Grasping in Primates: Mechanics and Neural Basis

Lucia Castellanos

Machine Learning Department Carnegie Mellon University

DATA ANALYSIS PROJECT Advisor: Robert E. Kass LUCIA@CMU.EDU

KASS@STAT.CMU.EDU

Abstract

We investigate the encoding of grasping kinematics in the primary motor cortex of rhesus monkeys (*Macaca mulatta*). Previous work in the Schwartz lab has demonstrated cortical control of an anthropomorphic arm and a simple gripper. We aim to replace the gripper with a 'primate-like hand' — this work is a study towards that goal. In our experiment two rhesus monkeys are trained to perform reach-to-grasp movements of several objects that have a variety of shapes and sizes. During these movements, the activity from single motor cortical neurons is captured with five movable electrodes, and the positions of several markers in the monkeys' hands are recorded. In this work we analyze the specific time points of kinematic data of grasping and investigate its relationship with the activity of M1 neurons. The natural hand has over 20 degrees of freedom. We define a synergy as a low dimensional representation of the kinematic variables that conform to the grasping. We use linear and nonlinear dimensionality reduction techniques to obtain kinematic synergies, and we apply classical regression techniques on the neural activity to determine the encoding of these kinematic data in the recorded neural activity.

Contents

1	Introduction and related work							
	1.1	Related work: grasping mechanics	3					
	1.2	Related work: neural basis of grasping in primates	5					
		1.2.1 Brief overview of spike train analysis	7					
		1.2.2 Neural basis of grasping	8					
	1.3	Overview of our work	10					
2	Exp	perimental Design	11					
3	ta exploration	11						
	3.1	Kinematic data	13					
	3.2	Neural data	16					
4	Obt	tainment of synergies or eigengrasps: kinematic analysis	20					
	4.1	Dimensionality reduction: methods	20					
	4.2	Synergies of grasping	25					
		4.2.1 All time versus time of interest	25					
		4.2.2 Outlier definition	28					
		4.2.3 Sampling versus averaging	28					
		4.2.4 Obtaining the low dimensional representations	29					
	4.3	Measures of goodness of eigengrasps	29					
		4.3.1 Naive bayes classifier	30					
		4.3.2 Multi-Class Support Vector Machine	30					

	4.4	Experiments	31
	4.5	Results for kinematic analysis	31
		4.5.1 Conclusions of kinematic analysis	35
5	Net	ural coding of eigengrasps	35
	5.1	Neural coding of finger kinematics: methods	38
		5.1.1 Preselection of neurons	38
		5.1.2 Model for finger kinematics: multiple linear regression	38
		5.1.3 Are the results due to chance?	39
	5.2	Analysis design	39
	5.3	Results for neural analysis	40
		5.3.1 Preselection of neurons	40
		5.3.2 Linear coding of kinematic variables	41
		5.3.3 Conclusion of neural analysis	45
6	Dise	cussion	45

1. Introduction and related work

In the United States, there are approximately 1.7 million people living with limb loss (Ziegler-Graham et al., 2008). Limb loss can be caused by trauma, cancer, congenital limb differences or due to complications of the vascular system. According to the National Limb Loss Information Center, over 68% of the total number of limb losses caused by trauma in 2008 corresponded to upper limb, as opposed to lower limb. 58.5% of amputations caused by congenital anomalies and 23.9% of the cancer related amputations were also of upper limbs.

The creation of functional and aesthetic protheses is a promising way to contribute to improve the quality of life of people who have undergone an amputation or limb loss. Motor neural prostheses are artificial extensions of the body that restore or supplement motor functions of the nervous system lost during injury or diseases causing amputation. The big motivation for this study is to contribute toward the creation of reliable motor neural prostheses that restore the function of human hands.

In order to create a fully functional neural motor prosthesis of the human hand it is necessary to understand the kinematics of grasping and the way the Central Neural System (CNS) controls those mechanics.

Our study aims at these two goals. As a first step, we start by studying the kinematics of grasping. Understanding the mechanics of the human hand and, in general, of primate grasping provides guidelines for robotic hand (Cutkosky, 1989), and prosthesis design.

1.1 Related work: grasping mechanics

There has been a large volume of work carried out in the area, ranging from the anatomy of primate hands and the mechanics of grasping in different settings, to the design and control of robotic models of the hand.

Taylor and Schwarz (1955) presented a summarized study of the anatomical and mechanical features of the human hand. They emphasized the anatomical and mechanical differences between the four digits and the thumb. With regards to anatomy, the four digits are built by three phalanges, whereas the thumb has only two of them (Figure 1). As to mechanics, all five fingers can perform extension-flexion movements, together with abduction and adduction (figure 2). In addition, the thumb has the ability to *oppose* the digits, thanks to the large mobility of the carpometacarpal joint (figure 1), that allows one to rotate the plane in which flexion and extension of the thumb takes place. The critical component of human grasp lies precisely in the thumb; its rotary movement, by which it swings about its own axis, allows gripping and grasping modalities that are unique among all existing species except for certain primates.

One tool to understand the mechanics of grasping is the creation of a *taxonomy of grasps*. In theory, a grasp taxonomy should represent all the possible grasps for a hand device. There are several proposals for taxonomies of human grasps.

One of the most influential classifications of human grasps was presented by Napier (1956). He proposed to categorize grasps into *power grasps* and *precision grasps*. According to his definition, *power grasps* are characterized by large areas of contact between the grasped object and the surfaces of the fingers and palm, and by little or no ability to impart motions with the fingers. Whereas in a *precision grasp* the object is held with the tips of the fingers and thumb. Based on Napier's work, Cutkosky (1989) defined a taxonomy of human grasps based firstly on function (or task to perfom), as opposed to object shape (Taylor and Schwarz, 1955). As a first step, he considered whether the grasp is a precision or a power grasp, then the taxonomy was developed taking into consideration details of the task and of the object shape.

Recently, Bock and Feix (2008) presented a unified taxonomy (figure 3). In their work they incorporated virtually all previous grasping taxonomies, including those originated in robotics, and developmental medicine, among others. They started by defining a grasp as each static hand pose with which an object can be held securely with one hand. Secondly, they eliminated previous hand poses that depended on gravity, and that described intrinsic movements (as opposed to movements in relation to holding an object). In order to develop the taxonomy they considered as main features to classify: the precision and power grasps from Napier (1956); the opposition types (Iberall, 1987), that is the basic direction in which a human hand can apply forces; the concept of virtual fingers (Iberall, 1997) needed to grasp the object; and whether the thumb is abducted or adducted with respect to its carpometacarpal joint. They obtained in total 33 types of grasps, which they reduced to 17 by completely disregarding the shape of the object held.

Within the robotics community, Goldfeder et al. (2009) created the *Columbia Grasp Database* with the vision of providing tools to improve robotic grasping. Their database was created according to two hand models: a simulated human hand, and the three-fingered robotic Barrett hand. As opposed to the *Human Grasping Database* (Bock and Feix, 2008), the Columbia counterpart emphasizes the different shapes and sizes of graspable objects. Indeed, their database is indexed according to the 3D geometry of a large range of objects.

Regardless of the premise of grasping taxonomies being shape-object dependent or not, there seems to be the possibility of including virtually all the possible grasps configurations into a finite taxonomy. This suggests that while the configuration space of dexterous hands is high dimensional, most useful grasps can be found in the neighborhood of a small number of discrete points (Ciocarlie et al., 2007). This observation certainly provides insights into the design of functional robotic hands, and it might help to explain how the CNS codes for hand and finger movements. Indeed, a fundamental question is how the CNS controls grasping. A first idea is that the position and mechanics of a single joint are controlled by neurons in the motor cortex. However, it has been shown that there is clear mechanical coupling on the hand's joints by which our fingers do not move in isolation from contiguous fingers (Ross and Schieber, 2000). We therefore pose the following question: *is it possible then that the neurons use a simplifying strategy to deal with the redundancy produced by this mechanical coupling*?

The hand is able to grasp different objects through the cooperative action of the bones, and muscles that comprise it. The cooperation of various elements towards a common goal can be called a $synergy^1$. One can represent a hand synergy through a number of variables that represent position of the bones, or amplitude of the joint angles. In order to fully describe a hand synergy, a large number of variables is needed. Three questions can be clearly posed: (a) is it possible to describe accurately the hand grasping through a set of low dimensional synergies?, (b) does the CNS code those synergies?, (c) can those synergies be used as input for robotic devices for robotic control and manipulation or for neural prosthesis?

As a matter of fact, the extraction of synergies in human grasping has been performed in different experimental settings ranging from static hand poses grabbing imaginary (Santello et al., 1998) and real (Todorov and Ghahramani, 2004) objects; reach-to-grasp tasks (Mason et al., 2001); and skilled activites (Soechting and Flanders, 1997); to unconstrained haptic exploration tasks (Thakur et al., 2008). In these works it has been shown that it is possible to obtain a small number of *synergies* that describe most of the variance of the different possible hand configurations. The classical way of extracting hand *synergies* has been through linear dimensionality reduction techniques, in particular through Principal Component Analysis (Table 1).

Ciocarlie et al. (2007) has, for example, shown that the synergies obtained through linear dimensionality reduction methods can indeed serve as input for several existent robotic devices (Figure 4), and that this procedure yields good results.

^{1.} In fact, the term synergy has several different meanings (Gantchev et al., 1999; Todorov and Ghahramani, 2004).



Figure 1: Anatomy of the bones and relevant joints of the (left) human hand. The metacarpophalangeal (MCP), the proximal interphalangeal (PIP) and the distal interphalangeal (DIP) joints are indicated for the index finger, but all fingers cointain their respective joints. (Figure extracted from *Wikipedia* and modified.)



Figure 2: Mechanics of digits. The mechanics of the different joints in the hand can be exemplified with the consideration of the coordinate system defined for a particular joint (A). The extensionflexion movement of a joint is depicted in figure (B), where the movement is with respect to the x-axis. The abduction-adduction movement, which happens with respect to the z-axis, is illustrated in panel (C). (Figure extracted from the Computational Bioengineering Laboratory, National University of Singapore website.)

Finally, a working hypothesis in literature has been that indeed, there exists a set of synergies that simplifies coding and control for the CNS ((Santello et al., 1998; Mason et al., 2001; Todorov and Ghahramani, 2004; Thakur et al., 2008), among others).

1.2 Related work: neural basis of grasping in primates

In order to study human processes, it is common to use animal models. For example, *Macaca mulatta* monkeys have arms, hands and brain that share anatomical and functional features with human parts. Indeed, like humans, *Macaca mulatta* monkeys (or rhesus monkeys) have the ability of precision and power grips, because they have opposable thumbs.

With regards to the brain, a large body of research exists on the validity of taking the rhesus monkey's brain as a model for the human brain. Zilles et al. (1995) presented a detailed study where it was confirmed that: (a) the topology of the somatosensory and motor areas is identical in both species (see Figure 5); (b) there exists 'great' architectonic similarities between the somatosensory and motor areas in macaques and humans; and, (c) the distribution of patterns of many receptors in the somatosensory and motor areas



Figure 3: A unifying Hand Taxonomy. (Figure from Bock and Feix (2008).)

Model	DOF	Eigengrasp 1			Eigengrasp 2		
Model	DOFS	Description	min	max	Description	min	max
Gripper	4	Prox. joints flexion	>-	+ 🗘	Dist. joints flexion	-	• 💱
Barrett	4	Spread angle opening	% -	+	Finger flexion	\checkmark	• 4
DLR	12	Prox. joints flexion Finger abduction	¥	+	Dist. joints flexion Thumb flexion	K -	►
Robonaut	14	Thumb flexion MCP flexion Index abduction	¥-	•	Thumb flexion MCP extension PIP flexion		•
Human	20	Thumb rotation Thumb flexion MCP flexion Index abduction	W-	+	Thumb flexion MCP extension PIP flexion	989	+

Figure 4: Synergies or eigengrasps that were obtained as in (Santello et al., 1998) but from four robotic hand models and a human hand model. (Figure from (Ciocarlie et al., 2007)).

Task	Subjects	Coordinate	Method	Reference
		system		
• hold imaginary object	human	JA	PCA	(Santello et al., 1998)
• specific manipulation	human	JA	PCA	(Todorov and Ghahramani, 2004)
of actual objects				
\bullet reach-to-grasp	human	3D	SVD	(Mason et al., 2001)
\bullet reach-to-grasp	monkey	3D	SVD	(Mason et al., 2004)
• skilled activity	human	JA	PCA	(Soechting and Flanders, 1997)
\bullet unconstrained	human	JA	PCA	(Thakur et al., 2008)
haptic exploration				

Table 1: Previous work on dimensionality reduction for extracting hand synergies. It is important to mention Principal Component Analysis and Singular Value Decomposition are essentially the same. JA stands for joint angles, and 3D for the three dimensional position of markers. (For more detailed information of the data they used and more details on their eigengrasp obtention, refer to Table 8 of the Appendix.

of both species have many features in common. The conclusion is, then, that there is enough evidence in support of taking the macaque motor cortex as a reliable model of the human motor cortex (Figure 5).

Grasping is an intentional or voluntary movement, which is typically defined through three characteristics: it is purposeful, it improves with experience and learning, and external stimulus need not preceed it (Kandel et al., 1991). In order for voluntary movements to take place, three processes must occur: target identification, plan of action, and execution. In primates, each of these processes is controlled by distinct regions of the cerebral cortex, namely: the posterior parietal cortex, the premotor areas of the frontal cortex, and the primary motor cortex (M1). These areas are organized somatotopically, that is to say, adjacent areas of the cerebral cortex control movements of adjacent parts of the body. Body parts that need more precision and fine control, such as the face and hands, have large representations in the motor map (Figure 6). In addition, other areas of the brain like cerebellum, and the basal ganglia have roles related to different aspects of motor control. For instance, the cerebellum is associated with coordination, precision and fine tuning of movements through sensory feedback. And damage in the basal ganglia produces motor disorders like Parkinson or Huntington's disease. In summary, all these brain areas are involved in the control of grasping.

There has been considerable progress in the understanding of human and non-human primate grasping control during the last decade (Castiello, 2005). In humans, neuroimaging techniques have been applied, whereas in non-human primates electrophysiological, approaches have predominated. In the following paragraphs we provide a brief survey that contains, to our view, the most relevant advances in the understanding of the neural basis of grasping in non-human primates. First, we give an overview of the different approaches for neural analysis of spike trains that exist in literature, then we summarize the main findings of different studies. We begin with a study of neurons in the posterior parietal area, followed by a study on cerebellar neurons, and we conclude with studies done in premotor and primary motor areas.

1.2.1 Brief overview of spike train analysis

The analysis of spike trains obtained from a neurophysiological study is done either considering the aggregated neural activity of a neuron across trials of a single condition, or considering single trials. The aggregated activity of a neuron can be visualized through raster plots and peristimulus time histograms (PSTHs). Raster plots display time in the abscissa, and number of trials in the ordinate, and contain dots at the time that the neuron fired in a specific trial (examples of raster plots with our data are shown in



Figure 5: The figure shows the correspondence of Motor Areas in human and macaque monkey brain. Premotor cortex (PM), Supplementary motor area (SMA), Primary motor cortex (M1), Prefrontal cortex (PF), Primary somatic sensory cortex (Pssc), Posterior Parietal Area (PP), Corpus callosum (Cc), Central sulcus (Cs). (Figure extracted from (Kandel et al., 1991).)

Figure 15 in the upper panel of the graphs). In this way, raster plots display the timing of the preferential activity of the neuron for a large number of trials (Awiszus, 1997). A PSTH requires the choice of a parameter called the *binwidth*, which is the size of the bin used to partition the time axis. The PSTH is the histogram of the sum of action potentials of the neuron across trials for specific time bins. From this histogram, a density estimation method can be used to obtain a smooth aggregated firing rate profile for the neuron. Some disadvantages of using raster plots or PSTHs are that they are trial alignment dependent (that is, they depend on the way of aligning the trials), also PSTHs are bandwidth dependent, and furthermore, they sum out the information given by the variation of a specific trial. In Figure 15 we show an example of the sensitivity of PSTHs to alignment type. Figure 15 panel (A) shows the different shape that the aggregate neural activity takes when different alignments are chosen.

The analysis of single trials tends to accurately reflect the variability caused by the signal separated from that of the noise. When single trials are considered, a first option is to estimate a smooth firing rate profile (with a sliding window, or convolving each action potential with a normal kernel), another is to define a binwidth and estimate the firing rate within bins of time.

Once the neural activity is quantified through one of the aforementioned methodologies a statistical analysis relating the patterns of neural activity with the behavioural task can be performed.

1.2.2 NEURAL BASIS OF GRASPING

Posterior parietal cortex. It is thought that posterior parietal cortex contains hand manipulation neurons, that is, cells that respond strongly to grasp of objects and to purposeful exploratory movements of the hands. Chen et al. (2009) trained two monkeys to grasp and lift knobs of different shapes and locations, while single unit recordings in the posterior parietal cortex (Brodmann area 5) were being recorded through a chronically implanted electrode. In order to visualize the data they obtained raster plots and PSTHs aligned to the onset times of hand contact with the knob. Then they performed the following analysis. They divided the timing of a trial into *epochs* describing: pretrial interval, movement period (subdivided in four stages), hold interval and relax period; and they obtained the firing rates for each epoch. They defined a neuron to be *task related* if (a) its firing rates demonstrated significant stage dependent modulation across the seven task stages and the pretrial interval with an F-test, and (b) it



Figure 6: Somatotopic organization in *Primary motor cortex* (M1). (Figure extracted from (Kandel et al., 1991).)

showed significantly increased or decreased firing rates during at least one task stage compared with the pretrial rate in paired means comparisons. Each neuron also had to have at least 50 trials. They grouped trials according to the shape, the location of the knob to be grasped, and the approach style to the target². Their analysis consisted in correlating the firing rates of the neurons with so-called *task factors*. These factors were the type of approach, the shape and location, all treated as categorical variables. To study the correlation they used analysis of variance (ANOVA) with its corresponding F-test. Their main result is that 77% (58/75) neurons showed significant effects of approach style on firing rates (showing higher rates and longer durations for specific approaching styles). Moreover, the approach style produced a much stronger effect on firing rates than *object features*, namely shape and location, which were distinguished only by half of the neural population. This is relevant since it helps to make a more accurate description of the role of neurons in posterior parietal cortex on grasp.

Cerebellum. Mason et al. (2006) performed a series of experiments with the same behavioral paradigm but recording at different areas of the brain. The paradigm is the same as in the study of grasping kinematics in monkeys (Mason et al., 2004), but it was extended to study correlations of the kinematics with the activity of Purkinje cells in cerebellum, and neurons in premotor and primary motor cortices. Their experiment consisted in training two monkeys to reach-and-grasp, with five different levels of force, sixteen objects that varied in size and shape (cubes, rectangular prisms, polysided prisms and cylinders). In (Mason et al., 2006) the authors recorded single cell activity from Purkinje cells in the cerebellum while the monkeys performed the task. For their analysis Mason et al. (2006) only considered the simple spikes of Purkinje cells as opposed to the *complex* spike type also characteristic of these type of neurons. A total of 77 cells were analyzed (37 for one monkey and 40 for the other). They divided the time of each trial into epochs: baseline, reach and grasp. They performed an epoch-based analysis and a finner time analysis. In the former they normalized the firing rates of the reach and grasp epochs with the firing rate of the baseline period. Then they pre-selected neurons to determine whether the discharge of the studied cells was significantly modulated during the task as compared to the baseline period (t-tests). The majority of cells showed significant simple spike modulation relative to the baseline during reach and grasp. They also performed an ANOVA of the normalized firing rates of individual trials on each of the epochs to

^{2.} The *approach style* refers to the trajectory the monkey followed with its hand to grab the knob. It was classified as forward, lateral, local approaches or regrasp. For details see (Chen et al., 2009).

determine whether the cell firing profile was significantly modulated in relation to *object shapes-sizes*³, or to force. They concluded that indeed the firing rates were significantly modulated in relation to both *object shapes-sizes* and to force. Finally, they performed a finer analysis on time by applying temporal multiple regression analysis with the two categorical variables: force and *object shapes-sizes*. They found that object related modulation occurred preferentially during reach or early in the grasp, that the firing was positively correlated with grasp force during both reach and grasp, and that there were no significant interactions between object and grasp force modulation.

Premotor cortex (PM) and primary motor cortex (M1). Hendrix et al. (2009) used the same paradigm as (Mason et al., 2004) and (Mason et al., 2006) but studied neurons located in dorsal premotor cortex (PMd) and in primary motor cortex. The analysis was very similar to (Mason et al., 2006). Their defined epochs where different – they considered: baseline, premovement, reach and grasp. And they continued by selecting task-selected neurons, and applying an epoch analysis and a finer temporal regression analysis using the categorical variables: level of force and object size. In this analysis, they considered neurons with between 60 to 100 trials per condition, that is, the data set was more than twice as big as the data set of Purkinje cells in (Mason et al., 2006). Their main findings were: (a) the firing profiles of most of the neural discharge of neurons in PMd and in M1 was significantly correlated with *object shapes-sizes*, and also to grasping force, but not to the same extent than object shape-size. Also, all neurons whose firing was correlated with grasping force presented correlation between their firing and grasping dimension, but not the vice-versa. (c) the correlations between grasping force and grasp dimension were stronger in the firing of M1 than in PMd, and that modulation with these parameters increased as reach to grasp progressed.

Concluding remarks. All of these studies are similar in their attempt to correlate kinematic variables during grasping tasks with neural activity in relevant regions in the primates brain. All of the approaches use a definition of time epochs determined by moments in the behavioral task. They also use multiple linear regression, but only on variables coded as categorical variables like: object shape-size, position, and level of force. There is a lack of studies that attempt to find correlations between variables such as actual hand configuration, or its low dimensional representation to motor control neuronal sites. We attempt to tackle these problems in our work.

1.3 Overview of our work

We have found some areas in the literature review that have not been explored and we have tackled some of them in this project. Namely, the application of non linear dimensionality reduction methods to obtain grasping synergies at specific time points, the comparison with linear dimensionality reduction methods, and the exploration of the possibility of their encoding in mean firing rate of neurons collected in the primary motor cortex. Questions that we want to answer is whether there is evidence that the kinematic embeddings obtained from classic static dimensionality reduction techniques are encoded or not in neurons from the primary motor cortex.

In the following, we outline our main contributions and provide a roadmap of the rest of the work.

- In Section 2 the experimental design, experimental conditions and data collection are explained in detail.
- In Section 3 we summarize the main features of our data sets, and emphasize the aspects of the data that are relevant for selecting the approaches and methods used in the data analysis.
- In Section 4 we deal with the kinematic data analysis fixing specific time points according to a function that summarizes the movement of the fingers, the *energy*. The kinematic analysis consists

^{3.} *Object shapes-sizes* is a categorical variable that specifies the shape and size of the object. For details see Mason et al. (2006), where they refer to it as object dimension.

in extracting static *synergies* that explain the linear and non linear correlation of the kinematic variables. We briefly discuss several dimensionality reduction methods for static data and apply them to our setting.

- In Section 5 we deal with the static neural data analysis, in which, through classical multiple linear regression we determine whether the embeddings obtained from the dimensionality reduction methods encode the information contained in the mean firing rate of neurons in the primary motor cortex. We fix specific numbers of *principal components* for each of the dimensionality reduction methods, and study how well the neurons' firing rates are explained by them.
- Section 6 corresponds to concluding remarks.
- Not included in the document, but available electronically, we have some animations that reconstruct the movement of the hand during specific trials based on the dimensionality reduction methods that we applied.

2. Experimental Design

The experiment design and data collection was made independently from us at the University of Pittsburgh's Motor Lab directed by Dr. Andrew Schwartz. The data collection was performed by Chance Spalding and Sagi Perel, whereas the first steps of data cleaning and preprocessing were performed by Samuel Clanton.

Two *Macaca mulatta* monkeys were trained to perform a reach-to-grasp task. The names of the monkeys are *Baxter* and *Vinny*. During the reach-to-grasp task two data sets were recorded: one corresponding to the hand position of the monkey during the task, and another corresponding to the activity of *some* neurons in the primary motor cortex of the monkeys (the area that codes for movement execution).

The kinematic data was recorded with a Vicon motion analysis system, recording the position of twentythree reflective markers sewn onto a custom made glove that the monkey wore on the movable hand [Figure 7A]. The neural data was obtained from a 5-Channel Electrode Mini Matrix System, from Thomas Recording GmbH [figure 7B]. The data was sampled every 5 milliseconds.

In the experiment the monkeys were positioned on a machine as illustrated in Figure 8. This device immobilizes the monkeys except from the hand with which they will perform the task.

A trial is described in figure 9. At the beginning of the trial, the monkey positioned its hand on the start pad. The industrial robot presented an object in a specific position to the monkey. A cue light turned on and the monkey reached and grasped the object. If the monkey exerted enough force, squeezing top and bottom mounted pressure sensors on the object as to surpass a threshold, then the trial was saved and the monkey was given a water reward.

During a session ten different objects were presented to the monkeys [Figure 10]. Each object was presented in one of seven different positions [Table 2]. The goal was to record five successful trials for each (object, presentation) pair during each session.

3. Data exploration

In this section we summarize the main features of our datasets.

For each monkey we have two sets of data per session: a kinematic data set, and a neural data set. Within a session the monkey performs the reach and grasp task several times. In each of these trials an object and a position are selected randomly. The goal was to have five repetitions per session of the reach and grasp for each (object, position) pair.

In Appendix 6 we show some tables summarizing the amount of data we have. In Table 9 we provide a summary of how many sessions per monkey we analyzed, together with the total number of trials, the



Figure 7: (A) Custom made glove with 20 reflective markers, whose position is tracked by an optical marker tracking system. (B) Picture of a 5-Channel Electrode Mini Matrix System from Thomas Recording GmbH. The magnified microdrive head shown on the left was inserted every session into the primary motor cortex of the macaques with the device shown at the right.



Figure 8: Experiment setting.



Figure 9: A trial.

Objects	Position
1. Button	1. -45° degrees
2. Small Handle	2. Horizontal rotation
3. Large Handle	3. 45° Rotation
4. Small Rectangle	4. 45° Flexion
5. Large Rectangle	5. 45° Extension
6. Disk	6. 45° Abduction
7. Small Precision	7. 45° Adduction
8. Large Precision	
9. Small Cone	
10. Large Cone	

Table 2: Objects used in the experiment and the different positions in which they were presented.



Figure 10: Objects presented to the monkeys.

average number of trials per session, the average number of time samples across trials and the average trial length in milliseconds. For each monkey, we display the total number of successful recorded trials per (object, position) pair (Tables 10 and 12), and also the mean number of trials per session (Tables 11 and 13). The latter information will be relevant when considering the neural analysis. The important point of these tables is that: (a) we have considerably more data for Vinny than for Baxter, that (b) for each monkey there are different numbers of repeats for each (object, position) condition, and that (c) there are no repeats at all for some conditions.

3.1 Kinematic data

The kinematic data set represents the hand posture along time for every trial, and is indexed by three parameters, the first dimension contains trials, the second dimension, kinematic variables, and the third dimension, time. The monkeys, however, do not take exactly the same amount of time to perform the reach and grasp movements for each trial, and therefore the data cube does not have all its entries. The specific trial determines the length of the data along the time dimension.

The orginal kinematic data consists of twenty-three 3D points representing the position of each of the reflective markers swen in the glove. Therefore, a total number of 69 numbers yields the hand configuration *per time point*. There were three markers for each of the following fingers: index, medium, ring and little (or pinky); four markers for the thumb; three markers for the hand; and four markers for the wrist. The total number of variables *per time point* that correspond only to the finger configuration is, consequently, fourty eight. See Figure 7 A for the exact location of the markers.

The team in the lab transformed the sixteen 3D points into twenty joint angles that represent the finger kinematics. The idea was to define an environment-independent coordinate system- that still contained most of the information of the hand position, but with a lower dimensionality. The final outcome describes each finger with four degrees of freedom: two for the metacarpophalangeal joint (MCP), one corresponding to flexion-extension and the second one corresponding to abduction-adduction; and two for the flexion-extension of the proximal interphalangeal (PIP) and the distal interphalangeal (DIP) (see Figure 1). According to Chang and Matsuoka (2006) the human thumb requires five DOF to be fully described because the axes of rotation in the thumb are neither perpendicular nor parallel and are non intersecting (Veber and Bajd, 2006). Nevertheless, in order to simplify the thumb representation, the Schwartz lab has considered the thumb to be described with the same number and type of variables than the other four fingers. This implies a loss of information for the thumb kinematics, but the information contained in the joint angles (JA) data set for the other four fingers is, essentially, the same as the information contained in the 3D position data set. It is worth to point out that the transformation from 3D marker coordinates to joint angles is nonlinear.

The joint angles data set is thus described by four variables per finger: MCP abduction-adduction, MCP flexion-extension, PIP flexion-extension and DIP flexion-extension (refer to Figure 1). We name each of these variables as a, f1, f2 and f3, preceded by the first letter of the finger they belong to: T, I, M, R and P, for thumb, index, medium, ring and pinky (Figure 11).

We now show different visualizations of the data. Figure 12 is a visualization of static data. It contains two matrices of scatter plots where the relationship of some variables are shown in a specific point in time (that is, when the hand is static). In this particular case, we considered the joint angle (JA) data set from Vinny when grasping object 7 in all possible presentations. The interesting fact from these plots is that we can draw some basic conclusions. First, some of the relationships are clearly linear. In particular the variables corresponding to the same marker in different fingers tend to be strongly linearly correlated. The ring and pinky fingers in the flexion of the marker closest to the wrist show an almost perfect line or correlation, with essentially no scatter. In the second matrix we see evidence of nonlinear relationship, specifically in relation to the variable corresponding to the marker of the thumb.

A dynamic example of the kinematic data is shown in Figures 14 and 13 where data from the joint angle data set from Baxter was selected. A specific session, and the trials corresponding to the pair (object, position) = (10, 6) were chosen. This pair corresponds to the large cone object presented with 45° of adduction. In Figure 13, variables corresponding to the joint angles of the ring finger of Baxter are shown; each trial is depicted with a different color. The raw data is shown in the left panel. The data also exhibits some *jiggling* originating from the experimental data recollection on the optical marker tracking system. Therefore, we decided to smooth the data. In this particular example the smoothing procedure was the *lowess* smoother. In the rigth panel the smoothed data is shown. Two things must be noted: first, each trial is different in length, that is, the monkey does not perform each grasping repetition exactly in the same time interval. In this case, the data is displayed after *aligning* the trials according to its last time point, that is, when the monkey is grasping the object. The second fact to note is that given a specific condition, namely (object, presentation) the trials present similarities: there is a correspondence between curves in terms of peaks and valleys, but they do not agree in their timing. These two facts will be important for our analysis.

Another example of dynamic visualization of the data is shown in Figure 14. Here a portion of the 3D data set is shown, but not in a completely raw fashion. In this display, the curves have been normalized in time in a way such that their length spans the (0,1) interval. In the figure, the smoothed 3D position and



Figure 11: Scheme of the joint angles calculated at the Motor Lab. In this work each finger is denoted by the first letter of its name. The corresponding marker is either a for abduction, or f for flexion. The flexion variables are enumerated, and their value increases as the marker they denote becomes more distal from the wrist. In this figure the index joint angles are shown. In the illustration, the thumb has four joint angles, but, in the transformation made in the Motor Lab the rightmost marker is not represented in the data. (Figure adapted from (Veber et al., 2007))



Figure 12: Vinny scatter plots. Figures that exhibit data aggregated across presentations, and show relationship between joint angles of Vinny at the time point where its hand is just about to touch object 7. (Left panel) Evidence of linear relationships. The variables plotted are the markers on the joints closest to the wrist (number 1): Tf1 = thumb flexion, If1 = index flexion, Rf1 = ring flexion, Pf1 = pinky flexion. (Right panel) Some evidence of non linear relationship: note plots between (a) Tf1 (thumb flexion 1, that is, marker closest to wrist) and Mf3 (medium flexion 3, the marker most distant from the wrist); (b) Tf1 and Rf2 (ring flexion 2, the medium marker); (c) Tf1 and Rf3. This panel show evidence that at a specific time point, there exist linearities and also non linearities. In these plots, looks like the joint angles for the four fingers are very similar, but at the same time different from the thumb. In fact, the non linear relationships happen between variables compared with the thumb.



Figure 13: Plots of the joint angles corresponding to the ring finger of Baxter when grasping the large cone (object 10) presented with 45 degrees of adduction (position 6) in a specific session (Baxter000481). The left panel shows the raw data, and the right panel the smoothed curves. Each trial is plotted in a different color, and they are aligned with respect to the last time point. Note that each trial is of different length. However the trials look similar in terms of their amplitude features (peaks and valleys), the biggest apparent difference being the timing of the features.

time normalized data from Baxter is shown. Each column corresponds to a finger; each row corresponds to a marker, with the lower row corresponding to the closest to the wrist, and the upper one to the most distal markers. In the same plot we are visualizing the forty-eight variables corresponding to the fingers position. The (x, y, z) coordinates are color coded. The data was smoothed with smoothing splines and normalized in time to correspond to the interval (0, 1).

In summary:

- we have two large kinematic data sets describing the grasping of objects from two monkeys. These data sets are almost equivalent in the information they contain, except for a small loss of information that occurs in the transformation from 3D position to joint angles.
- There is evidence of both linear and nonlinear correlations between kinematic variables.
- In addition, when considering the time component, the kinematic curves have a certain consistency: that is, they present similar features such as peaks or valleys. However, these features occur in different timings for each trial, and furthermore, each of the trials is of different length.
- It is worth noticing that the data is quite noisy due to the experimental setting, including factors like the glove and the empirical noise from the tracking system. Also there is no guarantee that both monkeys present the same behavioral features, so the analysis needs to be done independently for each monkey and results must be compared afterwards.

3.2 Neural data

Due to the nature of the electrodes used for the experiment, the array of electrodes had to be inserted and extracted for every single session and thus, for each session, different neurons were recorded. The neural data therefore consists of the activity of several neurons per session. Their activity is, in theory, correlated in some way with the kinematic variables recorded in the same session. Table 3 includes the total number of neurons recorded per session per monkey.



Figure 14: Matrix of plots of the forty-eight variables that correspond to the positions of the fingers and thumb along trials of the 3D position data set of Baxter. The columns of the array correspond to different fingers: thumb, index, middle, ring and pinky. The rows correspond to the markers. Due to the location of the markers in the glove, in the 3D data set, the thumb has information of four reflective markers, whereas the rest of the fingers have only three reflective markers. All the trials across sessions are plotted here as different curves, and where normalized to the [0,1] interval, so that the rescaled trials have identical lengths. The x-coordinate is depicted in red, the y-coordinate in green and the z-coordinate in blue.

Session	Num. valid neurons	Session	Num. valid neurons
Vinny000639	5	Baxter000467	7
Vinny000658	6	Baxter000475	3
Vinny000661	8	Baxter000478	4
Vinny000669	4	Baxter000481	5
Vinny000673	4		
Vinny000676	1		
Vinny000678	4		
Vinny000680	3		
Vinny000682	4		
Vinny000683	4		
Vinny000687	1		
Vinny000690	5		
Vinny000691	5		
Vinny000693	5		
Vinny000694	5		
Vinny000695	3		
Total number	67	Total number	19

Table 3: Number of valid neurons per session per monkey. In the last row we show the total number of neurons analyzed per monkey

We constructed raster plots and peristimulus time histograms (PSTH) for each of the studied neurons. These plots display aggregated activity and timing of discharge of the neurons in relation to the reach and grasp task. We built these type of plots aligning the trials in two ways: with respect to (a) the beginning of the trial, and (b) to the end of the trial. Two main qualitative observations resulted from this approach. First, within the population of studied neurons there exist neurons whose peak of activity is towards the beginning of the trial, that is, towards the beginning of the reach; but there are also neurons whose peak of activity is towards the end of the trial, that is, during the grasp. Figure 15 shows an example of each. It is sensible to hypothesize that the former type of neurons modulate reach, and the latter, grasp. And second, there exist neurons whose discharge is object dependent. An example is shown in Figure 15 panel (B) this neuron fires preferentially to trials of all objects, except for small handle. In effect, as time progresses, the neuron becomes silent in all trials associated with the small handle. The silent period is emphasized in the graph with an orange striped rectangle.

Neural commands of voluntary movement are generated in the brain before the kinematic behavior is observed. That is, there exists a *lag* between the neural activity representing a command and the kinematic response. As an example, observe Figure 16 where neuronal activity is displayed simultaneously with a summary of the kinematic behavior (that will be introduced in Section 4.2, but that represents amount of motion of the fingers). Notice, for instance, neuron spk003a which consistently shows a higher frequency of spikes several dozens of mS before the largest peak in the energy function happens, one might hypothesize that this neuron is signaling finger motion in the displayed trials. A challenge is to identify the *time lag* for each neuron.

Finally, the number of neurons that contain more than 100 trials for Baxter are 13 and for Vinny 37. In summary:

• The nature of the electrodes used for the experiments, required them to be inserted and removed for each experimental session. As a consequence, the number of neurons in our population which have



(A) Neuron predominantely active towards the beginning of trial

(B) Neuron predominantely active towards the end of trial, displaying object preferences Aligned to go cue



Figure 15: Raster plots and peristimulus time histograms showing the activity of the two neurons: Panel (A) Session Vinny 000661, neuron 5b. 292 trials. Panel (B) Session Vinny 000669, neuron 4a. 216 trials. On the (left side) the raster plot and the PSTH are aligned to the go cue (t=0), whereas the (right side) shows the plots aligned to the end of trials. The different colors of the spikes in the raster plot denote different objects, the spikes color code is shown in the graph. Magenta triangles denote beginning and end of trials. Panel (A) shows how the shape of the aggregated activity of the neuron changes as a function of alignment type. The PSTHs in panel (A) (and also in panel (B)) shows two peaks of activity: a major one, and a minor one. Observe that, regardless of the alignment, the order of peaks of activity does change as a function of alignment, the larger peak always preceeds the lower peak in panel (A) regardless of the alignment. This is also the case for the neuron shown in panel (B) where the larger peak always follows the lower peak (alignment with respect to end of trial is not displayed). In a nutshell, regardless of alignment, we can observe the timing of the highest peak of activity of the neurons and hypothesize which part of the reach-grasp they are modulating.



Figure 16: Spike trains and measure of movement of fingers for different trials. The kinematics of the fingers was summarized in the *energy function* that will be introduced in section 4.2). The upper pannels display a cross in the (time, neuron) position if the neuron presented an action potential at the given time. Neural activity was binned in 5mS intervals, to correspond to the frequency at which the kinematics data was sampled. Crosses in magenta denote that more than one action potential happened in the corresponding 5mS bin. In this case, all the energy functions were centered at zero, which explains why some energy values are displayed as negative.

more than 100 kinematic trials associated is quite small, specially for Baxter. This data limitation might restrict the scope of our conclusions.

• Through qualitative analysis we observed in our population the existence of neurons that fire preferentially towards the beginning of the trials, and neurons that fire preferentially towards the end.

4. Obtainment of synergies or eigengrasps: kinematic analysis

Given the clear correlations that exist during hand movements, and the evidence for linear relationships (Figure 12), it is sensible to try linear methods. However, there is no evidence that the relationships between variables are only linear. To the best of our knowledge the possibility that non linear low dimensional representations of the data can explain the grasping kinematics more accurately than linear methods, has not been explored (see Table 1). And thus we aim at exploring alternative nonlinear dimensionality reduction methods in this section.

There are different decisions to be made to extract a grasping synergy. In this section we first review the considered dimensionality reduction methods, secondly we describe the specifics of the problem of obtaining the grasp synergies, then we discuss alternatives and justify the decisions taken for our analysis. Finally, we describe the metrics of goodness, and we show the results of the analysis.

4.1 Dimensionality reduction: methods

For the sake of clarity consider the following notation: $X \in \mathbb{R}^{N \times K}$ is a data matrix where N is the number of observations, and K is the number of variables.

Dimensionality reduction techniques deal with the problem of finding or building a *small* number of components that describe most of the information that the original data set has. In other words, dimensionality reduction consists of transforming a dataset X with dimensionality K into a new dataset

Y with dimensionality k where K > k, while retaining some properties of the data, like variance structure or geometry, as much as possible. In general, neither the geometry/topology of the low dimensional representation of the data, nor the number of parameters needed to account for the observed properties of the dataset X are known. Thus, dimensionality reduction can only be solved by assuming certain properties of the data (such as intrinsic dimensionality or specific geometry/topology).

According to the geometric assumptions made on the data, there are linear and non-linear techniques for dimensionality reduction. Linear techniques assume that the data lie on or near a linear subspace of a high-dimensional space. Non-linear techniques do not rely on the linearity assumption, and thus a more complex embedding of the data in the high-dimensional space can be identified.

Principal Component Analysis (PCA). We begin with a linear approach: Principal Component Analysis. The specifics of the procedure are shown in Algorithm 1, and the main ideas are explained as follows. PCA finds a low dimensional representation of the data points that best preserves their variance as measured in the high dimensional input space, this is equivalent to assuming that the original data set lies on or near a linear hyperplane. Through PCA a new orthogonal basis is found by diagonalizing the centered covariance matrix of the dataset. The principal components are the eigenvector basis of the covariance matrix. PCA provides a *generalization property* (or *projection property*) where new data points (which do not belong to a set of training data points) can be embedded in a low dimensional space through a mapping computed by PCA (see Step 4 of Algorithm 1). Although this step is straightforward for PCA, such is not necessarily the case for other methods.

If the structure of the dataset is not linear, PCA will not account for all of the structure. A *kernelized* version of PCA proposed by Schölkopf et al. (1998) can extract more information by using suitable non linear features.

Kernelized versions: non-linear approach Kernelized methods is the term used for procedures that have been generalized through the *kernel trick*. Procedures that are succeptible of being generalized in this manner are those where the relationships between the data inputs can be written solely in terms of dot products. The *kernel trick* consists of substituting Euclidean dot products in the space of input patterns by generalized dot products in a large dimensional *feature space*.

In the large dimensional feature space the inter relationships of input patterns is measured through the Gram matrix of abstract inner products $\langle \phi(x_i), \phi(x_j) \rangle$ where $\phi : X \to \mathcal{H}$ is a (usually non linear) map from the data space X into the feature space \mathcal{H} . The feature space \mathcal{H} is assumed to be a Hilbert space of real valued functions defined on X. Given arbitrary points x_i, x_j in a training set X, what is needed is a non-linear function (or kernel) $\mathcal{K}(x_i, x_j)$ that represents the similarity between x_i and x_j . In practical terms, one just needs to substitute the *a priori* chosen kernel function $\mathcal{K}(w, z)$ for all occurrences of $\langle \phi(w), \phi(z) \rangle$, and the choice of \mathcal{K} implicitly determines the mapping ϕ and the feature space. This function \mathcal{K} will produce as many values as the square of the number of elements that there are in X: $k_{ij} = \mathcal{K}(x_i, x_j)$, i.e. $O(N^2)$ kernel evaluations. Using a kernel instead of a dot product in the input space effectively corresponds to mapping the data into a possibly high-dimensional space by a map ϕ , and taking the dot product there (Schölkopf et al., 1998).

To sum up, given a priori chosen kernel function \mathcal{K} and a data set X the above procedure results in that one only needs to compute dot products between mapped patterns (and not the mapped patterns explicitly) and write the algorithm in terms of these relationships. In this way it is easy to generalize the original procedure through the \mathcal{K} -defined dot product in the *feature space*.

Kernel PCA therefore finds principal components which are not linearly related to the input space. It does so by applying principal components analysis in the space produced by a non linear mapping, a feature map. The idea is that the low dimensional hidden structure might be easier to discover in the feature space (Algorithm 2).

There are some smaller difficulties with applying PCA to the kernel matrix. One of them is that, in general, it is not necessarily feasible to build the covariance matrix in the feature space, because it would need to be infinite-dimensional. Another difficulty is that PCA assumes that the input patterns have

input : $X \in \mathbb{R}^{N \times K}$ where N = number of observations, and K = number variables, [Optional: $x \in \mathbb{R}^{1 \times K}$ test example] output: $Y \in \mathbb{R}^{N \times k}$, [Optional: $y \in \mathbb{R}^{1 \times k}$ k-dimensional encoding of test example, \hat{x} reconstruction of test example] 1. Prepare data X for PCA 1.1 Center data X $\tilde{X} \leftarrow X - \texttt{mean}(X)$ 1.2 Obtain covariance matrix C of centered data $C \leftarrow \tilde{X}^T \tilde{X}$ where $C \in \mathbb{R}^{K \times K}$ 2. Compute the spectral decomposition of C $\Lambda \leftarrow \texttt{Eigenvalues}(C)$ where Λ is a diagonal matrix in $\mathbb{R}^{K \times K}$ $V \leftarrow \texttt{Eigenvectors}(C)$ 2.1 Obtain the intrinsic linear dimensionality of the data k $k \leftarrow \texttt{numberTopEigenvalues}(\Lambda)$ 2.2 Build the matrix with the k eigenvectors corresponding to the top eigenvalues, and the diagonal matrix of top k eigenvalues $\tilde{V} \leftarrow V^{K \times k}$ $\tilde{\Lambda} \leftarrow \Lambda^{k \times k}$ 2.3 Obtain the square root element by element of the diagonal matrix of top k eigenvalues $\tilde{\Lambda}$ $S_{\tilde{\Lambda}} \leftarrow \operatorname{sqrt}(\tilde{\Lambda})$ where $S_{\Lambda} \in \mathbb{R}^{k \times k}$ and is diagonal 3. Compute k-dimensional embedding of training data $Y \leftarrow \frac{1}{\sqrt{K}} X \tilde{V} S_{\tilde{\Lambda}}$ [Optional 4.] Encode test example 4.1 Obtain the matrix $J_{S_{\lambda}}$ that contains in the diagonal the inverse of the elements of the diagonal of $S_{\tilde{\Lambda}}$ $J_{S_{\tilde{\Lambda}}} \leftarrow \operatorname{diagInv}(S_{\tilde{\Lambda}})$ 4.2 Obtain encoding of the test example $y \leftarrow x V J_{S_{\tilde{\lambda}}}$

Algorithm 1: Principal Component Analysis (PCA)

zero mean, but after applying the feature map, it is not possible to assume that feature vectors have zero mean. There are ways around these complications: for instance, starting from an unnormalized kernel \mathcal{K} with unnormalized feature space, one can obtain a kernel $\tilde{\mathcal{K}}$ that is centered in the feature space through an additive normalization (Schölkopf et al., 1998). The result of this procedure is written in Step 1 of Algorithm 2. After this corrective step is performed, the usual PCA can be applied.

One issue we have not discussed is the selection of the kernel function. The selection of kernel is a tricky step in the application of kernelized methods. It is somehow an *art* in the sense that unless there is clear *a priori* information of what kind of structure the data has, the selection of a kernel is rather arbitrary. Three classical examples of kernels are the trivial (or linear) kernel $\mathcal{K}(x_i, x_j) = \langle x_i \cdot x_j \rangle$ (which is queivalent to linear PCA), the polynomial kernels $\mathcal{K}(x_i, x_j) = (1 + \langle x_i \cdot x_j \rangle)^p$ for $p \in \mathbb{Z}^+$ and the Gaussian kernels⁴ $\mathcal{K}(x_i, x_j) = exp\left(-\frac{\|x_i-x_j\|^2}{\sigma^2}\right)$. The trivial kernel identifies the feature space with the input space, the polynomial kernel maps the inputs into a feature space of dimensionality $O(K^p)$, and the Gaussian kernel maps the inputs onto the surface of an infinite-dimensional sphere (Weinberger et al., 2004). In our particular case, one intuition we have from the data is provided by Figure 12: we see plenty of linear relationships, but also some curvature given by the variable associated with the thumb. We observe a curvature that could be represented potentially with either a Gaussian kernel or a polynomial kernel of degree higher than 1.

Kernel PCA also exhibits the generalization property as stated in Step 4 of (Schölkopf et al., 1998). In general, it has been shown that kPCA is equivalent to the Nystrom approximation of eigenfunctions (Bengio et al., 2003). In this case, to extract the principal components with respect to kernel \mathcal{K} of a test point x what we need are the projections of the eigenvectors v_p in the feature space. Consider x a test point, then we can compute the dot product between the p-th eigenvector and $\phi(x)$ as follows $\langle v_p, \phi(x) \rangle = \frac{1}{\sqrt{\lambda_p}} \sum_{i=1}^n \alpha_i^p \mathcal{K}(x_i, x)$ where the factor $\frac{1}{\sqrt{\lambda_p}}$ ensures that $\langle v_p, v_p \rangle = 1$.

All the methods considered in this section (PCA, LDA and kernel methods) require the spectral decomposition of a matrix. The complexity of such a decomposition is cubic in the dimension of the matrix. Therefore, when the dimension is large the spectral decomposition is computationally expensive and can be practically unfeasible. The dimension of the matrix to be decomposed for PCA is exactly the number of variables, whereas for kernel methods, the dimension scales as the number of training examples. In the usual case, the number of observations is very large with respect to the number of variables. Therefore one drawback for kernel methods is that they require the spectral decomposition of a high dimensional matrix, which can be very expensive.

Linear supervised method: linear discriminant analysis PCA (and kPCA) are unsupervised methods of dimensionality reduction. If X is a data set whose points belong to a specific class, then it is possible to apply dimensionality reduction on X taking into account the classes of the data points, that is, to do dimensionality reduction in a supervised way. A classic method to reduce dimensionality in a supervised way is Fisher Linear Discriminant Analysis (LDA). In LDA the objective is to maximize the ratio of between class variability over within class variability.

Fisher Linear Discriminant analysis maximizes the function $J(w) = \frac{w' S_B w}{w' S_W w}$, where S_B is the between class scatter matrix and S_W is the within class scatter matrix. The scatter matrices are defined as $S_B = \sum_c (\mu_c - \bar{x})(\mu_c - \bar{x})^T$ and $S_W = \sum_c \sum_{i \in c} (x_i - \mu_c)(x_i - \mu_c)^T$ for c the classes where the input vectors belong, \bar{x} the overall mean of the data, and μ_c the mean of the data that belongs to class c. The scatter matrices are proportional to the covariance matrices, thus the solution using the scatter matrices or the covariance matrices is the same. The maximization problem is solved using the Lagrangian and KKT conditions, and it is reduced to a spectral decomposition in an analogous manner like PCA.

The advantage of this method is that it is supervised. We decided to apply it as a reference of how well the dimensionality reduction can be done in an unsupervised manner compared to the supervised manner. Since we only decided to use this method as a reference, we do not explore further other supervised

^{4.} $||x|| = \sqrt{\langle x \cdot x \rangle}$

input : $X \in \mathbb{R}^{N \times K}$ where N = number of observations, and K = number variables; k the low dimension, \mathcal{K} a kernel selected a priori, [Optional: $x \in \mathbb{R}^{1 \times K}$ a test example] output: $Y \in \mathbb{R}^{N \times k}$, [Optional: $y \in \mathbb{R}^{1 \times k}$ k-dimensional encoding of test example] 0. Compute kernel on data $\mathcal{K} \leftarrow \mathcal{K}(X, X)$ where $\mathcal{K} \in \mathbb{R}^{N \times N}$ 1. Compute normalized kernel $\tilde{\mathcal{K}}$ in the feature space $\tilde{\mathcal{K}} \leftarrow (I - \frac{1}{N} \bar{1} \bar{1}^T) \mathcal{K} (I - \frac{1}{N} \bar{1} \bar{1}^T)$ 2. Compute the spectral decomposition of $\tilde{\mathcal{K}}$ $\Lambda \leftarrow \texttt{Eigenvalues}(\mathcal{K})$ where Λ is a diagonal matrix in $\mathbb{R}^{N \times N}$ $V \leftarrow \texttt{Eigenvectors}(\mathcal{K})$ 2.1 Build the matrix with the k eigenvectors corresponding to the top eigenvalues, and the diagonal matrix of top k eigenvalues $\tilde{V} \leftarrow V^{N \times k}$ $\tilde{\Lambda} \leftarrow \Lambda^{k \times k}$ 3. Compute k-dimensional embedding of training data 3.1 Obtain the square root element by element of Λ $S_{\tilde{\Lambda}} \leftarrow \operatorname{sqrt}(\tilde{\Lambda})$ where $S_{\tilde{\Lambda}} \in \mathbb{R}^{k \times k}$ and is diagonal 3.2 Obtain the low dimensional representation of the training data $Y \leftarrow VS_{\tilde{\Lambda}}X$ [Optional 4.] Encode test example 4.1 Obtain the matrix that contains in the diagonal the inverse of the elements of the diagonal of $S_{\tilde{\Lambda}}$ $J_{S_{\tilde{\Lambda}}} \leftarrow \operatorname{diagInv}(S_{\tilde{\Lambda}})$ 4.2 Obtain normalized kernel evaluated in the test point 4.2.1 Compute kernel vector: apply input kernel function on training data and test point $\mathcal{K}_x \leftarrow \mathcal{K}(X, x)$ with $\mathcal{K}_x \in \mathbb{R}^{N \times 1}$ 4.2.2 Normalize kernel vector $\tilde{\mathcal{K}}_x \leftarrow \mathcal{K}_x - \bar{1} \cdot \sum_{i=1,\dots,N} \mathcal{K}_{xi} - B + \bar{1} \left(\frac{1}{N^2} \sum_{i=1,\dots,N} \sum_{j=1,\dots,N} \mathcal{K}_{i,j} \right) \text{ where } \bar{1} \in \mathbb{R}^{N \times 1}, B \in \mathbb{R}^{N \times 1}$ with its *i*-th coordinate: $B_i = \sum_{\alpha=1,\dots,N} \mathcal{K}_{\alpha,i}$ 4.3 Obtain encoding of the test example $y \leftarrow (J_{S_{\tilde{\Lambda}}} \cdot \tilde{V}^T \cdot \tilde{\mathcal{K}}_x)^T$ Algorithm 2: kernel Principal Component Analysis (kPCA)

methods nor the non linear extensions, like kernelized discriminant analysis((Mika et al., 1999),(Baudat and Anouar, 2000)).

4.2 Synergies of grasping

In order to formalize our problem, we define the following notation. Consider one specific monkey, and the joint angles of its hand during the reach-and-grasp task. Then for a specific trial (or repetition) rthere are (per time point) K = 20 joint angles corresponding to three flexion angles and one abductionadduction angle per finger (including thumb). Let $X_t^r \in \mathbb{R}^K$ be the row vector of joint angles at a specific time point t on the reach-grasp task for trial (or repetition) r. The time index t takes values from 1 (at the beginning of the trial), to T(r) that is the trial dependent length. Let N be the number of trials.

Within the context of our problem, the first decision is whether all the kinematic data or only a portion of it will be used to obtain the grasping synergies through dimensionality reduction methods.

4.2.1 All time versus time of interest

In order to frame the problem as a classical dimensionality reduction task we need to build a matrix Q which contains in its rows the trials and in its columns the variables.

Many authors ((Soechting and Flanders, 1997; Todorov and Ghahramani, 2004; Thakur et al., 2008)) have built \mathcal{Q} by considering each time point as an independent trial and have stacked the variables corresponding to different time points and different trials all together:

$$\mathcal{Q} = \begin{pmatrix} X_1^1 \\ \vdots \\ X_{T(1)}^1 \\ \vdots \\ \overline{X_{T(N)}^N} \\ \vdots \\ X_{T(N)}^N \end{pmatrix}$$

This approach (a) removes the temporal information in the data, (b) treats each time point as independent from the others, (c) builds a matrix of size $M \times K$, where $M = \sum_{r=1}^{N} T(r) + N$. In our case $M_{Baxter} \approx$ 104×10^3 and $M_{Vinny} \approx 614 \times 10^3$. This order of magnitude is feasible for linear dimensionality reduction methods, but it is computationally too expensive for some non linear methods (see Section 4.1), where an spectral decomposition of a $M \times M$ matrix should be performed.

An alternative is to select a specific time of interest (like in (Santello et al., 1998)). In that way, if we define the time of interest t_i^* for trial *i*, we can build *D* such that the *i*-th row of *D* corresponds to: $\mathcal{Q}_{i,\cdot} = X_{t_i^*}^i$ and $\mathcal{Q} \in \mathbb{R}^{N \times K}$. In our case, $N_{Baxter} \approx 670$ and $N_{Vinny} \approx 3,000$. These sizes are treatable by non linear methods.

Which time of interest?

Our goal is to understand the grasp kinematics. If we assume that the selection of objects samples the space of possible grasping configurations, a first reasonable time point of interest t_r^* for trial r is the moment when the monkey is softly holding the object, or equivalently, the point in time when the monkey is barely touching the object with the hand configuration ready to grasp the object.

In the experiments it is observed that the monkey exerts a lot of force when grasping the object, so much that the posture of the hand gets distorted. Therefore the lab reached the convention of selecting the time point corresponding to 150mS before the trial finishes. That is, 150mS before the threshold of force of the sensors placed on the objects is surpassed. This heuristic estimate of the time of interest given by the lab is reasonable. However we seeked a more principled way to estimate the time of interest.

As stated above, the time of interest corresponds to the time when the configuration of the hand is fully shaped to the object, and before it gets distorted due to excess of strength. In other words, the time point of interest is a time point at the end of the reach where the movement of the fingers is very close to zero. In order to identify this time point, a function that quantifies the amount of motion of the fingers is needed. We have two datasets to quantify the amount of motion of the fingers: the 3D position of the markers or the joint angles. Given that during the joint angle transformation there is a slight loss of information (see Section 3.1), the most accurate estimation of motion comes from the position of the sixteen 3D markers on the fingers (including thumb).

Let $Z^r(t) \in \mathbb{R}^{16\times 3}$ be the matrix containing the 3D measurements of the 16 marker positions for trial r. Let $\dot{Z}^r(t)$ denote the velocities of the markers, then define $G^r(t) = [\dot{Z}^r(t)][\dot{Z}^r(t)]'$. $G^r(t)$ is the matrix of inner products of marker velocities. The sum of the squared magnitudes of the velocities accross markers is: $E^r(t) = tr(G^r(t)) = tr([\dot{Z}^r(t)][\dot{Z}^r(t)]')$. The function $E^r(t)$ summarizes the magnitude of motion during a trail, and it is invariant under rotations of the 3D variables.

In general, the energy profile of a specific trial starts close to zero, presents one or several bumps, and goes back to be *close* to zero when the hand is already grasping the object, because relatively little motion is happening. In addition, in every case there is always a well defined maximum for the energy. In Figure 17 we show three different examples of energy profiles for three trials.

We mentioned that, in general, when the trial finishes the motion is close to zero. The notion of *close*ness to zero is trial dependent. In order to define a reasonable threshold that indicates when the motion of the fingers in a trial is close to zero, we define: $\epsilon_{max}^r = max_t \{E^r(t)\}$ and $\epsilon_{min}^r = min_t \{E^r(t)\}$, and define the threshold $\eta^r(\rho) = \rho \cdot (\epsilon_{max}^r - \epsilon_{min}^r)$ where ρ is the proportion of the difference we want to consider. We can further define the set of time points that intersect $\eta^r(\rho)$ as follows: $\mathbb{J}_{\rho}^r =$ $\{t|t \in \{1, \ldots, T(r)\}$ and $E^r(t) = \eta^r(\rho)\}$.

In the case when the energy function is unimodal, the energy intersects $\eta^r(\rho)$ at two points: when the hand is increasing velocity from starting the reach, and when the hand is very close to or actually grasping the object. In this case, the set \mathbb{J}_{ρ}^r has two elements. If the energy function has more than one bump, then \mathbb{J}_{ρ}^r can have more than two elements.

In general we can define the following time points (see Figure 17 Example 2 for illustration): the time when the maximum of the energy occurs:

$$t_{max}^{r} = \{t \in \{1, \dots, T(r)\} \mid E^{r}(\tau) \le E^{r}(t) \,\forall \, \tau \in \{1, \dots, T(r)\}\},\$$

the time when the energy crosses the threshold $\eta^r(\rho)$ for the first time:

$$t_{firstLess}^{r}(\rho) = \left\{ t \in \mathbb{J}_{\rho}^{r} \, | \, \text{there is no } \tau \in \mathbb{J}_{\rho}^{r} \, \text{and} \, \tau < t \right\},\$$

the last time the energy crosses the threshold before reaching its maximum:

$$t_{lastLess}^r(\rho) = \left\{ t \in \mathbb{J}_{\rho}^r \, | \, \text{there is no } \tau \in \mathbb{J}_{\rho}^r \, \text{and} \, \tau < t < t_{max}^r \right\},$$

the first time the energy gets close to zero after reaching its maximum:

$$t^r_{firstHigher}(\rho) = \left\{ t \in \mathbb{J}_{\rho}^r | \, \text{there is no} \, \tau \in \, \mathbb{J}_{\rho}^r \, \text{and} \, t^r_{max} < \tau < t \right\},$$

and, the last time the energy gets close to zero before the trial comes to an end

$$t^r_{lastHigher}(\rho) = \left\{ t \in \mathbb{J}_{\rho}^r | \text{there is no } \tau \in \mathbb{J}_{\rho}^r \text{and} \tau > t \right\}.$$

Note that it is true that: $t_{firstLess}^r(\rho) \leq t_{lastLess}^r(\rho) < t_{max}^r < t_{firstHigher}^r(\rho) \leq t_{lastHigher}^r(\rho)$. The time of interest we propose as an estimate for the time when the hand is about to hold the object

The time of interest we propose as an estimate for the time when the hand is about to hold the object is the time of the largest intersection between $\eta^r(\rho)$ and $E^r(t)$, that is: $t^r_{lastHigher}(\rho)$. We use the other time points to study the preshaping of the grasp along the reach movement.



Example 1: Energy profile. Vinny: session 673, trial 238. (Small handle, -45 rotation)

Example 2: Energy profile. Baxter: session 475, trial 147. (Large cone, 45 abduction)



Example 3: Energy profile of an outlier. Baxter: session 467, trial 25. (Button, -45 rotation)



Figure 17: Energy profiles. Three trials have been selected to illustrate the energy profiles during a reachgrasp movement. On the left hand column the total energy is shown; two horizontal lines show $\eta^r(\rho)$ for $\rho = 0.05$ and $\rho = 0.1$. On the rigth hand column the energy profile is decomposed into the energy for each of the fingers; the purpose of these plots is to illustrate the contributions to the total energy and the coupling of different fingers during the preshaping of the hand and the grasping. In the total energy plot of Example 2 we show the definition of our times of interest: (a) $t^r_{firstLess}(\rho)$, (b) $t^r_{lastLess}(\rho)$, (c) t^r_{max} , (d) $t^r_{firstHigher}(\rho)$, and (e) $t^r_{lastHigher}(\rho)$.

4.2.2 OUTLIER DEFINITION

In addition, the energy function provides us with the means of defining an outlier. There are trials where the energy function is not close to zero at the end of the trial. This could mean either that (a) there was an error in defining the end of the trial, or (b) the configuration of the hand at the defined end of trial is not stable, that is, there is still a lot of finger movement.

We defined a trial as an outlier when the motion of the fingers as the end of the trial has not gotten close enough to zero, that is when there is no $t \in \{1, ..., T(r)\}$ such that $\eta^r(\rho) = E^r(t)$, or equivalently, there is no $t^r_{lastHigher}$. Refer to Figure 17 for an example.

Under this definition Baxter had a 10% and Vinny 2% of outliers when $\rho = 0.05$, and 5% and 1.5% when $\rho = 0.10$. This indicates that Vinny reached and held a steadier hand configuration compared to Baxter at the end of the grasp.

4.2.3 Sampling versus averaging

Once a time point of interest is fixed, and outliers are removed, we end up with a matrix of size $N \times K$ where $N_{Baxter} \approx 600$ and $N_{Vinny} \approx 3,100$. Given the time/computation complexity of the non-linear methods (see Section 4.1), the analysis of Vinny's data takes a large amount of time. In order to be able to run several experiments with 10-fold cross validation, we considered ways of summarizing the data. In particular, we considered (a) sampling a specific number of trials, preserving the distribution of frequencies of (object, position) pairs; and (b) averaging across trials fixing an (object, position) pair, to obtain a representative trial from each condition.

In the case of sampling we considered all the trials for Baxter because the number of trials for him is not prohibitive for applying kernelized dimensionality reduction methods. However, for Vinny we sampled 1,000 trials. Averaging, on the other hand, drastically summarized the data into a matrix that contained only one representative mean trial per (object, position) pair for each of the monkeys. To obtain the representative mean trial, we averaged across trials fixing (object, position).

Figure 18 shows an squematic view of the analysis after outliers removal.



Figure 18: An schematic diagram of the analysis after outliers removal. Summarizing refers either to (a) sampling certain amount of trials preserving the distribution of number of trials that belong to a specific (object, position) condition; or to (b) averaging across trials fixing (object, position). The latter yields a drastic reduction of the number of trials in the matrix that will be used to obtain the mapping. One of the objectives of the analysis is to determine which strategy yields better results acording to your measures of goodness.

4.2.4 Obtaining the low dimensional representations

The data matrix Q that we built in the manner described in the last few sections was used to obtain the mapping that transforms the data from the high dimensional representation to the low dimensional representation. That is, Q is the input matrix to the dimensionality reduction methods we considered, namely, PCA, kPCA and LDA. This matrix was spectrally decomposed in the case of PCA and LDA; and it was used to build the kernel matrix for kPCA (in all the variants we decided to use). The eigenvalues and eigenvectors from these decompositions are the elements used to define the mapping to embed new datapoints, as explained in Step 4 of the algorithms presented in Algorithm 1 and 2. All the trials for Baxter and Vinny were projected into a low dimensional representation through those formulas, and further analyzed in our work.

4.3 Measures of goodness of eigengrasps

The kinematic data set consists of a set of vectors describing the configuration of the hand at the selected time point. Each of these vectors is associated with an object the monkey is grasping. This framework corresponds naturally to a supervised classification task, where the aim is to learn a function that assigns each vector to one of a finite number of discrete categories, in this case, the objects. Therefore we can take advantage of the framework we have, and use classification accuracy to *indirectly* quantify the goodness of the dimensionality reduction in the sense that we get an idea of how much *information* we preserve regarding the object being grasped. However, it is important to mention that good dimensionality reduction does not imply good classification accuracy, and good classification accuracy does not imply good dimensionality reduction. Two classic examples are shown in Figure 19. Furthermore, the results might be classifier dependent, that is limited by the classifier assumption. Therefore, the results of classification should only be taken as an indication of information being preserved, but not proof.

In order to perform a classification task, a classifier must be trained. We chose two classifiers to be trained: a discriminative and a generative one, and we briefly explain them below.



Figure 19: Good dimensionality reduction does not imply good classification, neither the other way around. In the example at the top the green line yields the best dimensionality reduction, but the classification is bad. Thus, good dimensionality reduction does not imply good classification. In the pannel at the bottom the classification accuracy of the red line is excellent, but clearly the dimensionality reduction should yield a sine function, as oppose to the straight line which determines the best classification. Thus, good classification does not imply good dimensionality reduction.

4.3.1 NAIVE BAYES CLASSIFIER

The Naive Bayes classifier is a supervised generative classifier based on Bayes rule. It makes the assumption that the considered variables are conditionally independent given the class to which the input data belongs. This assumption is made to reduce the complexity of general Bayesian classifiers from $O(2^n)$ to O(n) (Mitchell, 1997).

Consider \mathcal{O} a random variable indicating the object that is being grasped in a specific trial, and $X \in \mathbb{R}^{N \times K}$ the matrix of joint angles. The Naive Bayes classifier models the joint probability of observed variables given the object class as:

$$P(X_1,\ldots,X_K|\mathcal{O}) = \prod_{i=1}^K P(X_i|\mathcal{O}).$$

Suppose we estimate the distributions $P(\mathcal{O})$ and $P(X_i|\mathcal{O})$ from the training data in the way we will explain below, and consider a new reach-grasp trial, that is, new observed variables $\{X_i\}_{i=1}^K$. Then, by Bayes rule, the Naive Bayes classifiation procedure indicates that the object to be assigned to the new reach and grasp trial is given by:

$$O \leftarrow argmax_{o_j} P(O = o_j) \prod_{i \in Objects} P(X_i | \mathcal{O} = o_j).$$

Since the joint variables are continuous variables, we make the assumption that each X_i is normally distributed, and is defined by the mean and the variance specific to the variable X_i and the class o_j . The way of obtaining the mean and the variances of each Gaussian is through the expressions:

$$\mu_{ij} = \mathbb{E}(X_i | \mathcal{O} = o_j)$$
$$\sigma_{ij}^2 = \mathbb{E}((X_i - \mu_{ij})^2 | \mathcal{O} = o_j).$$

And the priors on \mathcal{O} can be estimated as: $\pi_j = P(\mathcal{O} = o_j)$

We selected this classifier because of its simplicity, and low complexity. Also beacuse it has been applied successfully in several applications many times in spite of the conditionally independent assumption being invalid.

4.3.2 Multi-Class Support Vector Machine

Suport Vector Machines (Boser et al., 1992) were developed from statistical learning theory and do not assume any probabilistic model for the data. They were theoretically motivated, and posed as an optimization problem where the idea is to maximize the margin between the class boundary and the training patterns (which is equivalent to minimizing the maximum loss). The function to minimize is an additive combination of training error and a complexity term. A convenient property is that the optimization problem to be solved is convex (thus there is no local minima), and is solved using classical optimization theory: the Lagrangian, the Karush Kuhn Tucker conditions, and duality theory. The resulting classification function only depends on the (few) training examples that are closest to the decision boundary, the so-called *support vectors*, and thus the solution is sparse.

The classical SVM theory was developed for binary problems. We have a multi-class problem. A traditional extension from a binary classifiaction framework to a multi-class framework is to decompose the problem into multiple independent binary class tasks (or the one-versus-all approach). In our case we chose to use Tsochantaridis et al. (2005) approach where the notion of separation margin is generalized (Crammer and Singer, 2001), the multiclass classification is framed as a constrained optimization problem with a quadrative objective function with a potentially prohibitive number of constraints, but where the problem is solved through a cutting plane algorithm that solves the problem in polynomial time. We used their implementation.

The advantages of using SVM is that it is a well studied classification method with the convenient theoretical properties afore-mentioned, and that it has been successfully applied in areas like computational biology, natural language processing, optical character recognition, and others. There are also several implementations of the methods. The main disadvantage is speed, but given the size of our datasets in our case it is not hindering.

4.4 Experiments

We considered the joint angles data set for each monkey. And performed the following steps:

- 1. Fix:
 - (a) A specific time of interest. To study the hand configuration when object is being softly grasped, we tried the heuristic given by lab, and $t_{lastHigher}(\rho)$. In addition, to study the preshaping of the hand we considered $t_{firstLess}(\rho)$, $t_{lastLess}(\rho)$, t_{max} , $t_{firstHigher}(\rho)$, $t_{lastHigher}(\rho)$.
 - (b) An outlier definition. We tried $\rho = 0.05$, and $\rho = 0.10$.
 - (c) A summarizing strategy.
 - i. Sampling:
 - A. Baxter: we used all the data, since the total number of trials does not exceed 700.
 - B. Vinny: we sampled 1,000 trials preserving the distribution of conditions.
 - ii. Averaging across trials fixing (object, position) condition.
- 2. Build a matrix D of kinematic information with the options selected in the previous step.
- 3. Use D to obtain a mapping to project the high dimensional data to a low dimensional data. The mapping was obtained using: Principal Components Analysis, Linear Discriminant Analysis, and kernel Principal Components Analysis with polynomial kernel (degree 2 and 3) and with Gaussian kernel ($\sigma = 0.3, 0.5, 0.8, 1, 1.5, 2, 2.5$)
- 4. Obtain the low dimensional embeddings of the original data through the mappings obtained in the previous step.
- 5. We applied the classification methods to: the original data, and the reduced data. We considered the following number of components for the reduced data:
 - (a) PCA, kPCA: 2, 3, 4 to 20 in steps of size two.
 - (b) LDA: 2, 3, 4, 6, 8.
- 6. Finally, to get a sense of the preservation of the data, we obtained the 10-fold cross validation accuracy from all the D matrices with:
 - (a) Multi-class support vector machines
 - (b) Naive Bayes classifiers

We report the main results in the following section.

4.5 Results for kinematic analysis

We compared the performance of the kPCA-Gaussian reduced data for all the different number of components for the different values of σ . And after 10 fold cross validation we selected $\sigma = 2$ because this value yielded the best results for both Baxter and Vinny.

Also, we found that the performance of the reduced data when kPCA-polynomial degree 2 and 3 was very similar, so we focused on the results of kPCA-polynomial degree 2.

Time of interest. We compared the accuracies obtained from classifying the original kinematic data (not reduced) at the heuristic time given by the lab and at our proposed time $t_{lastHigher}$. We can conclusively show that the classification accuracy of the original data at (a) the time given by the lab, and at (b) $t_{lastHigher}$ was on average better than all the other times we considered. When comparing these times between each other, both of these times yielded similar accuracy, but interestingly, under all experimental conditions, $t_{lastHigher}$ in average yielded better classification accuracy for Baxter, and the time given by the lab yielded better classification accuracy for Vinny. One possible explanation is that Baxter takes in average shorter time to perform the reach and grasp, so time given by the lab goes beyond (back) the actual hand configuration to a point where lots of movement is happening, Vinny on the other hand, has reached a more stable configuration at that point in time. This suggests that potentially we could have chosen ρ to be lower than 5%.

We use the selected times before $t_{lastHigher}$ as means to study the *preshaping* of the hand. One observation is that the classification accuracy of all the methods increases as the time of interest approaches the end of the trial, and thus the final hand configuration. This can be observed in Figure 22, where we show the performance of all the dimensionality reduced data at different time points along the evolution of the reaching. The plots correspond to $t_{firstLess}$, $t_{lastLess}$, t_{max} , and $t_{firstHigher}$, respectively. A very interesting finding is that the nonlinear reduced data with kPCA polynomial 2 is able to predict with much higher accuracy than other methods the object to be grasped since the time when the fingers are at their maximum velocity during the grasp (t_{max}) .

Variance explained with Linear Principal Components. We show the mean number of principal linear components needed to explain 85% and 95% of variance averaged across all the time of interests we consider in Tables 4 and 5 (the details for each time of interest are in tables: 14, 15, 16, 17 for more detail). The first fact to notice is that the mean number of components needed to explain a specific amount of variance is relatively robust to the definition of outliers through the threshold defined by ρ . Secondly, the number of principal components needed to explain the averaged data are lower than the number of principal components needed to explain the sampled data.

Various authors have reported the number of components needed to explain specific percentages of the variability of the data sets they considered. The number of components they reported are shown in Table 8 of the Appendix.

Our results are only comparable to those of Santello et al. (1998), since they also considered only one time point as opposed to all time points of the movement. The differences are that (a) our subjects are monkeys, and theirs, human beings; (b) they only considered 15 out of the 20 joint angles that describe the hand configuration, we included all. The condition where our results were closest to them was in Baxter, when $\rho = 0.10$ and when the eingenvalues of the covariance matrix was obtained from the averaged data across trials fixing (object, pair) condition. But, in general, the numbers we report are slightly higher particularly for Vinny. A plausible explanation for this is that the number of variables we considered is larger than the number of variables (Santello et al., 1998) used.

The amount of variance explained by kernel PCA is not directly comparable, since the covariance that is being obtained is the covariance matrix in the feature space, not in the space of variables.

Baxter	Sampling		Averaging	
	85%	95%	85%	95%
Outliers $\rho = 0.05$	4.2	7.6	2.5	5.1
Outliers $\rho = 0.10$	4.1	7.3	2.5	4.8

Table 4: Baxter: mean number of components needed across all times of interest considered, for the different alternatives of the analysis.

Vinny	Sampling		Averaging	
	85%	95%	85%	95%
Outliers $\rho = 0.05$	4.8	8.1	3.1	5.8
Outliers $\rho = 0.10$	4.8	8.3	3.5	5.8

Table 5: Vinny: mean number of components needed across all times of interest considered, for the different alternatives of the analysis.

Robustness to outliers definition criterion. Not only is the mean number of components needed to explain 85% and 95% of the variance in the PCA reduced data practically the same when considering different outlier criteria. But also, the difference in accuracy given by removing outliers with $\rho = 0.05$ and with $\rho = 0.10$ is not meaningful. We found thus, that our results are relatively robust to the specific threshold for defining outliers, and we fixed $\rho = 0.05$.

Sampling versus averaging to obtain mappings. We contrasted the strategy of sampling versus averaging the data set to obtain the mappings, and found that as a whole the classification accuracies are not affected for any of the unsupervised dimensionality reduced methods. However, the performance of the embedding using Linear Discriminant Analysis is deteriorated when averaging is used as opposed to sampling. In Figure 20 we show the phenomenon: for all number of components, the accuracy of LDA obtained with the mappings constructed from averaged data is lower than the accuracy of LDA obtained with the mappings built from sampled data. This is a representative example of all the comparisons between averaging and sampling in all the different experimental setups.



Figure 20: Contrasting summarizing strategies: sampling versus averaging. Using the averaged data across trials fixing (object, position) pair deteriorates considerably the performance of Linear Discriminant Analysis. Whereas the performance on the classifiers remains relatively unchanged for the unsupervised dimensionality reduced embeddings. The performance of LDA when evaluated with Naive Bayes classifiers is also worse when using averaged data as opposed to sampled data to obtain the reduction mapping. In this plot: Vinny. Classification method: SVM. Time point: *lastHigher*.

Classification accuracies of reduced data. The dimensionality reduced method naturally depends on the kernel. The two main options we tried were polynomial and Gaussian kernels, and they have a very different behavior when evaluated with the classification methods. Classification accuracy also depends on the specific classifier. In our experiments we observed that kernel PCA with kernel polynomial of degree 2 performs consistently the best as compared to other dimensionality reduction methods, when Support Vector Machines are used. In the same setting, the Gaussian kernel-reduced data performs consistently the worst, even compared against PCA.

On the other hand, linear and non linear dimensionality reduction methods do not present meaningful differences in classification accuracy when evaluated through a Naive Bayes Classifier. With the SVM classifier, the PCA embeddings rarely go higher than the accuracy yielded by the Original Data (Figure 20). In contrast, the PCA embeddings coupled with a Naive Bayes Classifier yield higher accuracy. One possible explanation to why Principal Component Analysis perform better when trained with Naive Bayes as opposed to Support Vector Machines is that: PCA might rotate the data in a way that it more closely satisfies the conditional independent assumptions.

One last interesting observation is that the performance of the Gaussian kernel-reduced data improves when evaluated through Naive Bayes classifiers, though we have yet to understand why this is the case.



Figure 21: Contrasting classification methods: Support Vector Machines (Figure 20 Part A) versus Naive Bayes classifier. The accuracy obtained from linear and non linear reduced data evaluated through Naive Bayes classifiers is not meaningfully different. In this plot: Vinny. Classification method: Naive Bayes. Time point: lastHigher.

Confusion matrices We show now qualitative results that bring insight at which objects are easier to classify, and we elaborate on possible explanations for this.

In Figure 23 we show graphical representations of confusion matrices for the three unsupervised dimensionality reduction methods coupled with a respective Support Vector Machines classifier for Baxter⁵. We show from left to right an increasing number of components. A confusion matrix C cointains in its (i, j) coordinate a count of observations known to be in group *i* but predicted to be in group *j*.

The first trend to note is that from left to right there is an increasing organization of the matrices towards the diagonal. The method that brings the elements faster to the diagonal is kPCA with polynomial kernel of degree two. Furthermore, observing the right most column of matrices which corresponds to taking *all* the components, the method that presents the cleanest pattern of elements only in the diagonal is kPCA polynomial of degree 2. The performance of kPCA with Gaussian kernel is the worst, and

^{5.} Baxter was not trained to grasp the large handle, and thus the number of objects considered for his classification task is only nine.



Figure 22: Preshaping of the hand: at the left the oldest time point in the reach is plotted, going further to the right the hand is closer to the object to be grasped. In this figure data from Baxter with outlier definition of $\rho = 0.05$

objects that PCA is not able to differentiate even with all the components (like the small precision versus the large precision) are successfully classified by kPCA with polynomial kernel (see Figure 23, first two matrices of the last column of matrices, coordinates corresponding to (small precision, large precision)).

On the left most column we observe the confusion matrices obtained with only two components. The object that is easiest to classify by all classifiers, and only with *two* components, is the button. Through observation of movies representing the hand movements we observe that the pattern of *grasping* is indeed quite different. The monkeys use their middle finger to push the button, and the rest of the fingers are basically extended. The hand configuration for all the other objects is very different than that pattern, and more similar between the other objects in which more classic power or precision grips are observed. The clear differentiation of hand configuration from button to all other objects can also be observed in the scatter plots of the LDA projections of the data (see Figure 24).

In the first column we also observe blobs of dark color outside the diagonal. Common troubles for all classifiers when only two components are considered are: (a) the large rectangle is classified as the disk (and viceversa, except for kPCA polynomial degree 2); (b) the large precision is classified as the small precision (and viceversa); (c) the large cone is classified as the small cone (and viceversa), and (d) surprisingly the small handle is classified as the large cone.

These qualitative results show evidence of how under specific conditions a unsupervised non linear dimensionality method can extract information that unsupervised linear methods cannot, and also can extract faster information that linear methods take longer to extract.

4.5.1 Conclusions of kinematic analysis

To conclude the kinematic analysis, we have experimentally identified several useful preprocessing strategies for kinematic hand configuration data and evaluated dimensionality reduction techniques using two types of classification schemes. The performance of nonlinear versus linear dimensionality reduction depends on the specific classifier used, as we have seen from experiments. However, it does seem that in certain situations, modelling the nonlinear relationships in the data is helpful for classifier performance. Out of all the experimental settings, for all monkeys the setting which achieves the highest accuracy with 10 components is always the Support Vector Machines with polynomial kernel degree 2. That said, the combination of PCA and Naive Bayes classifier yielded surprisingly good accuracies for lower number of components in Baxter (but not in Vinny).

5. Neural coding of eigengrasps

In this section we first pose statistically the problem of neuronal coding in the context of our work and introduce the methods we used. Secondly, we explain the performed experiments, and finally we present results and draw conclusions.



Figure 23: Confusion matrices obtained from SVM classification on Baxter data at time point: $t_{lastHigher}$ from the three non supervised dimensionality reduction methods. On each matrix the y-axis the true object is shown, on the x-axis the predicted object is shown. A sequence of increasing number of components considered is shown from left to right. See text for further details.



Figure 24: First two and three components obtained from the Linear Discriminant Analysis projection of Baxter at $t_{lastHigher}$.

5.1 Neural coding of finger kinematics: methods

The goal of the neural analysis is to investigate the suitability of a linear model for the neural representation of finger kinematics in the primary motor cortex, as well as of the low dimensional representations of the kinematics (eigengrasps) obtained in the previous section. In this section we propose a model towards this end, and propose a measure of evaluation for it. We then explain the approach to verify that the obtained results were not due to chance.

5.1.1 Preselection of neurons

Before applying any analysis on the neural data, a preselection of neurons was performed. Two criteria had to be fulfilled by a neuron in order to be considered elegible for the analysis: (a) the neuron had to be associated with at least 100 trials; (b) the neuron had to be *task related*.

In order to define task relatedness we divided each trial in three epochs: premovement, reach and grasp configuration. The first one comprised from the beginning of the trial until $t_{firstLess}$; reaching included the period between $t_{firstLess}$ and $t_{lastHigher}$; and, grasp-configuration was the period from $t_{lastHigher}$ until the end of the trial. Roughly these three epochs correspond to the monkey having the hand on the starting pad, the monkey moving and postioning the fingers on the object, and the monkey holding the object. An analysis of variance was done of the firing rates during premovement, reach and grasp-configuration. We tested the null hypothesis that states that the mean of the firing rates in each of the epochs are equal. We considered the neuron to be task related if the the p-value of the F-test was lower than 0.001, which meant that the null hypothesis was rejected.

5.1.2 Model for finger kinematics: multiple linear regression

A simple linear model is one possible way to approximate the firing rate of cells in M1 as a function of finger configuration. This approach has been tried in the case of cursor control (see for instance (Gao et al., 2003; Wu et al., 2003, 2005)).

The linear generative model of neural firing can be written as:

$$z_k \sim \mathcal{N}(Xb_k, \sigma_k^2),$$

where $z_k \in \mathbb{R}$ is a scalar representing the spike counts of neuron k within a time bin of size Δ , X contains p kinematic variables describing fingers kinematics, σ_k^2 is the unknown variance of the firing of neuron k, and $B_k = (\beta_0, \beta_1, \ldots, \beta_p)$ contains the coefficients that linearly relate the finger kinematics with the observed spike counts.

This model is the classical linear regression model and it makes the assumption that the spike counts are normally distributed about XB_k . This assumption does not strictly hold because z_k represents spike *counts*. However, in studies of cursor control, the normality assumption has been shown to yield good results in terms of coding and decoding (Wu et al., 2003; Gao et al., 2003). In addition, a square root transformation of the spike counts can be applied in order to make it better modeled by a normal distribution (Maynard et al., 1999; Moran and Schwartz, 1999). However, this transformation does not lead to great improvements neither in coding, nor in decoding (Wu et al., 2005) thus it is not essential.

In this case we are considering the spike counts of a neuron k to be binned in non-overlapping time bins of length Δ . Note that the physical relationship between neural firing and kinematic behavior implies the existance of a time lag between them (Moran and Schwartz, 1999; Wu et al., 2003; Paninski et al., 2004a). If an optimal lag exists for motor commands, it has not been found nor has a systematic way of finding one been established. In fact, it is quite plausible that lags are neuron dependant. Thus, one way to select a *reasonable* lag is to try different alternatives and choose the best one according to a predefined goodness measure.

Measure of goodness. A measure of goodness-of-fit of multiple linear regression is the *coefficient of* determination R^2 , which is a number between zero and one that indicates how well the least square

hyperplane fits the data. The interpretation is that when $R^2 = 0$ there is no *linear* relationship between the explanatory variables and the predicted variable; and as R^2 approaches one, the relationship between the dependent and independent variables is closer to be linear. The coefficient of determination gives no information of non-linear relationships.

In the general frame of linear regression, the total variability of the data can be decomposed as the sum of the variability explained by the model, and the variability in the error terms. R^2 is defined as the ratio between the former, and the total variability of the data. That is, R^2 measures the proportion of total variation in the dependent variable that is accounted for by variation in the regressors.

Even though the correlation coefficient yields a measure of goodness for the regression, it has the drawback that it is non-decreasing as the number of explanatory variables considered in the model increases. To avoid this caveat, the adjusted coefficient of correlation is defined as:

$$R_{adj}^2 = 1 - \frac{\frac{ESS}{n-k}}{\frac{TSS}{n-1}}$$

where n is the number of observations, and k is the number of explanatory variables included in the model. R_{adj}^2 increases only if the new term improves the model more than would be expected by chance, however R_{adj}^2 can be negative, and furthermore, it does not have the same interpretation as the coefficient of correlation.

Given the difficulty of interpretation we chose R^2 as the measure of goodness for our regressions, keeping in mind (a) the relationship between the number of explanatory variables in the regression and the value of R^2 , (b) that the coefficient of correlation provides information only of linear relationships, as opposed to non-linear relationships.

In this way, a regression of z_k on X with a high R^2 indicates that the coding of the kinematic variables of X in the firing rate of neuron z_k is linear, and that the linear regression is a good model to explain their relationship.

5.1.3 Are the results due to chance?

Finally, it is desirable to verify that the obtained results are not due to chance. A way to verify this is using a version of the bootstrap while breaking the relationship (if it exists) between the explanatory variables and the independent variable.

In order to do this, consider the matrix of kinematics $X \in \mathbb{R}^{n \times p}$ that contains *n* trials, and perform the following procedure *M* times (*M* is large): draw a sample with replacement of size *n* from among the trials contained in *X*, call the resulting sample $X_{res} \in \mathbb{R}^{n \times p}$, perform the regression of z_k on X_{res} , and save the resulting R^2 .

In this way, M coefficients of determination will be obtained, and the mean of them is an approximation of the R^2 obtained by chance. These values can be used to compare the R^2 obtained from the original regressions.

5.2 Analysis design

For the neural analysis we considered both monkeys, Vinny and Baxter, and their respective neurons. The first step performed was the preselection of neurons as explained in Section 5.1.1.

With respect to the considered kinematic variables: in addition to joint angles (like in the kinematic analysis), we considered joint angle velocity, 3D position -normalized to the wrist position-, and 3D marker velocity. The latter two data types consisted of 48 variables as explained in Section 3.1 (and Figure 7). The reason for considering velocity in addition to position was that there have been studies in literature that have found velocity⁶ a relevant variable coded in M1 (see for example (Paninski et al., 2004a,b; Wu

^{6.} Velocity of cursor, or velocity of arm.

et al., 2003) and others). Velocity of the kinematic variables was obtained through the smoothing B-spline basis functions we used to smooth the data in the first place.

The analysis was based on three of the points of interest defined as before, based on the energy function (Section 4.2). The points of interest were: $t_{givenByLab}$, $t_{firstHigher}$, and $t_{lastHigher}$. To define these data points we considered $\rho = 0.05$, since it is the most conservative definition for outliers. Dimensionality reduction methods were applied as explained in Section 4.2.4, using the sampling method to obtain the mappings to reduce the data.

In this manner, the design matrix $X \in \mathbb{R}^{n \times p}$ used for the multiple linear regression contained the kinematic information: for a specific monkey, a specific type of kinematic variable, time of interest, and raw or low-dimensional representation, all of these associated with a specific preselected neuron.

Table 6 shows (in each column) the different choices needed to create a matrix $X \in \mathbb{R}^{n \times p}$. The number of columns p was determined by the number of components to be considered. When considering the low representation of the data we considered 2, 3, 5, 8 and 15 principal components for the unsupervised dimensionality dimension methods (PCA and kernel PCA), and the same number of components except for 15 for LDA. In the case of the orginal data, p = 20 for joint angles position and velocity, and p = 48for 3D marker position and velocity.

Monkey	Kinematic variable	Time of interest	Dataset
• Vinny	• joint angles	• $t_{givenByLab}$	
• Baxter	• Joint angles velocity • 3D marker position	• $t_{firstHigher}$ • $t_{lastHigher}$	• PCA reduced • kPCA polynomial ker-
	I I I I I I I I I I I I I I I I I I I		nel reduced
	• 3D marker velocity		• kPCA gaussian kernel
			• LDA reduced
			• LDA reduced

Table 6: Summary of the choices made to create a single design matrix to investigate the suitability of a linear regression model for explaining the firing pattern of neurons in M1. * In the case of the Gaussian kernel, we considered $\sigma = 2$ as determined by the cross validated classification accuracy in Section 4.5

The neural data was binned considering a bin size of 70mS as in (Wu et al., 2003, 2002). First the the time of interest t^* was fixed, then the neural activity was binned in intervals of length $\Delta = 70mS$ as shown in Figure 25; we considered up to five lags backwards from the time of interest, in which the firing rate was calculated. Firing rate was obtained as sum of spike counts divided by the length of the bin.

A multiple linear regression analysis was performed for each of the design matrices built paired with each of the vectors of neural activities. Since the neural activity was binned in intervals of 70mS the lag which yielded the largest R^2 was selected.

5.3 Results for neural analysis

In this section we first present the results of the neurons' preselection, then the results regarding the linear coding of kinematics in primary motor cortex, and finally we conclude by enumerating the findings of the neural analysis.

5.3.1 Preselection of Neurons

The number of neurons that fulfilled the criterion of having at least 100 trials associated to them were 13 (out of 19) for Baxter and 37 (out of 67) for Vinny. An ANOVA was applied to the activity of those neurons as explained in Section 5.1.1. In Table 7 we show the number of neurons whose p-value of the



Figure 25: Illustration of how the neural data was binned based on a specific time point of interest t^* . Given a bin size $\Delta = 70mS$ the neural activity was compartmentalized in intervals of length Δ counting backwards from t^* . The firing rate of the neuron within each of the intervals was obtained.

F-test is lower than 0.05, 0.01 and 0.001. The null hypothesis that we tested was that the mean of the firing rates in each of the epochs (premovement, reach, and grasp-configuration) are equal. We considered the neuron to be *task related* if the the p-value of the F-test was lower than 0.001, which left 31 neurons for Vinny to be analyzed and 4 neurons for Baxter.

Monkey	p < 0.05	p < 0.01	p < 0.001
Vinny	91.89% (34/37)	83.78%~(31/37)	83.78% (31/37)
Baxter	$38.46\% \ (5/13)$	30.77%~(4/13)	30.77%~(4/13)

Table 7: Number of neurons whose p-value of the F-test in the ANOVA is less than the specified value in the columns (See section 5.1.1 for more explanation). Note that the two last columns look the same, but refer to different thresholds for the p-values.

5.3.2 Linear coding of kinematic variables

This section contains the results from the multiple linear regressions performed on neural activity binned in intervals of 70mS and the kinematics built as explained in Table 6. For each of the choices of kinematics, we counted how many neurons had a regression with an R^2 higher than a threshold of 0.15. We focused on the results from Vinny because there is a population of 31 neurons, as opposed to only four.

First we show the results for the regressions on the joint angles and joint angle velocities, followed by the results on the 3D marker positions and velocities.

Linear coding of joint angles kinematics. We compared the linear coding of joint angles versus joint angles velocities. The number of neurons (out of 31) that yielded an R^2 higher than 0.15 are in Figure 26. Three points in time were contrasted: $t_{givenByLab}$, $t_{lastHigher}$, and $t_{firstHigher}$.

The number of neurons that yielded R^2 above threshold for joint angles was always higher than the number of neurons surpassing threshold for joint angle velocity except for three cases: LDA with 2 components at time $t_{givenByLab}$, and at time $t_{firstHigher}$ kernel PCA polynomial with 15 components, and kernel PCA gaussian with 15 components. However, in these three cases the difference of number of neurons was only one. To quantify the difference of the number of neurons whose R^2 surpassed threshold for joint angles as opposed to joint angle velocity, we calculated the average of the absolute value of the difference between the number of neurons. This value was 7.75 for $t_{givenByLab}$, 7.2 for $t_{lastHigher}$, and 1.3 for $t_{firstHigher}$. The last value indicates that no matter what kinematic variable we consider (joint angle or joint angle velocity) the number of neurons whose firing rate can be explained (given our threshold) by a linear model of the kinematics is the same. Or, in other words, at time $t_{firstHigher}$ there is the same number of neurons in our population that code for joint angle and for joint angle velocity.

	givenByLab		lastH	lastHigher		igher
	Joint angles	JA velocity	Joint angles	JA velocity	Joint angles	JA velocity
PCA - 2	38.7% (12)	9.7% (3)	35.5% (11)	0.0% (0)	35.5% (11)	29.0% (9)
PCA - 3	38.7% (12)	9.7% (3)	45.2% (14)	6.5% (2)	35.5% (11)	32.3% (10)
PCA - 5	54.8% (17)	22.6% (7)	67.7% (21)	32.3% (10)	38.7% (12)	38.7% (12)
PCA - 8	74.2% (23)	35.5% (11)	83.9% (26)	67.7% (21)	54.8% (17)	48.4% (15)
PCA - 15	100.0% (31)	80.6% (25)	93.5% (29)	87.1% (27)	77.4% (24)	74.2% (23)
kPCA poly - 2	29.0% (9)	6.5% (2)	35.5% (11)	0.0% (0)	32.3% (10)	29.0% (9)
kPCA poly - 3	35.5% (11)	9.7% (3)	48.4% (15)	6.5% (2)	35.5% (11)	32.3% (10)
kPCA poly - 5	51.6% (16)	22.6% (7)	58.1% (18)	32.3% (10)	38.7% (12)	38.7% (12)
kPCA poly - 8	74.2% (23)	35.5% (11)	77.4% (24)	64.5% (20)	48.4% (15)	48.4% (15)
kPCA poly - 15	96.8% (30)	80.6% (25)	93.5% (29)	83.9% (26)	77.4% (24)	80.6% (25)
kPCA Gauss - 2	35.5% (11)	9.7% (3)	38.7% (12)	0.0% (0)	38.7% (12)	29.0% (9)
kPCA Gauss - 3	45.2% (14)	12.9% (4)	48.4% (15)	22.6% (7)	38.7% (12)	32.3% (10)
kPCA Gauss - 5	58.1% (18)	22.6% (7)	74.2% (23)	32.3% (10)	45.2% (14)	38.7% (12)
kPCA Gauss - 8	83.9% (26)	32.3% (10)	80.6% (25)	67.7% (21)	58.1% (18)	48.4% (15)
kPCA Gauss - 15	96.8% (30)	80.6% (25)	93.5% (29)	83.9% (26)	74.2% (23)	77.4% (24)
LDA - 2	22.6% (7)	25.8% (8)	45.2% (14)	22.6% (7)	22.6% (7)	22.6% (7)
LDA - 3	41.9% (13)	32.3% (10)	61.3% (19)	41.9% (13)	29.0% (9)	25.8% (8)
LDA - 5	64.5% (20)	48.4% (15)	83.9% (26)	58.1% (18)	38.7% (12)	35.5% (11)
LDA - 8	80.6% (25)	61.3% (19)	83.9% (26)	77.4% (24)	54.8% (17)	51.6% (16)

Figure 26: Percentage (and number) of neurons whose regression yields an $R^2 \ge 0.15$ considering joint angles and joint angles velocities as regressors. Each entry in the matrix corresponds to one kinematic data set, a dimensionality reduction method, a specific number of components, and a specific point in time. The rows indicate which method of dimensionality reduction was used, and how many components were considered, the column indicates two selections: (a) which time of interest was considered as described in Section 4.2, and (b) what type of kinematic variable was considered: joint angles or joint angles velocity. For each dataset several regressions were run on the firing rate of a specific neuron on different lags, the lag that yielded the highest R^2 was selected and the R^2 recorded. Each entry of the matrix contains the percentage of cells whose regression of the kinematic variables on the selected lag yielded $R^2 \ge 0.15$.

We also contrasted the different dimensionality reduction methods, that is, how the results varied as a function of the dimensionality reduction method used to project the data. Figure 27 shows the results in a format more suitable for this comparison. We concluded that when considering joint angles, the behaviour of *all* dimensionality methods is very similar; so is the case for when considering joint angles velocity and time point $t_{firstHigher}$. However, when considering joint angles velocity at $t_{givenByLab}$ or $t_{lastHigher}$, LDA performs better. Perhaps because at that point, the monkey is already grasping the object and so the information of the object that is definitely coded in the kinematics (since LDA is supervised) implies some specific firing pattern of the neurons. This hypothesis would need further testing.



Figure 27: The information in this figure is the same than the one contained in Figure 26, but it is displayed in a way were the comparison between dimensionality reduction methods is easier to make.

Linear coding of 3D markers kinematics. Figures 28 and 29 display the same information than those for joint angles, but for the 3D markers positions and velocities. Note that in this case, the design matrices consisted of 48 variables, as opposed to 20 variables as in the joint angles case.

In this case, the number of neurons whose regression yielded an R^2 higher than threshold were higher using 3D marker positions than 3D marker velocity except cases where a Gaussian kernel was used for the dimensionality reduction. There were also three other exceptions, but the difference was only of one neuron: LDA with 8 components at time $t_{lastHigher}$, and LDA with 8 components and PCA with 3 components at time $t_{firstHigher}$. The average absolute difference in number of neurons was 7.5 for time $t_{givenByLab}$, 7.55 for time $t_{lastHigher}$ and 5.6 at time $t_{firstHigher}$. Most of the difference was concentrated in the experiments where the Gaussian kernel was used to reduce the data, in fact, the average difference in those cases was 10.8.

Are these results due to chance? To verify whether the results were due to chance or not we performed a version of the bootstrap as explained in 5.1.3. The number of regressions to perform was large: $M \times numberNeurons \times numberConditions \times numberTimePoints \times numberLags \times numberKinematicVariables$ where M denotes the number of iterations of the bootstrap; numberNeurons = 31 for Vinny; numberConditions =19 referring to dimensionality reduction method with specific number of components (or number of rows in Figures 26 and 28); numberTimePoints = 3 for $t_{givenByLab}$, $t_{lastHigher}$, and $t_{firstHigher}$; numberLags = 5for the different lags tried; and numberKinematicVariables = 4 referring to joint angles, joint angles velocity, 3D marker position and 3D marker velocity.

We defined M = 10, that is, for each neuron we obtained M values of R^2 (each of them, the maximum R^2 across the *numberLags* lags) and took its average. Then, we counted how many neurons yielded an R^2 higher than the defined threshold. It was never the case that the number of neurons yielding an

	givenByLab		last H	ligher	firstH	ligher
	3D position	3D velocity	3D position	3D velocity	3D position	3D velocity
PCA - 2	32.3% (10)	19.4% (6)	48.4% (15)	29.0% (9)	29.0% (9)	29.0% (9)
PCA - 3	45.2% (14)	32.3% (10)	51.6% (16)	45.2% (14)	29.0% (9)	32.3% (10)
PCA - 5	61.3% (19)	38.7% (12)	77.4% (24)	58.1% (18)	54.8% (17)	41.9% (13)
PCA - 8	80.6% (25)	61.3% (19)	87.1% (27)	67.7% (21)	67.7% (21)	54.8% (17)
PCA - 15	96.8% (30)	80.6% (25)	100.0% (31)	87.1% (27)	83.9% (26)	77.4% (24)
kPCA poly - 2	25.8% (8)	12.9% (4)	45.2% (14)	9.7% (3)	32.3% (10)	9.7% (3)
kPCA poly - 3	35.5% (11)	12.9% (4)	48.4% (15)	19.4% (6)	41.9% (13)	12.9% (4)
kPCA poly - 5	58.1% (18)	19.4% (6)	77.4% (24)	29.0% (9)	54.8% (17)	29.0% (9)
kPCA poly - 8	80.6% (25)	32.3% (10)	87.1% (27)	58.1% (18)	67.7% (21)	35.5% (11)
kPCA poly - 15	96.8% (30)	58.1% (18)	100.0% (31)	74.2% (23)	77.4% (24)	58.1% (18)
kPCA Gauss - 2	0.0% (0)	16.1% (5)	0.0% (0)	9.7% (3)	0.0% (0)	22.6% (7)
kPCA Gauss - 3	0.0% (0)	22.6% (7)	0.0% (0)	25.8% (8)	3.2% (1)	29.0% (9)
kPCA Gauss - 5	6.5% (2)	38.7% (12)	6.5% (2)	51.6% (16)	6.5% (2)	45.2% (14)
kPCA Gauss - 8	16.1% (5)	61.3% (19)	9.7% (3)	74.2% (23)	9.7% (3)	54.8% (17)
kPCA Gauss - 15	35.5% (11)	80.6% (25)	32.3% (10)	93.5% (29)	22.6% (7)	64.5% (20)
LDA - 2	38.7% (12)	19.4% (6)	45.2% (14)	32.3% (10)	32.3% (10)	25.8% (8)
LDA - 3	51.6% (16)	19.4% (6)	64.5% (20)	45.2% (14)	32.3% (10)	25.8% (8)
LDA - 5	58.1% (18)	41.9% (13)	77.4% (24)	77.4% (24)	38.7% (12)	32.3% (10)
LDA - 8	77.4% (24)	67.7% (21)	83.9% (26)	87.1% (27)	48.4% (15)	51.6% (16)

Figure 28: Percentage (and number) of neurons whose regression yields an $R^2 \ge 0.15$ considering 3D marker position and 3D marker velocity as regressors. For specific explanation of the information displayed in this figure see Figure 26.



Figure 29: Number of neurons whose regression yields an $R^2 \ge 0.15$ considering 3D marker position and 3D marker velocity as regressors. This figure contains the same information than the one shown in Figure 28, but it is displayed in a way were the comparison between dimensionality reduction methods is easier to make.

 $R^2 \ge 0.15$ through this procedure was higher than the number of neurons reported in Figures 26 and 28. And therefore, these results are not due to chance.

5.3.3 Conclusion of neural analysis

To conclude this section the first thing to notice is that the number of neurons to be analyzed was greatly reduced by the imposed requirements of having at least certain number of trials and of being *task related*. We argue though that these constraints are necessary to ensure: (a) the reliability of the results of the linear regressions, and (b) that the signal we are analyzing is, in fact, related to the reach-to-grasp task. However, at the same time that these requirements are necessary, they do reduce drastically the data we can analyze. For instance, four neurons of Baxter are actually insufficient to make any reliable inference of primary motor cortex neurons behavior.

Perhaps the most important observation of the analysis is that, in most cases, varying the dimensionality reduction method used to obtain the low dimensional representation of the data does not seem to significantly affect the encoding of the firing pattern. This said, in our opinion, non linear dimensionality reduction methods are not fully developed; they need further research and perhaps new theory. Thus, the fact that the non linear dimensionality reduction methods we tried did not yield significantly better results, does not mean that another way of extracting non linear embeddings that yield better results in our setting cannot be developed.

An interesting feature, that deserves further analysis is that the Gaussian kernel PCA yields such different results when applied to regressions involving 3D marker position and 3D marker velocity.

Finally, it is intriguing that Linear Discriminant Analysis can yield in several settings much better results than unsupervised methods. This implies that *reduced* information about which object is being grasped can help to predict neural firing. In some sense, this is related to results from Mason et al. (2006), but it actually goes further because in our case we are not using merely the size of the object, but we are dealing with mathematically reduced hand configurations informed by the object to be grasped.

6. Discussion

In this work we analyzed the kinematics of the fingers of two monkeys (species *Macaca mulatta*) during a reach-to-grasp task. We considered two data sets: joint angles and 3D marker positions that describe the movement of the hand along time. We defined an energy function summarizing the amount of motion of the fingers. This function was useful to derive definitions for outliers, and to define time points of interest during the trials. We investigated low dimensional representations of the hand configurations at those specific time points during the reach-to-grasp movement. In order to perform the dimensionality reduction we tried linear supervised (Linear Discriminant Analysis) and unsupervised (Principal Components Analysis) methods, and non linear unsupervised methods (kernel Principal Components Analysis with various choices of kernels). The low dimensional representations were then evaluated according to a classifiation task: predicting what object was being grasped at a specific trial. Two classifiers were trained and tested: a generative one (Naive Bayes) and a discriminative one (Support Vector Machines). The main finding was that under certain conditions, modelling the nonlinear relationships in the data resulted in better classifier performance.

The analysis of the low dimensional representation of the hand configurations was extended to investigate whether these representations were encoded in neurons recorded from the primary motor cortex (M1) of the monkeys. In order to perform the neural analysis, neurons were required to have a specific number of trials associated with them, and to be task related, as defined through differentiated firing rate patterns in different epochs during the reach-to-grasp trial (analysis of variance). A classical multiple linear regression model was proposed to explain the firing rate of the neurons. The explanatory variables were the low dimensional representation of the kinematic variables of the fingers. The measure-of-goodness was defined to be the number of neurons that yielded an R^2 higher than a specific threshold. Using a variation of bootstrap in regression, results were verified to not be due to chance. During the course of the analysis many choices needed to be made with the goal of getting a meaningful signal. These choices can have a lot of impact in the results of the analysis. We tried many choices, but an exhaustive study is virtually impossible. A natural extension of our analysis is the modification of the model for the firing pattern of the neuron, for instance, with a generalized linear model.

In this work we dealt with the static case of kinematic low dimensional representations and their representations in neurons in primary motor cortex. It appears to be the case that non linear dimensionality reduction methods do not necessarily yield better results than linear methods when associating finger kinematics to neural activity in primary motor cortex. Further studies of the dynamic case, might provide evidence of the opposite, but this remains to be proved. The study of state space models for encoding finger kinematics and its relation with the firing of primary motor cortex neurons is the very next step in our path, and we are currently working on it.

Acknowledgments

Thanks to Rob and Vince Vu for theoretical advice, and ideas; Sagi Perel, and Chance Spalding for sharing their data, and for many clarification sessions; Samuel Clanton, for the transformation from 3D positions to joint angles; and, of course, thanks to Andy Schwartz for allowing me to collaborate with his team. Thanks to Stanislav Funiak for lending me his terminal to run experiments; and many, many thanks to Jonathan Huang for proof reading, computer power and moral support.

Appendix: Previous Work.

Reference	Task	Subject	Dataset, coordi-	Method	Results
			nate system		
(Santello et al.,	•hold imaginary	human	15 JA: flexion	PCA	Number com-
1998)	object		of all fingers		ponents for
			and thumb. No		explaining 85%
			abduction.		variance: 2; for
					95%, 4.
(Todorov and	• specific manipu-	human	• 15 JA: as before.	PCA	Number compo-
Ghahramani,	lation		Consider three dif-		nents for explain-
2004)			ferent angle stan-		ing for 85% of
			darizations. Re-		variance: 6.6; for
			port average of		95%, 9.6.
			number of compo-		
			nents across these		
			standarizations.		
	• specific manipu-	human	• 20 JA: position.	PCA	Number com-
	lation		Standarized in the	-	ponents for
			three ways.		explaining 85% of
					variance: 7.3: for
					95% 11
(Mason et al	• reach-to-grasp	human	3D marker posi-	SVD	1st principal com-
2001)	a reactine of Strapp	liaman	tions	~ . 2	popent explained
2001)			010110		97.3% of variance
					The 2nd compo-
					nont 1.0%
(Mason ot al	• reach to grasp	monkov	2D marker pogi	SVD	let principal com
(Mason et al., 2004)	• reach-to-grasp	monkey	5D marker posi-	510	nonont ambained
2004)			UOIIS		ponent explained 0.207 of constants
					93% of variance.
					Ine 2nd compo-
	1.11 1	1		DCIA	nent 5%.
(Soechting and	• skilled activity	human	II JA: no thumb	PCA	Number com-
Flanders, 1997)			included. MCP		ponents for
			flexion, DIP flex-		explaining 90% of
			ion, and abduc-		variance: 4.
			tion between 4 fin-		
			gers.		
(Thakur et al.,	•unconstrained	human	3D marker posi-	PCA	Number com-
2008)	haptic exploration		tions		ponents for
					explaining 90% of
					variance: 7.

Table 8: Previous work on dimensionality reduction for extracting hand synergies. It is important to mention Principal Component Analysis and Singular Value Decomposition are essentially the same. JA stands for joint angles, and 3D for the three dimensional position of markers.

Appendix: Data Exploration.

	Baxter	Vinny
Num. sessions	4	16
Total num. trials	672	3,166
Trials per session (mean \pm std)	168 ± 25	198 ± 71
Mean number of time samples $(\pm \text{ std})$	155 ± 35	211 ± 33
Mean trial length [mS] (\pm std)	772 ± 176	$1,052 \pm 164$

Table 9: General information of kinematic data collected from Baxter and Vinny

$\downarrow \text{Obj} / \text{Pos} \rightarrow 1 2 3 4 5 6 7 \text{Sum}$	_
1 10 10 10 10 0 0 0 0	
1 19 13 18 20 19 0 0 89	
2 4 4 2 0 7 7 8 32	
3 0 0 0 0 0 0 0 0 0	
4 5 8 1 7 6 5 2 34	
5 17 14 15 15 16 16 14 107	
6 18 19 20 17 19 19 15 127	
7 10 11 9 1 12 13 10 66	
8 11 13 8 9 15 14 9 79	
9 15 0 0 17 18 14 11 75	
10 15 0 0 13 15 13 7 63	

Table 10: Baxter: total count of (Objects, Position).

$\fbox{$\downarrow$ Obj / Pos \rightarrow $}$	1	2	3	4	5	6	7
1	4.75	3.25	4.50	5.00	4.75	0.00	0.00
2	1.00	1.00	0.50	0.00	1.75	1.75	2.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	1.25	2.00	0.25	1.75	1.50	1.25	0.50
5	4.25	3.50	3.75	3.75	4.00	4.00	3.50
6	4.50	4.75	5.00	4.25	4.75	4.75	3.75
7	2.50	2.75	2.25	0.25	3.00	3.25	2.50
8	2.75	3.25	2.00	2.25	3.75	3.50	2.25
9	3.75	0.00	0.00	4.25	4.50	3.50	2.75
10	3.75	0.00	0.00	3.25	3.75	3.25	1.75

Table 11: Baxter: Mean count of (Objects, Position) across sessions.

\downarrow Obj / Pos \rightarrow	1	2	3	4	5	6	7	Sum
1	60	55	51	55	58	0	0	279
2	31	28	41	0	33	23	17	173
3	54	28	58	13	50	61	24	288
4	63	39	69	60	60	67	41	399
5	59	14	52	52	47	57	54	335
6	72	67	69	65	70	65	69	477
7	62	53	62	60	64	60	52	413
8	65	56	64	56	65	64	62	432
9	12	26	5	7	20	0	0	70
10	67	64	43	56	70	0	0	300

Table 12: Vinny: total count of (Objects, Position).

\downarrow Obj / Pos \rightarrow	1	2	3	4	5	6	7
1	3.75	3.44	3.19	3.44	3.63	0.00	0.00
2	1.94	1.75	2.56	0.00	2.06	1.44	1.06
3	3.38	1.75	3.63	0.81	3.13	3.81	1.50
4	3.94	2.44	4.31	3.75	3.75	4.19	2.56
5	3.69	0.88	3.25	3.25	2.94	3.56	3.38
6	4.50	4.19	4.31	4.06	4.38	4.06	4.31
7	3.88	3.31	3.88	3.75	4.00	3.75	3.25
8	4.06	3.50	4.00	3.50	4.06	4.00	3.88
9	0.75	1.63	0.31	0.44	1.25	0.00	0.00
10	4.19	4.00	2.69	3.50	4.38	0.00	0.00

Table 13: Vinny: Mean count of (Objects, Position) across sessions.

Time of Interest	Baxter 85%	Vinny 85%	Baxter 95%	Vinny 95%
minMenor	4	3	7	6
\max Menor	3	3	5	6
max	2	4	4	6
\min Mayor	2	3	5	5
\max Mayor	2	3	5	6
givenByLab	2	3	5	6

Table 14: Number components needed to explain specific percentage of variance. Criterion for outliers:1.Mapping obtained from: averaging across trials fixing (object, position) condition.

Time of Interest	Baxter 85%	Vinny 85%	Baxter 95%	Vinny 95%
minMenor	4	5	8	8
\max Menor	5	5	8	9
max	4	5	7	9
\min Mayor	4	4	7	7
\max Mayor	4	5	8	8
givenByLab	4	5	8	8

 Table 15: Number components needed to explain specific percentage of variance. Criterion for outliers:1.

 Mapping obtained from: sampling

Time of Interest	Baxter 85%	Vinny 85%	Baxter 95%	Vinny 95%
minMenor	4	3	6	6
\max Menor	3	4	4	6
max	2	4	4	6
\min Mayor	2	3	5	5
\max Mayor	2	4	5	6
givenByLab	2	3	5	6

Table 16: Number components needed to explain specific percentage of variance. Criterion for outliers:2. Mapping obtained from: averaging across trials fixing (object, position) condition.

Time of Interest	Baxter 85%	Vinny 85%	Baxter 95%	Vinny 95%
minMenor	5	5	8	8
\max Menor	4	5	7	9
max	4	5	7	9
\min Mayor	4	4	7	8
\max Mayor	4	5	7	8
givenByLab	4	5	8	8

 Table 17: Number components needed to explain specific percentage of variance. Criterion for outliers:2.

 Mapping obtained from: sampling

References

- Friedemann Awiszus. Spike train analysis. <u>Journal of Neuroscience Methods</u>, 74(2):155 166, 1997. ISSN 0165-0270.
- G. Baudat and F. Anouar. Generalized discriminant analysis using a kernel approach, 2000.
- Yoshua Bengio, Pascal Vincent, and Jean-Franois Paiement. Spectral clustering and kernel pca are learning eigenfunctions. Cirano working papers, CIRANO, 2003.
- Otto Bock and Thomas Feix. Emergence of cognitive grasping through emulation, introspection, and surprise (grasp). taxonomy. http://www.csc.kth.se/grasp/, 2008.

- Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. In <u>Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory</u>, pages 144–152. ACM Press, 1992.
- Umberto Castiello. The neuroscience of grasping. Nature Reviews Neuroscience, 6(9):726–736, 2005.
- Lillian Y. Chang and Yoky Matsuoka. A kinematic thumb model for the act hand. In <u>ICRA</u>, pages 1000–1005. IEEE, 2006.
- Jessie Chen, Shari D. Reitzen, Jane B. Kohlenstein, and Esther P. Gardner. Neural Representation of Hand Kinematics During Prehension in Posterior Parietal Cortex of the Macaque Monkey. <u>J Neurophysiol</u>, page 90942.2008, 2009. doi: 10.1152/jn.90942.2008.
- Matei Ciocarlie, Corey Goldfeder, and Peter Allen. Dimensionality reduction for hand-independent dexterous robotic grasping. In Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on, pages 3270–3275, 2007. doi: 10.1109/IROS.2007.4399227. URL http://dx.doi.org/10.1109/IROS.2007.4399227.
- Koby Crammer and Yoram Singer. On the algorithmic implementation of multiclass kernel-based vector machines. Journal of Machine Learning Research, 2:265–292, 2001.
- M. R. Cutkosky. On grasp choice, grasp models, and the design of hands for manufacturing tasks. <u>Robotics and Automation, IEEE Transactions on</u>, 5(3):269–279, 1989. doi: 10.1109/70.34763. URL http://dx.doi.org/10.1109/70.34763.
- Gantcho N Gantchev, Shigemi Mori, and Jean Massion. <u>Motor Control: today and tomorrow</u>. Academic Publishing House, 1999.
- Yun Gao, Michael J. Black, Elie Bienenstock, Wei Wu, and John P. Donoghue. A quantitative comparison of linear and non-linear models of motor cortical, 2003.
- Corey Goldfeder, Matei Ciocarlie, Hao Dang, and Peter K. Allen. The columbia grasp database. 2009. URL http://app.cul.columbia.edu:8080/ac/handle/10022/AC:P:29598.
- Claudia M. Hendrix, Carolyn R. Mason, and Timothy J. Ebner. Signaling of grasp dimension and grasp force in dorsal premotor cortex and primary motor cortex neurons during reach to grasp in the monkey. <u>J Neurophysiol</u>, pages 00016.2009+, April 2009. URL http://jn.physiology.org/cgi/content/abstract/00016.2009v1.
- T. Iberall. The nature of human prehension: Three dextrous hands in one. In <u>Robotics and Automation</u>. Proceedings. 1987 IEEE International Conference on, volume 4, pages 396–401, 1987.
- Thea Iberall. Human prehension and dexterous robot hands. <u>The International Journal of Robotics Research</u>, 16(3):285–299, June 1997. doi: 10.1177/027836499701600302. URL http://dx.doi.org/10.1177/027836499701600302.
- E. R. Kandel, J. H. Schwartz, and T. M. Jessel. Principles of neural science, third edition edition. 1991.
- C. R. Mason, J. E. Gomez, and T. J. Ebner. Hand synergies during reach-to-grasp. <u>The Journal of</u> Neurophysiology, 86(6):2896–2910, 2001.
- Carolyn R Mason, Claudia M Hendrix, and Timothy J Ebner. Purkinje cells signal hand shape and grasp force during reach-to-grasp in the monkey. J Neurophysiol, 95(1):144-58, 2006. ISSN 0022-3077. URL http://www.biomedsearch.com/nih/Purkinje-cells-signal-hand-shape/16162833.html.

- C.R. Mason, L.S. Theverapperuma, C.M. Hendrix, and T.J. Ebner. Monkey hand postural synergies during reach-to-grasp in the absence of vision of the hand and object. <u>J Neurophysiol</u>, 91(6):2826–37, 2004.
- E. M. Maynard, N. G. Hatsopoulos, C. L. Ojakangas, B. D. Acuna, J. N. Sanes, R. A. Normann, and J. P. Donoghue. Neuronal interactions improve cortical population coding of movement direction. <u>J.</u> Neurosci., 19(18):8083–8093, September 1999.
- Sebastian Mika, Gunnar Rtsch, Jason Weston, Bernhard Schlkopf, and Klaus-Robert Mller. Fisher discriminant analysis with kernels, 1999.

Tom Mitchell. Machine Learning. McGraw Hill, 1997.

- Daniel W. Moran and Andrew B. Schwartz. Motor Cortical Representation of Speed and Direction During Reaching. J Neurophysiol, 82(5):2676-2692, 1999. URL http://jn.physiology.org/cgi/content/abstract/82/5/2676.
- J. R. Napier. The Prehensile Movements of the Human Hand. J Bone Joint Surg Br, 38-B(4):902-913, 1956. URL http://www.jbjs.org.uk/cgi/content/abstract/38-B/4/902.
- Liam Paninski, John P. Donoghue, Matthew R. Fellows, Matthew R. Fellows, Nicholas G. Hatsopoulos, and Nicholas G. Hatsopoulos. Spatiotemporal tuning of motor cortical neurons for hand position and velocity. J. of Neurophysiology, 91:515–532, 2004a.
- Liam Paninski, Matthew R. Fellows, Nicholas G. Hatsopoulos, and John P. Donoghue. Spatiotemporal Tuning of Motor Cortical Neurons for Hand Position and Veloc-J Neurophysiol, 91(1):515-532,2004b. doi: 10.1152/jn.00587.2002. URL ity. http://jn.physiology.org/cgi/content/abstract/91/1/515.
- C. Häger Ross and M. H. Schieber. Quantifying the independence of human finger movements: comparisons of digits, hands, and movement frequencies. <u>The Journal of neuroscience : the official</u> journal of the Society for Neuroscience, 20(22):8542-8550, November 2000. ISSN 1529-2401. URL http://view.ncbi.nlm.nih.gov/pubmed/11069962.
- M. Santello, M. Flanders, and J.F. Soechting. Postural hand synergies for tool use. <u>J Neurosci</u>, 18(23): 10105–15, 1998.
- B. Schölkopf, A. Smola, and K. Müller. Nonlinear component analysis as a kernel eigenvalue problem, 1998. URL citeseer.ist.psu.edu/sch98nonlinear.html.
- J. F. Soechting and M. Flanders. Flexibility and repeatability of finger movements during typing: analysis of multiple degrees of freedom. Journal of computational neuroscience, 4(1):29–46, January 1997. ISSN 0929-5313. URL http://view.ncbi.nlm.nih.gov/pubmed/9046450.
- C. L. Taylor and R. J. Schwarz. The anatomy and mechanics of the human hand. <u>Artificial limbs</u>, 2(2): 22-35, May 1955. ISSN 0004-3729. URL http://view.ncbi.nlm.nih.gov/pubmed/13249858.
- Pramodsingh H. Thakur, Amy J. Bastian, and Steven S. Hsiao. Multidigit Movement Synergies of the Human Hand Unconstrained Haptic Task. $_{in}$ anExploration 2008.J. Neurosci., 28(6):1271-1281,doi: 10.1523/JNEUROSCI.4512-07.2008. URL http://www.jneurosci.org/cgi/content/abstract/28/6/1271.
- E. Todorov and Z. Ghahramani. Analysis of the synergies underlying complex hand manipulation. In 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pages 4637–4640, 2004.

- I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun. Large margin methods for structured and interdependent output variables. Journal of Machine Learning Research, 6:1453–1484, 2005.
- M. Veber and T. Bajd. Assessment of human hand kinematics. In <u>Robotics and Automation, 2006. ICRA</u> 2006. Proceedings 2006 IEEE International Conference on, pages 2966–2971, 2006.
- M. Veber, T. Bajd, and M. Munih. Assessing joint angles in human hand via optical tracking device and calibrating instrumented glove. Meccanica, 42:451–463, 2007.
- Kilian Q. Weinberger, Fei Sha, and Lawrence K. Saul. Learning a kernel matrix for nonlinear dimensionality reduction. In <u>ICML '04</u>: Proceedings of the twenty-first international conference on <u>Machine learning</u>, page 106, New York, NY, USA, 2004. ACM. ISBN 1-58113-828-5. doi: <u>http://doi.acm.org/10.1145/1015330.1015345</u>.
- W. Wu, M. J. Black, Y. Gao, E. Bienenstock, M. Serruya, and J. P. Donoghue. Inferring hand motion from multi-cell recordings in motor cortex using a kalman filter. In <u>SAB02Workshop on Motor Control</u> in Humans and Robots: On the Interplay of Real Brains and Artificial Devices, pages 66–73, 2002.
- W. Wu, M. J. Black, Y. Gao, E. Bienenstock, M. Serruya, A. Shaikhouni, and J. P. Donoghue. Neural decoding of cursor motion using a kalman filter. In <u>in Advances in Neural Information Processing</u> Systems 15, pages 133–140. MIT Press, 2003.
- Wei Wu, Yun Gao, Elie Bienenstock, John P. Donoghue, and Michael J. Black. Bayesian population decoding of motor cortical activity using a kalman filter, 2005.
- Kathryn Ziegler-Graham, EJ MacKenzie, PL Ephraim, TG Travison, and R Brookmeyer. Estimating the prevalence of limb loss in the united states 2005 to 2050. <u>Archives of Physical Medicine and</u> Rehabilitation, 89(3):422–429, 2008.
- K. Zilles, G. Schlaug, M. Matelli, G. Luppino, A. Schleicher, M. Qü, A. Dabringhaus, R. Seitz, and P. E. Roland. Mapping of human and macaque sensorimotor areas by integrating architectonic, transmitter receptor, mri and pet data. Journal of Anatomy, 187(3):515–537, December 1995.