}^2$ ,  $\mathbf{V}^3$  and  $\mathbf{V}^4$ , in (b) the synergies  $\mathbf{V}^1$ ,  $\mathbf{V}^5$ ,  $\mathbf{V}^6$  and  $\mathbf{V}^7$ . Continuous edges indicate the synergies used in the action representation

capture common aspect shared between many actions, whereas synergies at lower tree's levels will capture distinctive features shared between few actions (**hierarchical representation**); 3) as it is usual in a multiple and hierarchical representation, different subtrees will identify different types of hand actions (**multiple and hierarchical**).

To compute TPSs with the tree-structured organisation, i.e., TPSs satisfying the above introduced tree-constraint, we used an unsupervised dimensionality reduction method proposed by (Jenatton et al 2010). Here we call this method *Tree-Structured Synergies Method (TSSM)*, and the action representation obtained in this way is called *tree-based action representation*. TSSM is framed within the larger class of *sparse coding* problems (see, for example, Aharon et al 2006; Engan et al 1999) where any element is represented as a linear combination of atoms in which a small set of coefficients are different from zero. Consequently, any actions represented by a tree-based action representation has a sparse representation, involving a small subset of a pre-specified set of TPS.

To assess whether and to what extent our tree-based action representation was able to provide us with a hierarchical and multiple representation of action, we collected three data sets of eight different types of grasping actions, processing them by means of five different analyses: 1) *Tree-structured TPSs*. The first analysis aimed to compute tree-structured TPSs and to show that this kind of synergies organisation was more robust to kinematic noise than a standard synergies approach devoid of any specific organisation. 2) *Action type representa-*

*tion in a tree-structured synergy organisation*. Given the tree-structure TPSs found in the previous analysis, we highlighted which subtrees are mainly involved in the representation of the different action types. 3) *Action representation error*. Here we measured to which degree the TPSs belonging to a given tree level contribute to represent actions. 4) *Shared and selective TPSs*. We analysed the presence of TPSs which were involved in reconstructing a large number of actions of different kind (which we shall call *action type-shared* TPSs) and the TPSs which were used in various actions of the same kind (which we shall call *action type-selective* TPSs). 5) *Action type similarity*. In the last analysis we considered the question as to whether and to what extent action type-shared synergies enable us to distinguish between different types of actions.

There were three main findings. Firstly, a tree-based action representation resulted to be more robust than a standard synergies approach. Secondly, different types of grasping actions were represented by different subtrees of a pre-established tree. Thirdly, we showed that a tree-based action representation allowed us to distinguish between *action type-shared* TPSs and *action type-selective* TPSs, the former being located at the higher levels of the tree-structure and capturing common action features, whereas the latter are located at the lower ones and hold distinctive action features. We therefore concluded that a tree-structured synergy organisation allows a multiple representation of action whose levels are hierarchically organised in terms of different synergies.

## 2 Methods

### 2.1 Experimental task and procedure

Subjects were instructed to reach, grasp and hold different objects several times. They were seated at a table with two clearly visible surface marks ( $m_1$  and  $m_2$ ) placed at a distance of roughly 40 cm from each other. For each target object, the subject was asked to put the right hand on the mark  $m_1$  in a prone position. Then, the subject had to reach and grasp the target object which was placed on the mark  $m_2$  or an experimenter handed it to the subject close to the mark  $m_2$ . A computer generated beep was used to signal the start and the end of the action. Moreover subjects were asked to perform the grasping actions rapidly and to hold the object after grasping until they hear the stop signal. Objects with different size and shape were used according to the type of grasp used.

Eight different grasping action types were selected from the classification made in (Feix et al 2009) (see









Grasp name	Object	Grasp Picture
Prismatic-2-Finger	pen	
Palmar-Pinch	pendrive cup	
Tripod	ping-pong ball	
Writing-Tripod	marking pen	
Power-Sphere	tennis ball	
Extension-Type	compact-disk	
Sphere-4-Finger	tennis ball	
Sphere-3-Finger	tennis ball	

Table 1: *Grasping action types*. The table shows the eight grasping action types used in our experiments as defined in (Feix et al 2009).

Table 1), and each action type was executed by all the subjects. We chose these types of actions insofar as they represent a large enough repertoire of grasping actions, and, at the same time, some of them exhibit kinematic similarity such as tripod and writing-tripod or sphere-3-finger and sphere-4-finger.

For each type of grasping actions a total of 50 actions were recorded. Thus for each subject a total of 400 actions were recorded.

A HumanGlove (Humanware S.r.l., Pontedera (Pisa), Italy) endowed with 16 sensors (see Figure 3) to record joint angles at a maximum frequency of  $100Hz$  and with a  $12bit$  resolution was used. Wrist related sensors were not considered for this work whereas 10 hand related sensors are considered according to (Vinjamuri et al 2010a). In particular sensors which measure angles of the carpometacarpal (CMC) and metacarpophalangeal (MCP) joints of the thumb and the metacarpophalangeal (MCP) and proximal interphalangeal (PIP) joints of the other four fingers were considered, for a total of  $d = 10$  sensors. We truncated all actions in order to preserve only their relevant parts where the hand was actually moving. We then resampled each action in order to have the same length  $s$ . Note that we worked with raw-data normalized by linear mapping each sensor value in the range  $[-1, \dots, 1]$ . Thus, a grasping action can be expressed as a  $p$ -dimensional vector  $\mathbf{x} \in \mathbb{R}^p$  composed of



Fig. 3: *DataGlove*. HumanGlove (Humanware S.r.l., Pontedera (Pisa), Italy) endowed with 16 sensors was used to record all the grasping actions.

a sequence of  $s$  hand-joint configurations  $\mathbf{hc} \in \mathbb{R}^d$ , i.e.,  $\mathbf{x} = [\mathbf{hc}(1), \mathbf{hc}(2), \dots, \mathbf{hc}(s)]$  with  $p = s \times d$ .

Three right-handed male subjects (age ranging from 24 to 30 years and without neurological disorders) took part in the experiments. All these subjects were informed about the nature of the study and signed institutionally approved consent forms.

Summarizing, three datasets  $DS_1$ ,  $DS_2$ , and  $DS_3$ , one for each subject, were constructed. Each dataset contains a total of 400 grasping actions, 50 actions for each one of the 8 action types previously described. The actions belonging to the three datasets  $DS_1$ ,  $DS_2$ , and  $DS_3$  are represented as sequence of  $s = 28, 30, 33$  hand-joint configurations, respectively.

## 2.2 Tree-Structured Synergies Computation

In order to find an action organisation in terms of TPSs which fulfils our tree-constraint, we followed the method proposed in (Jenatton et al 2010) which we called here TSSM. Let us define a matrix  $\mathbf{X} \in \mathbb{R}^{n \times p}$  of  $n$  rows in  $\mathbb{R}^p$ , each one corresponding to an experimental observation of a grasping action expressed as a sequence of  $s$  hand-joint configurations with  $p = s \times d$ . The problem we addressed can be solved by finding a matrix  $\mathbf{V} \in \mathbb{R}^{p \times r}$  such that each row of  $\mathbf{X}$  can be approximated by a linear combination of the  $r$  columns of  $\mathbf{V}$ , i.e.,  $\mathbf{X}_i = \sum_{j=1}^r u_{ij} \mathbf{V}^j$ .  $\mathbf{V}$  is the matrix of TPSs which

are disposed column-wise.  $\mathbf{V}$  is also called *dictionary* in machine learning context. Let us call  $\mathbf{U} \in \mathbb{R}^{n \times r}$  the matrix of the linear combination coefficients, i.e., the  $i$ -th row of  $\mathbf{U}$  corresponds to the  $r$  coefficients of the linear combination of the  $r$  columns of  $\mathbf{V}$  in order to approximate the  $i$ -th row of  $\mathbf{X}$ . Consequently,  $\mathbf{UV}^T$  is an approximation of  $\mathbf{X}$ . Following (Jenatton et al 2010) the problem can be formulated as the following minimization problem:



$$\min_{\mathbf{U}, \mathbf{V}} \frac{1}{2np} \|\mathbf{X} - \mathbf{UV}^T\|_F^2 + \lambda \sum_{i=1}^n \sum_{j=1}^r \|\mathbf{D}_j \circ \mathbf{U}_i\| \quad (1)$$

*s.t.*  $\forall j \|\mathbf{V}^j\|_2 \leq 1$

where  $\mathbf{U}_i$  is the  $i$ -th row of  $\mathbf{U}$ ,  $\mathbf{V}^j$  is the  $j$ -column of  $\mathbf{V}$ , and the matrix  $\mathbf{D} \in \mathbb{R}^{r \times p}$ , encoding the tree  $T$ , is defined so that  $d_{ij}$  is equal to 1 if the  $j$ -th node in  $T$  is a descendant of the node  $i$ , and 0 otherwise. The vector  $\mathbf{D}_j \circ \mathbf{U}_i$  is the element-wise product of  $\mathbf{D}_j$  and  $\mathbf{U}_i$ . The first term in equation (1) is the usual reconstruction error, the second term  $\Omega(\mathbf{U}_i) = \sum_{j=1}^r \|\mathbf{D}_j \circ \mathbf{U}_i\|$  is a penalization term which favours solutions of (1) such that the tree-constraint is fulfilled. The parameter  $\lambda \geq 0$  controls to which extension the penalization term is used. The more  $\lambda$  increases, the more one obtains a sparse representation of the actions, i.e., the action is represented using a subset of the synergy set.

In order to solve the problem (1), we followed the usual approach of finding the minimum by alternating optimizations with respect to the coefficients  $\mathbf{U}$  and to the dictionary  $\mathbf{V}$ . Most methods are based on this alternating scheme of optimization (Basso et al 2011). Therefore the algorithm used here is composed of two alternate stages: 1) *Tree-Structured Coding Stage*. In this stage the dictionary  $\mathbf{V}$  is fixed and the matrix  $\mathbf{U}$  is updated. 2) *Synergy Dictionary Stage*. Here the matrix  $\mathbf{V}$  is updated while keeping the  $\mathbf{U}$ 's values fixed. For more details see Appendix (Section 5).

It is worth to note that different choices of the tree  $T$  and the sparsity parameter  $\lambda$  induce different solutions of the minimization problem. This dependence implies the need for a careful choice of  $T$  and  $\lambda$ .

## 2.3 Data Analysis

In the following five data analyses, all the actions recorded from the tree subjects were considered, i.e., we used all the three datasets  $DS_1$ ,  $DS_2$  and  $DS_3$ , each one composed of 400 grasping actions, 50 actions for each one of the 8 action types as reported in Section 2.1. Note that in the analysis described in Subsection 2.3.1 each dataset  $DS_i$ , has been split into two subsets, each one composed of 200 actions obtained choosing in a random way 25 actions for each one of the 8 action types. For each  $DS_i$ , the first subset obtained was used as training set for a linear classifier, and we will call this subset  $TS_i$ , whereas from the second subset we built three corresponding noisy test sets  $NT_i^1$ ,  $NT_i^2$  and  $NT_i^3$  adding to the recorded actions a zero-mean Gaussian noise with  $\sigma = 0.2$ ,  $\sigma = 0.4$  and  $\sigma = 0.6$ , respectively. Thus, these datasets are composed of grasping actions noised by "small" kinematic variations.

### 2.3.1 Tree-structured synergies

In this analysis we computed tree-structured TPSs and tested the representational capacity of this kind of synergies comparing them with TPSs computed by a standard approach such as the Principal Component Analysis (PCA) which denies any specific synergies organisation (Hastie et al 2003). In particular we compared the tolerance to kinematic noise of the action representation obtained by these two different approaches. To this aim we tested, for each kind of action representation, the ability of a classifier to discriminate between the eight action types, described in Table 1, when the classifier is fed with actions corrupted with kinematic noise. PCA, or the equivalent method of singular value decomposition, has been widely used in computing hand action synergies (see Tessitore et al 2010; Vinjamuri et al 2010a,b; Mason et al 2001; Jerde et al 2003; Santello et al 2002, 1998). As it is known, the output of PCA is a set of ordered orthogonal vectors, called *Principal Components (PCs)*, which enables one to express input data as a linear combination of these *PCs*. Thus, PCA allows one to express actions as linear combination of TPSs which correspond to the first *PCs*.

As the tree-based action representations depend on the choice of the tree  $T$  which defines the organisation of the TPSs we considered three different trees  $T_1$ ,  $T_2$  and  $T_3$  as showed in Table 2. These trees differed for the number of both levels and nodes.  $T_1$ ,  $T_2$  and  $T_3$  have a number of levels (nodes) equal to 2 (5), 3 (13) and 4 (29), respectively. We chose these three rooted-trees insofar as they allowed to investigate the soundness of a tree-based action representation on different degrees of detail. To obtain a PCA-based action representation comparable with tree-based action representations, the number of principal components to be selected was chosen varying in the set  $\{5, 13, 29\}$ , i.e., the number of nodes of each of the three trees  $T_1$ ,  $T_2$  and  $T_3$ . It is worth to note that in a PCA-based action representation any actions is typically represented using all the selected TPSs.

For each training set  $TS_i$ , with  $i = 1, 2, 3$ , we solved the minimization problem as expressed in (1) by TSSM using in turn each of the three trees  $T_1$ ,  $T_2$  and  $T_3$ , and a regularization parameter  $\lambda$  (see eq. 1) ranging in  $[0.01, 0.1]$  at step 0.005. The regularization parameter  $\lambda$  was chosen so that to obtain an high sparsity value<sup>1</sup>. This choice enabled us to obtain both compact

<sup>1</sup> We consider high sparsity a mean sparsity value roughly equal to 30% of the total number of TPSs of each tree. More specifically the mean sparsity of the tree-based action representations belonging to each training set was computed as  $1 - \frac{1}{pn} \sum_{j=1}^n \|\mathbf{U}_j\|_0$ , where  $\mathbf{U}_j$  are the coefficients of the tree-based action representation for  $j$ -th action of the training set. Note that the mean

and meaningful action representations. Thus, for each training set  $TS_i$  we found three different sets of TPSs,  $S_{T_1}^i$ ,  $S_{T_2}^i$  and  $S_{T_3}^i$ .

Each action belonging to the training set  $TS_i$  was represented by three different tree-based action representations, one for each tree. Similarly, for each action belonging to the training set  $TS_i$  we obtained three different PCA-based action representations, one for each choice of the maximum number of principal components ( $\{5, 13, 29\}$ ).

A regularized linear regression model was used as multi-class classifier (Hastie et al 2003). The regularization parameter for the linear regression model  $\lambda_{rm}$  was varied in the range  $\log_{10}(\lambda_{rm}) \in [-22, -21, \dots, 1]$ .

Hence, by feeding the multiclass classifier with the coefficients of the obtained action representations, a 5-fold cross-validation method was used to choose the classifier regularization parameter  $\lambda_{rm}$ . Thus, at the end of the training phase, the TPSs computed by TSSM and PCA, and the classifier’s parameters were determined.

In the test phase the actions belonging to the noisy test sets  $NT_i^1$ ,  $NT_i^2$  and  $NT_i^3$  were used. Here, while leaving unchanged the TPSs computed previously in the training phase, for each action belonging to each  $NT_i^j$  we computed, similarly to the training phase, three different tree-based action representations and three PCA-based action representation. The classifier as determined in the training phase was fed, for each action belonging to each noisy test set, with the corresponding two tree-based action representations and the corresponding PCA-based action representations. Finally, the performance of the multi-class classifier was measured by *classification accuracy* which is defined as the ratio between correctly classified actions over the total number of actions. Thus, for each noisy test set, we obtained: 1) for the tree-based action representations, three classification accuracy values corresponding to the three different trees, 2) for the PCA-based action representations, three classification accuracy values.

### 2.3.2 Action type representation in a tree-structured synergy organisation

Given the sets of TPSs  $S_T^i$  previously found, we analyzed which subsets of TPSs in  $T$  are mainly used in the representation of the different 8 action types. Thus, given a set  $S_T^i$  we proceeded according to the following two consecutive steps: first, for each synergy  $\mathbf{V}^k \in S_T^i$

value of the sparsity multiplied by the number of synergies gives the mean number of synergies used to represent each action. For the corresponding  $\lambda$  we found that the reconstruction error  $\frac{1}{2np} \|\mathbf{X} - \mathbf{UV}^T\|_F^2$  was always lower than  $4 \times 10^{-3}$ .

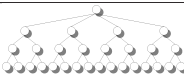


	# levels	split at each level	total nodes
$T_3$		4, 2, 2	29
$T_2$		4, 2	13
$T_1$		4	5

Table 2: *Tree-structured synergies*. The Tree-Structured Synergy Method (TSSM) has been applied using 3 different rooted-trees  $T_1$ ,  $T_2$  and  $T_3$  which have heights 1, 2 and 3, respectively. The tree  $T_1$  is composed of 5 nodes with 4 nodes at the level 1, the tree  $T_2$  is composed of 13 nodes with 4 nodes at the level 1 and 8 nodes at the level 2, and the tree  $T_3$  is composed of 29 nodes with 4, 8 and 16 nodes at the levels 1, 2 and 3, respectively. To each tree a number of TPSs is associated, which is equal to the number of tree nodes, one for each node.

we compute the percentage of actions belonging to the  $h$ -th action type  $A_h$  which uses the  $k$ -th synergy as

$$a_k^h = usage(\mathbf{V}^k, A_h) = \frac{1}{card(h)} \sum_{j \in A_i} \|u_{jk}\|_0 \quad (2)$$

where  $card(h)$  is the number of actions belonging to the  $h$ -th type of actions; second, for each action type  $A_h$ , with  $h = 1, 2, 3, \dots, 8$ , we weighted the  $T$  edge going from the synergy  $l$  to the synergy  $m$  with the value  $a_m^h$ . Thus, we obtained, for each rooted-tree  $T$ , 8 different weighted trees, one for each action type. Each weighted tree gives a representation of an action type in terms of synergy usage. Recall that by construction if a node  $m$  is involved in the representation of an action, then all the ancestor nodes of  $m$  are also involved in the representation. In order to obtain a multiple and hierarchical action representation we expected that each action type used a distinct subtree involving almost all the levels of the rooted-tree  $T$ .

### 2.3.3 Action representation error

Using the sets of TPSs  $S_T^i$ , which were computed as described in the previous subsection, we obtained tree-based action representations of the actions belonging to the datasets  $DS_1$ ,  $DS_2$  and  $DS_3$ , i.e, all the actions collected by the three subjects involved in the present experiment. Afterwards, on the basis of these representations we measured how the TPSs belonging to a given level of the rooted-tree  $T$  contribute to representing the

collected grasping actions. More specifically, for each set of TPSs  $S_T^i = \{\mathbf{V}^h\}_{h=1}^r$  and each action belonging to the corresponding dataset  $DS_i$ , we computed the reconstruction error between the recorded actions and the reconstructed actions using only the TPSs belonging to the first  $j$  levels of the the rooted-tree  $T$ . Given a dataset  $DS_i$  and a rooted-tree  $T$  this reconstruction error is defined as  $err_{i,T}^j = \frac{1}{2np} \|\mathbf{X} - \mathbf{U}(\mathbf{V}_{|L_j})^T\|_F^2$ , where  $L_j$  is the set of node indexes belonging to the first  $j$  levels of the tree  $T$ , and  $\mathbf{V}_{|L_j}$  is the matrix composed of the  $\mathbf{V}$ 's columns belonging to  $L_j$  only. The index  $j$  runs in the range from 0 to  $p_T$ , where  $p_T$  is the maximum number of levels in  $T$ .

### 2.3.4 Shared and selective TPSs

Using the sets of TPSs  $S_T^i$  computed as described in subsection 2.3.1, we measured to what extension a TPS is either shared between different kinds of actions or used in just one specific action type. To this aim, we introduced two measures: *commonality* and *selectivity*. A TPS with a high commonality is strongly used in more than one type of action. A TPS with a high selectivity is mainly used in just one type of actions. The commonality of a given TPS  $\mathbf{V}^k$  is defined as:

$$commonality(\mathbf{V}^k) = \frac{M_{\mathbf{V}^k}}{1 + S_{\mathbf{V}^k}} \quad (3)$$

where  $M_{\mathbf{V}^k}$  and  $S_{\mathbf{V}^k}$  are the mean and standard deviation of the usage values for the  $k$ -th synergy over all action types.

Note that for any given  $\mathbf{V}^k$ ,  $0 \leq commonality(\mathbf{V}^k) \leq 1$ . A value near to 1 means that the corresponding synergy is widely used by almost all the action types.

We now turn to the selectivity property which is defined as follows:

$$selectivity(\mathbf{V}^k) = \max_i usage(\mathbf{V}^k, A_i) - \frac{1}{C-1} \sum_{j \neq i_k} usage(\mathbf{V}^k, A_j) \quad (4)$$

where  $i_k$  is the index of the action type for which  $usage(\mathbf{V}^k, A_i)$  assumes the maximum value. Note that also *selectivity* ( $\mathbf{V}^k$ ) lies between 0 and 1. The maximum value 1 is reached when a given synergy is used by all the actions belonging to just one type of actions.

Once computed these measures, as the sets of TPSs  $S_T^i$  that we are using have a tree-structured organisation  $T$ , we computed the mean values of commonality and selectivity for each level of the trees  $T$ .

### 2.3.5 Action type similarity

For each set of TPSs  $S_T^i$ , we grouped together the types of actions on the basis of the usage of the synergies belonging to the first level of the rooted-tree  $T$ . We chose the first level of the tree as we expected that on this first level there were synergies which had a degree of commonality (see subsection (2.3.4)) higher than in other  $T$ 's levels. More in detail, given the set of TPSs  $S_T^i$  and the corresponding 8 weighted trees, one for each action type, which were computed as described in the previous subsection (2.3.2), we grouped together the types of actions which have the same most used synergy belonging to the first level of the corresponding weighted tree.

In order to assess the resulting groups we introduced the notion of *action-type temporal profile*. For each action  $\mathbf{x}(t)$  belonging to one of the 8 types of grasping actions, at a given time  $t$  we computed the *cosine* between the vector  $\mathbf{x}(t)$  and the  $p$ -dimensional unit vector  $\mathbf{1}_p$ . At each time  $t$ , the mean and standard deviation of these values over all actions of the same action type were computed. Hence, we obtained 8 different *action-type temporal profiles* which enable us to assess the similarity between the 8 types of grasping actions from a kinematic point of view.

## 2.4 Results

### 2.4.1 Finding tree-structured synergies

In Table 3, for each one of the noisy test sets  $NT_1^1$  ( $\sigma = 0.2$ ),  $NT_2^1$  ( $\sigma = 0.4$ ) and  $NT_3^1$  ( $\sigma = 0.6$ ), the three classification accuracy values, for the tree-based action representations, and the three classification accuracy values, for the PCA-based action representations, are shown. Classification accuracies for the tree-based action representations corresponding to trees  $T_3$  and  $T_2$  were always higher than for PCA-based action representations, but in one case (see Subject 2, noise  $\sigma = 0.2$  in Table 3). Moreover this difference is much more evident in case of higher noise, see second row ( $\sigma = 0.4$ ) and third row ( $\sigma = 0.6$ ) in Table 3. Conversely, when the tree-based action representation involves TPSs organized using the tree  $T_1$  the classification accuracy turns out to be always very low, therefore these TPSs were discarded for further analyses.

Accordingly, only the two different sets of TPSs,  $S_{T_2}^i$  and  $S_{T_3}^i$ , corresponding to the two rooted-tree  $T_2$  and  $T_3$  were considered. In Figure 4 and 5 the sets of TPSs  $S_{T_2}^1$  and  $S_{T_3}^1$  are shown. Note that the synergies become more smooth (in the sense of having fewer peaks) as the synergies come closer to the root of the tree.

Noise		Subject 1			Subject 2			Subject 3		
		$T_3$	$T_2$	$T_1$	$T_3$	$T_2$	$T_1$	$T_3$	$T_2$	$T_1$
$\sigma = 0.2$	Tree-based	<b>0.98</b>	<b>0.98</b>	0.58	<b>0.96</b>	0.92	0.56	<b>0.88</b>	<b>0.84</b>	0.63
	PCA-based	0.94	0.97	<b>0.84</b>	0.94	<b>0.94</b>	<b>0.80</b>	0.74	0.80	<b>0.83</b>
$\sigma = 0.4$	Tree-based	<b>0.92</b>	<b>0.94</b>	0.52	<b>0.85</b>	<b>0.89</b>	0.52	<b>0.69</b>	<b>0.75</b>	0.57
	PCA-based	0.80	0.87	<b>0.80</b>	0.72	0.84	<b>0.82</b>	0.41	0.61	<b>0.73</b>
$\sigma = 0.6$	Tree-based	<b>0.69</b>	<b>0.81</b>	0.51	<b>0.72</b>	<b>0.82</b>	0.50	<b>0.48</b>	<b>0.58</b>	0.53
	PCA-based	0.59	0.67	<b>0.68</b>	0.57	0.73	<b>0.67</b>	0.29	0.51	<b>0.57</b>

Table 3: *Action classification accuracy*. The table shows the accuracy of a linear classification when a PCA-based action representation or a tree-based action representation with a sparseness less than 0.3 is used. The classification task consists in classifying grasping actions belonging to 8 different action types. Classification accuracies are reported for different noise levels ( $\sigma$ ). For more details see text.

We used these TPS sets  $S_{T_2}^i$  and  $S_{T_3}^i$ , with  $i$  running on the three different subjects involved in the present experiment, in both the three analyses described in subsection 2.3.4, 2.3.2 and 2.3.5.

#### 2.4.2 Action type representation in a tree-structured synergy organisation

In figures 6 and 7 the use of the TPSs belonging to  $S_{T_2}^1$  and  $S_{T_3}^1$  (subject 1), respectively, for each one of the 8 different action types is shown. The thickness and the darkness of an edge going from the node  $l$  to  $m$  represents the usage of the  $m$ -th synergy as described in Subsection (2.3.2). It is worth noting that each action type uses a specific sub-tree including all the levels of the tree.

#### 2.4.3 Action representation error

Given the TPS's sets  $S_{T_2}^i$  and  $S_{T_3}^i$  selected in Section 2.4.1, with  $i = 1, 2, 3$ , for each dataset  $DS_i$  we computed the reconstruction errors  $err_{i,T_2}^j$ , with  $j = 0, 1, 2$ , and the reconstruction errors  $err_{i,T_3}^j$ , with  $j = 0, 1, 2, 3$ . The Table 4 shows these values for the three dataset  $DS_i$  (all the subjects). Note that the reconstruction error decreases when a tree level is added to the representation of action. The error decreasing assumes the maximum value when one adds to the representation of action the TPSs belonging to the first level of the tree only. On the contrary, the error decreasing assumes lower values when TPSs belonging to lower tree levels, that is, tree levels which are farther from their corresponding root, are included.

#### 2.4.4 Shared and selective synergies

On the basis of the selected TPS's sets  $S_{T_1}^i$  and  $S_{T_2}^i$ , for each subject, we computed the selectivity and commonality mean values for each level of the trees  $T_1$  and  $T_2$ . The top-left graphs in Figure 8 and 9 show these mean

values for the dataset  $DS_1$  (subject 1). The synergy associated to the root of the tree is not included in the analysis as it is always used for construction. The remaining graphs in Figure 8 and 9 show the usage of the synergies for each level (or depth) of the tree. In particular, for each tree depth the percentage of the synergy usage for each action type is reported. It is worth noting that for both trees the commonality (selectivity) mean values have a decreasing (increasing) trend over tree levels. A similar behaviour is obtained also for the other two subjects as showed in Figure 10.

#### 2.4.5 Action type similarity

On the basis of the analysis described in SubSection (2.3.5), in Table 5 and 6 the action type groups obtained considering the usage of the set of TPSs  $S_{T_1}^1$  and tree  $S_{T_2}^1$  (subject 1) are shown. Note that very similar results are obtained for both the sets of TPSs. In particular in both cases the *Palmar-pinch* type is put alone into a specific group, while the two precision grasps *Tripod* and *Writing-Tripod*, the two whole-hand grasping types *Power-Sphere* and *Sphere-4-Finger*, and the action types *Sphere-3-Finger* and *Prismatic-3-Finger* are grouped together.

In order to emphasize the similarity and the differences between each action type we have reported the action-type temporal profile for each one of the 8 action types in Figure 11.

It is worth noting that action types belonging to the same group, as shown in Table 5 and 6, have a very similar action-type temporal profile.

Very similar results were obtained also for other two subjects.

### 3 Discussion

The main aim of the present study was to assess whether and to what extent a tree-structured synergies organisation might provide a multiple and hierarchical repre-



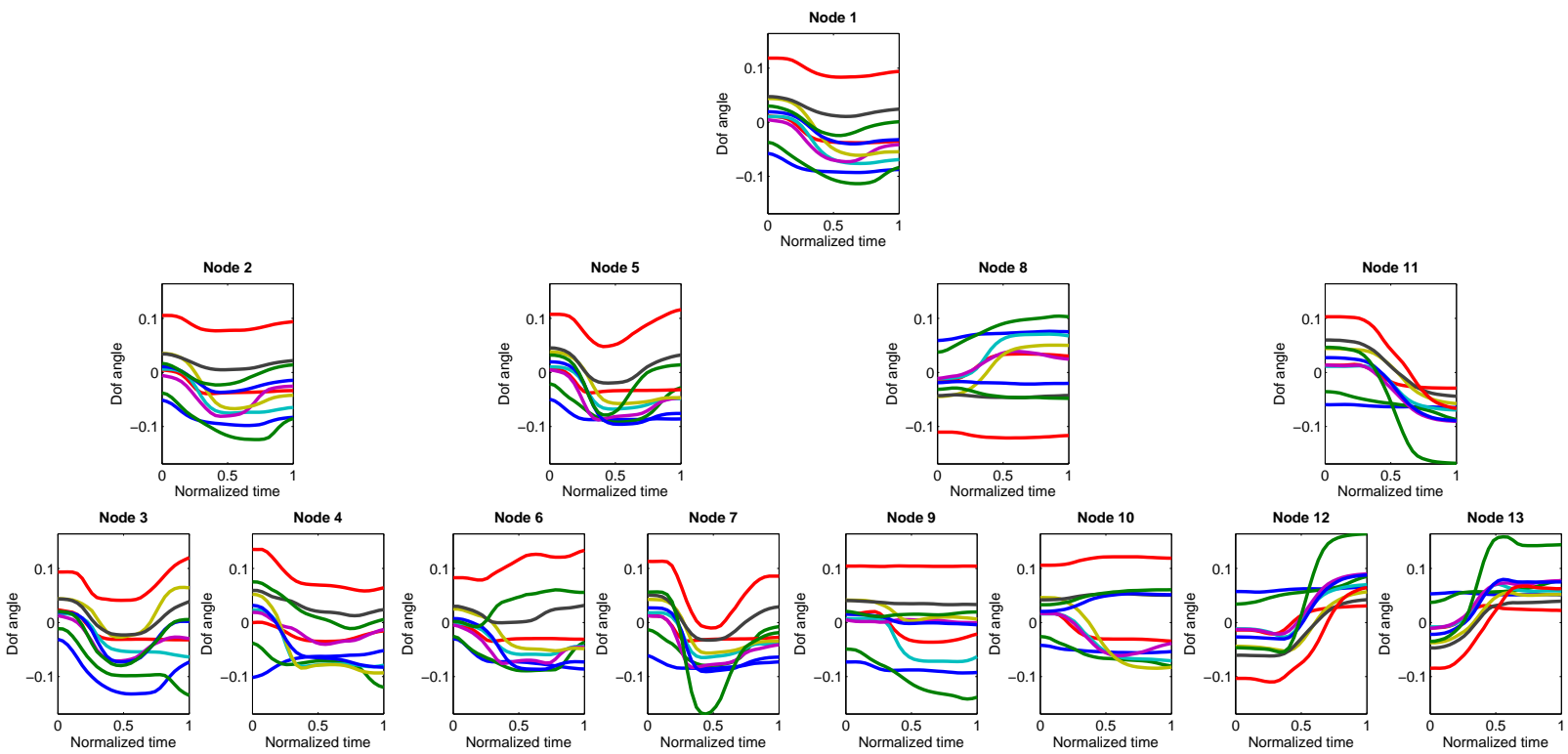
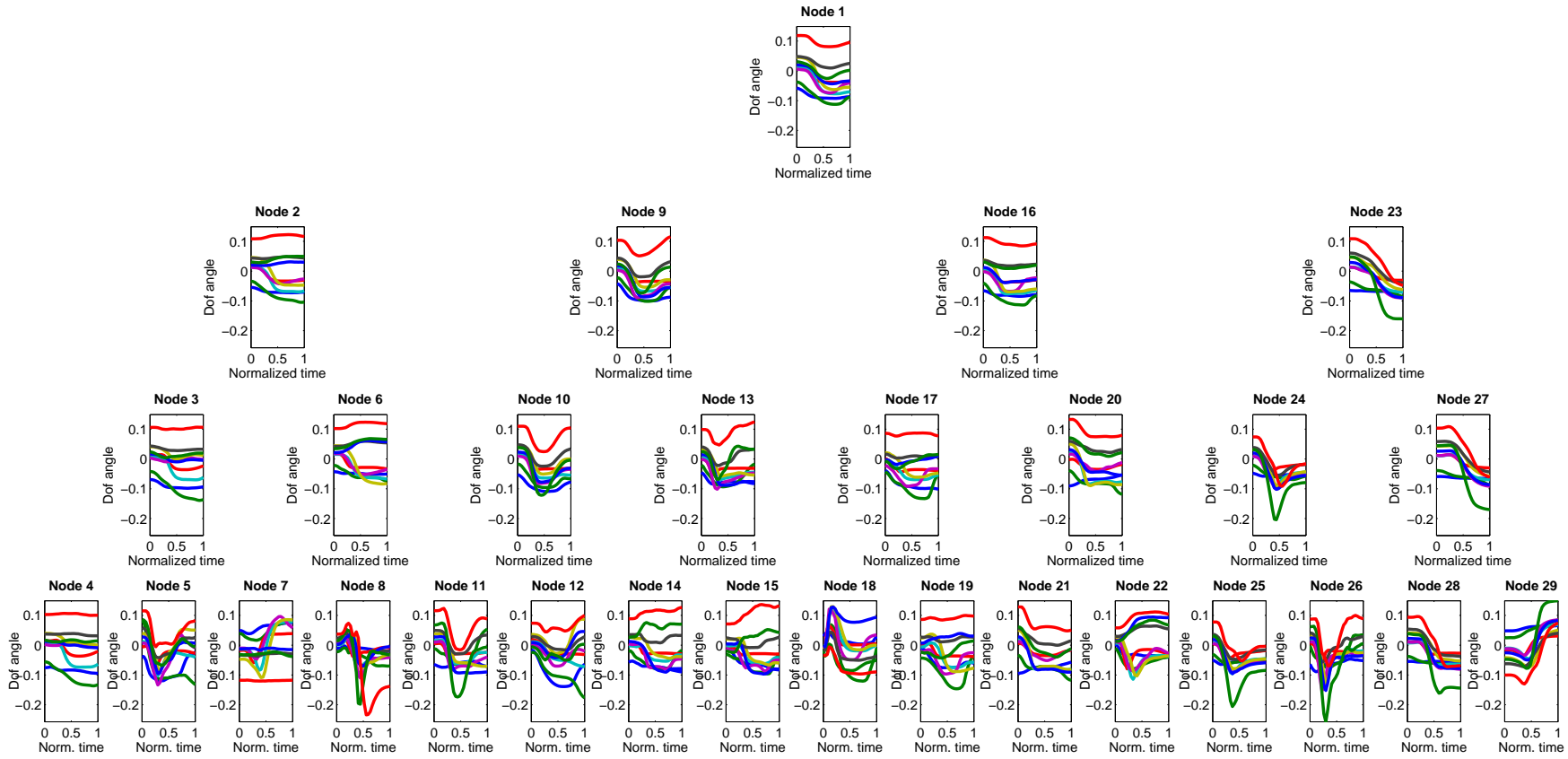


Fig. 4: TPPs computed by TSSM using the rooted-tree  $T_2$  (see text).

Fig. 5: TPPs computed by TSSM using the rooted-tree  $T_3$  (see text)



		Level 0	Level 1	Level 2	Level 3
<b>Subject 1</b>	$T_2$	0.051	0.018 (69%)	0.003 (31%)	
	$T_3$	0.059	0.026 (57%)	0.008 (31%)	0.001 (12%)
<b>Subject 2</b>	$T_2$	0.055	0.019 (72%)	0.005 (28%)	
	$T_3$	0.067	0.032 (54%)	0.011 (32%)	0.002 (14%)
<b>Subject 3</b>	$T_2$	0.046	0.015 (72%)	0.003 (28%)	
	$T_3$	0.044	0.018 (60%)	0.006 (28%)	0.001 (12%)

Table 4: *Reconstruction errors for levels.* We defined  $err^j(t)$  (see text) as the reconstruction error between the recorded actions and the reconstructed actions using only the TPSs belonging to the first  $j$  levels of a tree-structured TPS organisation. The table shows such reconstruction error for all the dataset  $DS_i$  and using the trees  $T_2$  and  $T_3$ . The contribution (in percent) of each tree level to the reconstruction error is reported in brackets.









Most used synergy belonging to level 1 of $T_2$			
	Prismatic 2 Finger	Extension Type	Sphere 3 Finger
2			
11	Palmar Pinch 		
8	Tripod 	Writing Tripod 	
5	Power Sphere 	Sphere 4 Finger 	

Table 5: *Grouping actions types by TPSs.* In a tree-based action representation using the tree  $T_2$  we have grouped all action types on the basis of the most used synergies belonging to the first level.









Most used synergy belonging to level 1 of $T_3$			
	Prismatic 2 Finger	Sphere 3 Finger	
16			
23	Palmar Pinch 		
2	Tripod 	Writing Tripod 	
9	Power Sphere 	Sphere 4 Finger 	Extension Type 

Table 6: *Grouping actions types by TPSs.* In a tree-based action representation using the tree  $T_3$  we have grouped all action types on the basis of the most used synergies belonging to the first level. The results are presented in the figure.

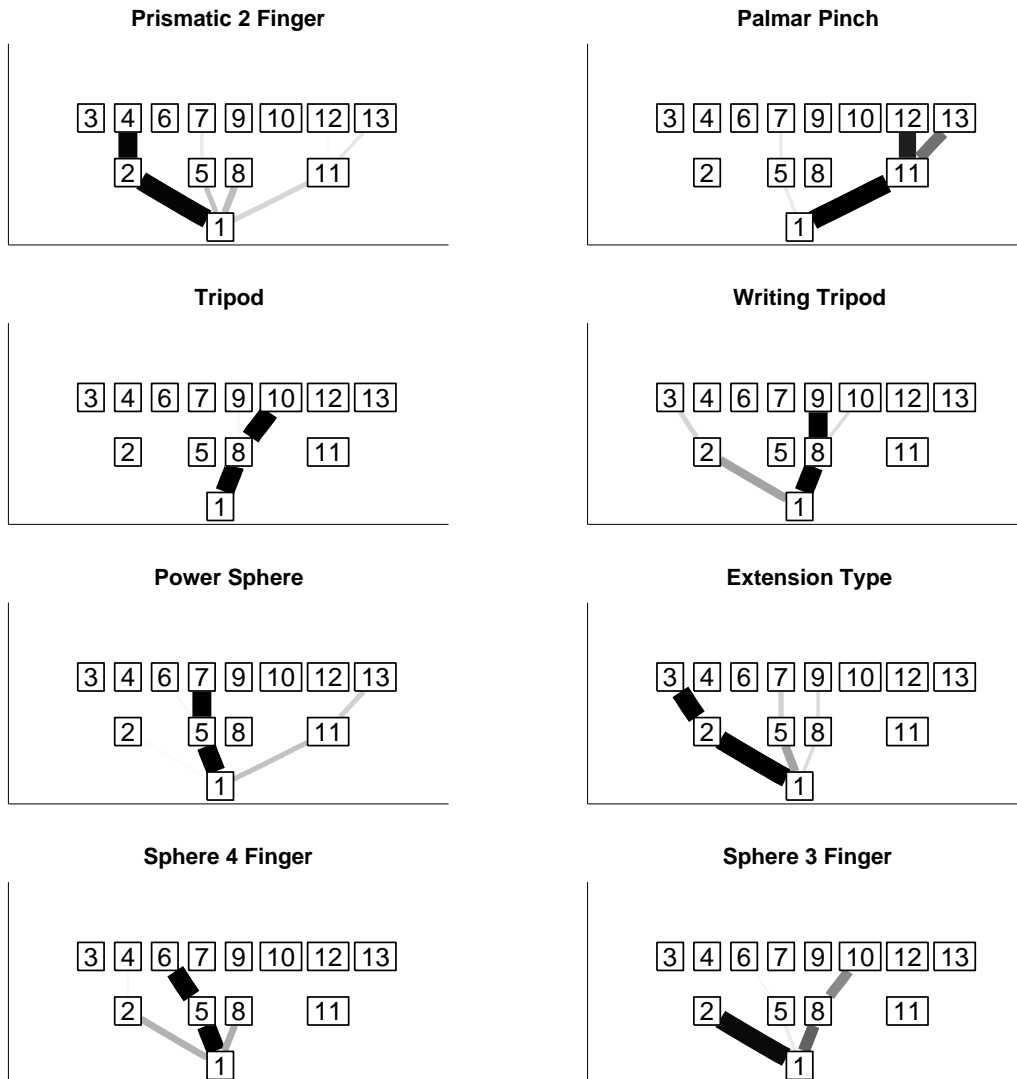


Fig. 6: *Synergy usage for the tree  $T_2$* . The figure shows the usage of the TPSs in a tree-based action representation for the eight action types of actions when the tree  $T_2$  is used. The numbered white squares organized in a tree refer to the computed TPSs. The gray level of an edge going from  $i$  to  $j$ , with  $j > i$ , represents the usage of the TPSs  $j$ . Black level indicates the maximum value. If the edge is absent the synergy is not used.

sensation of action. To this purpose, we processed three data sets of eight different types of grasping actions by means of five different analyses. In the first analysis we compared the representation of the different types of grasping actions based on our tree-structured TPS organisation computed by TSSM with that resulting from TPS computed by PCA and devoid of any specific organisation. The task consisted in classifying the

collected grasping actions noised by “small” kinematic variations. The results showed that tree-based action representations were more robust than PCA-based action representations with respect to the kinematic noise. In the second analysis, we highlighted that different subtrees involving a limited number of TPSs represent different action types. In the third analysis we measured to which degree the TPSs belonging to a given level of



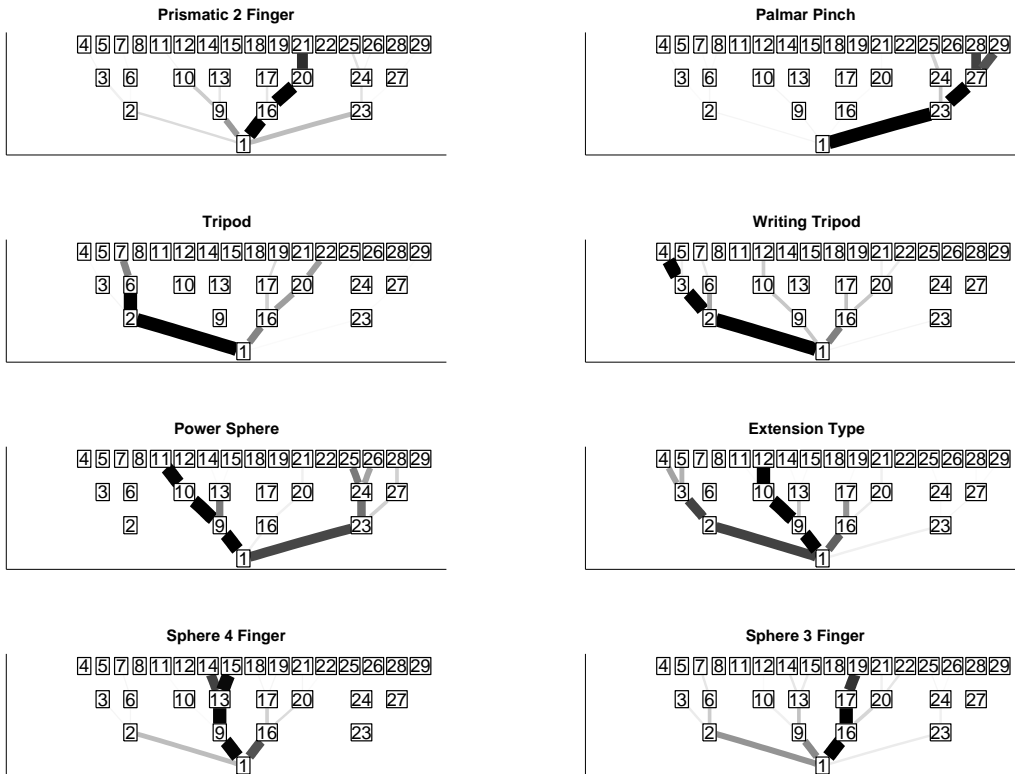


Fig. 7: *Synergy usage for the tree  $T_3$ .* The figure shows the usage of the synergies in a tree-based action representation for the eight types of actions when the tree  $T_3$  is used. The numbered white squares organized in a tree refer to the computed TPSs. The gray level of a edge going from  $i$  to  $j$ , with  $j > i$ , represents the usage of the TPS  $j$ . Black level indicates the maximum value. If the edge is absent the synergy is not used.

the rooted-tree contribute to represent grasping actions. We found that TPSs belonging to the higher levels of the tree-structure synergy organisation capture much of the action development, whereas TPSs belonging to the lower levels represent grasping action details. The presence of both synergies shared between many kinds of action, *action type-shared synergies*, and synergies selective to just one kind of action, *action type-selective synergies*, was tested in the fourth analysis. The result was that the former have been mainly located near the root of the tree whereas the latter near the leafs. This property reflects the hierarchical organisation of the synergies in tree-based action representation. Finally in the last analysis we showed that similar action types are represented by the same action type-shared synergies.

Taken together, these analyses suggest that basic actions such as grasping hand actions can be multiply and

hierarchically represented by means of TPSs, provided that the latter are organised in a suitable tree-structure.

As far as the multiple nature of action representation is concerned, Figure 6 and Figure 7 clearly show that, in our tree-based action representation, any grasping hand action was represented using almost all the levels of the rooted-tree, and grasping actions of the same type tend to use the same subset of TPSs. Interestingly, grasping actions of different type used different subsets of TPSs which could overlap. Consequently, the representation of the different grasping actions was distributed on a small subset of TPSs. In addition, we found that any TPS can be either selective for a specific action or for an action type or shared between different action types. Thus, grasping actions were represented with different degrees of generality by different TPSs. In other words, they were multiply represented.

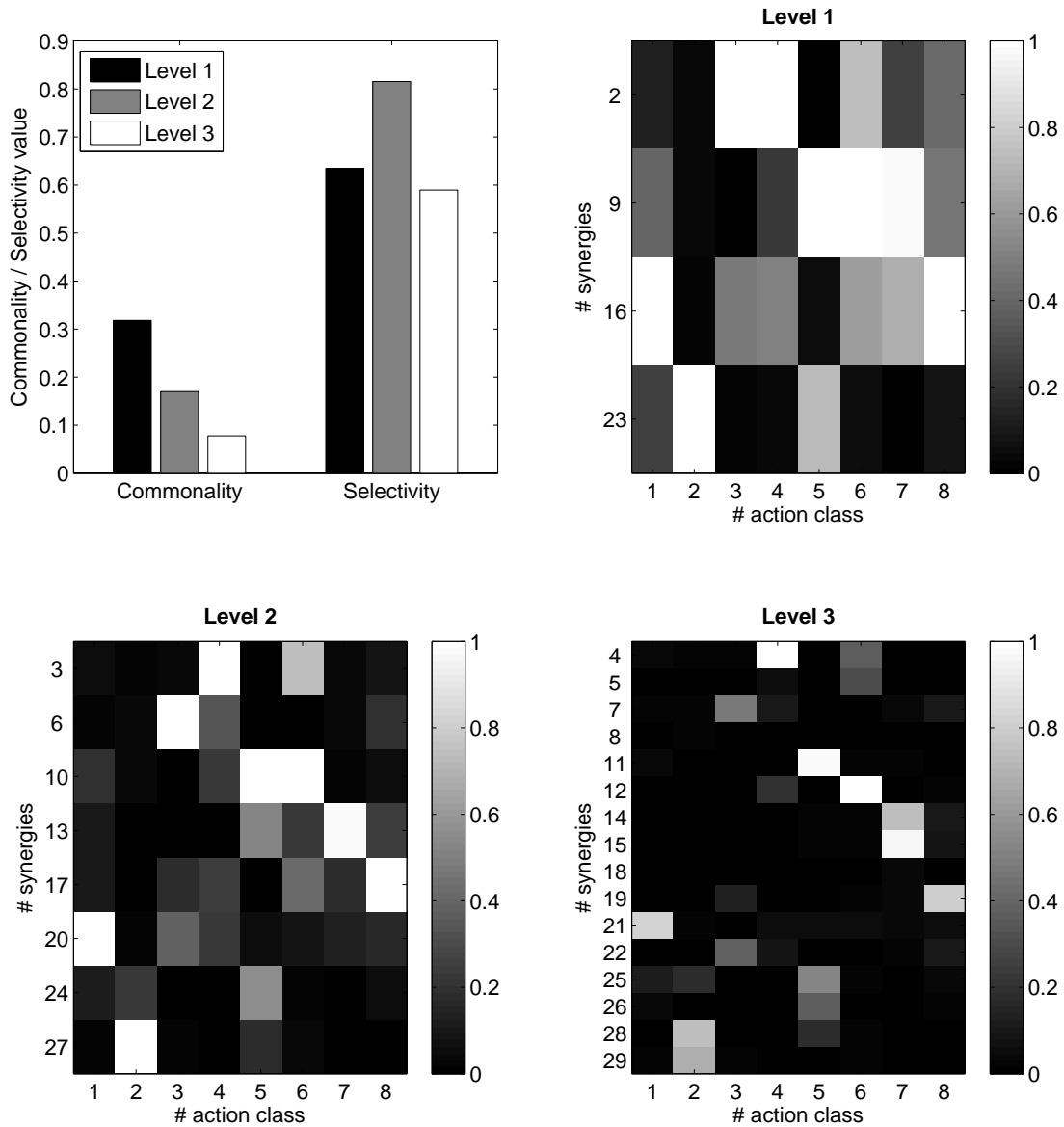


Fig. 8: *Commonality and selectivity using the tree  $T_3$* . The top-left graph shows selectivity and commonality mean values computed for each level of the rooted-tree  $T_3$  used in the tree-based action representation. The black bar stands for the first level, the gray bar stands for the second level, and the white bar stands for the last level. The remaining three graphs show the usage of the synergies for each level of the tree. For each tree level the synergy usage, in percentage, for each action type is reported.

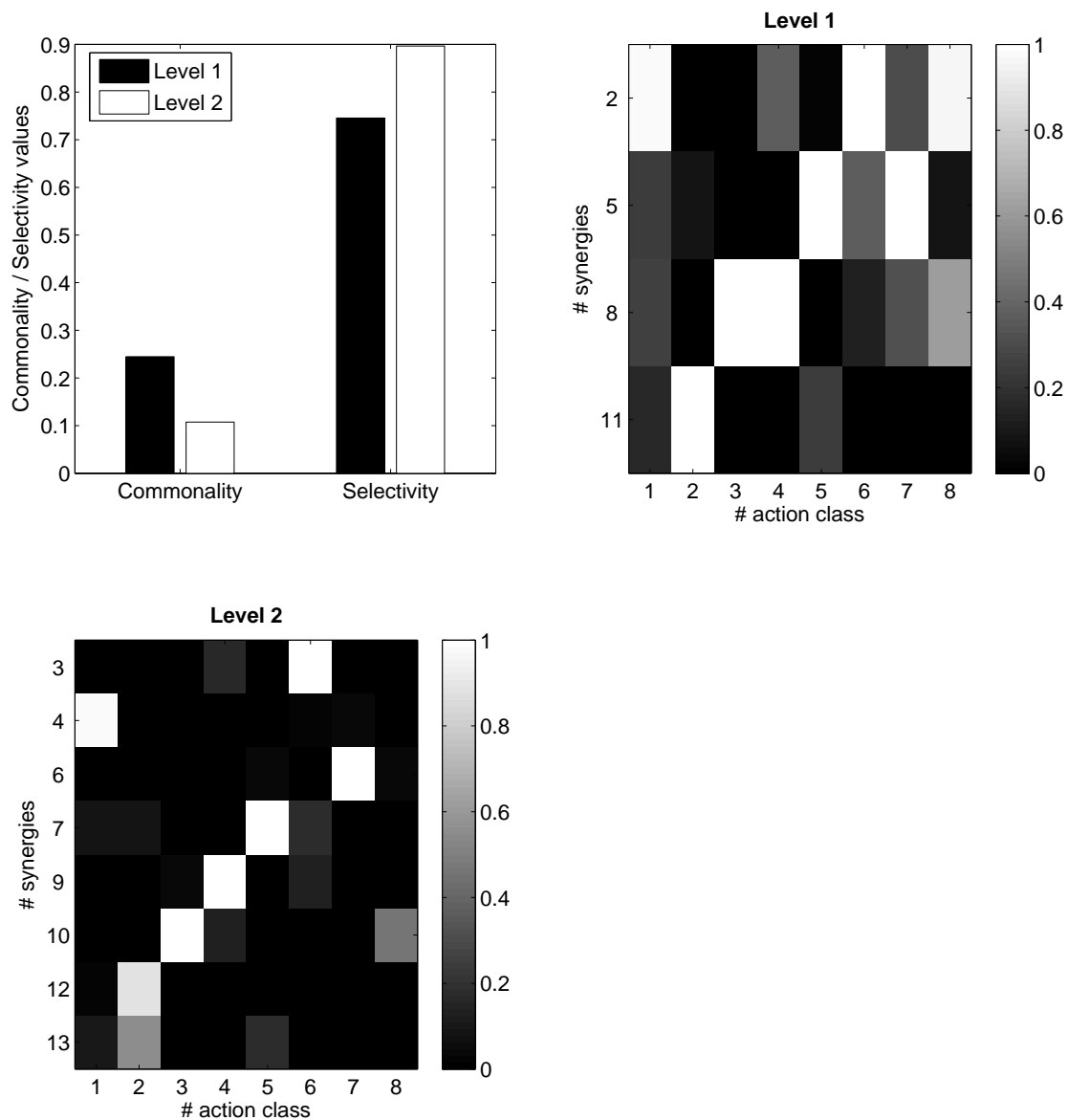


Fig. 9: *Commonality and selectivity using the tree  $T_2$* . The top-left graph shows selectivity and commonality mean values computed for each level of the tree  $T_2$  used in the tree-based action representation. The black bar stands for the first level and the white bar stands for the last level. The remaining two graphs show the usage of the synergies for each level of the tree  $T_2$ . For each tree level the synergy usage, in percentage, for each action type is reported.

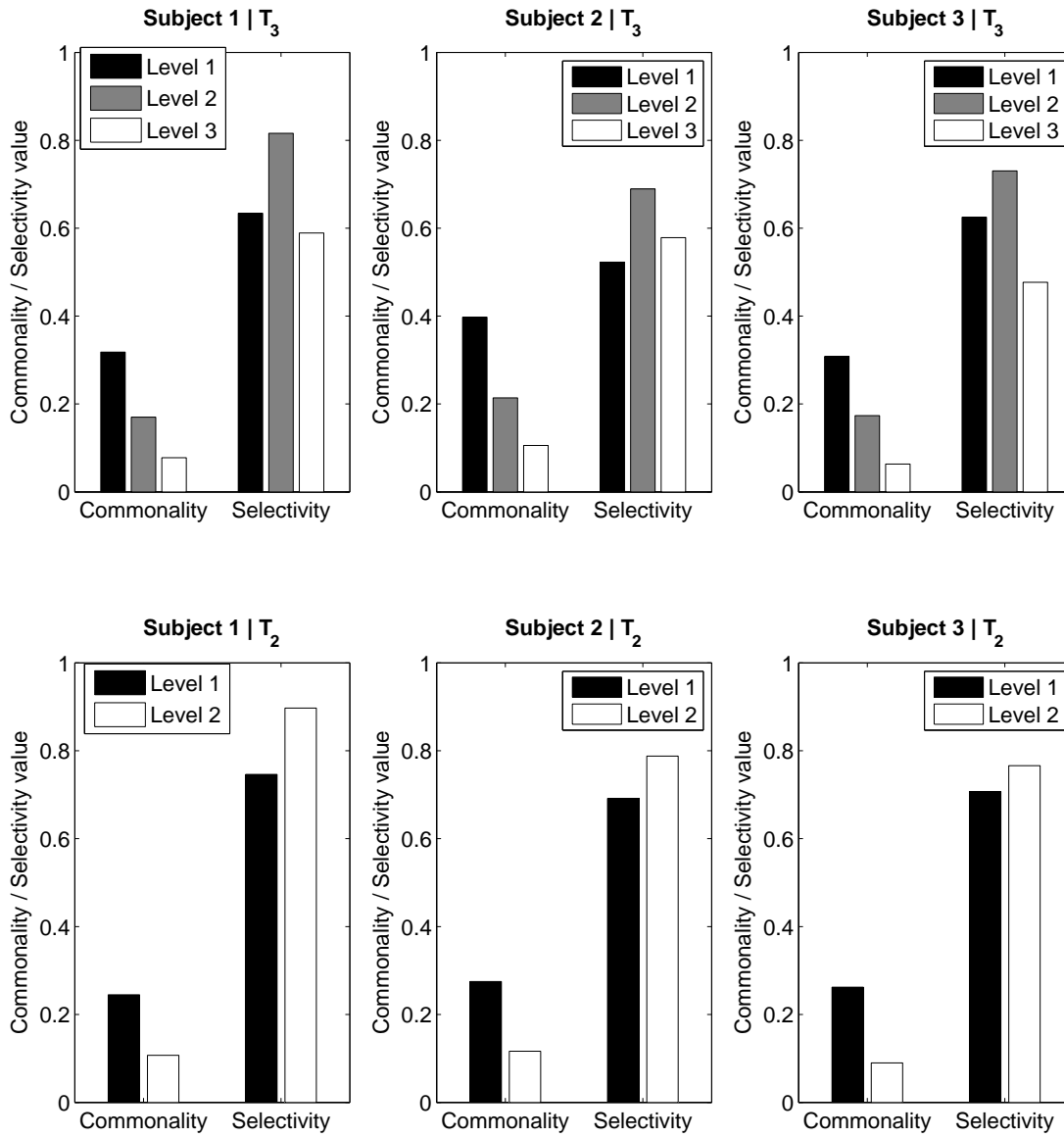


Fig. 10: *Commonality and Selectivity*. The figure shows selectivity and commonality mean values computed for each level of the tree  $T_3$  (first row) and  $T_2$  (second row) used in the tree-based action representation and for all the subjects. In the first row the black bar stands for the first level, the gray bar stands for the second level and the white bar stands for the last level of the tree  $T_3$ . In the second row the white bar stands for the second level of the tree  $T_2$ .



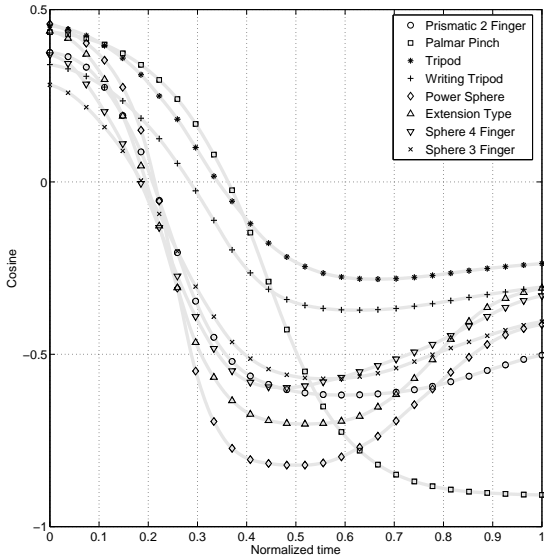


Fig. 11: *Comparing action-type temporal profiles.* In this figure the action temporal profile of the types of actions considered in the third test are compared. For each type of action a temporal profile is computed at each time  $t$  as the mean of the cosine between the hand configuration at the time  $t$  and the  $d$ -dimensional unit vector  $\mathbf{1}_p$ .

This multiple action representation turned out to be also hierarchical in nature. Indeed, Figure 8, Figure 9 and Figure 10 show that the TPSs exhibit a commonality value which has a decreasing trend over the tree levels, from the top to the bottom. On the contrary, the selectivity value reveals an increasing trend. In the case of tree  $T_3$  we obtained an increasing selectivity trends only for the first two levels. This was mainly due to the fact that some synergies of the last level tended to be much more specific for a small subset of grasping actions of a given action type. This result, together with the trend of the reconstruction error over different levels showed in Table (4), suggests that the TPSs on the higher levels of the tree captured action features which were shared from many grasping actions belonging to different action types. On the contrary, TPSs on the lower levels of the tree included action features which were selective for just one action type. Interestingly, Table 5, 6 and Figure 11 indicate that similar action types could be grouped together on the basis of the synergies of the first level only. Grasping actions were therefore represented with different degrees of generality on the different levels of the tree-structured synergy organisation.

The notion of a synergy based hierarchical representation of action is not completely new, of course. Several models of basic hand actions such as grasping actions have been provided over the last decade. Most of these models rely on a two-level synergy organisation (Gorniak et al 2009; Latash et al 2007; Shim et al 2003; Zatsiorsky et al 2003).

In spite of their theoretical relevance, however, these models do not seem to be general enough in taking into account the main hierarchical features of action representation. Indeed, the hierarchy of their action representation was determined only by two different synergy levels: the first level was formed by the synergies between thumb and a virtual finger (that is, an imaginary finger with mechanic effect equal to the sum of the mechanic effects produced by all the other four fingers), while the second level was constituted by the synergies of the individual four fingers forming the virtual one. As a consequence, their action representation does not naturally include further hierarchical levels. On the contrary, our tree-based action representation turns out to be flexible enough to represent action with hierarchical levels which may vary in number, being not a priori defined in terms of specific pattern of finger synergies. In addition, unlike the two-level synergy models, our tree-based action representation sheds new light on the multiple nature of the representation of action, by capitalising it in accounting for the hierarchical structure of action.

To this regard, it is worth noting that our tree-based action representation is in line not only with action theories emphasising the multiple and hierarchical structure of action representation (Goldman 1970), but also with several findings on the differentiated activation of the cortical motor system when planning and executing hand grasping actions. In particular, single cell recordings from the ventral premotor cortex (Rizzolatti et al 1988) and the inferior parietal lobule of the monkey brain showed that most neurons do not discharge in association with simple movements (like flexing the fingers), but are activated only and exclusively by the execution of movements accomplishing the representation of a given action such as grasping, tearing, holding or manipulating objects (Jeannerod et al 1995; Gallese 2000; Murata et al 2000; Fogassi et al 2005).

Most interestingly for the purposes of our study, both the ventral premotor and inferior parietal motor neurons represented the corresponding grasping actions with different degrees of generality: indeed, there were motor neurons that were sensitive to a motor action such as grasping a piece of food, regardless of whether the latter was accomplished with a specific grip (grasping with a whole hand prehension, grasping with a pre-

cision grip, etc.). However, even when selectivity was at its highest, the motor responses of neurons could not be interpreted in terms of single and merely physical movements. The neurons that discharged during certain movements (flexing of a finger, for example) performed with a specific motor goal (e.g. grasping), discharged weakly or not at all during the execution of those movements with a different motor goal (e.g. scratching) (Rizzolatti et al 2004, 2008). Similar findings have been obtained by brain imaging studies (Nelissen et al 2005) in monkeys and humans (Grafton and Hamilton 2007; Rizzolatti et al 2008).

In conclusion, the present study shows that tree-structured TPSs allow for both a multiple and hierarchical representation of basic actions such as grasping actions. Similar results might be obtained by using other notions of synergies such as dynamic or postural synergies, provided that they fulfil a suitable tree structure organisation. Moreover, our tree-based action representation exploits linear combinations of synergies, being the latter enough to represent fast hand movements such as reach-to-grasp actions performed in a natural way. However, grasping actions can be also represented in terms of non-linear superposition of synergies. Hence, a still open issue is to investigate whether and to what extent actions can be multiply and hierarchically represented by means of a non-linear superposition of synergies.

#### 4 Acknowledgments

This work was supported by a grant by Compagnia San Paolo. The authors wish to thank Giacomo Rizzolatti, Guglielmo Tamburrini, Giuseppe Tratteur for their comments and suggestions on a previous version of this paper. We are also grateful to Chiara Brozzo for her help in revising the last version of the paper.

#### 5 Appendix: Tree-Structured Synergy Method (TSSM)

We will use the following notations. Bold uppercase letters refer to matrices, e.g.,  $\mathbf{X}, \mathbf{V}$ , and bold lowercase letters designate vectors, e.g.,  $\mathbf{x}, \mathbf{v}$ . We denote by  $\mathbf{X}_i$  and  $\mathbf{X}^j$  the  $i$ -th row and the  $j$ -th column of a matrix  $\mathbf{X}$ , respectively. We use the notation  $x_i$  and  $v_{ij}$  to refer to the  $i$ -th element of the vector  $\mathbf{x}$  and the element in the  $i$ -th row and the  $j$ -th column of the matrix  $\mathbf{V}$ , respectively. Given  $\mathbf{x} \in \mathbb{R}^p$  we use the notation  $\|\mathbf{x}\|$  to refer to  $l_\infty$  norm. Given two vectors  $\mathbf{x}$  and  $\mathbf{y}$  in  $\mathbb{R}^p$ , we denote by  $\mathbf{x} \circ \mathbf{y} = (x_1y_1, x_2y_2, \dots, x_py_p) \in \mathbb{R}^p$  the element-wise product of  $\mathbf{x}$  and  $\mathbf{y}$ .

#### 5.0.6 Tree-Structured Stage

The update of the  $\mathbf{U}$ 's values is performed in this stage, and, more importantly, following the approach suggested by Jenatton et al (2010), a tree-structured representation of the rows in  $\mathbf{X}$  is found. The main difficulty is that the optimization of the  $\mathbf{U}_i$ ,  $i \in 1, 2, \dots, n$ , for a fixed  $\mathbf{V}$  involves the nonsmooth regularization term  $\Omega(\mathbf{U}_i) = \sum_{j=1}^r w_j \|\mathbf{D}_j \circ \mathbf{U}_i\|$ , where  $w_j$  are positive weights<sup>2</sup>. In this case the update of the vectors  $\mathbf{U}_i$  can be performed using a *proximal* method (Combettes and Wajs 2006). In general proximal approaches are used when one has to minimize a convex nonsmooth objective function which assumes the following general form:

$$f(\mathbf{u}) + \lambda\Omega(\mathbf{u})$$

where  $f(\mathbf{u})$  is the usual data-fitting term  $\frac{1}{2}\|\mathbf{x} - \mathbf{u}\mathbf{V}^T\|_2^2$  and  $\Omega(\mathbf{u})$  is a non-differentiable regularization term. In a nutshell, the proximal approach consists of two consecutive updating steps: first, the vector  $\mathbf{u}$  is updated using the standard gradient update rule w.r.t the first term of the objective function as follows:

$$\bar{\mathbf{u}} \leftarrow \mathbf{u} - \frac{1}{\sigma_{\mathbf{U}}} \nabla f(\mathbf{u}) = \mathbf{u} + \frac{1}{\sigma_{\mathbf{U}}} (\mathbf{x} - \mathbf{u}\mathbf{V}^T)\mathbf{V} \quad (5)$$

then, starting from the value  $\bar{\mathbf{u}}$  the new value for  $\mathbf{u}$  is computed by applying a proximal operator  $\Pi_{\mathbf{U}}$  defined by the following minimization problem:

$$\Pi_{\mathbf{U}}(\mathbf{u}) = \underset{\mathbf{v}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{u} - \mathbf{v}\|_2^2 + \lambda\Omega(\mathbf{v}) \quad (6)$$

Thus we obtain  $\mathbf{u}^{new} \leftarrow \Pi_{\mathbf{U}}(\bar{\mathbf{u}})$ . For a number of regularization terms the minimization problem expressed in (6) can lead to closed-form solutions. For example when  $\Omega(\mathbf{u})$  is the  $\ell_1$  norm of  $\mathbf{u}$  the corresponding proximal operator  $\Pi_{\mathbf{U}}$  is the well-known soft-thresholding operator. In the case of the regularization term used here this minimization problem can be solved by a *primal-dual* approach which enable us to implement the proximal operator defined in (6) by the procedure presented in Algorithm (1).

Summarizing, the optimization of the  $\mathbf{U}_i$  values is performed using the gradient descent rule expressed in (5) and, then, applying the proximal operator as defined previously.

<sup>2</sup> Note that all  $w_j$  are fixed to 1 in the experiments

---

**Algorithm 1** Proximal operator.  $\Pi_{\lambda w_j}^*$  is the orthogonal projection on the ball of radius  $\lambda w_j$  of the dual norm  $\|\cdot\|_*$ .

---

**Input:**  $\mathbf{u} \in \mathbb{R}^r$  and  $\mathbf{D} \in \mathbb{R}^{r \times p}$

**Output:**  $\mathbf{v} \in \mathbb{R}^r$

**for**  $i \leftarrow 0$  **to** *MaxNumberOfIteration*

**for**  $j \leftarrow 1$  **to**  $r$

$$\mathbf{P}_j \leftarrow \mathbf{u} - \sum_{h \neq j} \mathbf{P}_h$$

$$\mathbf{P}_j \leftarrow \Pi_{\lambda w_j}^*(\mathbf{P}_j \circ \mathbf{D}_j)$$

**end for**

**end for**

$$\mathbf{v} \leftarrow \mathbf{u} - \sum_{j=1}^r \mathbf{P}_j$$


---

### 5.0.7 Synergy Dictionary Stage

This stage consists in updating the  $\mathbf{V}$ 's values while keeping fixed the values of  $\mathbf{U}$ . Note that the objective function in (1) is composed of two terms to be minimized, and the second term does not depend on  $\mathbf{V}$ . Therefore, the optimization problem posed in (1) can be, in this stage, reformulated as follows:

$$\min_{\mathbf{V}} \frac{1}{2np} \|\mathbf{X} - \mathbf{UV}^T\|_F^2 \text{ s.t. } \forall i \|\mathbf{V}^i\|_2 \leq 1 \quad (7)$$

Due the fact that the columns of  $\mathbf{V}$  are constrained to lie inside the unit ball, the update of  $\mathbf{V}$  is performed in two consecutive steps. First, we apply a standard gradient updating rule as follows

$$\bar{\mathbf{V}} \leftarrow \mathbf{V} + \frac{1}{\sigma_{\mathbf{V}np}} (\mathbf{X} - \mathbf{UV}^T) \mathbf{U}^T \quad (8)$$

where  $\eta$  is a parameter. Then, we use the projection operator  $\Pi(\mathbf{v}) = \frac{\mathbf{v}}{\max\{1, \|\mathbf{v}\|_2\}}$  in order to project the columns of  $\bar{\mathbf{V}}$  on the unit ball in  $\mathbb{R}^p$ . Consequently the update of  $\mathbf{V}$  is computed as follows:

$$\mathbf{V} \leftarrow \Pi(\bar{\mathbf{V}} + \frac{1}{\sigma_{\mathbf{V}np}} (\mathbf{X} - \mathbf{UV}^T) \mathbf{U}^T) \quad (9)$$

The overall algorithm of TSSM is reported in algorithm 2. Note that a fixed step gradient descent procedure was adopted with the two learning rate  $\sigma_{\mathbf{U}}$  and  $\sigma_{\mathbf{V}}$  chosen equal to the Lipschitz constant of  $\nabla f(\mathbf{u})$  and  $\nabla f(\mathbf{v})$  respectively.

## References

Aharon M, Elad M, Bruckstein A (2006) K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing* 54(11):4311–4322

---

## Algorithm 2 Tree-structured synergies algorithm

---

**Input:**  $\mathbf{X} \in \mathbb{R}^{n \times p}$ ,  $\mathbf{U}^0 \in \mathbb{R}^{n \times r}$  and  $\mathbf{V} \in \mathbb{R}^{p \times r}$

$T_{max} \in \mathbb{Z}^+$ ,  $\lambda \geq 0$

**Output:**  $\mathbf{U} \in \mathbb{R}^{n \times r}$  and  $\mathbf{V} \in \mathbb{R}^{p \times r}$

**for**  $t \leftarrow 1$  **to**  $T_{max}$

**repeat until convergence**

$$\mathbf{U}^t \leftarrow \mathbf{U}^{t-1} + \frac{1}{\sigma_{\mathbf{U}}} (\mathbf{X} - \mathbf{U}^{t-1} \mathbf{V}^T) \mathbf{V} \text{ gradient descent}$$

  step

$$\mathbf{U}^t \leftarrow \Pi_{\mathbf{U}}(\mathbf{U}^t, \lambda) \text{ proximal operator step}$$

**end**

  possibly replace under-used atoms

**repeat until convergence**

$$\mathbf{V}^t \leftarrow \mathbf{V}^{t-1} + \frac{1}{\sigma_{\mathbf{V}np}} (\mathbf{X} - \mathbf{UV}^{(t-1)T}) \mathbf{U}^T \text{ gradient}$$

  descent step

$$\mathbf{V}^t \leftarrow \Pi_{\mathbf{V}}(\mathbf{V}^t) \text{ proximal operator step}$$

**end**

**end for**

**return**  $\mathbf{U}^t, \mathbf{V}^t$

---

Arbib M, Iberall T, Lyons D (1985) Coordinated control programs for movements of the hand. *Experimental brain research* pp 111–129

Basso C, Santoro M, Verri A, Villa S (2011) PADDLE: Proximal Algorithm for Dual Dictionaries Learning, *Lecture Notes in Computer Science*, vol 6791, Springer Berlin / Heidelberg, pp 379–386

Bernstein NA (1967) *The Co-ordination and regulation of movements*, first english edition edn. Pergamon Press Ltd.

Combettes P, Wajs V (2006) Signal recovery by proximal forward-backward splitting. *Multiscale Modeling and Simulation* 4(4):1168–1200

d'Avella A, Portone A, Fernandez L, Lacquaniti F (2006) Control of fast-reaching movements by muscle synergy combinations. *The Journal of Neuroscience* 26(30):7791–7810

Engan K, Aase S, Hakon Husoy J (1999) Method of optimal directions for frame design. In: *Proceedings of the Acoustics, Speech, and Signal Processing, 1999. on 1999 IEEE International Conference, IEEE Computer Society, ICASSP '99*, vol 05, pp 2443–2446

Feix T, Pawlik R, Schmiedmayer H, Romero J, Kragic D (2009) A comprehensive grasp taxonomy. In: *Robotics, Science and Systems: Workshop on Understanding the Human Hand for Advancing Robotic Manipulation*

Fogassi L, Ferrari PF, Gesierich B, Rozzi S, Chersi F, Rizzolatti G (2005) Parietal lobe: From action organization to intention understanding. *Science* 308(5722):662–667

Gallese V (2000) The inner sense of action. agency and motor representations. *Journal of Consciousness Studies* 7(10):23–40

Gallese V, Fadiga L, Fogassi L, Rizzolatti G (1996) Action recognition in the premotor cortex. *Brain*

- 119:593–609
- Goldman A (1970) *A theory of human action*, vol 2. Prentice-Hall Englewood Cliffs, NJ
- Gorniak S, Zatsiorsky VM, Latash ML (2009) Hierarchical control of static prehension: II. multi-digit synergies. *Experimental Brain Research* 194(1):1–15
- Grafton ST, Hamilton AF (2007) Evidence for a distributed hierarchy of action representation in the brain. *Human movement science* 26(4):590–616
- Grinyagin IV, Biryukova EV, Maier MA (2005) Kinematic and dynamic synergies of human precision-grip movements. *Journal of Neurophysiology* 94(4):2284–2294
- Hastie T, Tibshirani R, Friedman J (2003) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, corrected edn. Springer
- Iberall T, Fagg AH (1996) Neural Network models for selecting hand shapes, A.M. Wing, P. Haggard and J.R. Flanagan, Editors, *Hand and Brain: the Neurophysiology and Psychology of Hand Movements*, Academi, pp 243–264
- Iberall T, Bingham G, Arbib MA (1986) Opposition space as a structuring concept for the analysis of skilled hand movements. *Experimental Brain Research Series* 15:158–173
- Jeannerod M (1988) *The neural and behavioural organization of goal-directed movements*. Clarendon Press
- Jeannerod M, Arbib MA, Rizzolatti G, Sakata H (1995) Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends in Neurosciences* 18(7):314–320
- Jenatton R, Mairal J, Obozinski G, Bach F (2010) Proximal methods for sparse hierarchical dictionary learning. *Proc ICML*
- Jerde T, Soechting J, Flanders M (2003) Biological constraints simplify the recognition of hand shapes. *Biomedical Engineering, IEEE Transactions on* 50(2):265–269
- Latash ML, Scholz JP, Schönner G (2007) Toward a new theory of motor synergies. *Motor control* 11:276–308
- Mason CR, Gomez JE, Ebner TJ (2001) Hand synergies during reach-to-grasp. *J Neurophysiol* 86(6):2896–910
- Mele A (1997) *The philosophy of action*. Oxford university press New York
- Murata A, Gallese V, Luppino G, Kaseda M, Sakata H (2000) Selectivity for the shape, size, and orientation of objects for grasping in neurons of monkey parietal area aip. *Journal of Neurophysiology* 83(5):2580–2601
- Nelissen K, Luppino G, Vanduffel W, Rizzolatti G, Orban G (2005) Observing others: multiple action representation in the frontal lobe. *Science* 310(5746):332
- Pellegrino G, Fadiga L, Fogassi L, Gallese V, Rizzolatti G (1992) Understanding motor events: a neurophysiological study. *Experimental brain research* 91(1):176–180
- Rizzolatti G, Sinigaglia C (2010) The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nat Rev Neurosci* 11(4):264–74
- Rizzolatti G, Camarda R, Fogassi L, Gentilucci M, Luppino G, Matelli M (1988) Functional organization of inferior area 6 in the macaque monkey. *Experimental Brain Research* 71(3):491–507
- Rizzolatti G, Fadiga L, Gallese V, Fogassi L (1996) Premotor cortex and the recognition of motor actions. *Cognitive Brain Research* 3(2):131–141
- Rizzolatti G, Fogassi L, Gallese V (2001) Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat Rev Neurosci* 2(9):661–670
- Rizzolatti G, Fogassi L, Gallese V (2004) Cortical mechanisms subserving object grasping, action understanding, and imitation, *The New Cognitive Neurosciences*, 3rd Edition (ed. in chief Gazzaniga, M.S.), A Bradford Book, MIT Press, Cambridge, Ma, pp 427–440
- Rizzolatti G, Sinigaglia C, Anderson F (2008) *Mirrors in the brain: how our minds share actions, emotions*. Oxford University Press, USA
- Santello M, Soechting J (2000) Force synergies for multifingered grasping. *Experimental Brain Research* 133(4):457–467
- Santello M, Flanders M, Soechting JF (1998) Postural hand synergies for tool use. *Journal of Neuroscience* 18(23):10,105–10,115
- Santello M, Flanders M, Soechting JF (2002) Patterns of hand motion during grasping and the influence of sensory guidance. *Journal of Neuroscience* 22(4):1426–1235
- Shim JK, Latash ML, Zatsiorsky VM (2003) Prehension synergies: trial-to-trial variability and hierarchical organization of stable performance. *Exp Brain Res* 152(2):173–84
- Shim JK, Latash ML, Zatsiorsky VM (2005) Prehension synergies in three dimensions. *Journal of Neurophysiology* 93(2):766–776
- Tessitore G, Prevede R, Catanzariti E, Tamburrini G (2010) From motor to sensory processing in mirror neuron computational modelling. *Biological Cybernetics* 103(6):471–485
- Vinjamuri R, Lee H, Mao Z (2010a) Dimensionality reduction in control and coordination of the human hand. *IEEE Trans Biomed Eng* 57(2):284–295



- Vinjamuri R, Sun M, Chang CC, Lee HN, Sclabassi RJ, Mao ZH (2010b) Temporal postural synergies of the hand in rapid grasping tasks. *Information Technology in Biomedicine, IEEE Transactions on* 14(4):986–994
- Zatsiorsky VM, Gao F, Latash ML (2003) Finger force vectors in multi-finger prehension. *Journal of Biomechanics* 36(11):1745–1749