

NEURAL MODELING OF THE CORTICAL PROCESSES UNDERLYING IMITATIVE BEHAVIORS: A DYNAMIC NEURAL FIELD APPROACH

THÈSE N° 3959 (2007)

PRÉSENTÉE LE 14 DÉCEMBRE 2007

À LA FACULTÉ DES SCIENCES ET TECHNIQUES DE L'INGÉNIEUR
LABORATOIRE D'ALGORITHMES ET SYSTÈMES D'APPRENTISSAGE
SECTION DE MICROTECHNIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Eric SAUSER

ingénieur informaticien diplômé EPF
de nationalité suisse et originaire de Affoltern Im Emmental (BE)

acceptée sur proposition du jury:

Prof. D. Floreano, président du jury
Prof. A. Billard, directrice de thèse
Dr R. Cuijpers, rapporteur
Prof. W. Gerstner, rapporteur
Prof. G. Schöner, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Lausanne, EPFL
2008

ABSTRACT

IMITATION is the ability to recognize, learn and reproduce the actions of others. In addition to facilitating the transmission of knowledge and skills, it has been suggested that this fundamental cognitive capacity is at the origin of other human faculties such as action understanding, empathy, mind-reading and language. Although the behavioral processes of imitation have been studied for a century in ethology and psychology, neuroscience has only started being interested in, stimulated by the discovery of a common neural substrate devoted to the recognition and production of actions, namely, the mirror neuron system. However, the existence of this system does not provide a comprehensive answer to all questions about imitation. For instance, as mirror neurons, which fire both during the observation and the execution of actions, were found in the monkey, how could the behavioral differences reported in humans be explained in terms of cortical functions and connectivity? Moreover, because of the existence of this common neural substrate, which is activated irrespectively of who is performing an action, how do humans not confuse their own actions with those of others? And finally, how could different imitative strategies and imitation levels be produced by the brain according to behavioral and neurophysiological constraints? In order to shed some light on these key questions, the approach adopted in this work is to apply the principles of **computational neuroscience** by developing neural models of the cortical networks involved in imitation processes.

In this thesis, the use of artificial neural networks is considered a fundamental methodology. More specifically, **neural fields** were chosen to be the central modeling tools since this class of neural networks have been shown to be endowed with many computational abilities, most of which were actually observed experimentally in cortical regions. Although several computational models addressing imitation have already been proposed, rare are those which consider the previously mentioned issues and especially by means of neural modeling. To fill this gap, this thesis addresses the neural mechanisms underlying several important cognitive processes. They are: *a)* the principle of **ideomotor compatibility**, by which one's own performance is influenced by observing others; *b)* the problem of **frames of reference transformations** that make it possible to put oneself into someone else's place and to simplify the sensorimotor mappings required for imitation; and *c)* the ability to **distinguish oneself from the others** to

make the perceived visual signals related to oneself or to others less ambiguous. Importantly, in order to confirm or refute the neural models that are developed in this work, behavioral studies are proposed along with the predictions of the models. Finally, this thesis also provides a global model of cortical organization in which the brain pathways of imitation are highlighted. Through this synthesis, the neural processes of imitation in humans are suggested to not involve specific cortical areas only as usually believed, but instead, the sensorimotor cortex in its whole.

The contributions of this thesis are basically twofold. First, the developed neural models of the cortical networks involved in imitation contribute to a better understanding of the neural mechanisms underlying this human ability. As such, it suggests several new directions and hypotheses for research in the field of both experimental psychology and neuroscience. Secondly, this thesis also contributes directly to the field of artificial neural networks by providing new technical developments on the current knowledge of the computational power of the neural field approach.

KEYWORDS: Imitation, Cortical networks, Neural fields, Ideomotor principle, Behavioral interferences, Frames of reference transformations, Sensory discrimination.

RÉSUMÉ

L'IMITATION est généralement définie comme le résultat de la combinaison des capacités de reconnaître, d'apprendre et de reproduire les actions produites par d'autres individus. En plus de faciliter la transmission du savoir et des techniques manuelles, cette capacité cognitive semble aussi être à l'origine de certaines facultés sociales telles que l'empathie et le langage. Cependant, des comportements imitatifs non intentionnels et ne requérant aucune forme d'apprentissage peuvent aussi être observés chez l'homme. Même si ces derniers sont plus flagrant chez des personnes atteintes de trouble cérébraux, ils peuvent néanmoins être à l'origine de nombreux comportements vécus au quotidien.

Bien que les processus comportementaux de l'imitation ont déjà été étudiés durant le siècle dernier dans les domaines de l'éthologie et de la psychologie, l'attention des neurosciences ne s'est développée que récemment. Cette dernière a été stimulée par la découverte d'un substrat neuronal dévoué à la fois à la reconnaissance visuelle des actions ainsi qu'à leur exécution, à savoir le système miroir. Cependant, l'existence de ce système miroir ne peut, par lui seul, expliquer tous les phénomènes de l'imitation. Par exemple, étant donné que ces fameux neurones dits miroirs, qui sont activés à la fois lors de l'observation et lors de l'exécution d'actions, ont été découvert chez le singe, comment est-ce que les importantes différences comportementales entre ce dernier et l'homme pourraient être expliquées en termes de fonctions corticales ainsi qu'en termes de connectivité cérébrale? De plus, puisque ce substrat neuronal commun est recruté insensiblement de l'auteur d'une action observée, à savoir soi ou un autre individu, comment ce fait-il que les humains ne confondent-ils pas systématiquement leur propres actions avec celles des autres? Finalement, comment est-ce que les principales stratégies d'imitation reportés chez l'homme peuvent-elles être produites et contrôlées par le cerveau? Afin d'apporter quelques éléments de réponse à ces questions, cette thèse s'est concentrée à appliquer les principes des neurosciences computationnelles. C'est pourquoi, des modèles neuronaux simulant les réseaux corticaux impliqués dans les processus de l'imitation ont été développés.

Tout au long de cette thèse, l'utilisation des réseaux de neurones artificiels a été considérée comme une méthodologie fondamentale. Plus spécifiquement, des réseaux connus sous le nom de **champs de neurones** (ou *neural fields* en

anglais) ont été choisis comme outils de modélisation. En effet, cette classe de réseaux de neurones possède de nombreuses propriétés computationnelles, la plupart desquelles ont été observées expérimentalement dans le cerveau. Malgré le fait que plusieurs modèles computationnels se focalisant sur l'imitation ont déjà été proposés, rare sont ceux qui considèrent les problèmes soulevés précédemment, et plus particulièrement en utilisant des modèles neuronaux. Afin de combler ce vide, cette thèse se concentre sur plusieurs processus cognitifs importants. Ces derniers sont: *a)* le principe de **compatibilité idéomotrice**, par lequel l'observation des actions d'autres individus peut influencer la qualité de ses propres actions, *b)* le problème des **transformations de référentiels** qui permettent de se mettre dans la peau de quelqu'un d'autre, ainsi simplifiant l'imitation, et *c)* la capacité de **se distinguer des autres** qui est importante pour pouvoir déterminer si les signaux visuels perçus appartiennent à soi ou aux autres. Ensuite, afin de pouvoir confirmer ou réfuter les modèles développés dans ce travail, des propositions d'études comportementales accompagnées des prédictions des modèles sont fournies. Enfin, cette thèse propose aussi un modèle global de l'organisation du cortex, dans lequel les chemins cérébraux de l'imitation peuvent être mis en avant. A travers cette synthèse, les processus neuronaux de l'imitation sont suggérés de ne pas recruter uniquement certaines aires cérébrales, mais plutôt le cerveau sensorimoteur dans son ensemble.

Les contributions de cette thèse sont principalement de deux types. Premièrement, les modèles neuronaux représentant les réseaux corticaux impliqués dans l'imitation contribuent à une meilleure compréhension des mécanismes neuronaux responsable de cette capacité humaine. Par conséquent, ce travail suggère de nouvelles directions et des hypothèses destinées à la recherche dans les domaines que sont la psychologie expérimentale et les neurosciences. Deuxièmement, cette thèse contribue aussi au domaine des réseaux de neurones artificiels. En effet, de nouveaux développements techniques étendant les connaissances actuelles concernant les propriétés computationnelles de la classe de réseaux de neurones adoptée ici ont été proposés.

MOTS CLÉS: Imitation, Réseaux corticaux, Champs de neurones, Principe idéomoteur, Interférences comportementales, Transformations de référentiels, Discrimination sensorielle.

*A Maman, Papa, Julien, 正瑜, 惠美, Pati, à tout le reste de ma famille, et
ainsi qu'à tous mes potes . . .*

ACKNOWLEDGMENTS

FIRST of all, I would like to deeply thank my supervisor, Prof. Aude Billard for having offered me the great opportunity to work in a fascinating field of research as well as to evolve in a stimulating environment and enjoyable atmosphere. Your confidence for letting me developing mostly independently my own ideas turned this PhD thesis into a strong challenge which I had great pleasure to take up. Of course, this was not always a easy. I was often filled with uncertainty, but this forced me to learn a lot from the scientific way of thinking, as well as from myself. During the hard times, I am particularly grateful to your support, encouragement, and especially your patience for guiding my research in a wise and involved manner. Moreover, I had also several great opportunities to present my work at various conferences, what offered me the chance to meet other researchers from different universities and fields of research. This would never have been possible without your help, who devoted so much time in reading and correcting the the papers we published throughout my PhD studies.

I would like also to acknowledge the members of my thesis committee, Dr. Raymond Cuijpers, Prof. Dario Florerano, Prof. Wulfram Gerstner and Prof. Gregor Schöner for the time you spent reading my thesis as well as for the scientific and the more casual discussions we had during the day of my private defense. I would like also to thank those of you who I met prior to or during my PhD studies for having nurtured the seed of my interest for research.

To all my colleagues, Agnes, André, Annick, Basilio, Biljana, Daniel, Dave, Davis, Elena, Florent, Hoap, Ioana, Karim, Marie-Jo, Micha, Olivier, Robota, and Sylvain, thank you for the support and friendship that you offered me, for the exciting scientific discussions we had together, as well as for the few beers we shared at Satellite during these four years. Thanks also to the master students, Andres, Jean, Mara and Xavier, who dared to do their semester or diploma project under my supervision. I learned much from our interactions, and I hope you did as well.

I would like also to deeply thank my family and friends for your support and for the love and friendship you gave me. I knew that I could always count on you when my motivation was on the descending slope. When I was so exhausted that I had to change my mind for a while, you were always present. Big up!

Finally, thanks to the Ecole Polytechnique Fédérale de Lausanne for offer-

ing me the opportunity to work in a great scientific environment and enjoyable atmosphere¹ and also to the Swiss National Science Foundation for having supported work².

¹A special thanks goes to the Saltelite's team, who gently allowed me to move temporarily my office into their homelike and lively place.

²The financial support for this work was provided by the Swiss National Science Foundation, through grant 620-066127 of the SNF Professorships program.

TABLE OF CONTENTS

1	Introduction	1
1.1	Motivations	1
1.2	Aim of this Thesis	9
1.3	Road-map of this Dissertation	10
2	Biological and Modeling Background	13
2.1	Views on Imitation	13
2.1.1	Low-level Imitation	14
2.1.2	True Imitation	15
2.2	Imitation in Experimental Psychology	16
2.2.1	The Ideomotor Principle	16
2.2.2	Meaningless and Goal-Directed Imitation	19
2.2.3	Specular and Anatomical Imitation	22
2.2.4	Discrimination between the Self and Others	22
2.3	The Neurophysiology of Imitation	25
2.3.1	The Cognitive Processes in the Brain: An Overview	25
2.3.2	The Mirror Neuron System in Monkeys	48
2.3.3	The Mirror System in Humans	52
2.3.4	Shared Representations and the Discrimination Between the Self and Others	57
2.4	Related Neural Models of Imitation	63
2.4.1	Theories of Imitation	63
2.4.2	Modeling and Understanding the Mirror System	66
3	Artificial Neural Networks	79
3.1	Computational Neuroscience	79
3.1.1	Artificial Neural Networks	81
3.2	Neural Fields	84
3.2.1	Definition	85
3.2.2	Properties of Neural Fields	88
3.2.3	Modeling Cortical Pathways: Network of Networks	107
4	Brain Pathways of Imitation and the Ideomotor Principle	111
4.1	Related Experimental Study	111
4.2	Neural Models	113
4.2.1	Architecture of the Models	114
4.3	Results	121
4.3.1	Replication of the Related Experimental Study	121
4.3.2	Stimulus-Response Incompatible Mapping	122
4.3.3	Metric of Spatial Representation	124
4.4	Discussion	126

5	Frames of Reference Transformations	131
5.1	Introduction	132
5.1.1	The Problem	133
5.2	A Fundamental Neural Building Block	135
5.2.1	Vectorial Representations and Cosine-Tuning	136
5.2.2	A Continuous Attractor Neural Network	138
5.2.3	A Two-layer Neural Network	144
5.2.4	Synaptic Projections across Neural Populations	145
5.2.5	A Gain Field	145
5.3	Neural Models of Frame of Reference Transformations	146
5.3.1	Translations	147
5.3.2	Planar Rotations	147
5.3.3	Extension to 3D Rotations	149
5.3.4	3D Rotations	151
5.3.5	Projections on the Principal Axes	152
5.4	Simulation Results	153
5.4.1	Visuomotor Transformations for Reaching	153
5.4.2	Transformations of Moving Targets	154
5.4.3	Approximation Errors	156
5.5	Discussion	159
6	Interferences in the Transformation of Frames of Reference	165
6.1	Introduction	165
6.2	Experimental Setup	167
6.3	Neural Model	168
6.3.1	Network Architecture	170
6.4	Results	172
6.4.1	Reaction Times and Accuracy	172
6.4.2	Interference Patterns	173
6.4.3	Effects of Inhibitory Modulation	176
6.5	Discussion	177
7	Motion Integration, Motion Sensitivity, Sensory Prediction and Sensory Discrimination	181
7.1	Introduction	182
7.2	Neural Model	185
7.2.1	Architecture	185
7.2.2	Control of the Intrinsic Dynamics	188
7.2.3	Stimulus Encoding	190
7.2.4	Information Transmission across Neural Fields	193
7.3	Experiments and Results	193
7.3.1	Dynamic Velocity Integration	194
7.3.2	Transfer of Information across Neural Fields	194
7.3.3	Dynamic Velocity Tuning	196
7.3.4	Neural Locking and Unlocking to External Stimulus	199
7.3.5	Sensory Discrimination	200
7.4	Discussion	203
8	Movement Generation, Sensory Integration, Sensory Discrim- ination and Interferences	209
8.1	Introduction	209
8.2	Neural Model	213
8.2.1	Generation of Movements on a Spherical Surface: Forward Model	213

8.2.2	Generation of Velocity Commands: Inverse Model	214
8.2.3	Neural Architecture	220
8.3	Experiments and Results	223
8.3.1	Motor Adjustments caused by Multisensory Conflicts . . .	223
8.3.2	Interferences during Simultaneous Execution and Obser- vation of Movements	227
8.4	Discussion	232
9	Synthesis: Toward a Unified Cortical Model of Imitation . .	239
9.1	Contributions of this Thesis	239
9.2	Toward a Unified Cortical Model	240
9.2.1	Cortical Organization along Multiple Frames of Reference	241
9.2.2	The Fronto-Parieto-Temporal Model	243
9.2.3	The Cortical Routes Involved in Imitation	245
9.3	Discussion	251
10	Limitations and Future Work	255
11	Conclusion	261
A	Technical and Implementation Details	263
A.1	Spherical Representation	263
A.2	A Coupled Attractor Model	265
A.2.1	Velocity Integration	266
A.2.2	Stimulus Input	269
A.2.3	Synaptic Projections	272
B	Simulation Parameters	275
B.1	Brain Pathways of Imitation and the Ideomotor Principle	275
B.2	Interferences in the Transformation of Frames of Reference . . .	277
B.3	Motion Integration, Sensitivity and Sensory Discrimination . . .	277
B.4	Movement Generation, Sensory Discrimination and Interferences	278
C	Publications of the Author	279
	References	281
	Curriculum Vitae	301

INTRODUCTION

1.1 MOTIVATIONS

IMITATION is usually defined as the combined ability to recognize, learn and reproduce the actions of others. It is also a fundamental cognitive mechanism for the transmission of knowledge and skills. For decades, the behavioral processes of imitation have been the focus of studies in ethology (Tomasello, Savage-Rumbaugh, & Kruger, 1993) and in developmental psychology (Greenwald, 1970; Piaget, 1978; Meltzoff & Moore, 1997; Wohlschläger, Gattis, & Bekkering, 2003). Indeed, this cognitive skill is believed to have important implications with respect to other cognitive abilities such as action understanding (Zukow-Goldring, 2004), empathy, mind-reading (Gallese & Goldman, 1998) and language (Arbib, 2002). Moreover, research on imitation has also been driven by more practical goals, such as those devoted to develop pedagogical techniques (Piaget, 1978), design human-machine interfaces (Greenwald, 1970; Calinon & Billard, 2007), and understand and treat mental diseases such as apraxia or autism (Field, Field, Sanders, & Nadel, 2001; Iacoboni, 2006; Petreska, Adriani, Blanke, & Billard, 2007).

Another reason why the study of the mechanisms of imitation is constantly attracting more and more researchers, is simply the human desire to understand the brain, and consequently to be able to reproduce artificially its functioning. Importantly, the scientific community involved in robotics has great expectations in improving current control methods by developing computational techniques inspired from the insights provided by neuroscience and psychology (Schaal, 1999; Asada, McDorman, Ishiguro, & Kuniyoshi, 2001; Billard, Epars, Calinon, Cheng, & Schaal, 2004; Calinon, Guenter, & Billard, 2007). One important trend in this discipline consists in endowing robots with the ability to learn from demonstration. Indeed, rather than using abstract programming languages, this method could be a much easier and a more natural way to provide skills to robots, as humans do with their children. The work presented here is however more theoretical, and aims at understanding the underlying neural mechanisms and the cortical structures at the origin of the behavioral expressions of imitation.

The Behavioral Expression of Imitation

Despite the importance of the human ability for imitation, a better understanding of the cognitive and the neural mechanisms underlying it has only recently received significant new insights from both experimental psychology and neurophysiology. Interestingly, the former field of research often considers only the automatic side of imitative behaviors which do not require explicitly any form of learning (Bekkering, Wohlschlger, & Gattis, 2000; Brass, Bekkering, Wohlschlger, & Prinz, 2000; Strmer, Aschersleben, & Prinz, 2000; Kilner, Paulignan, & Blakemore, 2003; Heyes & Ray, 2004). Indeed, the cognitive mechanisms of imitation are not restricted to learning processes. Humans do often exhibit imitative behaviors which are unconsciously triggered and which simply reflect automatic sensorimotor processes. The approach adopted in this thesis assumes that understanding the mechanisms underlying these sensorimotor processes will provide useful insights which may further help get a glimpse of the neural processes of imitation as a whole.

In the developmental and experimental psychology literature, automatic imitative behaviors are often referred as a manifestation of the principle of *ideomotor compatibility*, by which the observation of an action performed by another individual is suggested to activate the motor images responsible for the execution of that same action (Greenwald, 1970; Piaget, 1978; Brass, Bekkering, & Prinz, 2001). As reported by experimental studies, this ideomotor principle has been shown to affect one’s own behavior at different levels. At the motor level, the observation of an action has been shown to facilitate or interfere with movement initiation when both movements are identical or different, respectively (Brass et al., 2000; Bertenthal, Longo, & Kosobud, 2006). In addition, at a more emotional level, a significant increase of the breathing rhythm has been reported in subjects who only observed people performing effortful actions (Paccalin & Jeannerod, 2000). Finally, at a more cognitive level, it has also been shown that the context in which action observation occurs may influence how imitation is performed. For instance, the strength with which one may manipulate an object can be affected by the observation of someone else manipulating either a heavy or a light object (Hamilton, Wolpert, & Frith, 2004). Together, these results suggest that the neural mechanisms associated with action observation and action execution are tightly linked, and subsequently, that several neural structures should be shared by these two processes.

Similar automatic sensorimotor behaviors involving visual stimuli of different kinds are very common in experimental psychology. A famous example is the so-called Simon-effect. It describes that the spatial compatibility between an instruction stimulus and the associated required motor responses produces faster response, whereas spatially incompatible stimulus-response pairs are more prone to interfere with movement initiation (Simon, Sly, & Vilapakkam, 1981). Due to their conceptual similarity, ideomotor and spatial compatibility processes were

often intermingled (Brass et al., 2000). Only recently, studies tried to disambiguate them. Results of behavioral and brain imaging experiments suggest that both processes do contribute concurrently to the selection of a motor response, but they seem to be controlled by different cognitive pathways and brain regions (Brass, Derrfuss, & von Cramon, 2005; Rumiati et al., 2005; Bertenthal et al., 2006). However, the neural mechanisms as well as the exact cognitive routes still remain unclear. Brain imaging studies usually only highlight the cortical areas that are activated during a given task, but they do not give much clues concerning the neural processing occurring at the neuron level. The present work will try to investigate this issue and those that will be described further in this introduction. It will also suggest computational answers as well as methods to validate or refute the subsequently raised hypotheses.

Furthermore, another important behavioral aspect of imitation is its **goal-directedness** (Meltzoff, 1995; Bekkering et al., 2000; Wohlschläger & Bekkering, 2002; Gergely & Csibra, 2003). This means that, when humans are asked to imitate, they are biased to primarily reproduce the goal of observed actions rather than the means of their achievement. However, when the goal of an observed action is absent, the reproduction of purely intransitive or meaningless gestures becomes more accurate (Wohlschläger & Bekkering, 2002; Heyes & Ray, 2004; Rumiati et al., 2005). Interestingly, ethological studies comparing monkey and human behavior showed that the imitative strategies adopted by monkeys are strictly goal-directed, whereas humans usually try to copy both goals and means (Tomasello et al., 1993). This behavioral finding clearly suggests that humans are endowed with a more complex cortical structure for imitation which is also capable to process intransitive actions.

Another issue concerns the human natural tendency for **mirror or specular** imitation with respect to **anatomical** imitation (Bekkering et al., 2000; Wohlschläger & Bekkering, 2002; Heyes & Ray, 2004). Indeed, when a demonstrator is facing a person and executes a movement to imitate, the imitator shows a preferential tendency to mirror the observed movements, i.e., (s)he preferentially uses the left hand to imitate a movement performed with the right hand and vice-versa, rather than to reproduce it as if (s)he was located in the shoes of the demonstrator (Bekkering et al., 2000; Wohlschläger et al., 2003; Heyes & Ray, 2004). From this, one may ask how these two different imitative strategies are encoded in the brain. Are there separate pathways mediating each of them, and how could one strategy be selected against the other?

Neural Substrates

Following from these findings, one may ask for the existence of neural substrates which may explain the causes of these behavioral expressions of imitation. For a long time, the knowledge of the neural substrates and mechanisms of this key human cognitive ability was very limited, but recent collaborative research in neuroscience and experimental psychology has made crucial discoveries at both the brain and the single cell levels (Rizzolatti, Fadiga, Gallese, & Fogassi, 1996; Iacoboni et al., 1999; Decety, Chaminade, Grezes, & Meltzoff, 2002). The most important finding of these studies is the neurological evidence of the existence of a common neural substrate devoted to the recognition and the production of movements, namely, the **mirror neuron system**. However, the existence of such a mirror system may not be sufficient to explain all the mechanisms of imitation in humans. Indeed, since this mirror system was discovered in both humans and monkeys, a direct association between imitation and the mirror system could be seen as contradictory with respect to the behavioral differences between these akin species (Gallese & Goldman, 1998; Zentall, 2003; Lyons, Santos, & Keil, 2006). Nevertheless, the mirror system is still strongly believed to be part of the imitation system. Its probable implications in mental simulations, forward control and the prediction of actions argue in this direction (Gallese & Goldman, 1998; Miall, 2003; Wolpert, Doya, & Kawato, 2003).

Moreover, while this system offers an exciting line of study, it has yet to be shown how this circuit, in connection with other better-known neural circuits for visual representation of movements and for motor control, may explain the behavioral data on imitation. Indeed, imitation is not the behavioral expression of a stand-alone system which would be localized in a specific brain region. It is rather distributed and embedded in the whole cortical sensorimotor circuit. Consequently, the mirror system is suggested to be an emerging product of the combination of lower-level brain mechanisms, which were originally devoted to other functions, e.g., the control and the visual perception of movements (Arbib, Billard, Iacoboni, & Oztop, 2000; Oztop & Arbib, 2002; Wolpert et al., 2003; Keyser & Perrett, 2004). Additionally, the mechanisms of imitation may also be carried out with the help of other and less considered cognitive abilities, such as those mediating the recognition of conspecifics, the distinction between the self and the others, and the attribution of others' intentions or states of mind (Gallese & Goldman, 1998; Decety & Sommerville, 2003; Jeannerod, 2003; Oztop, Kawato, & Arbib, 2006).

Frames of Reference for Imitation

Central to the vast majority of neurophysiological, behavioral and computational studies addressing imitation is the mirror neuron system. It is however located relatively far from the brain areas receiving primary sensory information. As indicated by neurophysiological studies in both humans and monkeys,

this system involves parts of the premotor cortex, posterior parietal cortex and superior temporal sulcus (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Iacoboni et al., 1999; Rizzolatti, Fogassi, & Gallese, 2001). The neurons located within these regions react thus to highly processed stimuli. They were shown to encode information in a goal-centered frame of reference, which suggests a mirror system tuned for action. In addition to this neurophysiological characteristic of the monkey's mirror system, the human ability to imitate intransitive actions may also imply an encoding of the visual inputs related to other individuals within a body- and body-parts- centered frames of reference.

In order to obtain such representations, a common belief, corroborated by experimental data, is that the series of transformations across reference frames, required for transferring visual information initially sensed within the retina into goal-centered representations, is gradually performed by different groups of cells along the visual pathways (Perrett, Harries, Mistlin, & Chitty, 1989; Burnod et al., 1999). This nevertheless requires appropriate neural mechanisms. Moreover, since many neurons, especially those named as mirror, exhibit a sensitivity to sensory information coming from multiple modalities such as proprioceptive, motor, visual and auditory, one may also wonder whether the neural mechanisms in charge of transforming each of these modalities are similar or different. For instance, while proprioceptive information is mostly encoded in a local frame of reference, i.e., each sensory receptor responds to the limb orientation which it is associated with, visual information is encoded within a global reference frame, i.e., all the visual inputs are perceived within the same reference frame centered on the retina. Therefore, it is likely that the nervous system may take advantage of these discrepancies to perform the transformations, but this remains to be proven.

Shared Representations and the Self versus the Others

Another issue which is rarely considered while modeling the mechanisms of imitation arises because of the shared nature of the mirror system. Indeed, this common neural substrate is activated irrespectively of who is performing an action, i.e., either the self or other individuals. From this, one may ask why humans do not confuse their own actions with those executed by others. This ability for recognizing the self from the others, which can already be observed in neonates (Rochat & Hespos, 1997), is fundamental in order to deal with shared representations (Decety & Sommerville, 2003; Jeannerod, 2003). While performing an action, if one simultaneously observes someone else executing a different action, and one is capable to avoid considering this visual information as a self-sensory feedback, one's own action would not be perturbed. However, interferences reported by an increase of movement variability can be noticed in such conditions (Kilner et al., 2003; Chaminade, Franklin, Oztop, & Cheng, 2005). This suggests that the inhibitory effect of self-recognition processes is not as strong as it could be. One may even argue that this weak, but sufficient

inhibition may be at the origin of the behavioral expression of the ideomotor principle, and subsequently of imitative behaviors.

This calls for an explanation as to how the brain is capable, within its shared representations, whose role is also to perform multisensory integration, to minimize the conflicts coming from potentially contradictory sources of information, such as those related to the self and those related to other individuals. First, the visual attributes of a given body seem to contribute to a large extent to self-recognition (Van Den Bos & Jeannerod, 2002; Haggard & Clarke, 2003; Jeannerod, 2003). However, when this information is ambiguous, a mechanism has been suggested whereby an internal prediction of the consequences of motor acts is compared with their real sensory outcomes. Depending on their similarity, the brain may determine the ownership of the observed movement (Van Beers, Baraduc, & Wolpert, 2002; Decety & Sommerville, 2003). Since this issue is rarely considered in computational studies addressing imitation, an important part of this thesis will be devoted to its investigation. But before summarizing the main research questions which will be tackled in this work, a rapid outlook of the related modeling studies on imitation is provided. Note that more details on each topics addressed in this introduction will be presented along this thesis.

Related Computational Models

A growing number of modeling studies have addressed the mechanisms of imitation as well as those underlying the mirror neuron system. These models can be separated within two broad and partly overlapping classes. First, the large majority of the models addressing imitation is not primarily intended to explain the cognitive mechanisms behind the mirror neuron system or imitation, but rather tries to understand the computational prerequisites of imitation and apply engineering techniques for endowing robots with imitative abilities (Schaal, 1999; Ijspeert, Nakanishi, & Schaal, 2002; Calinon & Billard, 2007; Guenter, Hersch, Calinon, & Billard, 2007). Although these models may provide important computational landmarks and key steps toward a better understanding of this human skill, this work is more concerned with the second class of models, which is fundamentally directed toward the understanding of the cognitive and the neural processes underlying imitation as observed in humans.

Interestingly, the authors of this second class of models come from many different fields of research, such as experimental psychology, neurophysiology, robotics and computer sciences, to cite some examples. Consequently, across these studies, several levels of modeling have been adopted. At the highest level, there are general theories of imitation, one of which suggests for instance that this ability is grounded within the sensorimotor loops and may then develop through their maturation (Piaget, 1978). This has then led to the simulation theory which proposes that imitation is driven by sensorimotor resonance, which is triggered by observation of other individuals. When one observes somebody

else performing actions, one is internally simulating these acts as if one was actually executing them (Piaget, 1978; Gallese & Goldman, 1998; Rizzolatti et al., 2001). Then come more specific models such as those putting forward the goal-directedness of imitation (Bekkering et al., 2000; Wohlschläger et al., 2003). High-level studies also emphasize the associative nature of the mechanisms underlying imitation, both at the behavioral level (Heyes, 2001) and at the neural level (Keysers & Perrett, 2004). Going one step down further, there are also computational or connectionist approaches explaining the mirror system, some of which directly aim at reproducing neural activation patterns (Arbib et al., 2000; Oztop & Arbib, 2002; Bonaiuto, Rosta, & Arbib, 2007), whereas the others try more to take advantage of their properties in order to endow their model with imitative behaviors (Billard & Mataric, 2001; Erlhagen, Mukovskiy, et al., 2006; Cuijpers, Van Schie, Koppen, Erlhagen, & Bekkering, 2006; Petreska & Billard, 2006). While these studies mainly focus on a system level, other approaches consider the mirror system as an integrative part of the motor system and even show how mirror properties may naturally emerge from the structure of the motor system (Demiris & Johnson, 2003; Wolpert et al., 2003; Oztop, Wolpert, & Kawato, 2005). Finally, at the lowest level, several studies adopted a developmental methodology as highlighted by Piaget (1978), where the models are embodied, and initially endowed with only simple motor and visual abilities. While exploring their environment, these systems learn their own visuomotor mapping, and by extending their sensitivity to the perception of the body of other agents, imitative behaviors emerge through a form of visual confusion between the visual feedback of these agents and their own internal representation (Andry, Gaussier, Nadel, & Hirsbrunner, 2004; Weber, Wermter, & Elshaw, 2006).

A Computational Neuroscience Approach

This thesis will take inspiration from most of the computational studies presented above in order to investigate the issues raised throughout this introduction. However, this research work departs from these studies from the fact that it primarily investigates important mechanisms required for imitative behaviors to exist, which were often forgotten or bypassed while modeling this phenomenon. They include the neural mechanisms of transformations across frames of reference and the ability to discriminate the self from the others. Nevertheless, this work still follows the same approach as several recent studies by adopting the **computational neuroscience** methodology (Arbib et al., 2000; Keysers & Perrett, 2004; Erlhagen, Mukovskiy, et al., 2006; Oztop et al., 2006; Petreska & Billard, 2006; Weber et al., 2006). Since the major aim of this thesis is to provide a better understanding of the brain circuits of imitation, this approach, which serves as an important theoretical method for investigating the function and mechanism of the nervous system, is well suited. Computational neuroscience allows to bridge experimental psychology, neurophysiology

and theoretical models, by taking inspiration and merging the results obtained in these disciplines. Although computational modeling could never provide exact explanations of the processes occurring in real biological systems, it can however produce simplifying theories and fundamental principles. In addition, this approach is influenced by different constraints than those of experimental studies, and hence may develop hypotheses from different points of view, which could be challenging and fruitful for both scientific communities (Lazebnik, 2002).

Therefore, the neural models developed throughout this thesis will try to stick as close as possible to biology in order to be capable of pretending to model biological systems and giving sound experimental predictions that could be verified by future biological experiments. As modeling tools, the class of neural network known as the **neural fields** or the **continuous attractor neural networks** has been adopted (Wilson & Cowan, 1973; Amari, 1977; Erlhagen & Schöner, 2002; Trappenberg, 2005). Indeed, in the past decades, these mathematical tools were intensively used in order to model and implement brain cognitive functions. In addition to their biologically plausible structural relationship with real cortical neural ensembles, their numerous computational properties have made them very attractive to the computational neuroscience community. For instance, these models, originally proposed to explain the formation patterns of cortical representations, were then applied to other research topics such as visual processing (Ben-Yishai, Bar-Or, & Sompolinsky, 1995; Mineiro & Zipser, 1998; Giese, 2000), visual attention (Rougier, 2006; Vitay & Rougier, 2006), spatial navigation (Zhang, 1996; Redish, 1999; Xie, Hahnloser, & Seung, 2002; Stringer, Rolls, & Trappenberg, 2004), memory (Compte, Brunel, Goldman-Rakic, & Wang, 2000; Rolls, Stringer, & Trappenberg, 2002), motor control (Kopecz & Schöner, 1995; Lukashin, Amirikian, Mazhaev, Wilcox, & Georgopoulos, 1996), decision making (Kopecz & Schöner, 1995; Erlhagen & Schöner, 2002; Schöner, 2002), sensorimotor transformations (Deneve, Latham, & Pouget, 2001; Salinas & Thier, 2000; Meñard & Frezza-Buet, 2005), stimulus binding (Wersing, Steil, & Ritter, 2001), parameter estimation (Deneve, Latham, & Pouget, 1999; Pouget, Dayan, & Zemel, 2003), and even higher-level cognitive functions such as imitation (Andry et al., 2004; Erlhagen, Mukovsky, & Bicho, 2006).

1.2 AIM OF THIS THESIS

The aim of this thesis is to develop **neural models** capable to shed some light on several **key neural mechanisms** related to human ability to **imitate**. As outlined in the previous section, despite the recent and important discoveries related to the behavioral and neurophysiological mechanisms of imitation, several research questions still remain open. Those of which that will receive a special attention throughout this thesis are summarized below.

Brain Pathways of Imitation:

- How are automatic imitative processes grounded on the classical information pathways of sensorimotor transformation? How do these different processes interact, at which cortical level, and how do these interactions affect human behavior?

Frame of Reference Transformations and Imitative Strategies:

- How may the brain transform sensory information across different frames of references? In particular, how could the visual representation of a person's posture gathered in a eye-centered frame of reference be transformed in a reference frame useful for imitation?
- How are the different imitative strategies, i.e., spatial and anatomical imitation, encoded? Is there a single neural structure mediating them or do they involve different neural pathways?

Sensory Integration, Prediction, and Discrimination:

- Are there neural mechanisms that allow internal representations to simultaneously integrate and predict their sensory inputs?
- Then, would these mechanisms be capable of differentiating between the visual signals associated with self-generated movements and those produced by other individuals? And finally, could these neural processes explain the imitative interferences reported during simultaneous action observation and action execution?

In order to address these questions, this thesis will follow two major lines of research grounded on different levels of modeling. First, at the **microscopic level**, this work addresses the mechanisms of information processing within neural circuits, which are composed of a relatively small number of neurons grouped within a compact brain region and subserving some fundamental functionalities. In the light of the main thread of this work, neural representations and

the dynamics of sensory integration and transformation will be analyzed and discussed. As mentioned earlier, although this work focuses mainly on the mental processes related to imitation, the biological inspiration comes from a wide range of subfields of neuroscience. Indeed one could argue that the fundamental neural mechanisms are similar across cortical regions. Complementarily, during the modeling of the neural circuits related to imitation, the resulting models may also, by extension, explain more general processing strategies, common to other cortical areas, than initially thought.

Secondly, at the **macroscopic level**, large network models of these neural circuits will be considered. According to brain imaging and neurophysiological studies, these circuits will be associated with specific cortical areas. The analysis of the interaction between the subparts of such large networks will then help shedding some light on the cortical information pathways of imitation. At both levels, the validation of the models with biological data will allow to provide predictions and to raise new research questions.

1.3 ROAD-MAP OF THIS DISSERTATION

The topics addressed by each chapter as well as their associated contributions are briefly described below.

Chapter 2: BIOLOGICAL AND MODELING BACKGROUND

This chapter reviews experimental findings related to the biological mechanisms of imitation. It includes a description of the results gathered by **experimental psychology** and **neuroscience**, on which the present modeling study is grounded. In addition, some relevant **computational models** that address the brain mechanisms of imitation are also presented.

Chapter 3: ARTIFICIAL NEURAL NETWORKS

In this chapter, the modeling approach, and more specifically the modeling tools, adopted throughout this thesis are introduced. After a brief review concerned with artificial neural networks in general, this chapter then focuses on the class of neural networks known as the **neural fields**. A description of their dynamics and their fundamental computational properties is provided.

Chapter 4: BRAIN PATHWAYS OF IMITATION AND THE IDEOMOTOR PRINCIPLE

This chapter begins the description of the modeling work accomplished throughout this thesis. This study first shows that the neural field approach is adapted to model the dynamics of automatic imitative behaviors. It presents two connectionist models

accounting for a related behavioral study, which addresses the principle of **ideomotor compatibility** in a stimulus-response task involving imitative and abstract visual cues. Interestingly, distinct cortical pathways can be associated to each model, which assumes, respectively, a direct sensorimotor route for imitative responses, or a common decisional pathways for both imitative and goal-directed actions. Thus, in order to confirm or refute either of the two models, a novel experimental paradigm is proposed along with its experimental predictions.

Chapter 5: FRAMES OF REFERENCE TRANSFORMATIONS

In the study presented in Chapter 4, the visual representation of observed movements is assumed to be encoded within the same frame of reference as the representation of self-actions. In order to fill this gap, this chapter addresses the problem of **transformations across different frames of reference**. It basically presents several biologically plausible neural mechanisms capable of performing this transformation, and discusses their properties in the light of neurophysiological data. In particular, this work proposes that different mechanisms may be in charge of transformations involving different sensory modalities.

Chapter 6: INTERFERENCES IN THE TRANSFORMATION OF FRAMES OF REFERENCE

Following from the models developed in Chapter 5, a model involving two parallel imitative strategies based on different mechanisms of frames of reference transformations is described. More precisely, **anatomical and spatial imitation** are considered. In addition, an experimental task, based on an interference paradigm, is also proposed, which requires imitation of meaningless body postures with respect to these two different strategies. The predictions provided by this simulation study are finally suggested to be confronted to real behavioral data in order to validate or refute the model.

Chapter 7: MOTION INTEGRATION, MOTION SENSITIVITY, SENSORY PREDICTION AND SENSORY DISCRIMINATION

This chapter presents an extension of the classical neural field models, that allows the dynamical integration of velocity information for updating dynamically neural representations. By means of a detailed encoding of the inputs to the network, the developed model is then analyzed in the light of its implications to some fundamental neural mechanisms such as stimulus velocity tuning , mental simulation, and sensory discrimination.

Chapter 8: MOVEMENT GENERATION, SENSORY INTEGRATION, SENSORY DISCRIMINATION AND INTERFERENCES

This chapter proposes a model derived from that described in Chapter 7 in order to confront it to real experimental paradigms. The issues addressed here concerns the problem of multisensory integration in situations where biases are imposed on self-sensory feedbacks, and where movements performed by others may interfere with one's own internal representation. In this modeling study, a neural structure capable of generating movements is first described. Then, when coupled with the model described in Chapter 7, a predictive closed-loop system is produced. Finally, by confronting this model to behavioral experiments, several hypotheses and predictions are raised relatively to the neural mechanisms of imitation.

Chapter 9: SYNTHESIS: TOWARD AN UNIFIED CORTICAL MODEL OF IMITATION

This chapter first summarizes the main contributions of this thesis. Then, it presents a global cortical model which encompasses and merges together all the modeling studies accomplished in this work. This synthetic model is finally suggested to provide a coherent view of the neural processes and cortical pathways underlying imitation in general.

Chapter 10: LIMITATIONS AND FUTURE WORK

This chapter provides a discussion related to the main limitations of the work presented in this thesis as well as future directions for research.

Chapter 11: CONCLUSION

This chapter finally concludes this thesis by summarizing its main contributions.

In the **Appendices**, supplementary information such as simulation parameters, additional implementation details, and mathematical developments can be found.

BIOLOGICAL AND MODELING BACKGROUND

THIS chapter presents a general review of the behavioral, neurophysiological and computational literature addressing imitation. The studies herein provide the basis for the modeling work presented throughout this thesis. First, different types of behaviors related to imitation are described from both ethological and behavioral points of view. Next, a review of the fundamental aspects of imitation, as reported by experimental psychology, is given. In order to be able to deeply explore the brain mechanisms underlying imitation, the major neural correlates of imitation are also presented. This neurophysiological review is deliberately broad. Since imitation is believed to be the result of a tight relationship between several sensory and motor processes, presentation of these sub-systems is necessary in order to get a glimpse of the global processes underlying the emergence of the mirror system, a neural substrate thought to be part of the core system for imitation in animals. Following that discussion will be a synthesis of the cortical pathways suggested to be responsible for imitative behaviors in monkeys and humans. The chapter ends with a description of some relevant attempts to model imitative skills and the mirror system.

2.1 VIEWS ON IMITATION

Imitation has been the focus of research interests for a very long time. Early works on this subject can be found in the psychological and ethological literature at the end of the 19th century (Thorndike, 1898; Baldwin, 1902). Without going into too much detail, imitation was, in those days, described as a mechanism providing better chances of survival for an individual within a social group (Baldwin, 1902), by learning to do an act from observing another (Thorndike, 1898). Later, through a longitudinal study of the development of his children Piaget (1978) proposed that imitation may be the mechanism allowing the cognitive transition from low-level sensorimotor processes to higher semantic representations. In parallel, Nadel, Guerini, Peze, and Rivet (1999) have stressed the importance of imitation in the development of social communication skills which begin already in very young children.

During this century of research, imitation has been shown to have different forms, corresponding to different levels of cognitive complexity. Importantly,

some of them can be seen as producing imitative behaviors from the point of view of an observer, but do they really express the use of a special cognitive strategy for imitation? When two friends drink a glass of water at the same time, does it mean that they are imitating each other? The following paragraphs specially aim at giving some definitions on imitation, which contemporary scientists, including behaviorists (Tomasello, 1990), psychologists (Piaget, 1978; Meltzoff & Moore, 1997; Heyes, 2001) and cognitive scientists (Schaal, 1999; Billard, 2002; Demiris & Hayes, 2002) use to distinguish between different forms of imitation .

2.1.1 LOW-LEVEL IMITATION

A general distinction between the levels of imitative acts has been made. At the highest level, what is called true imitation is a behavioral strategy which is usually only attributed to human beings. As will be described in more details further in the text, it requires the combination of complex cognitive processes, such as the attribution of intention to observed individuals, and the ability to faithfully reproduce sequences of movements aimed at reaching a behavioral goal. In contrast, low-level imitation is more concerned with basic sensorimotor mechanisms linking perception with action. Such low-level mechanisms, displayed in both human and monkey, are nevertheless of fundamental importance. Indeed, in a bottom-up approach to the development of behaviors, they are suggested to be prerequisites on which true imitation can develop.

Stimulus Enhancement and Emulation

In one of its simplest forms, imitation is considered as the result of a **stimulus enhancement**, which elicits a specific motor response to a stimulus. Importantly, the trigger of the imitative act is not the observation of the action per se, but the presence of an object, i.e., the stimulus, which has become salient due to the actions of others (Byrne & Russon, 1998). By focusing the attention on the salient object, the probability of reproducing the same action as that which has made the object relevant, is increased. In this case, the chance to observe an imitative act is more the result of a biased lottery. Indeed, here, imitation is not intentional. Similarly, the mechanisms of **emulation** are based on the same principles (Byrne & Russon, 1998). The subtle difference between stimulus enhancement and emulation is that, emulation depends only on the satisfaction of a specific goal (Tomasello, 1990). For instance, when observing someone acting in order to reach a goal, the mechanisms of emulation can initiate any behavioral strategy which can successfully reproduce that goal. Again, an imitation of the means of achieving the goal gets more chance to be observed, but is still not intentional.

Response Facilitation

Next, a third form of imitation is related to the means of achieving an action-goal, namely **response facilitation** (Byrne & Russon, 1998; Schaal, 1999). The probability of imitating the means of an action is increased by its prior observation. The underlying mechanisms of response facilitation are proposed to be the result of a visuomotor resonance. The representation of an action, which has to be known by the observer, is stimulated by the observation of that same action. As a consequence, its reproduction becomes more likely. Examples of such imitative behaviors have been reported in several species such as in rats (Heyes & Dawson, 1990), chimpanzees (Tomasello et al., 1993; Whiten, Custance, Gomez, Teixidor, & Bard, 1991) and humans (Meltzoff & Moore, 1977). In humans, behaviors involving response facilitation have been described in young infants. For instance, they were shown to be capable of imitating tongue protrusion while observing an experimenter doing the same action in order to trigger this specific behavior (Meltzoff & Moore, 1977). Interestingly, the authors of this study suggested a model in which the observation of actions performed by others triggers an internal representation responsible for reasoning about self-actions as well as for their execution (Meltzoff & Moore, 1997). Indeed, this form of imitation is the first one which considers the concept of shared representations between the observation of action and its execution. Importantly, this shared representation hypothesis has recently been shown to possess neural correlates in the primate brain (Rizzolatti, Fadiga, et al., 1996; Iacoboni et al., 1999; Decety & Sommerville, 2003), which will be addressed in details in Section 2.3 of this chapter.

Although the forms of imitation described above are similar because they can all be unintentional, they also share the property that they evolve at the level of an atomic act, i.e., a single action, and not at the level of complex sequences of movements. The observation of an action or an object is directly mapped into a motor representation within the motor primitives of the observer. Consequently, a novel action, not already present in the motor repertoire can not be imitated. Moreover, since memory is not considered in these imitative strategies, sequences of movements leading to a satisfaction of a goal can also not be reproduced. These issues are precisely what distinguishes low-level imitation from true imitation.

2.1.2 TRUE IMITATION

In its most complex form, **true imitation** requires the ability to reproduce and learn new motor skills which are not part of the imitator's current motor repertoire. The imitator must also be capable of extracting the purpose of a given sequence of movements, and reproduce the means of its achievement faithfully in the correct order (Byrne & Russon, 1998). In addition to the intentionality

of true imitation, the ability to decompose observed movements into primitives, and then to structure them in a novel hierarchy is fundamental (Byrne, 1999). Moreover, true imitation is also often endowed with generalization capabilities in time and space. Spatiotemporal generalizations correspond to the ability to imitate in a different moment in time and in a different context. Cognitive processes for understanding the steps of how to achieve a goal by observation in a particular situation, should allow either complete or partial repetition, when the goal or sub-goals are perceived in a different context.

This paragraph ends the brief review of how types of imitation have been distinguished during the past century. This section was not meant to be extensive, but tried to concentrate on some of the fundamental aspects of imitative behaviors on which this work is based. However, several open questions remain with respect to these behavioral aspects. They concern the cognitive processes which enable imitative behaviors to develop as well as the possible cortical discrepancies between humans and monkeys which results in the behavioral differences reported above between these species. In addition, quantitative results are missing from these studies which mostly describe imitation qualitatively. This last methodological issue nevertheless been partly resolved thanks to the growing interest of experimental psychology to imitation. Importantly, several other important features of imitation have been discovered, such as the goal-directedness or the meaninglessness of this behavior. These new fundamental issues and principles which have been developed in experimental psychology are described in the next section.

2.2 IMITATION IN EXPERIMENTAL PSYCHOLOGY

Studies in experimental psychology have shown several important behavioral aspects of human imitative abilities and related functions. Central to the present thesis are the principle of ideomotor compatibility, the distinction between goal-directed and meaningless imitation, and the discrimination between the self and the others. Definitions of these aspects and related experimental findings are reviewed below.

2.2.1 THE IDEOMOTOR PRINCIPLE

The ideomotor principle states that observing the movements of others influences the quality of one's own performance (Greenwald, 1970; Brass et al., 2000; Kilner et al., 2003; Heyes, Bird, & Haggard, 2005). This theory means that the execution of a given action or gesture should be facilitated by a prior observation of that movement. In contrast, the observation of a different movement should produce interferences. In psychological terms, if a perceptual event is similar to the response image that is used to control a response, then the perceived event

should activate the response image and, hence, influence the initiation of the response (Greenwald, 1970; Brass et al., 2000; Wohlschläger et al., 2003). This principle, also referred as the common-coding theory, is tightly related to the low-level forms of imitation reviewed above (Hommel, Musseler, Aschersleben, & Prinz, 2001; Knoblich & Flach, 2001). It implies a motor resonance mechanism, through which the observation of an action performed by others may activate, at least partially, the same motor control centers responsible for its planning and execution (Greenwald, 1970; Hommel et al., 2001; Knoblich & Flach, 2001). Consequently, action reproduction would become easier and unconscious and automatic imitative behaviors, which are sometimes referred as the result of a motor contagion, may be observed (Greenwald, 1970; Gallese & Goldman, 1998; Paccalin & Jeannerod, 2000; Blakemore & Frith, 2005). This mechanism of motor resonance or motor simulation has already been suggested to be used during behavioral tasks requiring motor imagery, which consists in mentally simulating a movement without actually executing it (Jeannerod & Decety, 1995). Interestingly, the timing at which these mental movements are performed correspond to that at which real movements are executed (Jeannerod & Decety, 1995). The link between this finding and the ideomotor principle, is that they both suggest an involvement of visuomotor representations shared by action observation and by extension, mental observation and actual motor execution (Jeannerod & Decety, 1995; Gallese & Goldman, 1998; Porro et al., 1996; Jeannerod, 2003).

Several studies involving an experimental stimulus-response paradigm have measured the behavioral expression of the ideomotor principle (Brass et al., 2000, 2001; Kilner et al., 2003; Chaminade et al., 2005; Bertenthal et al., 2006). They experimentally confirmed the existence of an unconscious and automatic tendency for imitation. For example, in the studies by Brass et al. (2000) and Bertenthal et al. (2006), human subjects were asked to respond as fast as possible with a movement of a specific finger to the presentation of different stimuli. As shown in Figure 2.1a, the stimuli consisted of finger movements on which a visual or symbolic cue was superimposed. The authors of these studies showed that when the visual cue is the only relevant feature to compute the finger response, the observed finger movement does *interfere* with the instructed response (see Fig. 2.1b). An additional finding following from these studies, is that the time course of the inhibition of the imitative automatic tendency was different to that of the other types of cue. They concluded that visuomotor and imitative processes may be mediated by two different control strategies (Brass et al., 2000; Bertenthal et al., 2006). However, the exact cortical pathways through which these control strategies are processes, as well as the mechanisms mediating their interactions still remain unclear. Moreover, do these strategies recruit only purely sensorimotor pathways, or are higher-level cognitive processes also involved?

In a different setup, illustrated in Figure 2.1c, Kilner et al. (2003) instructed

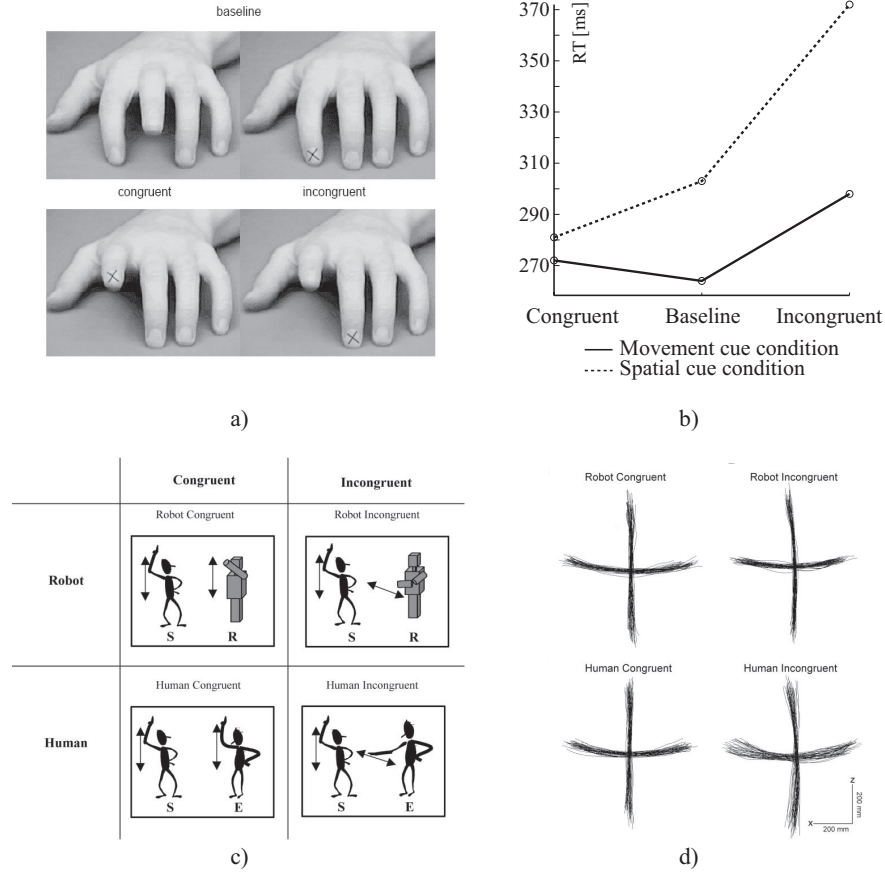


Figure 2.1: **a)** Examples of stimuli used in the experiment of Brass et al. (2000) who investigated the influence of spatial and movement cues in the initiation of imitative response. The baseline conditions only involved one of the possible cues. When the spatial cue was located on the moving finger, this condition was considered as congruent, and otherwise incongruent. **b)** The results of this experiment mainly showed that responses to a movement cue are faster than to a spatial cue and that the movement cue produces more interferences. **c)** Experimental setup of the interference paradigm proposed by Kilner et al. (2003) where human subjects were requested to perform horizontal or vertical movements while observing either another human or a robot arm performing similar or different movements. **d)** The recorded trajectories in all conditions revealed a clear effect of ideomotor compatibility. Congruent observed-executed movement pairs produce less variance in the executed movements than incongruent pairs. Moreover, observing a robot arm does not lead to such interferences. (*Adapted from Brass et al. (2000); Kilner et al. (2003)*)

subjects to perform vertical and horizontal arm movements while simultaneously observing a human executing similar movements. They showed that when observed and executed movements are **congruent**, i.e. similar, the variability of the recorded movements is smaller than during conditions where the movements are **incongruent**. Additionally, they also showed that the observation of a robot arm performing the same motions does not reproduce this interference effect, suggesting the influence of a mechanism specialized for the processing of biological motions (see Fig. 2.1d). These results clearly suggested that the reported movement interferences should be caused by cortical areas responsible for motor control, but which also receive sensory signals associated with the perception of others' movements. Consequently, one may wonder how multisensory signals are integrated within these shared representations and also how interferences may emerge from these multisensory interactions.

2.2.2 MEANINGLESS AND GOAL-DIRECTED IMITATION

Another important behavioral aspect is that imitation has been shown to be primarily goal-directed, especially in young infants (Meltzoff, 1995; Bekkering et al., 2000; Wohlschläger & Bekkering, 2002; Gergely & Csibra, 2003). This emphasizes the fact that, when imitating, copying the goal of an action seems to be the primary aim, whereas the means of its achievement is secondary. However, when the goal of an observed action is absent, a good reproduction accuracy of meaningless gestures is still observed (Wohlschläger & Bekkering, 2002; Heyes & Ray, 2004; Rumiati et al., 2005).

For instance, in an experiment involving the imitation of finger movements, Wohlschläger and Bekkering (2002) requested subjects to simply imitate what they were seeing. The stimuli shown in Figure 2.2a, consisted of the left and right hands of an experimenter placed next to each other on a table, in a closed posture but with the index fingers extended. As a movement to imitate, one of the two fingers executes a tapping motion, which is either strictly vertical or which crosses the midline to end under the other finger. Wohlschläger and Bekkering (2002) also devised two additional conditions: one with and one without an object-like cue located below each finger. Compared to the first condition, the second involved marks placed on the table, which were considered as the object cues. The results of their experiment, reported in Figure 2.2b, showed that, when there are no objects (no marks), gestures are generally well reproduced, and that the reaction times between the two finger conditions (involving a crossing trajectory or not) are fairly equivalent. However, with the object cues, this tendency changed. More movement errors were observed and the reproduction of what they called the goal of the action, i.e., reaching the corresponding mark on the table, was favored against the exact reproduction of the movement. Consequently, one may be interested in knowing which cognitive

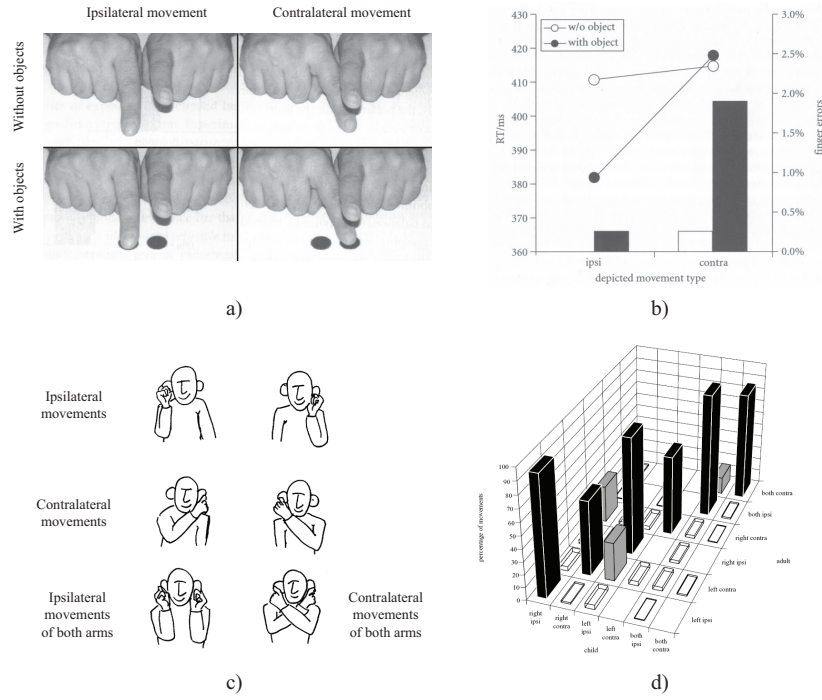


Figure 2.2: **a)** Set of movements to imitate that Wohlschläger and Bekkering (2002) used in an experiment investigating the effect of goals in imitative strategies. **b)** Their results indicate that when object-goals are present, they facilitate the initiation of the movement response as compared to conditions without objects. This suggests a cognitive switching to a goal-directed imitative strategy. Indeed, more errors in the exact reproduction of the observed movements were reported when the latter movements are contralateral, i.e., are not the most effective movements to reach the goal. **c)** Another experiment reported by Wohlschläger and Bekkering (2002) is shown, where children were asked to imitate the illustrated movements performed by an experimenter. **d)** Clear mirror imitative tendencies were observed, in that most of the movements performed with the left hand were reproduced with the right hand, and vice-versa. Moreover, a goal-directed strategy was also observed. Indeed, children also had a tendency to imitate contralateral ear-reaching movements with ipsilateral gestures.

processes are at the origin of this natural bias toward the reproduction of goals.

Other behavioral findings also provide causal evidence for the dominance of goal-directed imitation. During the observation of movements, humans tend to primarily fixate on the end effect of the demonstrator. More importantly, a bias toward the end-point of the trajectory, if predictable, was reported (Mataric & Pomplun, 1998; Flanagan & Johansson, 2003). Thus, an attentional mechanism seems to limit the information conveyed by the demonstrator's movements by providing the brain with the most relevant part of the observed gestures, i.e., the goal. Since the focus is on the end-point of the trajectory and, by extension, on the goal, this may explain why goal-directed imitation sometimes appears to overshadow the means of achieving goals (Heyes, 2001). Indeed, the imitation of the goal of an action is easier than to copy the complete means of achieving it. In addition, the former strategy also avoids the cognitive problem of being capable of corresponding the gestures of others into self movements since the

only relevant aspect of the action to reproduce is its interaction with the goal (Nehaniv & Dautenhahn, 2002).

There is currently a debate regarding whether imitation of meaningless and goal-directed gestures are really mediated by the same mechanism in humans, i.e., by a mirror system which has been suggested to be a shared representation recruited by both action observation and motor execution. Indeed, as will be described in more details further in Section 2.3, the activity of that neural complex in monkeys, which were not reported to imitate the means of observed actions, correlates only with goal-directed actions (Rizzolatti, Fadiga, et al., 1996). On one hand, it is argued that the human mirror system extends that of its ancestor by receiving additional connections from brain regions subserving the observation of others' body movements and those representing self motion. This extension may allow the mirror system to be more flexible through multiple levels of granularity (Lyons et al., 2006). Hence, the reproduction of meaningless gestures becomes possible, and by extension, the means of performing actions. This hypothesis was strengthened by a comparative study of imitation in chimpanzees and human children (Horner & Whiten, 2005), where the subjects looked at an experimenter removing a salient object from a puzzle-box through a series of actions, some of which were necessary, some of which were not. The experiment showed that human children have an overwhelming tendency to imitate all the observed actions, whereas the monkeys imitate only the useful actions to open the box so as to reach the goal. Human children thus imitated all the particular means of attaining the goal (Whiten et al., 1991; Horner & Whiten, 2005). In addition, in the experiments of Wohlschläger and Bekkering (2002) which was previously described, the behavioral results showed that the effect of the goal had only a facilitatory effect on reaction times as compared to the case where no goal was present. Since no clear interferences were observed, this suggests a single mechanism for imitation of both meaningless and goal-directed movements.

On the other hand, recent findings suggest that the processing pathways for meaningless and goal-directed imitation are separate (Stürmer et al., 2000; Rumiati et al., 2005). In an experiment contrasting the facilitation of imitative responses to static and moving movement stimuli of the hand, Stürmer et al. (2000) investigated whether the imitation of gestures and postures were mediated by different cognitive mechanisms. They effectively showed a significant difference in response reaction times to the processing of both types of stimuli. In order to link this finding with the possible distinction between the process of imitation of meaningless and goal-directed gestures, one may consider that static images represent final postures corresponding to goals, while videos of moving hands would correspond to the means of performing actions. Consequently, as suggested by Rumiati et al. (2005), observing a qualitative difference between the imitative processing of static and moving hands may reveal the existence of different cognitive strategies for goal-directed and meaningless imitation.

2.2.3 SPECULAR AND ANATOMICAL IMITATION

Another interesting issue concerns the human natural tendency for **mirror** or **specular** imitation with respect to **anatomical** imitation, which has been observed in several behavioral experiments (Bekkering et al., 2000; Wohlschläger & Bekkering, 2002; Koski, Iacoboni, Dubeau, Woods, & Mazziotta, 2003; Heyes & Ray, 2004). When a demonstrator is facing a person, and executes a movement to imitate, the imitator shows a tendency to mirror the observed movements, i.e., he/she uses the left hand to imitate a movement performed with the right hand and vice-versa (Bekkering et al., 2000; Heyes & Ray, 2004). An experiment was conducted by Wohlschläger and Bekkering (2002), where they instructed children to imitate the movements of an experimenter facing them. These movements are shown in Figure 2.2c. Their results clearly showed the great influence of the mirror imitative tendency (see Fig. 2.2d). A common explanation for this effect to occur is based on the spatial compatibility between the visual location of the stimulus and that of the hand. This effect has been known for a long time as the Simon effect (Simon et al., 1981; Eimer, Hommel, & Prinz, 1995; Heyes & Ray, 2004). However, the intriguing point here is that the Simon effect is generally not associated with imitation. Indeed, imitation is supposed to involve different cortical mechanisms than those involved in classical visuo-motor transformations (Brass et al., 2005). Thus, there should be a cognitive mechanism mostly devoted to imitation which exhibits a preference to mirror imitation (Koski et al., 2003).

During goal-directed imitation, in contrast to the imitation of meaningless gestures, the mirroring tendency slightly differs. For instance, when reaching for objects, a bias toward the use of the dominant hand is mostly observed (Wohlschläger & Bekkering, 2002). But still, it seems that the selection process of the effector to use considers simultaneously the hand which is the closest to the target, the dominant one and that which mirrors the demonstrator (Bekkering et al., 2000).

2.2.4 DISCRIMINATION BETWEEN THE SELF AND OTHERS

The final psychological aspect which is considered here concerns the problem of recognizing the self from the others (Jeannerod, 2003), an ability which can already be observed in neonates (Rochat & Hespos, 1997). In addition to the implications of such a process in almost pure philosophical issues such as the mechanisms of consciousness, it is the focus of several hypotheses that are raised from the study of imitation (Decety & Sommerville, 2003; Jeannerod, 2003). Indeed, the most common hypothesis related to the mechanisms of imitation concerns the existence of a neural substrate shared by the mechanisms of movement perception and motor execution. This substrate has been shown

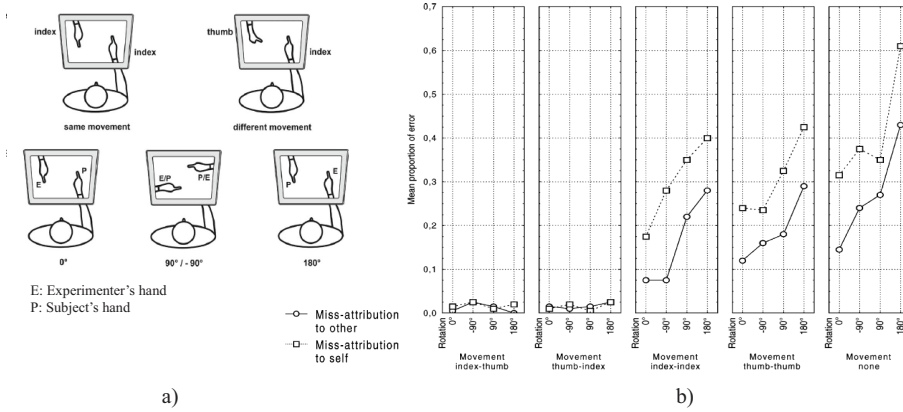


Figure 2.3: a) The experimental setup proposed by Van Den Bos and Jeannerod (2002) is shown, where the capacity of humans to recognize their movements in ambiguous situations was tested. Two hands wearing identical gloves were displayed on a screen. One was that of the subject, and the other was that of the experimenter. The location of the hand was varied by a rotation of the displayed image. Subjects and experimenter were requested to execute a given finger movement simultaneously. After the completion of the movements, subjects had to determine the ownership of a hand selected by the experimenter. b) The results of this experiment indicate that when the movements of both hands are different, the task is always correctly solved. However, for different movements, errors were observed, which did also increase as a function of the amount of rotation of the display.

to be similarly activated when observing an action and executing that same action¹. This calls for an explanation as to how one may keep track of the ownership of the "mentalized" information, especially when self execution and observation of others occur simultaneously. Are there interferences? And if yes, how does the brain minimize the conflicts coming from these contradictory sources of information?

Current findings consider two forms of sensory cues (Van Den Bos & Jeannerod, 2002; Decety & Sommerville, 2003; Haggard & Clarke, 2003; Jeannerod, 2003). First, when one has to recognize one's own limb, body cues such as the spatial and visual attributes of the limb are primarily used. This matching of visual, tactile and proprioceptive signals originating from a specific body part constitutes a multimodal sensory image of the body, often referred to as the body image (Gallagher, 2000; Holmes & Spence, 2004). An experiment involving a rubber arm positioned in place of a subject arm hidden by a screen, showed that, until a certain level of discrepancy between the position of the fake and of the real arm, humans tend to perceive an illusory touch at the location of the stimulation of the rubber arm rather than at that of their real arm (Botvinick & Cohen, 1998; Farné, Pavani, Meneghello, & Ládavas, 2000). These studies showed that self-recognition relies primarily on visual cues. Vision dominates over proprioception. However, experimental results indicate that

¹at least in monkeys. In humans this property was derived from brain imaging studies in homologous areas showing a equivalent activation during observation and execution of an action (Iacoboni et al., 1999).

when these attributes are ambiguous, one can rely on action or movement cues such as the time course of movement, velocity, and acceleration (Van Den Bos & Jeannerod, 2002; Jeannerod, 2003). In an experiment depicted in Figure 2.3a, subjects were shown two hands wearing identical gloves on a screen (Van Den Bos & Jeannerod, 2002). One hand was their own whereas the other was that of the experimenter. This setup also allowed variation of the visual location of the hands on the screen. The subjects were requested to perform a finger movement simultaneously with the experimenter and then to determine the ownership of a given hand. The results of this experiment, reported in Figure 2.3b, clearly showed that when the movement of both hands was different, subjects correctly identified the ownership of a given hand. However, when the movements were identical, the hands were often misattributed. This misattribution also increased when the discrepancy between the real position of the hand and its corresponding location on the screen was increased. Nevertheless, the errors were generally below the level of chance, indicating that humans can still recognize their movements despite visual ambiguities. A mechanism has been suggested whereby an internal prediction of the consequences of a motor act performed by neural processes associated with motor imagery, is compared with real sensory outcomes. Depending on their similarity, the brain may determine the ownership of the observed movement (Frith, Blakemore, & Wolpert, 2000; Decety & Sommerville, 2003; Wolpert et al., 2003). Although this mechanism of disentangling sensory inputs corresponding to the self and to the others is efficient in well-being people, when deficient, it can produce serious perturbation in the representation of the self (Blanke & Mohr, 2005). Nevertheless, in normal conditions, interferences at the level of shared representations still occur even if one is clearly conscious of self with respect to others (Kilner et al., 2003; Chaminade et al., 2005). This interference effect builds a clear bridge between the mechanisms of self perception and the ideomotor principle previously described. Observing the movements of others influences the quality of one's own performance (Greenwald, 1970; Brass et al., 2000). This link between self perception and imitation consequently raises questions regarding the neural mechanisms which allow humans to disambiguate sensory signals related to either the self or others. How could a neural model be capable to implement the sensory prediction of one's own movements and then suppress others-related incompatible sensory feedback. Moreover, how would such a model account for the processes of self-recognition as well as for the interferences observed during simultaneous action observation and execution?

Before addressing, along this thesis, the research questions raised through this review of the data related to imitative behaviors as reported by experimental psychology, next section will present experimental results gathered at the brain level which will allow to ground the modeling studies developed here on a strong biological basis.

2.3 THE NEUROPHYSIOLOGY OF IMITATION

Neural correlates of imitation as well as other cognitive abilities have been gathered by a large set of experimental studies which used different brain recording techniques. In monkeys, data are principally collected through single-cell recordings and more recently, with multi-electrode arrays. These methods allow monitoring of the activity of neurons *in vivo* while the animal is performing various tasks, and then to correlate neural responses with specific controlled parameters of the task. In humans, since such invasive techniques are rarely possible, lesions studies and more recently, brain imaging techniques allow the identification of the neural networks involved during specific behaviors. Positron Emission Tomography (PET), functional Magnetic Resonance Imaging (fMRI) and electro-encephalography (EEG) are the most common, non-invasive tools that are adopted today. They allow the visualization of which brain areas which are activated during given behavioral tasks, and hence they are useful tools to determine the functions of cortical regions. A last technique used on both species is known as Transcranial Magnetic Stimulation (TMS), which generates a high energy magnetic field focused on a precise, targeted brain area. The consequence of such a stimulation is an external modification of the neural activity, which usually produces observable alterations of normal behavior. The combination of the data gathered through cell recordings in monkeys and brain imaging in humans is expected to provide a coherent and global picture of the neural mechanisms of the human cognition. However, one must still pay attention to the homologies and differences across these kin species (Arbib, 2002; Arbib & Bota, 2003). This work clearly relies on a strong similarity between the neural processes occurring in monkeys and humans and hence assumes, except when explicitly mentioned, that the neurophysiological findings obtained in monkeys also apply to the human brain.

2.3.1 THE COGNITIVE PROCESSES IN THE BRAIN: AN OVERVIEW

This section presents an overview of the principal findings related to the neural mechanisms underlying human imitative abilities. But before going into a detailed description of the neural correlates of imitation and related functions, a quick summary will first introduce the global functional organization of the primate brain.

FUNCTIONAL ORGANIZATION OF THE PRIMATE BRAIN

The brain is the principal component of the nervous system. It can be delineated into several subparts according to functional and anatomical properties.

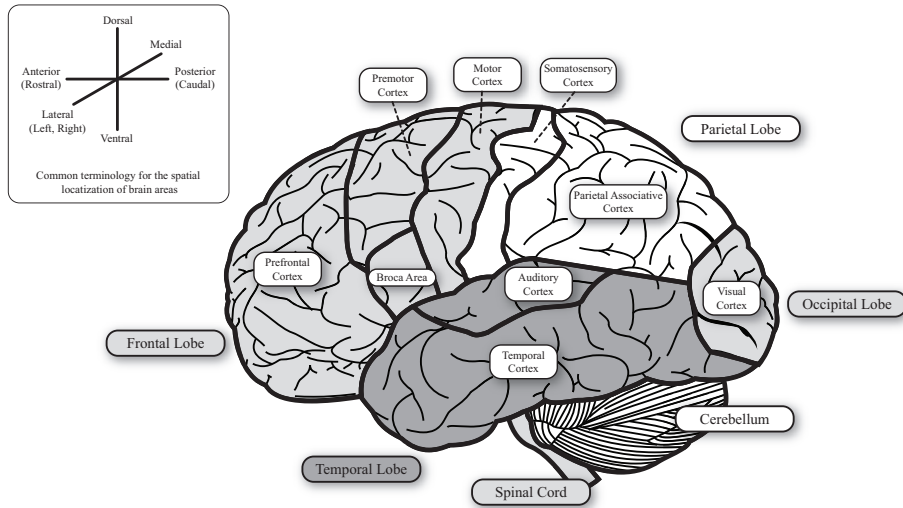


Figure 2.4: Lateral view of the human brain. The cerebral cortex comprises four major lobes which can in turn be split into areas responsible for specific sensory and motor processes.

As illustrated in Figure 2.4, the most prominent part of the central nervous system (CNS) is the cerebral cortex whose functionalities will be discussed at length further in the text. Another important nervous center is the cerebellum, which participates in sensorimotor control and in learning of movements. Located ventrally and between the two hemispheres of the cerebral cortex, the limbic system is composed of several sub-regions such as the hippocampus, the basal ganglia, and the hypothalamus. These regions are suggested to be the locus of the processing of emotion, motivation, learning by reinforcement, long-term memory, and the representation of space (Doya, 1999; Redish, 1999; Hirokawa, Nakamura, Sakai, & Nakahara, 2002). Finally, the sensorimotor relay between the spinal cord and the brain is the thalamus, which is involved in the routing of sensory and motor information to and from the cerebral cortex, the limbic system and the cerebellum.

Each hemisphere of the cerebral cortex can be divided into four principal lobes: the occipital, parietal, temporal and frontal lobes. Each of these lobes is involved in different cognitive processes and can also be split into more specific sub-regions. The occipital or visual cortex, located in the posterior part of the cerebral cortex, is primarily concerned with the analysis of visual information. Its neighbors are the parietal lobe, located dorsally, and the temporal lobe, located ventrally. The temporal lobe can be primarily considered as an extension of the visual cortex, in the sense that it also processes visual information. It is however, also involved in higher cognitive abilities such as the recognition of objects and people present in the environment. Next, the parietal cortex integrates and processes multisensory information such as vision, proprioception, audition, taste, and olfaction. These sensory modalities are first processed sepa-

rately along the borders of the parietal cortex. As information flows toward the center of the lobe, a multisensory association area merges these sensory inputs to form a coherent view of the body and its environment. Finally the frontal lobe can be divided into two main subregions, namely the motor and prefrontal cortices. While the motor cortex, located in the posterior part of the frontal lobe near the parietal cortex is in charge of the control of movements, the prefrontal cortex is involved in higher cognitive functions such as the control of the sensorimotor processes occurring within the other lobes, planning, reasoning, and top-down attention.

This global view of the brain is however too simplistic in order to have a detailed view of the neural mechanisms underlying human cognition. Therefore, the next sections will provide a more precise description of the major streams of brain information processing as well as the functions of the cortical areas which are relevant for the present study on the neural mechanisms underlying imitation (See Figure 2.5). Indeed, the goal of this section is not to provide an extensive review of the current knowledge of the neural processes occurring in the whole cerebral cortex and lower brain regions. Therefore, this review only concentrates on the main cognitive prerequisites of imitation, which are:

- the visual perception of the interactions between people and objects,
- the visual perception of objects for self action,
- the integration of multisensory information at the core of the coherent view of the self, i.e., the body schema,
- the planning and production of movements
- the cognitive control of specific parts of the previously mentioned systems for normal behavior.

Just before beginning this review with the description of the perception of visual information, an important point must be clearly stated. This work is primarily concerned with the neural processing occurring at the cortical level. Therefore, although this work acknowledges the fundamental role of the other regions in motor, spatial and visual processing, the following review will not explicitly address the precise neural activity patterns found within these regions. Nevertheless, if the implication of a given region is fundamental for a clear understanding of the mechanisms of a specific cortical process, it will be mentioned.

VISUAL PERCEPTION

The processing of visual information is one of the fundamental abilities of brain. Central to imitation, the brain must extract the relevant visual features

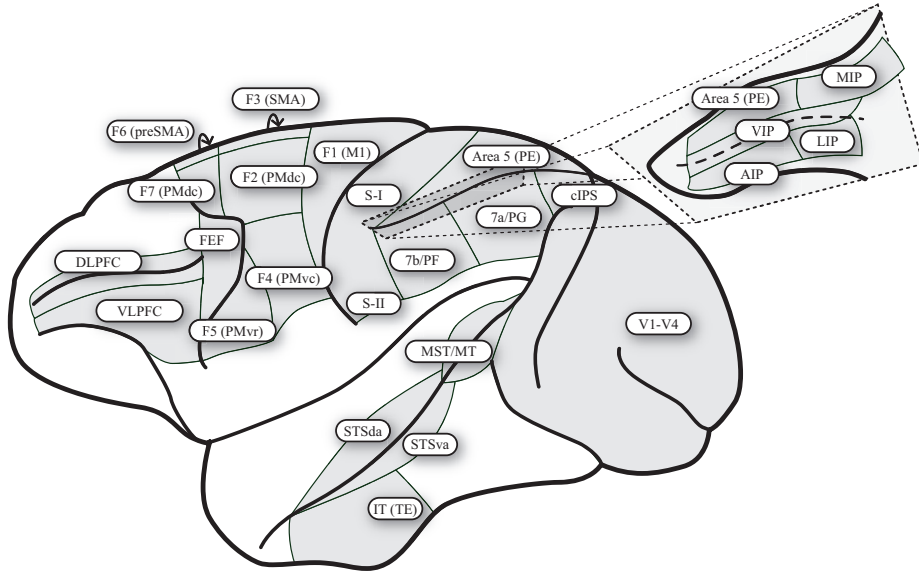


Figure 2.5: Functional organization of the cerebral cortex: On the lateral view of a macaque brain, the principal brain areas involved in the sensorimotor process of visual perception for action, motor control, recognition of actions performed by others, and by extension, imitation are shown. (*Adapted from Rizzolatti and Luppino (2001); Sakagami et al. (2006)*).

to identify objects, people and their respective interactions. Based on this internal sensory representation of the environment, imitative responses to observed actions can then be computed. However, imitation does not only rely on visual information, but still, visual perception is of critical importance.

First, visual information enters the brain through the eyes. The sensory receptors distributed on the retina transform light colors and intensities into electrical signals which are further processed by the brain. Relayed by the lateral geniculate nuclei (LGN) in the thalamus, visual information then enters the cerebral cortex through the primary visual cortex (V1). The visual cortex is composed of several areas (Gattass et al., 2005). In the occipital lobes areas V1, V2, V3 and V4 have mainly been shown to process visual information by extracting, through increasing levels of complexity, orientations, contours, structures of motions, and patterns of colors (De Yoe & Van Essen, 1988; Goodale & Milner, 1992). These features are represented in retinotopic maps respecting the topography of the image. Moreover, along the visual pathway, the receptive fields of the neurons increases in size, indicating that, as visual information flows into the cortex, the neurons encode more and more complex visual features. For instance, whereas V1 neurons have been shown to detect local orientations of a very small subset of the visual scene, those of area V4 can specifically respond to the orientation of objects of larger size, such as bars of several degrees on the visual field (De Yoe & Van Essen, 1988; Goodale & Milner, 1992).

After these few processing stages, the visual processing pathway splits in two different routes, one directed toward the temporal cortex and the other toward

the parietal cortex. These pathways are suggested to process different aspects of visual information. Initially believed to separately process forms and motions, the ventral and dorsal streams were called the "what" and the "where" streams (De Yoe & Van Essen, 1988). However, following from new discoveries, they were renamed as the "what" and the "how" streams, respectively (Goodale & Milner, 1992). Basically, the dorsal "how" stream is suggested to process visual information useful for action, in the sense of how objects could be manipulated, where they are, and in which direction they move (Johnson, Ferraina, Bianchi, & Caminiti, 1996). In contrast, the ventral "what" stream is assumed to be the locus of the recognition of objects, which allows the precise identification of what is present in the visual field (Goodale & Milner, 1992; Booth & Rolls, 1998). These both pathways are suggested to be involved in the cognitive process of imitative behaviors. While the latter stream would allow, for instance, to identify people and the way they interact with objects (Perrett, Harries, Bevan, et al., 1989), the former would rather make possible the correct reproduction of actions by monitoring visually their execution (Johnson et al., 1996). In the next paragraphs, a description of the general neural properties found within these streams is given.

Recognition of Bodies and Objects: The Ventral "What" Stream

As mentioned above, along the ventral pathway, visual information flows from the primary visual cortex to the temporal lobes. Of particular interest, the temporal lobes includes the inferotemporal area (IT) and the superior temporal sulcus (STS). IT is a pure visual area which is devoted to the recognition of objects and faces (Booth & Rolls, 1998; Tsunoda, Yamane, Nishizaki, & Tanifuji, 2001; Afraz, Kiani1, & Esteky, 2006). By direct projections to the frontal areas partly responsible for reasoning, IT provides useful information about the visual environment for high-level cognitive control. In the posterior part of IT, neurons were shown to be sensitive to one or more specific features of visual stimuli such as size, shape, color, orientation, and direction of motion (Gross, Bender, & Rocha-Miranda, 1969). While moving further toward the most anterior part of this region, IT contains populations of neurons that exhibit an increasing sensitivity to complex combinations of visual features forming objects and faces (Tsunoda et al., 2001; Afraz et al., 2006). These neural responses to visual objects were also found to be encoded in different reference frames. While some neurons fire for a specific set of features of an object (snapshot neurons), others respond invariantly to its orientation, size, position and distance (Booth & Rolls, 1998). These patterns of activity seem to respectively correspond to a viewer-centered and an object-centered frame of reference.

Another important brain area responsible for visual perception is located on the dorsal part of the inferotemporal cortex, namely the superior temporal sulcus (STS). The neural activity patterns within this brain region receiving projections from IT and from the primary visual areas, has been shown to cor-

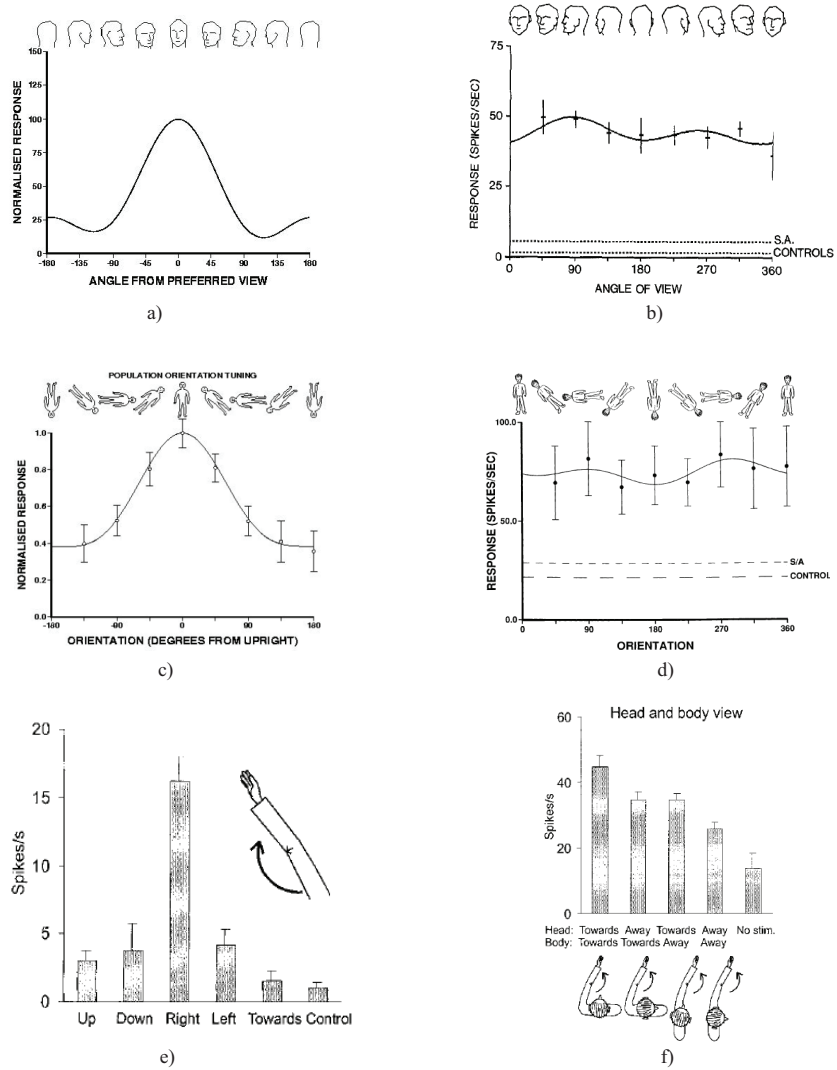


Figure 2.6: Response of STS neurons to several types of stimuli are shown where the data reported on each panel correspond to a different cell. The first row shows neurons sensitive to face having **a)** a preferential tuning to front view, and **b)** an invariant response across the depicted orientations. **c-d)** On the second row, similar activity patterns are displayed by other cells sensitive to bodies. **e)** STS neuron sensitive to the observation of an arm extending to the right of the monkey. **f)** Another neuron sensitive to arm extension toward the animal. The shoulder and head orientations have a significant effect on the firing of that neuron, which shows highest activity when the movement is directed toward the observer. (Adapted from Ashbridge *et al.* (2000); Oram and Perrett (1996)).

relate with the recognition of biological motions and body postures. STS is the primary locus for the perception of actions in visual terms. Importantly, as illustrated in Figure 2.6, there are cells sensitive to the sight of complex body movements such as walking, articulation of body parts and even goal-directed interactions with objects (Perrett, Harries, Mistlin, & Chitty, 1989; Perrett et al., 1990; Oram & Perrett, 1996; Jellema, Maassen, & Perrett, 2004). The sensitivity of these neurons seems to be encoded in various frames of reference, from purely viewer-centered to goal-centered representations. The majority of the STS cells are sensitive to viewpoint, but object-centered neurons, whose activity is invariant with respect to the orientation of the observed body or object, were also reported (See Figure 2.6). Such invariant representation may allow an abstract representation of objects, postures, movements and actions. This explain why STS has been suggested to be the locus of the transformation of visual information related to biological cues through several stages (Perrett, Harries, Mistlin, & Chitty, 1989). Next, when considering the neural responses to goal-directed actions, different groups of neurons have been reported to encode different type of hand actions such as grasping, dropping and tearing. These neurons also differentiate between mimed actions and real ones showing a greater activity in the latter condition (Baker, Keysers, Jellema, Wicker, & Perrett, 2001). Indeed, the activity of some STS neurons has been reported to be context-dependent. For instance, a neuron sensitive to the sight of walking may keep firing even if the walking person goes behind a wall, but will not fire to the sight of the wall alone (Baker et al., 2001). This suggests a mechanism which allows the neurons within STS to complete missing visual information. This finding produced a change in the way STS neurons were considered. Rather than being purely visual neurons, they also exhibit conditional responses in the sense that the same visual stimulus does not always elicit the same response.

Another non purely visual property has also been reported within several populations of STS cells. It concerns their responses to the observation of self generated movements or movements performed by others. While the majority of these neurons fire invariantly according to the ownership of the observed movement, groups of cells within the upper bank of STS (the dorsal anterior part) do not respond to the sight of one's own motions (Hietanen & Perrett, 1993). This finding was first suggested to reflect the sensory abilities of these neurons to recognize the specific visual aspect of the arm of the monkey they belong to. However, a further study using computer generated visual gratings which motion was either induced by an experimenter or the monkey, showed that these cells were activated only when the motion of the visual stimulus was not predictable, i.e., when it was induced by the experimenter (Hietanen & Perrett, 1996). Corroborated by the presence of anatomical connections with parts of the parietal and frontal lobes reported to represent the body schema of the animal (Holmes & Spence, 2004), these findings highly suggest that a mechanism can predict the visual consequences of self-executed movements and

hence inhibit the responses of specific neurons in STS (Keysers & Perrett, 2004).

To summarize this section, the ventral visual stream processes visual information for the recognition of objects, bodies, and body parts. Through the combination of these internal representations, it is capable of determining their spatial relationships. As such, it can also recognize body postures and movements, and even the interactions between people and objects. These highly specialized representations are thus encoding visual information in various reference frames, including purely viewer-centered, goal-centered and invariant frames of reference. Finally, the dorsal anterior part of STS has also been shown to be not a purely visual area. Its responses also reflect the presence of a predictive mechanism which can inhibit visual responses which are attributed to self generated movements. These predictive abilities are supposed to arise from the parietal lobe which roles include the representation of the body in space and the processing of visual information for self-action. This latter aspect is thought to be processed along the second visual stream, namely the dorsal "how" stream.

Recognition of Objects for Action: The Dorsal "How" Stream

The multisensory relay of visual information for the reaching and the manipulation of objects was reported to be located in the posterior parietal cortex (PPC), which is known to be composed of many sub-regions subserving different functions. Its dense and selective reciprocal connectivity with the premotor cortex can explain its involvement in the control and representation of the large palette of possible arm and hand actions (Wise, Broussaoud, Johnson, & Caminiti, 1997; Johnson et al., 1996; Burnod et al., 1999). Moreover, its reciprocal connections with the somatosensory and the visual cortices make the PPC a good candidate for the representation of the body in its environment (Stein, 1989; Holmes & Spence, 2004).

Considering the visual areas for action, the dorsal "how" stream goes through the intraparietal sulcus (IPS) (Andersen, Asanuma, Essick, & Seigel, 1990; Goodale & Milner, 1992). Close and within IPS, several areas are involved in different specific functions. They are the cIPS, located in the caudal part of IPS, MIP, the medial intraparietal sulcus, VIP, the ventral intraparietal sulcus, LIP, the lateral intraparietal sulcus, and AIP, the anterior intraparietal sulcus (Stein, 1989; Johnson et al., 1996; Wise et al., 1997). Except the cIPS, all these areas are multisensory structures combining vision and the information related to the body state. Indeed, their putative functions related to visual monitoring for action imply a tight coupling with the perception of the animal own body (Johnson et al., 1996; Holmes & Spence, 2004).

At the beginning of the IPS, area cIPS receives projection from the early primary visual area (De Yoe & Van Essen, 1988; Goodale & Milner, 1992), and has been shown to process the three dimensional analysis of objects. Neural responses correlate with the various spatial features of objects including their po-

sition, distance, orientation, principal axes and geometric shape (Sakata, Taira, Husunoki, Murata, & Tanaka, 1997; Sakata et al., 1999). cIPS then further sends direct connections to area AIP. Similarly to cIPS, the activity patterns of AIP neurons are also object specific, and more importantly are tuned to the geometrical features of object shape for manipulation, namely the affordances of objects (Sakata et al., 1997; Fagg & Arbib, 1998). This area also exhibits visual responses to interactions between one's own hands and objects (Sakata et al., 1998). Following from these findings and the reported connectivity of this area with motor centers for grasping in the anterior part of the ventral premotor cortex (Wise et al., 1997), AIP has thus been suggested to be an fundamental region for the grasping and the manipulation of objects (Sakata et al., 1997; Fagg & Arbib, 1998).

In parallel to this visual pathway for representing objects for manipulation and grasping in viewer-centered and goal-centered reference frames, areas LIP, MIP and VIP represent also fine combinations of visual and proprioceptive information for action in multiple frames of reference (Andersen et al., 1990; Johnson et al., 1996; Colby & Goldberg, 1999). In a reductionist view, these area were reported to code, respectively, the far visual space, the space at reaching distance, and the ultra-near space (Colby & Goldberg, 1999). In the posterior part of IPS is located area LIP, which represent the spatial location of salient objects (Gottlieb, Kusunoki, & Goldberg, 1998). Concordant with the hypothesis of the presence of a saliency map in the parietal cortex (Itti & Koch, 2001), LIP is suggested to be the locus of attentional mechanisms (Gottlieb et al., 1998), to hold a visual memory of potential targets in retinocentric coordinates (Andersen, Snyder, Bradley, & Xing, 1997; Colby & Goldberg, 1999), and to reflect a decisional process for target selection (Platt & Glimcher, 1999; Shalden & Newsome, 2001). In turn LIP projects to both areas MIP and VIP. Similarly to LIP, MIP also encode the location of visual targets (Eskandar & Assad, 1999). However, while LIP neurons have been reported to encode primarily targets in pure visual terms for gaze control through saccadic eye movements (Snyder, Batista, & Andersen, 1997; Batista, Bruneo, Snyder, & Andersen, 1999; Shalden & Newsome, 2001), MIP neurons are more tuned to the direction of hand movements in visual coordinates which would allow to reach the target (Eskandar & Assad, 1999). Finally, area VIP is also concerned with the control of hand and eye movements in visual terms. However, in contrast to area LIP from which it receives projections, it encodes target in a more head-centered coordinate frame, or at least in a combination of eye- and head-centered frame of reference (Snyder et al., 1997; Batista et al., 1999; Bremmer, Graf, Ben Hamed, & Duhamel, 1999). In addition, complementarily to area MIP, the firing of VIP neurons has also been shown to correlate more with the location of the stimulus location rather than to that of the hand (Eskandar & Assad, 1999).

To finish this review of the neural processing in the dorsal "how" stream, one

important point is that the transformations across the frames of reference occurring in these areas and in their neighboring regions do contribute to the process of transferring purely visual information into multisensory representations which are the most appropriate for the control of motor responses. Through this distributed representation within different frames of reference, the posterior parietal cortex allows the parallel computation of sensorimotor strategies for eyes movements toward target objects, for their reaching and their manipulation such as grasping movements. Nevertheless, although the role of this cortical pathway seems clear, the neural mechanisms mediating its associated computational processes are less obvious. This thesis will attempt to provide some insights about this issue but more details about the research questions addressed in this work will be given at the end of this section.

MULTISENSORY REPRESENTATION OF THE BODY IN THE PARIETAL CORTEX

In the previous section, the neural correlates of the visual perception of objects for action were addressed. Since the knowledge of where one's own body parts are located in space is fundamental to correct plan any execution of movements, the brain must hold an internal representation of the body it belongs to, namely the body schema. A current view in neuroscience suggests that the body schema is an emergent property of a network of interacting cortical and subcortical centers, where each of these areas involves multisensory representations of body parts and their relationships in various reference frames. Each of these used reference frames is not arbitrary, but rather appropriate to the sensory inputs their corresponding representations receive, and to the behavioral responses they do control (Rizzolatti, Luppino, & Matelli, 1998; Holmes & Spence, 2004; Avillac, Denève, Olivier, Pouget, & Duhamel, 2005). The cortical areas suggested to participate in such a body representation are widespread in both parietal, motor and premotor cortices (Stein, 1989; Graziano, Cooke, & Taylor, 2000; Holmes & Spence, 2004). While the description of the properties of the motor related regions will be given later in this chapter, the following paragraphs review the neural correlates of such internal representations of the body and space in the parietal cortex. As mentioned earlier, the emergence of the body schema implies a combination of sensory inputs corresponding to different modalities. Although auditory and olfactory and taste signal are acknowledged to be part of this process, the present discussion will only consider the proprioceptive, tactile and visual modalities, as they are those which are the most involved in imitation.

Proprioception

Proprioception is the sensory input driving the kinesthetic sense of the body. It allows the brain to know where each limb is located without visual feedback. Human sense of kinaesthesia have been shown to be mediated by muscle spin-

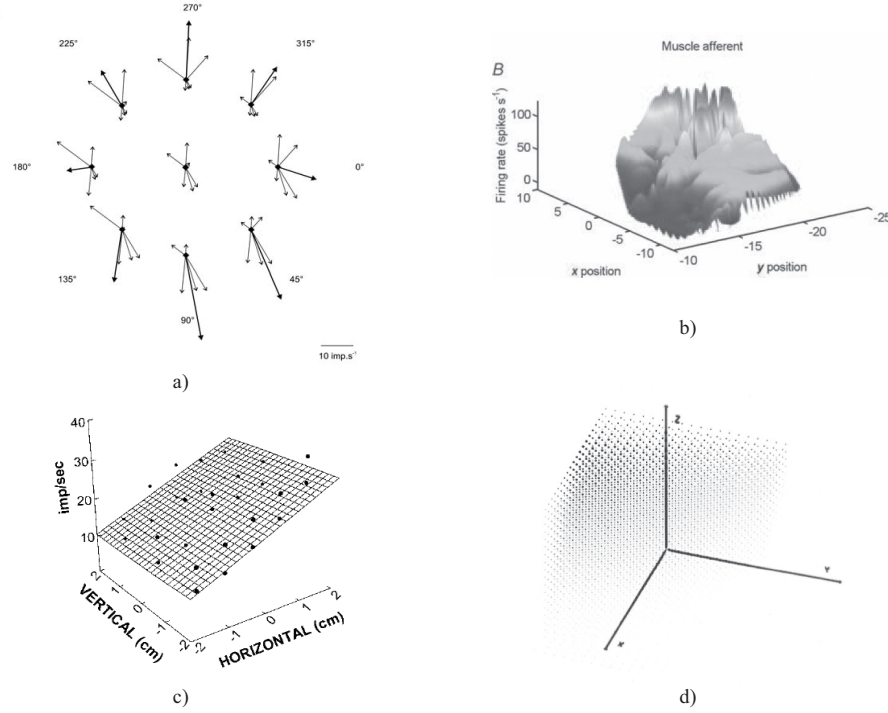


Figure 2.7: Neural basis of proprioceptive representations. **a)** Measured response of afferent nerves of muscles acting around the hip joint of a human subject, while the leg was passively moved in eight locations around a neutral rest position. Each light arrow corresponds to the response of an afferent nerve (arrow amplitude) which corresponds to a unique muscle exerting a force in a specific direction (arrow direction). The dark arrow represent the summation of the contribution of each afferent nerve. It can faithfully represents the actual position of the limb. **b)** Response of a muscle afferent of the cat leg to different passively moved leg positions on the horizontal plane. This afferent shows a sensitivity for a specific preferred position of the leg. **c)** Response of a cerebellar neuron sensitive to the limb position of a rat. A plane can be fitted to the neural response, showing a linear relationship between the cell activity and the limb position. **d)** Response of a neuron in the motor cortex to the arm position of a monkey. The darker is the dot, the stronger is the neural response. A similar linear relationship can be observed. (*Adapted from Kettner et al. (1988); Roll et al. (2000); Casabona et al. (2004).*)

dles, which are sensory receptors sensitive to the position, velocity and forces applied to body joints (Proske, Wise, & Gregory, 2000; Cordo, Flores-Vieira, Verschueren, Inglis, & Gurfinkel, 2002; Ribot-Ciscar, Bergenheim, Albert, & Roll, 2003). For example, as shown in Figures 2.7a and b, during passive movements of the hip, the joint activities of the spindles among the muscles acting around that joint can provide a robust estimator of the position of the leg (Ribot-Ciscar et al., 2003). During active movements, the spindles activity have been shown to be more sensitive to the limbs velocity, hence providing a distributed code of the direction of movements (Proske et al., 2000).

The information conveyed by the muscle spindles is relayed by the spinal cord to the primary somatosensory cortex (SI) (Prud'homme & Kalaska, 1994).

The neurons in this brain region are somatotopically organized, forming a map which follows the well known sensory homunculus. Along the medio-dorso-to-ventral axis of SI, the representation of the legs can first be found, followed by that of the body, arms, hands and finally of the head. The size of the topographic map of a specific body part corresponds to the actual density of sensory receptors located within this body part (Kandel & Jessell, 1991). Considering now the representation of limb location and joint orientation, SI neurons were mainly shown to be sensitive to the movements of the limbs and to be tuned to movement direction and to arm posture (Prud'homme & Kalaska, 1994; Tillery, Soechting, & Ebner, 1996). This indicates that sensations of movement and position are represented in an overlapping and distributed manner in SI. It was also reported that the firing patterns related to arm position varied monotonically with its spatial location, reaching the extremum values at the boundaries of the workspace. Similar properties do also apply to the local orientation of limb joints (Tillery et al., 1996). As shown in Figures 2.7c and d, it is worth noting that similar representations coding for the arm position in space can also be found in motor areas such as the primary motor cortex and the cerebellum (Kettner et al., 1988; Casabona et al., 2004).

Another somatosensory cortical area is located ventrally to SI, from which it receives projections, namely SII. SII neurons exhibits similar firing properties to those in SI, but it plays a more critical role in the tactile processing of objects (Fitzgerald, Lane, Thakur, & Hsiao, 2006). Interestingly, fMRI studies showed that, in addition to its sensitivity to touch, this area might also possess complex visual properties. Indeed, seeing someone else being touched also activates SII, as if one had been directly touched (Keysers et al., 2004). In addition to the direct connectivity between SI-II and the motor control centers in the frontal lobe, these areas also send projections to their neighboring regions in the parietal cortex, namely the parietal areas 5 and 7b/PF (Stein, 1989; Wise et al., 1997). These two regions, respectively located dorsally and ventrally to the intraparietal sulcus (IPS) are suggested to be involved with different but similar functionalities, which are described next, beginning with those of area 5.

Multisensory Integration

The two areas mentioned above, namely area 5 and 7b, are processing centers which integrate sensory signals arising from the primary sensory areas, the intraparietal sulcus and superior temporal sulcus (Wise et al., 1997). The major difference between the neural properties of these areas lies in the fact that area 5 is more prone to the strict representation and monitoring of self body (Stein, 1989; Holmes & Spence, 2004), whereas area 7b/PF is rather suggested to process sensorimotor representations for actions toward objects and responses to social signals (Fogassi et al., 2005). Nevertheless, both areas are still closely related and participate in the cortical network for representing and monitoring the body and its relations with its environment.

In the dorsal part of the parietal cortex, area 5 neurons were shown to fire according to the kinematic features of current posture and movements of the body (Mountcastle, Lynch, Georgopoulos, Sakata, & Acuna, 1975; Stein, 1989; Kalaska, Cohen, Prud'homme, & Hyde, 1990). Moreover, it seems that there is no conclusive findings related to the frame of reference in which these cells encode movements. Neither purely body part-centered responses, nor viewer-centered responses were reported (Holmes & Spence, 2004). Nevertheless, since several neural activity patterns were shown to correlate to two or more limb joints (Mountcastle et al., 1975), this may suggest that there is in fact multiple levels of body representation which are encoded in different reference frames. They may further gradually change from body part-centered to viewer-centered representations along the same path of proprioceptive information, which flows from the primary sensory cortices, to the posterior visual areas such as MIP (Burnod et al., 1999). This ability to represent sensory information in multiple frames of reference may allow this region to take part in the planning of movements, which was reported to correlate with an anticipatory firing of several groups of neurons prior to movement execution (Chapman, Spidalieri, & Lamarre, 1984; Kalaska et al., 1990). Another related cognitive property is that neurons within area 5 shows also predictive sensory responses to visual and movement cues (Mac Kay & Crammond, 1987). This ability to predict the upcoming sensory consequences of either self action or external events suggests that this area is endowed with forward models of body dynamics. This hypothesis seems to be corroborated by the reciprocal connectivity of this region with premotor areas (Johnson et al., 1996; Wise et al., 1997), which may send back motor efference copies of commands actually sent to the muscles. Indeed, area 5 neurons were shown to fire more vigorously to active movements initiated by the animal than to passive ones (Mountcastle et al., 1975).

As mentioned earlier, area 7b/PF, located ventrally to area 5, also participates in the representation of the body, but it is more concerned with the relationships between the body and objects, and also with social signals arising from the observation of actions performed by conspecifics (Fogassi et al., 2005). First, area PF is somatotopically organized, and exhibit an activation to multisensory forms of inputs, such as tactile, somatic and visual stimulation (Andersen et al., 1990; Dong, Chudler, Sugiyama, Roberts, & Hayashi, 1994; Fogassi & Gallese, 2002). Visual neurons are can be categorized into separate groups, each of which responds to different forms of visual stimuli. First, since area 7b receives projection from AIP which encodes the affordances of objects (Sakata et al., 1998), a subset of 7b neurons are indeed activated by the sight of a manipulable objects (Sakata et al., 1997). Other projections from the ventral visual pathway arise from STS, which encodes the visual perception of actions performed either by the self or by others (Perrett et al., 1990). Consequently, this explains the visual properties of another group of neurons within area PF which mimic those found in STS (Fogassi et al., 2005). They also fire during

the observation of someone else performing a movement. A second general feature of PF neurons is that they have been shown to display motor sensitivities corresponding to the putative role of this area (Hyvarinen, 1982). The group of motor-related neurons are sensitive to the execution of complex movements such as reaching or grasping movements, which are motor acts having a tight relationship with objects (Fogassi & Gallese, 2002). Importantly, this neural code is independent on precise movement kinematics or dynamics. These neurons seem thus to encode a high-level representation of actions. In addition, some of these neurons also reflect an preferential activation of relevant sequences of movements such as grasping-to-eat or grasping-to-place (Yokochi, Tanaka, Kumashiro, & Iriki, 2003; Fogassi et al., 2005).

The most interesting property of some of 7b neurons is that they do integrate the visual and motor properties of previously related neural populations. Importantly, a visuomotor neuron activated by the sight a specific type of object will also be activated by the execution of the grasping movement associated with that object, or at least some parts of the movement (Fogassi & Gallese, 2002). This congruency between visual and motor modalities reveals the role of this area in the coding of the relationship between the body and the objects in its surroundings. A final key experimental finding is the reported presence of the so-called mirror neurons (Rizzolatti et al., 2001; Fogassi & Gallese, 2002). Mirror neurons were defined as visuomotor neurons which respond both to the sight of a given action performed by another individual as well as by self motor execution (Rizzolatti, Fadiga, et al., 1996; Rizzolatti et al., 2001). Importantly, this may suggest that this neural substrate partly allows the animal to understand the goal of observed actions performed by others (Gallese & Goldman, 1998; Rizzolatti et al., 2001). Indeed, because the monkey may know the outcome of the motor act triggered by a given mirror neuron, when this neuron is activated by the observation of another monkey or human performing that same action, this would mean that the monkey recognizes the movement and consequently its goal. Moreover, since these neurons can start firing before the beginning of the movements, the ability for the recognition of intentional actions through this mirror activity is strengthened.

A last area of the posterior parietal cortex involved in the representation of the body schema has been omitted. Area 7a/PG, located posterior to area 7b/PF, shares some similarities with adjacent cortical regions. For instance, this area has also been suggested to contain, like LIP, a spatial map of salient visual stimuli which can remain active even if a stimulus disappears (Constantinidis & Steinmetz, 1996, 2001). In addition, area 7a was reported to monitor the spatial relationships and the coordination between body parts, and to exhibit predictive responses to upcoming sensory consequences (Stein, 1989; Rushworth, Nixon, & Passingham, 1997). The last interesting property of this specific area concerns the frame of reference in which information is coded. Indeed, this region has been reported to encode spatial information in world-centered coordinates,

which would allow the animal to situate its body in the environment (Snyder, Grieve, Brotchie, & Andersen, 1998). Through the direct connectivity of area 7a with the hippocampus (Clower, West, Lynch, & Strick, 2001), an important subcortical center encoding spatial maps of the environment for self localization (Redish, 1999), an extrinsic representation the self in the world can be built. This finding is of particular interest in that a global view of the self can be produced, which is directly linked with the local self-representations in the posterior parietal cortex.

To summarize this section, the posterior parietal cortex is a multisensory area which integrates visual and proprioceptive information in order to provide the brain with a coherent representation of the body within its environment. In addition to this purely representational property, PPC is also involved in the cortical network for action generation and planning. The multiple levels of representation across multiple frames of reference allow this region to merge different sensory modalities which are not initially encoded within the same coordinate frame, and to control body movements through different strategies. For instance, goal-directed movements requires the visual information of an object to be transferred in to a goal-centered representation. Then it has to be transformed into a body-part centered representation for the direct control of the limb joints. Finally, in the ventral part of the parietal cortex, area 7b/PF, which was reported to encode goal-directed actions, has been shown to exhibit mirror properties. As a consequence, when an action performed by another individual is perceived, a motor resonance mechanism may be initiated and propagated through the remaining of the parieto-premotor network for action understanding and imitation purposes. Indeed, parietal and premotor are suggested to form a core system for the recognition and reproduction of actions. The fundamental properties of the motor and premotor regions are discussed in the next section. As will be described, similar mirror properties were also observed in the premotor areas reciprocally connected to parietal area 7b/PF.

THE CONTROL OF MOVEMENTS: MOTOR AND PREMOTOR CORTICES

Despite being studied for a long time, the control of movements is still a important issue in neuroscience. The motor brain is composed of several cortical areas and of some regions of the lower brain. Its cortical part is reciprocally connected with the parietal cortex. In the view of the behavioral role of the motor regions in the control of movements, the motor cortex is fundamentally involved in the control of the activation of the muscles, either directly or through coordination patterns of muscle activation. It also involved in the control of higher-level and complex movements such as goal-directed actions and the sequencing of motor acts. In the cerebral cortex, the different motor regions are somatotopically and hierarchically organized (Rizzolatti, Camarda, et al.,

1996; Schieber, 2001). Each of the subregions were shown to subserve different functions, which are reviewed below.

The Direct Control of Movements

The last cortical relay in the motor processing chain is the primary motor cortex (M1 in humans and F1 in monkey). The majority of the neurons within this area sends motor commands through direct corticospinal projections to motoneuron pools that in turn activate the muscles (Georgopoulos, 1996). Results of neurophysiological experiments indicate that F1 neurons display a sensitivity to several, if not all, movements parameters such as position velocity, acceleration and force (Georgopoulos, Kettner, & Schwartz, 1988; Kettner et al., 1988; Schwartz, Kettner, & Georgopoulos, 1988; Aflalo & Graziano, 2007). By considering the neurons mostly sensitive to one parameter such as the velocity, a preferred direction of movements can usually be associated to each of them, in which the neuron fires maximally. An illustration of a typical firing pattern and tuning curve of a F1 neuron can be found in Figures 2.8 a and b. In order to measure the macroscopic effect of the joint activities of the neurons participating in this distributed representation of movements, Georgopoulos, Kalaska, Caminiti, and Massey (1982) initially proposed the **population vector**, which consists of a weighted summation of the firing activity of each neuron with its preferred direction. Later, a slightly different method using optimal linear estimators was proposed in order to account for the non-uniform distribution of preferred directions which has been reported in several populations of neurons (Salinas & Abbott, 1994; Deneve et al., 1999). Figures 2.8 c-e illustrate a population vector representation of the activity of a group of M1 neurons during the performance of a straight movement and during lemniscate tracing.

It has been initially observed that M1 neurons were coding movements of the end-effector in a body centered-reference frame. However, this appealing and simple principle has been challenged by complimentary studies indicating that the activity of M1 neurons also depends on limb configuration (Kettner et al., 1988; Scott & Kalaska, 1997; Kakei, Hoffman, & Strick, 1999; Aflalo & Graziano, 2006a). The preferential activities of neurons within M1 seems in fact more distributed over several frames of references which are still close to the strict joint-centered reference frame compatible with the muscle activation patterns (Kakei et al., 1999). Figure 2.8f reports the results of a cell recording study showing that different F1 neurons may encode movement in different frame reference (Kakei et al., 1999).

More recent insights on the functional organization of the motor cortex were given by Graziano, Taylor, and Moore (2002). They showed that transcranial micro-stimulations of the premotor (PM) and primary motor cortices of the monkey produce the execution of behaviorally relevant movements. Depending on which site is stimulated, the movements are performed toward various spatial locations, suggesting a reciprocal mapping between the cortical surface overlap-

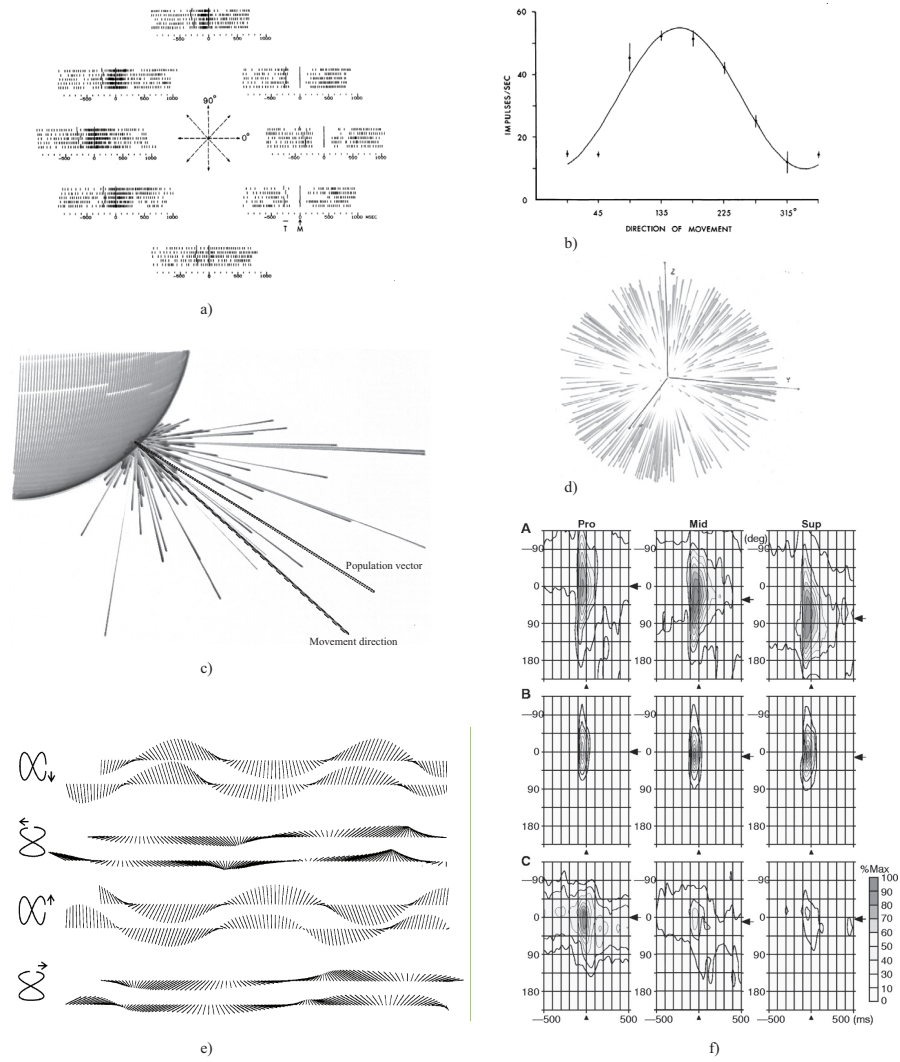


Figure 2.8: The neural basis of population vector coding in the motor cortex. **a)** Spike trains of a motor neuron recorded while a monkey was reaching from a central position to eight peripheral locations on the horizontal plane. The spike trains are displayed relative to the direction of movement. **b)** Mean firing activity of the same cell during movement time sorted according to the direction of movement. The cell exhibit a preferential broad tuning for a specific movement direction: its preferred direction. **d)** Nearly uniform distribution of preferred directions of motor neurons in three dimensional space. **c)** 3D representation of motor neurons tuned to three dimensional movements when the monkey moved its arm in a given direction. The firing activity of the recorded neurons during the movement are shown as a line with a direction corresponding to the preferred direction of the neuron and with length to its mean activity. The population vector computed from this distributed representation of movement is also displayed. **e)** The evolution of the population vector during lemniscate tracing exhibit a high correlation with the real movement direction. Only a temporal shift can be noticed. **f)** Multiple frame of reference of movement coding of the wrist in the primary motor cortex were found. Each row corresponds to the evolution of the firing activity of a neuron during the performance of a wrist movements in all allowed direction while the forearm was maintained in three different orientation. Each row corresponds to, (A) a cell tuned to muscle centered coordinates, (B) to extrinsic movement direction, and (C) to extrinsic movement direction but modulated by forearm orientation. (Adapted from Georgopoulos et al. (1982, 1988); Schwartz and Moran (1999); Kakei and Hoffman (2001)).

ping PM and M1, and the target position of movements (Graziano et al., 2002; Aflalo & Graziano, 2006b). This finding suggests a distributed representation of movements across the motor regions forming a map of the surrounding and reachable space. This discovery may conflict with the observed activity patterns of motor cortical areas, which exhibit only a small sensitivity to the position of the end-effector. However, it must be noted that TMS is non-physiological, i.e., it does not occur during normal behavior. It may nevertheless uncover some neural processes which can not be recorded in term of neural activation, such as sub-threshold activity.

Visuomotor Control of Proximal and Distal Movements

Compared to M1, the premotor cortex (PM) has been suggested to be involved in higher motor function. The ventral region of PM coding for the arm is separated into two subregions, namely, F4 (PMvc) and F5 (PMvr), which respective activity correlates with the proximal and distal control of the arm (Rizzolatti, Camarda, et al., 1996; Gentilucci et al., 1988; Rizzolatti, Camarda, Fogassi, Luppino, & Matelli, 1988). F4 is located in the caudal part of the ventral premotor cortex (PMv). It is somatotopically organized and its neurons respond mostly to the direction of movement of the arm, but also to other movement parameters as mentioned above while considering F1 (Gentilucci et al., 1988; Aflalo & Graziano, 2007). Another interesting property of these neurons is that they also exhibit visual response in a body part-centered manner (Fogassi et al., 1996; Graziano, Hu, & Gross, 1997). The ability of F4 to respond to visual cues is supposed to originate from posterior parietal area VIP processing the location of visual objects (Johnson et al., 1996). When a visual stimulus get closer to a specific part of the arm, some neurons may start firing in response to this approaching stimulus. Importantly, these neurons correspond to those which would be activated by a movement of this particular body part. Finally, as mentioned above, since these neurons respond to visual cues in a body part-centered frame of reference they thus fire to a visual stimulus irrespectively to the arm posture, as long as the object is located in a similar fashion relative to the arm (Graziano et al., 1997). This reflect a complex visuomotor mapping within this premotor region, which is suggested to be the basis of the representation of the peripersonal space (Holmes & Spence, 2004).

In contrast to F4, F5 encodes the distal control of the arms and hands (Rizzolatti et al., 1988). F5 neurons discharge mostly for specific goal-related movements, such as grasping and tearing objects. Both hemispheres seem to control either the left or the right arm. The precise shaping of the hand during such movements involving object manipulation (precision grip, whole hand prehension,...), seems to be controlled by different populations of premotor neurons (Rizzolatti, Fadiga, et al., 1996; Fogassi & Gallese, 2002). Through its interconnections with area AIP in the parietal cortex, some neurons within F5 exhibit visual responses to the sight of objects (Murata et al., 1997). Similarly

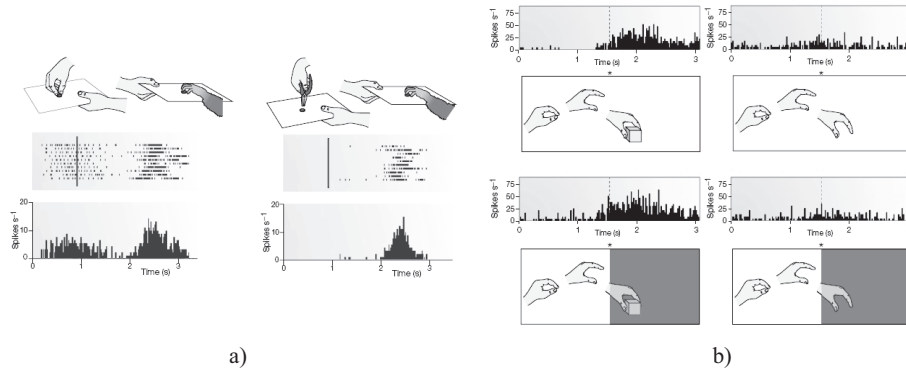


Figure 2.9: a-left) Response of a mirror neuron to the sight of an experimenter grasping a piece of food on a tray and while the monkey do the same action. **a-right)** This neuron remains silent when the experimenter uses a tool. **b-top)** Mirror neurons do not respond to mimed actions. **b-bottom)** These neurons also exhibit a context-dependent activity. When the end of the movement is hidden, the neuron keeps firing if an object has been shown to be placed behind the occluder. It remains silent on the sight of the same visual information, if there is no object-goal. (*Adapted from Rizzolatti et al. (2001)*).

to F4, the neurons sensitive to a particular object are also activated by the execution of the precise movement which would manipulate it. This property may reflect an ability of F5 to map object shapes in motor terms, and to encode sets of potential actions to be executed (Rizzolatti et al., 2001).

Interestingly, F5 also exhibits additional visual properties that are suggested to be the result of reciprocal indirect connections with the superior temporal sulcus relayed by parietal area 7b/PF (Johnson et al., 1996; Wise et al., 1997). For recall, some STS neurons encode a visual representation of hand actions performed by either the animal or another individual. The presence of canonical and mirror neurons in area F5 is thus suggested to be the consequence of this relationship with STS (Rizzolatti, Fadiga, et al., 1996; Rizzolatti et al., 2001; Fogassi & Gallese, 2002). As already mentioned when considering area 7b of the parietal cortex, mirror neurons are neurons sensitive to both the observation and execution of the same motor act. Figure 2.9a shows an illustration of a typical activity pattern of a mirror neuron. In contrast, canonical neurons reflect more causal relationships between an executed action that could be performed in response to an observed one. Indeed, such neurons were found to fire both for the observation of a human experimenter placing an object on a tray and the subsequent execution of a grasp by the monkey (Rizzolatti, Fadiga, et al., 1996; Fogassi & Gallese, 2002). Importantly, these neurons are motor-related. Their discharge has been shown to be not due to self-vision only (Gallese et al., 1996). The visual responses of mirror neurons have also been reported to be goal-centered, in that the intensity of their firing is independent of the point of view of the animal, i.e., their visual responses are invariant to the distance, position and orientation of the observed act (Gallese et al., 1996).

The discovery of such mirror neurons has led to several exciting hypothesis concerning the cognitive process of action understanding (Gallese & Goldman, 1998; Rizzolatti et al., 2001; Fogassi et al., 2005). Similarly to their homologue neurons in parietal area 7b/PF, F5 mirror neurons are suggested to be part of a complex network responsible for the understanding of intentions of others. Since their firing has been shown to be capable to predict the outcome of a given action before its completion, and even if parts of the movement or goals are occluded (Umiltà et al., 2001) (See Figure 2.9b), the F5 mirror neurons were proposed to be part of a forward model mechanism which could predict the consequences of a observed motor act. By motor resonance, i.e., by covertly activating the same cortical areas as if the animal was performing itself the same action, this predictive mechanism may allow monkey, and by extension humans, to understand the purpose of others actions (Gallese & Goldman, 1998; Rizzolatti, Craighero, & Fadiga, 2002; Miall, 2003).

Conditional Control of Movements and Overt Execution

The conditional control of movements refers to the brain ability to select the most appropriate motor response depending on the context. For instance, if a specific movement is required in response to an abstract visual stimulus such as a colored cue, a stimulus-response mapping has to be performed in order to select the correct response. Neural correlates of such a decisional process within the premotor cortex have been reported by several cell recording studies (Crammond & Kalaska, 1994; Cisek & Kalaska, 2005). They showed that dorsal premotor areas encode, prior to task execution, the potential response to visual cues. Then, upon movement execution, their activity patterns change and finally reflect only the actual movement performed by the monkey. These related regions on the dorsal part of the premotor cortex are areas F2 and F7. While area F7 is mostly reciprocally connected with area LIP of the parietal cortex, area F2 has rather been shown to be linked with area MIP (Johnson et al., 1996). Similarly to the parietal areas from which they receive projections, F2 and F7 are, respectively, more selectively responsible for the conditional control of arm movements and of eye saccades, and do represent information in almost the same coordinate frames, i.e., a mostly eye-centered reference frame (Geyer, Matelli, Lupino, & Zilles, 2000). In addition, these areas are also connected to the prefrontal cortex, which is known to process context-dependent information for action in a relatively high-level fashion, i.e., independently of the required motor response (Sakagami et al., 2006). Therefore, areas F2 and F7 may be seen as the interface between such high-level task abstraction with actual motor responses.

Further, since all the premotor areas described until here appear to encode potential motor responses to objects and stimuli, this suggests that a control system in charge of their transformation into actual movement must exist. This cognitive process has been proposed to be localized in the supplementary motor

areas (SMA) which can be split into two subareas, namely F3 (proper-SMA) and F6 (pre-SMA) (Luppino & Rizzolatti, 2000; Rizzolatti & Luppino, 2001). They are both interconnected and have direct projections to most of the other premotor and motor areas. Indeed, this functional role of F3 and F6 is corroborated by an important fact. Projections from the SMA to the primary motor and ventral premotor areas seem indeed to have a modulatory influence on the control of movement initiation, since F6 and F3 neurons discharge well in advance of movement initiation, which contrast with the data obtained, for instance, in F1 (Alexander & Crutcher, 1990; Rizzolatti & Luppino, 2001).

To summarize this section addressing the cortical control of movements, the important point is that the control of movements, similarly to the representation of one's body, involves hierarchically distributed motor representations. The dorso-medial part of the premotor cortex (areas F3 and F6) computes the control of which motor acts should be initiated. Indeed, through reciprocal connections with the parietal cortex processing visual information for action, other motor regions (areas F2 and F7) have been shown to provide potential motor responses according to the current task and to the environmental cues. The control of arm movement per se has then been suggested to involve the ventral part of the premotor cortex, areas F4 and F5, which respectively control the proximal and distal aspects of arm and hand movements. Interestingly, these areas also show visuomotor responses to visual cues, which may simplify the computations in higher motor control regions. Indeed, F4 and F5 are suggested to provide sets of motor primitives, where each of them is associated with the objects relevant to that given motor act. Further, each of these premotor areas project more or less directly to the primary motor cortex, which has been shown to be the principal source of corticospinal projections to the spinal cord for the direct control of the muscles. Finally, similarly to the neurons in area 7b/PF of the parietal cortex with which they are interconnected, some F5 neurons also exhibit mirror properties. This finding suggest that observing the actions of others, in addition to activating purely visual recognition areas, also trigger parts of the motor control areas. This comfort the hypothesis that a motor resonance mechanism is involved in the perception of the actions of others by activating the same regions as if one was actually performing the same actions.

COGNITIVE CONTROL OF NEURAL PROCESSING PATHWAYS

As mentioned earlier, the parieto-premotor cortices are reciprocally connected to the prefrontal cortex (PFC), to which the control of high-level cognitive mechanisms such as reasoning and decision making has mainly been assigned (Miller, 2000). This cortical area is a very large region which can be separated into several subdivisions. This review will however not address all of them but only those which are suggested to have an implication in the con-

trol of movements, and by extension to imitation. The areas known to have a modulatory control over motor strategies are basically the ventrolateral and the dorsolateral prefrontal cortices, VLPFC and DLPFC, respectively (Decety & Sommerville, 2003; Brass et al., 2005; Sakagami et al., 2006). The primary input to VLPFC is area IT of the temporal lobe, whereas DLPFC is reciprocally connected to the parietal and premotor cortices (Petrides & Pandya, 2002). Sakagami et al. (2006) suggested that these two regions, i.e., VLPFC and DLPFC, form respectively an extension to the ventral "what" and dorsal "how" visual streams. They proposed that the extended dorsal pathway makes stereotyped decisions about the control of action, whereas the ventral pathway is more responsible for deliberate decisions through an inhibitory control over the automatic decisions taken dorsally (Sakagami et al., 2006).

Neural responses in these areas were shown to primarily reflect go or no-go signals during the performance of various stimulus-response tasks (Watanabe, 1986; Sakagami et al., 2001). These neurons are temporally activated few milliseconds after stimulus onset, but not at the time of motor execution. This suggests that the firing of these neurons correlates more with the behavioral meaning of the observed stimulus rather than to a motor response per se (Lauwereyns et al., 2000; Sakagami et al., 2006). Interestingly, the detailed analysis of the firing patterns of VLPFC neurons suggests that they indicate what the animal should not do, rather than what it should do (Sakagami et al., 2001). This finding strengthened the hypothesis of the inhibitory mechanism by which the prefrontal cortex exerts control on the parieto-frontal network controlling the execution of actions.

Lesions studies in human prefrontal cortex have shown to result in several deficiency in the ability to inhibit stereotyped responses and to make decisions (Lhermite, Pillon, & Serdaru, 1986; Goldman-Rakic, 1996; Brass et al., 2005). For instance, following prefrontal lesions, patients have been shown to exhibit degenerative imitative behaviors such as apraxia or echopraxia (Lhermite et al., 1986; Shimomura & Mori, 1998). The suggested inhibitory control mechanism of PFC over all the cortex may raise the question as to whether different areas might control different motor strategies. Recently, a fMRI study investigated this question by comparing brain activity in the inhibition of non-imitative and imitative response tendencies (Brass et al., 2005). It was shown that response inhibition in the two tasks may involve different neural networks. While the inhibition of non-imitative responses seems to require a fronto-parietal network involved in interference control and task management, the inhibition of imitative responses involves cortical areas that are required to distinguish between self-generated and externally triggered motor representations (Decety & Sommerville, 2003; Brass et al., 2005). The prefrontal cortex is thus of critical importance when considering the neural mechanism controlling imitation as well as all the previously mentioned premotor, parietal and temporal areas.

Summary

In this section the functions and major properties of several cortical areas and brain pathways were reviewed. In particular, the mechanisms of sensory perception, multisensory integration, and motor and cognitive control were discussed. Several interesting common neural properties and principles were examined, particularly the brain's ability to hold information on several representational levels or frames of references. Although the range of these reference frames is relatively broad - it can be eye-centered, body part-centered, goal-centered or even world-centered - the nervous system is capable of coherently maintaining all of its information. This coherency is produced with the help of a cortical mechanism which endows neural structures with predictive abilities. In addition to the fact that this ability can help the brain to complete missing information, it can also be used to keep internal representations up to date without needing to wait for slow sensory feedbacks.

This thesis investigates issues related to how the nervous system transfers information from one frame of reference to another, produces predictive activation patterns, and how these action patterns can be useful for multisensory integration. These abilities play an important role in imitation. In order to be capable of imitating someone executing movements, one should be capable of transferring information perceived in eye-centered coordinates into a body-centered representation. In order to avoid confusing one's own visual feedback with those of others, one should be capable of predicting the outcome of one's own actions and, consequently, of inhibiting sensory inputs that are too discrepant.

In line with the research questions raised in Section 2.2, which considered imitation as described by experimental psychology, this thesis will also discuss the neural mechanisms and brain pathways at the cortical level that mediate imitation. Although until now each functional property of each cortical area has been considered separately, understanding the mechanisms involved in the control of movements requires knowledge of the interactions between each area: the neural mechanisms underlying imitation are not exempt from the need for a global view of the brain's networks for perception and action. The next section provides a synthesis of what is known about imitative processing pathways. Experimental data gathered in monkeys, followed by similar data obtained from human subjects, will be examined.

2.3.2 THE MIRROR NEURON SYSTEM IN MONKEYS

As pointed out by the present review, a network of brain regions showing visuomotor properties has been identified by numerous experimental studies in the monkey, namely the mirror neuron system. In monkeys, mirror neurons were found in area F5 of the premotor cortex and in area PF of the parietal cortex (Rizzolatti, Fadiga, et al., 1996; Fogassi & Gallese, 2002). Mirror neurons, partly defined as exhibiting mandatorily visual and motor responses, are activated both when the animal performs or sees a goal-directed action² (Rizzolatti et al., 2001). Although STS does not contain neurons firing for both action execution and observation, it may still be considered as a mirror area since it also shows visuomotor properties, but in an inverted fashion. First, STS appears to be the locus of the recognition of individual performing actions and is hence highly suggested to be the main visual input of the mirror system (Perrett, Harries, Mistlin, & Chitty, 1989; Fogassi & Gallese, 2002). Second, considering its motor properties, a group of neurons in this area has been reported to be insensitive to self initiated actions, which was further shown to be dependent on the visual stimuli (Hietanen & Perrett, 1993, 1996). Thus STS has been suggested to receive inhibitory back-projections from area PF, causing it to have this inverted mirror patterns of activity (Hietanen & Perrett, 1996).

In this section, a synthesis of the principal neural information flows which may explain why mirror neurons in the reported brain regions do display their intriguing visuomotor properties is provided. Two important processes are described: the execution of goal-directed actions and the observation of similar action performed by other individuals. Indeed, since these two behaviors are those which were primarily used to identify the mirror system that mainly consists of the overlapping areas required during both types of tasks (Rizzolatti et al., 2001).

Execution of Goal-Directed Actions: Visuomotor Transformations for Reaching

Figure 2.10 illustrates the principal identified and supposed neural pathways responsible for the execution of goal-directed actions such as grasping a piece of food. The graphical representation of the involved brain areas is grounded on the division of the monkey brain as shown in Figure 2.5.

First, visual information of an object to grasp is perceived through both eyes and processed by the early visual cortical areas. Then, they project sequentially to areas cIPS and AIP, which compute the visual 3D features of the object and extract its affordances (Sakata et al., 1997). Next, area PF integrate this information in order to develop a goal-centered sensory representation of

²In addition, it should be noted that these cells may also display a sensitivity to other sensory modalities. For instance Kohler et al. (2002) showed mirror neurons in F5 which are also triggered by auditory stimuli such as the sound of breaking a nut.

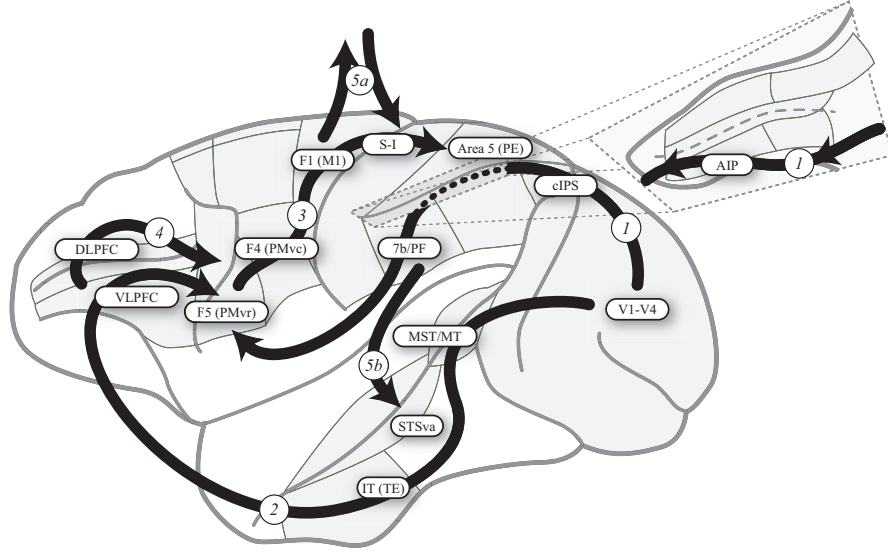


Figure 2.10: Principal cortical information pathways during the execution of goal-directed action: **1)** The presence of an object in the visual field is analysed by early visual areas which then project to area cIPS and AIP of the extractions of the object affordances. Receiving inputs from AIP, area 7b computes the multisensory representation of the object with respect to the hand and further sends this information to premotor area F5 which compute the potential motor commands to manipulate the object. **2)** In parallel, the ventral visual pathway extracts the identity of the object which is further used by the VLPFC to select the correct action to execute among those encoded in F5. **3)** Motor preparatory activity is then dispatched into premotor and primary motor cortices responsible for arm and hand movements. This requires communication with the posterior parietal areas in order to monitor the location of one's own limb. **4)** The go signal is then sent to the motor regions for execution. **5a)** Motor commands are sent to the muscles. **5b)** Simultaneously, the parietal cortex monitors the actual movement execution and sends inhibitory signals to STS in order to suppress predicted self-related visual information.

the object in relation with the effectors of the animal (Fogassi et al., 2005). Connected to area PF, area F5 then processes this multisensory representation of the object and activates the potential high-level, goal-centered motor commands to execute the possible grasps of this specific objects as indicated by area PF (Rizzolatti, Fadiga, et al., 1996). In parallel, through the temporal cortex, area IT extracts the identity of the object for the VLPFC to select which motor command should be actually executed (Sakagami et al., 2006). Next, the dorsal part of the parieto-frontal network is activated according to potential actions to be executed encoded within area F5. It primarily computes the joint-centered movements of the limbs to be performed through a series of frames of reference transformations from the goal-centered representation in F5. No overt movements are yet performed. The go signal is then sent to the motor control centers and the selected action is performed. For an accurate execution, the parietal cortex simultaneously monitors the proprioceptive feedback which should corresponds to the internal model of the movements, and hence maintains a coherent

representation of the body state (Johnson et al., 1996). Finally, re-afferent signals are sent back to STS in order for its neurons to remain silent to the visual perception of self-generated movements (Hietanen & Perrett, 1993).

Observation of Goal-Directed Actions: The Mirror Neuron System

As shown in Figure 2.11, it has been proposed that the cortical processes of action observation significantly overlap with those concerned with action execution. First of all, during action observation, the same processes which compute the possible actions toward the object located in the environment is also involved, i.e., the dorsal and ventral visual streams. For recall, the former passes through cIPS, AIP and PF and ends in premotor area F5, whereas the latter flows through area IT and VLPFC which in turn projects to F5 to select the most appropriate action which might be executed. However, simultaneously, the presence of another individual manipulating an object triggers STS which is assumed to identify the action actually performed by the other agent (Perrett et al., 1990; Jellema et al., 2004). STS then further projects its goal-centered representation of the observed action to the parietal area PF. By visuomotor resonance, PF may attribute to this visual description of the action the sensory responses which would be perceived during self-execution of that same action (Perrett et al., 1990). This mentioned motor resonance mechanism is not the outcome of PF only. Interconnected with PF, F5, which is more responsible of the processing of motor commands within a goal-centered or distal motor representation, may hence display a mirror activity of a specific group of neurons.

From this description of the brain pathways responsible for action execution and observation, one may have noticed that they have several brain areas and pathways in common. Apart from the groups of areas responsible for the analysis of visual information for object manipulation and for motor execution, three areas remains in between. These are those suggested to be part of the mirror neuron network, namely, F5, PF and STS. Indeed, although they primarily subserve different functions, all of these brain regions receive, and thus encode visual and motor description of actions in a goal-centered frame of reference. Along this line, a gradual sensitivity to different modalities may be considered. STS is primarily visual, PF proprioceptive, and F5 motor. The main property, which may be at the origin of the tight link these three areas, is that they do all represent actions in a goal-centered frame of reference. Interestingly, no direct neural link between STS and F5 have been reported so far, which may seem intriguing at the first glance. Indeed, why visual representation of a motor act can not be transferred directly to a motor one? Why should the mirror information flow be relayed by area PF, as indicated by experimental studies? An hypothesis has nevertheless been raised. Area PF may be a mandatory step for visual information to be associated with motor commands. Basically, while F5 encodes the causes of a action, i.e., the motor commands, STS rather encodes

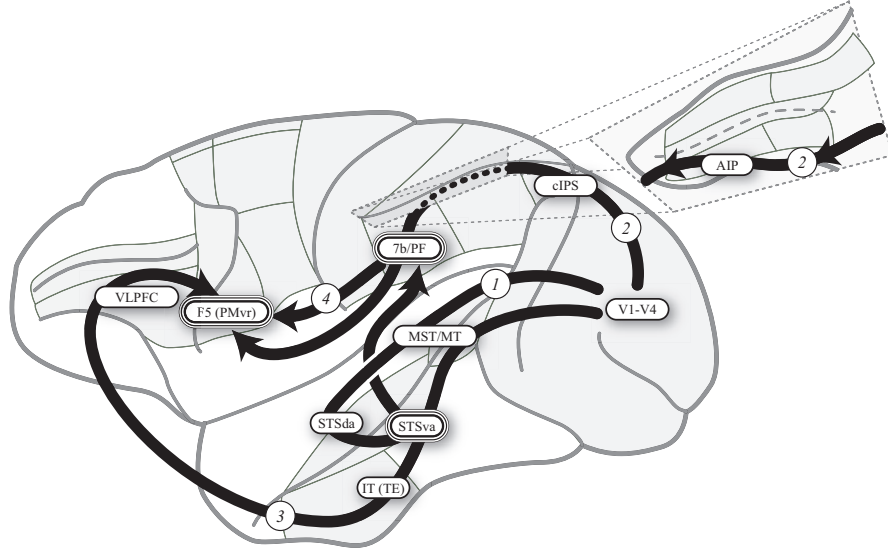


Figure 2.11: Principal cortical information pathways during the observation of action performed by others: **1)** The presence of an individual in the environment triggers the temporal centers for the recognition of biological motions. During the observation of a goal-directed action performed by that individual, STS neurons identify that action and send this information to area 7b. This area now contains a representation of that action in terms of the sensory responses which would be perceived during self-execution. **2,3)** However, simultaneously, since an object is present in the visual field, both dorsal and ventral pathways for object recognition are activated. As a result the observer is ready for acting upon the object by the activation of area F5. **4)** The projection of the information processed by STS is then relayed to area F5 through area 7b, which produces the reported mirror activity. The highlighted areas are the mirror areas 7b and F5 which are sensitive to both action execution and action observation. Moreover, even if STS does not display the same activity pattern, it can also be considered as mirror since its activity is modulated in both conditions. Indeed, during action observation, it reacts to movements performed by others, whereas during action execution, its visual response is inhibited by the consequences of self action.

its consequences, i.e., the visual feedback. Then, since PF, and more generally the parietal cortex, is suggested to hold an internal model of the body which consists of storing a multisensory representation of the body and of maintaining it updated through motor efference copies, PF may be located in between, i.e., PF may represent both the causes and the consequences of an action. This may therefore explain its mandatory role in the transfer of information from the primarily visual area to the motor region, and vice-versa.

The reciprocal connectivity across STS-PF-F5, which is responsible for the mirror response of certain groups of neurons, thus allows a mapping between the visual perception of movements and the motor representations in charge of their execution. Because this network of brain regions could help understand the actions of others through a mechanism of visuomotor resonance, it has been suggested to be at the core of the cognitive processes by which monkeys are capable of behaving socially. However, as highlighted in this review, the monkey

mirror system is grounded on goal-centered representations only. This may thus explain why imitative abilities of monkeys only focus on actions involving objects, which is clearly different as to what humans can do. Therefore, one may ask for the cortical differences between these species. Cortical similarities as well as possible differences are presented in the next section.

2.3.3 THE MIRROR SYSTEM IN HUMANS

In human, studies of brain lesions resulting in degenerate imitative behaviors (apraxia or echopraxia) were the first to give some insight into the brain areas responsible for imitation, pointing to generic areas in the frontal and parietal cortices (Lhermite et al., 1986; Shimomura & Mori, 1998). More recently, despite important experimental constraints, evidence of a similar mirror system has been gathered by a series of brain imaging studies. (Decety et al., 1997; Iacoboni et al., 1999, 2001; Decety et al., 2002; Rizzolatti et al., 2002; Mühlau et al., 2005). The results of these experiments suggest that the human mirror system primarily involves Broca area (in the ventral premotor cortex (PMv)), parts of the posterior parietal cortex (PPC) and of the superior temporal cortex, which are supposed to be the human homologue regions of macaque areas F5, 7b/PF and STS (Rizzolatti & Arbib, 1998; Arbib & Bota, 2003). These mirror areas were primarily found during fMRI or PET scans, where subjects were alternatively asked to observe movements and then to imitate them (Iacoboni et al., 1999; Decety et al., 2002; Koski et al., 2003). In addition to their specific and somatotopic activation patterns during both action observation and execution (Buccino et al., 2001), these areas appear to be more strongly activated when there is an intention to imitate (Iacoboni et al., 1999; Chaminade & Decety, 2002). Together with the reported stronger activation of human STS, PMv and PPC during the observation of well practiced movements as compared to unfamiliar movements (Calvo-Merino, Glaser, Grèzes, Passingham, & Haggard, 2005), these findings highly suggest a motor resonance mechanism similar to that found in monkeys. Human may thus also understand others by means of their own motor vocabulary³ (Blakemore & Decety, 2001).

Despite this functional similarity with the macaque brain, an important issue about mirror neurons and their suggested implications in the mechanisms of imitation concerns the inability of monkeys to imitate tasks in which targets or goals are not involved⁴ (Myowa-Yamakoshi & Matsuzawa, 1999). In contrast, humans can handle both goal-targeted tasks and intransitive ones (Whiten et

³The concept of motor vocabulary together with the reported mirror activity of Broca's area, suggested to be the locus of language processing, have led to the hypothesis that the mirror system may be the precursor of the brain network responsible for language (Rizzolatti & Arbib, 1998; Arbib, 2002; Pulvermüller, 2002).

⁴These tasks in which no explicit goals are involved were often associated with the means of achieving actions, in contrast to the goal only (Tomasello et al., 1993; Wohlschläger et al., 2003).

al., 1991). One simple reason for this was proposed, suggesting that monkeys may lack intentions for imitation (Rizzolatti & Luppino, 2001). However, recent studies rather pointed out that, in humans, another processes recruiting different cortical areas are also involved. It not only because monkeys do not have an intentional system for imitation that they do not truly imitate, they also seem to lack a fundamental cognitive processes for imitation. Indeed, the human mirror system has been reported to be activated by the observation of both goal-directed and meaningless, or intransitive, movements (Iacoboni et al., 1999; Buccino et al., 2001; Koski et al., 2002; Grezes, Armony, Rowe, & Passingham, 2003). For instance, in the human primary motor cortex (M1), transcranial magnetic stimulation (TMS) studies showed that the motor excitation threshold needed to induce motor responses decreases during the observation of either meaningful or intransitive movements (Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995; Clark, Tremblay, & St-Marie, 2003).

At the level of the cortical functions, the main differences between humans and monkeys were primarily identified by means of brain imaging experiments. In humans, in addition to the areas homologous to those composing the monkey's mirror system, other cortical areas were also reported to be active during tasks resembling imitation. They include the primary motor cortex (M1) (Fadiga et al., 1995; Hari et al., 1998), the primary and secondary sensory cortices (SI-II) (Keysers et al., 2004), the supplementary motor area (SMA) (Ferstl & Von Crammond, 2002), the extra-striate body area (EBA) (Downing, Jiang, Shuman, & Kanwisher, 2001; Astafiev, Stanley, Shulman, & Corbetta, 2004; Chan, Peelen, & Downing, 2004), the inferior parietal lobule (IPL) also known as the tempo-parietal junction (TPJ) (Decety et al., 2002; Brass et al., 2005), and the anterior medial frontal cortex (AMFC) (Decety et al., 2002; Brass et al., 2005). Patterns of activation found in M1, SI-II and SMA may not be inconsistent with actual monkey data. Indeed, although direct recordings within these areas in monkeys did not reveal any mirror activity, these regions may nevertheless be involved, either by sub-threshold excitations or contextual inhibition (Graziano et al., 2002; Keysers et al., 2004). However, the presence of the others areas in the putative human mirror system requires further considerations relative to their potential role.

The extra-striate body area (EBA) is highly suggested to be specialized for the recognition of biological motions and postures (Downing et al., 2001; Astafiev et al., 2004; Chan et al., 2004). It has also been reported to exhibit greater activation prior to imitation as compared to a task requiring only observation (Decety & Chaminade, 2003; Jackson, Meltzoff, & Decety, 2006). Interestingly, EBA has also been shown to be activated during self motor execution (mirror activity), which has been further shown to be not the consequence of self observation only (Astafiev et al., 2004). Importantly, EBA is not especially sensitive to goal-directed actions, but to any form of movements and postures. However, the exact role of EBA in imitation is still not clear. It is nevertheless

suggested to encode a visual representation of one's own body and that of others (Berlucchi & Aglioti, 1997; Ruby & Decety, 2001; Jeannerod, 2003; Jackson et al., 2006). This ambivalent property of EBA in the processing of either self or others' body has led to the hypothesis that this area is involved in the transfer of visual information represented in the third-person perspective into the first-person point of view. Indeed, EBA shows slightly greater activation during the observation of body parts presented in a third-person perspective (Chan et al., 2004; Astafiev et al., 2004). This suggests that more computational resources are needed in this condition for this transfer to be performed. However, another hypothesis may be raised to explain this differential activation contrasting the observation of bodies in a first-person perspective with a third-person point of view. In brief, it assumes that EBA basically represents body-centered representations of bodies, irrespective of their ownership, but more details will be given in Section 2.3.4 which addresses the mechanisms of distinction between self and others.

A similar modification of cortical activity has also been reported in the tempo-parietal junction (TPJ) in the inferior temporal lobe, which seems to indicate that EBA is tightly connected to this area (Ruby & Decety, 2001; Arzy, Thut, Mohr, Michel, & Blanke, 2006). The recruitment of TPJ in the processing of visual information may be explained by the fact that the human parietal cortex is known for its implications for the integration of sensorimotor information (Sirigu et al., 1996). More particularly, IPL or TPJ were clearly identified as mediating the processes of bodily awareness (Berlucchi & Aglioti, 1997; Blanke, Ortigue, Landis, & Seeck, 2002; Farrer & Frith, 2002), the representation of self information in allocentric or world-centered coordinates (Farrer & Frith, 2002), ego-centric perspective taking (Ruby & Decety, 2001; Jeannerod, 2003), and imitation (Chaminade & Decety, 2002; Meltzoff & Decety, 2003; Chaminade et al., 2005). Another important property of this area is that it is also involved in the inhibition of automatic imitative behaviors (Brass et al., 2005). From its central location between the visual inputs and the internal sensorimotor processes, TPJ may be seen as a gate which mediates the flow of visual information entering the parieto-frontal network controlling the execution of actions. During this inhibitory process, AMFC, which has been reported to control the self-initiation of movements, is also activated, suggesting a close relationship between this area and TPJ (Decety et al., 2002; Brass et al., 2005; Brass & Heyes, 2005). This TPJ-AMFC network has thus been suggested to control imitative responses. While AMFC provides intentional cues to imitation, TPJ controls the flow of visual information entering the imitative resonance mechanism.

From these findings, a strong belief has developed in that, in contrast to that of the monkey, the human mirror system may involve at least two separate processes (Blakemore & Frith, 2005; Rumiati et al., 2005; Oztup et al., 2006). While a monkey-like *affordance-based* neural information flowing through IPS to Broca's area would be more concerned with goal-centered aspects of the

action (Koski et al., 2002; Chaminade & Decety, 2002; Brass & Heyes, 2005), a **motor contagion** pathway going from the extra-striate body area, through the inferior parietal lobule, to the dorsal part of the parietal cortex would be more specifically tuned to the processing of intransitive biological motions (Rumiati et al., 2005). This cortical differences with the monkey brain then suggested that imitation is not inherent in a macaque-like mirror system but instead depend on the embedding of circuitry homologous to that of the macaque in more extended systems within the human brain (Rizzolatti & Arbib, 1998; Oztot et al., 2006).

In the following paragraphs, a schematic view of the two suggested pathways underlying the human abilities for imitation are presented. Since, the affordance-based or goal-directed action pathway resemble the monkey mirror system, it will be first described.

Goal-Directed Action Pathway

Similarly to that of the monkey, the human goal-direct action pathway involves the cortical regions responsible for the analysis of visual information related to objects, individuals, and their interactions, i.e., IPS, STS and IT. As shown in Figure 2.12, since EBA is involved in the recognition of people, it is also included in this neural pathways as an input to STS. The ventral visual pathways for action recognition is then assumed to be gated by TPJ before entering the action stream which ends into PMv, the homologue regions of area F5 in the monkey. Further, the execution of either a self-motivated action toward an object or an imitative response is mediated by the frontal cortices which triggers the actual performance controlled by the parieto-frontal network for reaching and grasping. Importantly, it should be noticed that the action streams primarily involves object- or goal-centered representations.

Motor-Contagion Pathway

In contrast, the motor contagion pathways are primarily distributed within cortical regions involving body-centered representations. The schematic view of these pathways is depicted in Figure 2.13. Similarly to the goal-directed pathways, the flow of visual information passes through EBA and STS. IT is not involved here since no objects are considered in this case. Further, gated by TPJ, visual representation of bodies are projected to the posterior parietal cortex in particular in the human homologue of monkey area 5, integrating a multisensory representation of the body. This region can thus perform the mapping between observed movements with one's own movements. Then, this area projects to SMA in order to activate the corresponding motor commands. This dorsal parieto-frontal network is thus suggested to be the motor resonance system for intransitive movements. More or less simultaneously, the extended ventral visual stream passing through STS and LPFC control with more precision what type of movement has been observed and help SMA select the correct movement

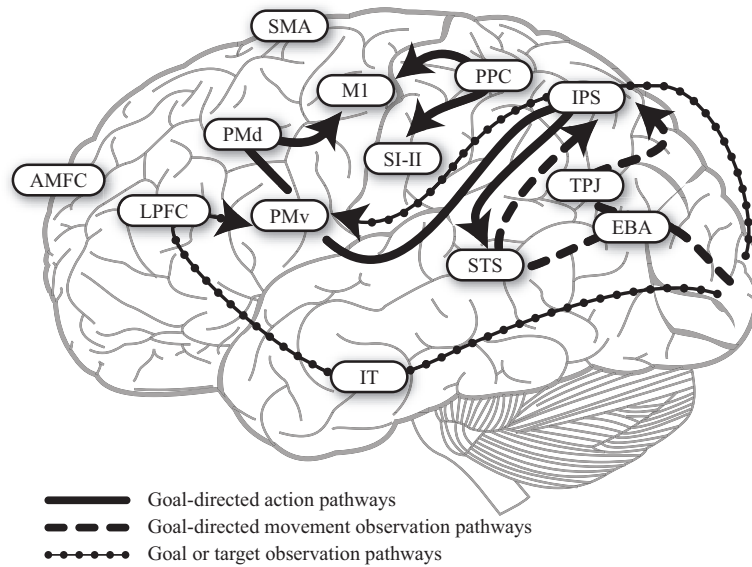


Figure 2.12: The goal-directed action pathways in a lateral view of a human brain. This figure mostly combines the pathways associated to action observation and execution as reported in the monkey and already shown in Figures 2.11 and 2.10. In addition, in order to lighten the figure, several areas present on previously mentioned figures were fused into single tags. Finally, the cortical areas more specific to the human brain were also added.

for reproduction. Finally, AMFC controls the initiation of the self-generated movements.

This analysis of the human cortical pathways for imitation clearly suggest that an additional processes for imitating intransitive movements is at play as compared to current knowledge of monkey neurophysiology. Importantly, this stresses that the possible lack of an intentional system for imitation in monkey should not be considered as the only problem by which monkeys do not truly imitate (Rizzolatti & Luppino, 2001). Indeed, the main difference between humans and monkeys seems that human have a more elaborated cortical network for imitation (Rumiati et al., 2005). This agrees with several current views concerning the role of monkey mirror system with respect to imitation. Mirror neurons may be involved in generating imitation without imitation being the function that favored their evolution by natural selection. In other words, imitation and other functions of mirror neurons could be exaptations rather than adaptations (Arbib et al., 2000; Oztop & Arbib, 2002; Keysers & Perrett, 2004; Brass & Heyes, 2005). The human ability for imitation would therefore be built on top of the mirror system found in its ancestor and have evolved similarly by exaptations of the primary role of the areas involved in the second processing stream of the reproduction of intransitive actions.

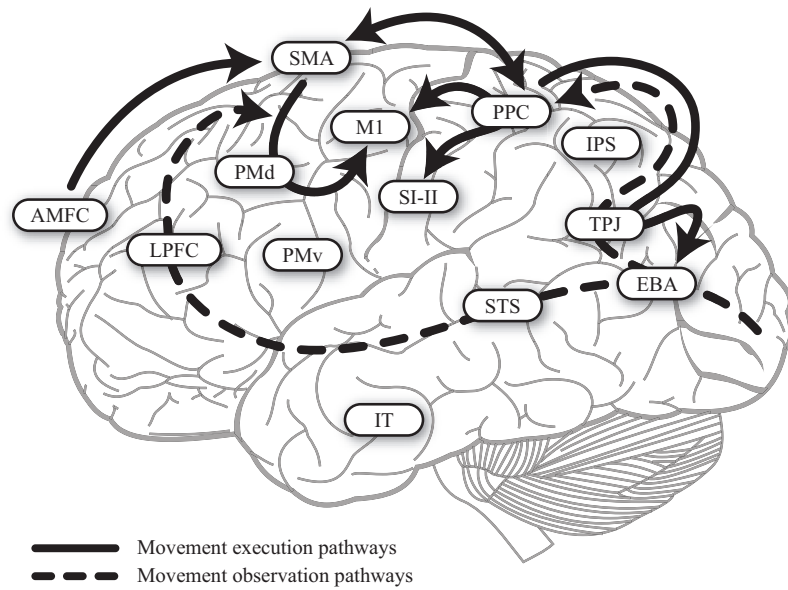


Figure 2.13: Illustration of the intransitive action pathways in a lateral view of a human brain. Similarly to Figure 2.12, both pathways mediating movement observation and execution were combined on the same figure.

2.3.4 SHARED REPRESENTATIONS AND THE DISCRIMINATION BETWEEN THE SELF AND OTHERS

An important issue related to the putative neural mechanisms at the basis of imitation concerns a problem arising from the principle of shared representations. Why don't we confuse observed actions with our own intentions and copy every movement that we see (Jeannerod, 2003; Brass & Heyes, 2005). Although interferences still occur during normal behaviors, suggesting conflicts at the level of these shared representations (Kilner et al., 2003; Chaminade et al., 2005), the human ability to discriminate between self actions and those of others is still efficient (Van Den Bos & Jeannerod, 2002; Jeannerod, 2003).

Several experiments have addressed the brain mechanisms of self-awareness and self-agency which imply the ability to discriminate between self and others (Chaminade & Decety, 2002; Decety et al., 2002; Farrer & Frith, 2002). A key finding of these studies concerns a probable lateralization of the processes underlying the representation of self and others. With respect to imitation, the most important differentially highlighted area in both hemispheres of the human brain is the tempo-parietal junction (TPJ). Interestingly, in a reciprocal imitation task, where human subjects were required in a first phase to imitate observed actions of an experimenter, and further to produce self-intended actions and observe the experimenter imitating them, Decety et al. (2002) showed that the right TPJ was more activated when the subjects were being imitated, whereas the left TPJ was more strongly involved during active imitation. This

study clearly puts forward the necessity to consider the two brain hemispheres as being involved in similar but different fundamental functions. Indeed, Decety et al. (2002) interpreted their results by stating that the lateralization of TPJ may indicate a separate processing of self-related versus others-related information, respectively in the left versus the right hemisphere. However, this hypothesis might be not correct. As will be explained below, converging lines of evidence rather suggest the opposite while staying in accordance with their experimental results.

In addition to this hypothetical separation between the left and the right hemispheres, studies focusing on other cognitive processes have also suggested functional roles which may apply differentially to each hemisphere. Considering the control of movements, lesions studies indicate that the left hemisphere is more associated with the control and planning of movement trajectory, i.e., the executive functions, (Hermsdörfer, Blankenfeld, & Goldenberg, 2003; Haaland, Prestopnik, Knight, & Lee, 2004), while the right hemisphere is more related to the online control of movements and closed-loop processing, i.e., the internal representation and monitoring functions (Haaland & Harrington, 1996; Hermsdörfer et al., 2003). Together, these hypotheses appear to converge toward an unified description of the general role of both sides of the brain. The left hemisphere seems to be the dominant one for action, more or less irrespectively of the agent performing the action, i.e., it may be more concerned in relating the self with the others and conversely by means of shared representations (Barresi & Moore, 1996; Decety & Chaminade, 2003; Mühlau et al., 2005; Jackson et al., 2006). In contrast, the right hemisphere would be more involved in self-awareness and self-representation by differentiating self from the others⁵ (Farrer & Frith, 2002; Decety & Chaminade, 2003; Brass et al., 2005). For instance, self-face recognition has been reported to primarily involve that right hemisphere and not in the left one (Uddin, Kaplan, Molnar-Szakacs, Zaidel, & Iacoboni, 2005). Additional specific evidences of this functional lateralization between self and others processing are further given with respect to the brain areas suggested to be involved in imitation.

Primarily sensory areas are first considered. The somatosensory area SII which is sensitive for feeling being touched, has been shown to display a mirror-like activation in the left hemisphere only (Keysers et al., 2004). In contrast, EBA, which also exhibit a differential hemispherical activation during the observation of bodies, has been shown to be more activated on the right side during the observation of body parts presented in a third-person perspective than in a first-person perspective (Astafiev et al., 2004; Chan et al., 2004). Similarly, representing mentally oneself from an allocentric point of view results in a higher activation of the right EBA with respect to an egocentric representation (Arzy

⁵It is however important to note that this separation is not strict. Indeed, since both hemispheres are highly interconnected, one should consider these points as indicating hemispherical dominance rather than exclusivity. For instance, the parietal cortex is bilaterally involved in the monitoring of self generated movements (Wolpert, 1998; Chaminade et al., 2005).

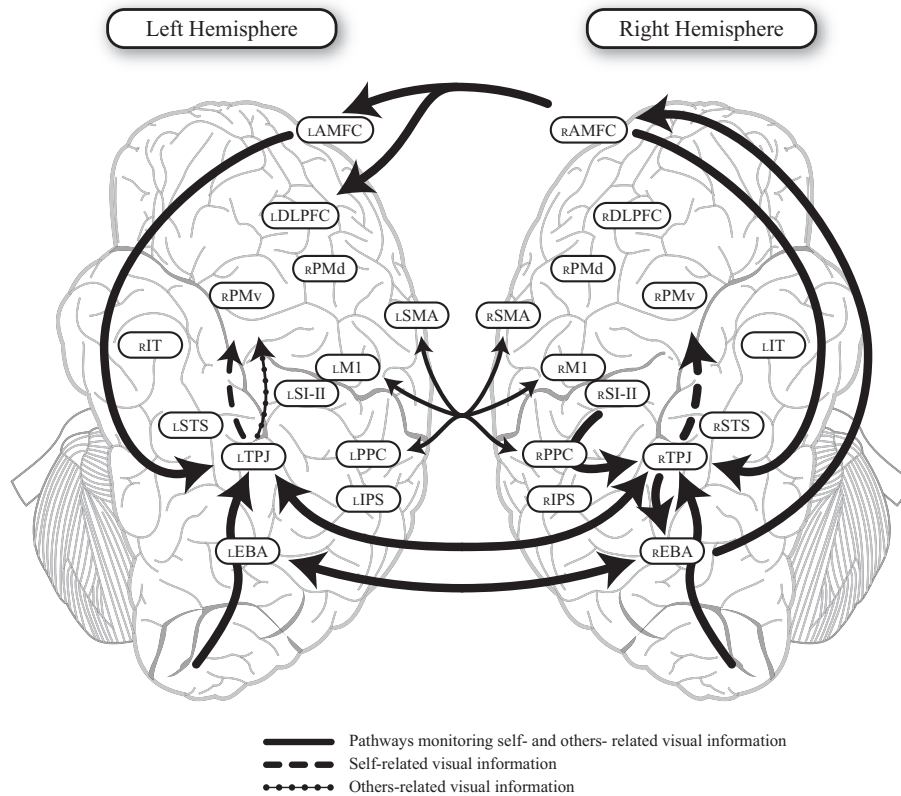


Figure 2.14: Illustration of the cortical pathways mediating the distinction of the visual information related either to the self of other individuals. The cortical circuit formed by EBA, TPJ, and AMFC is likely to be responsible for gating the flow of visual information entering either the left or the right hemisphere. While the right hemisphere may be restricted to represent self-related information only, the left hemisphere seems to allow shared representations to develop by combining both self- and others- related information.

et al., 2006). The left EBA is similarly activated in both conditions, suggesting that it encodes view-independent visual information (Chan et al., 2004; Arzy et al., 2006).

At first glance, the findings related to the right EBA, may contradict the proposed model of hemispherical lateralization. Indeed, how may one explain the stronger involvement of the right EBA during both the observation of the body parts of others, and the representation of self from an allocentric point of view, since that hemisphere is suggested to process the sense of self-awareness. Nevertheless, one may assume that EBA is not primarily concerned with disambiguating self and others, in that it may represent bilaterally a body-centered representations of bodies irrespectively of their ownership. Then, its reported greater activation in the third-person condition may be caused by an inhibitory

back-projection from the right TPJ⁶. Indeed, the right TPJ seems to be involved in the processing of self-awareness and self-agency by integrating multimodal sensorimotor inputs from its neighboring regions (Blanke et al., 2002; Farrer & Frith, 2002; Decety & Sommerville, 2003; Brass et al., 2005). Similarly, the left TPJ is also suggested to perform multisensory integration, but rather with the purpose to actively imitate others (Decety et al., 2002; Mühlau et al., 2005; Jackson et al., 2006) and to reason about the beliefs of others (Samson, Apperly, Chiavarino, & Humphreys, 2004). Here, an hypothesis relative to the functional role of the TPJ is adopted. The role of both left and right TPJ is proposed to gate the flow of view-independent visual information coming from EBA, such that this information may either be integrated into the parietal cortex as belonging to the self-body schema (mostly right TPJ), or be considered as to belong to another individual (mostly left TPJ). Within the left hemisphere, the integration of the visual description of another individual may further be useful for understanding its intentions or even for imitation. Then, in order to be capable to determine whether a given visual input belong to the self, the right TPJ may act as a comparator between an observed body part and the mental prediction of the visual aspect of that body part processed within the parietal cortex. When there is a perfect match, the right TPJ let the visual information conveyed by EBA enters the parietal cortex bilaterally. However, a mismatch may trigger the right TPJ to inhibit the right EBA. Interestingly, the inhibition of automatic imitative behaviors has been shown to also involve right TPJ (Brass et al., 2005; Brass & Heyes, 2005), which may bring one to postulate that this area has also a selective inhibitory effect on the left EBA. Indeed, by controlling the amount of visual information entering the left hemisphere, the brain may hence reduce the motor resonance suggested to lead to automatic imitative response tendencies. Thus, together with prefrontal areas responsible for movement inhibition and self-awareness, i.e., LPFC and AMFC, the network encompassing the tempo-parietal junction is suggested to both control the self body schema, and to modulate the amount of convergence between self-related and others-related sensory information for action understanding and imitation.

To summarize, the left hemisphere is believed to be the dominant one for action (Barresi & Moore, 1996). It can map the action of others into one's own sensorimotor representation and hence is more concerned in relating others with self and thus in imitation. (Decety et al., 2002; Decety & Chaminade, 2003; Jackson et al., 2006). In contrast, the right hemisphere seems to be more involved in self-awareness. One of its role is primarily to differentiate the self from the others. By this cognitive ability, it can also inhibit automatic imita-

⁶Although one may think that an inhibited region should not appear active during a PET or fMRI scan, it should be mentioned that the brain areas highlighted in a given task are not mandatorily directly involved in that task. Indeed, the activation patterns measured by these brain imaging techniques may also indicate that these areas receive strong synaptic inputs from other areas, which can therefore be either excitatory or inhibitory (Arbib, Bischoff, Fagg, & Grafton, 1995).

tive responses and motor contagion mechanisms initiated in the left hemisphere (Decety et al., 2002; Decety & Chaminade, 2003; Brass et al., 2005; Brass & Heyes, 2005).

Summary

The brain regions and neural pathways responsible for fundamental imitative behaviors in monkeys and humans were described in this section. These pathways are similar in both species, since they are relatively close in the evolutionary tree. However, in line with the behavioral discrepancies between humans and monkeys reported by ethological studies (see Section 2.1), there are also several cortical differences. This review suggested that both that monkeys may lack an intentional system for imitation as well as that a supplementary brain areas and a second cortical route are involved in the imitation of intransitive actions, and by extension, in the imitation of the means of achieving goals. While a goal-directed imitative pathway, common to both species, has been shown to recruit goal-centered neural representations, the second route may have evolved in humans through the use of cortical regions that represent information encoded in body part-centered reference frames. This thesis provides models that accomplish two goals: exploring how such representation could be obtained from purely sensory representations, and more precisely identifying the neural pathways that mediate these two imitative routes. The approach here consists of developing and reusing existing experimental paradigms developed in experimental psychology in order to analyze these seemingly separate processes. This allows for an investigation of how these processes could interact and also at what level of the cortical sensorimotor circuits this interaction occurs.

Using the same approach this thesis will also investigate the neural mechanisms responsible for discrimination between the self and others. The principle of shared representations, whose neural substrate is located within the mirror neuron system, implies a mechanism for disentangling the sensory feedback associated with the body of oneself or of another. This research is therefore also intended to illuminate which neural mechanisms, at both the neural structure and network level, are responsible for this fundamental cognitive process. The next section first presents some relevant attempts to model the mechanisms of imitation, including the mirror neuron system.

2.4 RELATED NEURAL MODELS OF IMITATION

Many modeling studies have addressed imitation and the mirror system. Although these two computational issues have often been intermingled, it is important to stress that they are not reciprocally inclusive. Indeed, a fundamental problem in this easy association that today scientists usually make between imitation and its putative neurophysiological substrate, is that monkeys, which are animals endowed with such a system, do not imitate in a strict sense. Therefore, building a system having mirror properties as reported by neurophysiology may not be sufficient to explain the mechanisms underlying human imitation. Additional networks or connectivity among existing ones must be considered in order to provide a computational model to perform true imitation. Modeling the mirror system or its development is one thing, modeling the mechanisms of imitation as performed by the human brain is far a more complex one.

In this section, a review of some models of imitation which provide important computational landmarks and key steps toward a better understanding of this human ability is given. But before starting, it is necessary to note that different approaches have been adopted in the literature. In this thesis, since the focus is on the understanding and on the modeling of the cortical processes of imitation, this review will only describe models which use computational neuroscience as a fundamental modeling basis⁷. First of all, generalist theories which try to provide global insights on the processes underlying imitation are described. Next, computational studies primarily concerned with the modeling of the mirror system as well as how to take advantage of its properties in order to endow these model with imitative behaviors are presented. Further, a description of an approach to imitation based on control theory in which inverse and forward control models are coupled is given. Developmental approaches will also be discussed in which imitation abilities are suggested to arise from sensorimotor development. And finally, this review will end with models which are intended to explain how high-level cognitive functions related to imitation can be produced, such as those involving reasoning.

2.4.1 THEORIES OF IMITATION

Sensorimotor Resonance and the Ideomotor Theory

Imitation and its underlying cognitive mechanisms has received a lot of attention from psychology in the past century. An early study was reported by Piaget (1978), who did a longitudinal analysis on his three children. He suggested that imitation develop through several stages in the childhood. He also postulated

⁷The reader interested in computational models developed along a more engineering-based methodology is referred to the following references (Kuniyoshi, Inaba, & Inoue, 1994; Schaal, 1999; Ijspeert et al., 2002; Calinon & Billard, 2007; Guenter et al., 2007).

that these stages closely follow the hierarchical growth and maturation of the internal sensorimotor loops for the control of the body. Interestingly, by what he already called sensorimotor resonance, the developing cognitive processes of sensorimotor awareness, in addition to produce low-level forms of imitation, allow children to establish social codes and communicate with the adults, and even to infer the mental states of others. In agreement with the more recent work of Nadel et al. (1999), Piaget (1978) also suggested that the early resonance processes or circular reactions, may be due to the inability of infants to differentiate the self from the others, an ability which arise further through the development of the children. The idea of a common representation originally recruited irrespectively of the ownership of input visual stimuli representing movements, was also a fundamental hypothesis of the ideomotor theory, by which observing others influence the quality of one's own performance (Greenwald, 1970).

The Active Intermodal Mapping Model

Thirty years ago, Meltzoff and Moore (1977) reported that neonates can already imitate a small set of facial expressions such as tongue protrusion. This discovery was particularly intriguing because it is difficult to explain how a child, who has not seen his/her own face yet, can map the visual information of someone protruding the tongue to his/her own motor commands for performing the same action. In order to explain this phenomenon, Meltzoff and Moore (1997) suggested the Active Intermodal Mapping hypothesis (AIM) that supposes an innate visuomotor mapping of the body mediated by a supra-modal representation. A sketch of this model is given in Figure 2.15a. Unfortunately, the authors of this hypothesis neither mentioned nor provided any evidence of what may be the content of this supra-modal representation and of in which part of the nervous system it may be located. Nevertheless, subsequent studies suggested that face-detection abilities of newborn infants may be processed through an innate subcortical route (Johnson, 2005). A problematic issue with the AIM theory is that it postulates the existence of an innate visuomotor mapping between visual perception of body parts with the motor codes making them move, and therefore, it implies that imitation is body part specific, which was shown to be wrong. Indeed, imitation is much more complex. First, it clearly involves more than one automatic cognitive processes, and second, it also depends on rational forms of reasoning (Bekkering et al., 2000; Gergely & Csibra, 2003; Wohlschläger et al., 2003).

The Theory of Goal-Directed Imitation

Another important characteristic of imitation is that it is primarily directed toward achieving goals. The theory of Goal-Directed Imitation (GOADI) proposed by Wohlschläger et al. (2003) considers imitation as being purposeful, a principle which can thus be valid in children, adults and animals. GOADI describes an imitative act in several steps. First, perceived acts are decom-

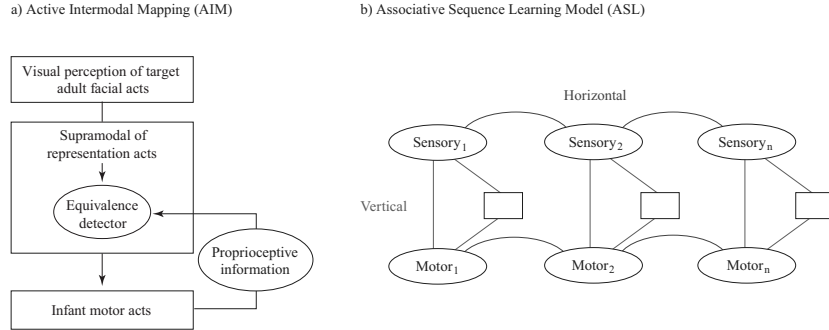


Figure 2.15: Cognitive Theories of Imitation: **a)** The Active Intermodal Mapping (AIM) theory assumes that visually perceived acts are actively mapped onto motor output via a supra-modal representation system. **b)** The Associative Sequence Learning (ASL) model assumes that visual (sensory) representations of action become linked to motor representations (encoding somatosensory information and motor commands) through Hebbian learning. In environments where the same action is simultaneously seen and executed, links are formed between visual and motor representations of the same action. As a result, action observation may trigger the corresponding motor representation. (*Adapted from Meltzoff and Moore (1997); Heyes (2001)*)

posed of into separate goal-centered aspects. One or more aspects, the most important ones, are then selected by a cognitive process with respect to capacity limitations. Next, the selected goals are hierarchically ordered according to the functionality and relevance of actions. For instance, ends, e.g., objects and treatments of the latter, are more important than means, e.g., effectors and movement paths. Finally, the selected goals elicit the motor programs to which they are strongly associated (ideomotor principle). These motor programmes do not necessarily lead to matching movements, although they might do so in many everyday cases (Bekkering et al., 2000; Wohlschläger et al., 2003). In contrast to AIM, GOADI not only explains several experiments on behavioral imitation in children and adults, but also gives this ability a more functional nature. The purpose of learning by imitation consists primarily to reach the same goals as the model, while copying the means may be helpful, but is not mandatory (Wohlschläger et al., 2003).

The Associative Sequence Learning Model

Complementarily to GOADI, Heyes (2001) proposed the Associative Sequence Learning model (ASL). In contrast with the AIM model, and in agreement with Piaget (1978), the ASL model considers that the development of imitative abilities is highly experience-dependent. Moreover, it also suggests that imitation is mediated by sets of bidirectional excitatory links between sensory and mo-

tor representations of movement units (See Figure 2.15b). While links between within motor or sensory representations represent possible sequences of actions, the so-called vertical associations connecting together sensory and motor representations are proposed to be either direct or indirect. The latter indirect mapping is suggested to be relayed by an intermediate representation of distinctive content. For instance, a language representation such as words or phrases could be established when the verbal stimulus co-occurs with either the sight or execution of a movement. Despite being strongly based on experience-dependent sensorimotor associations, this model also acknowledges the possible existence of innate vertical connectivity for some facial behaviors such as smiling, yawning or tongue protrusion (Heyes, 2001; Brass & Heyes, 2005). Finally, the ASL model also attempts to explain the imitation of opaque actions such as facial expressions. It suggests that the corresponding sensorimotor representations are normally generated in specific environments such as those created by optical mirrors, imitative social partners and explicit training regimes.

2.4.2 MODELING AND UNDERSTANDING THE MIRROR SYSTEM

This section presents modeling studies which aimed at understanding the neural mechanisms imitative behaviors as well as their suggested core neural substrate, i.e., the mirror system.

CONNECTIONIST APPROACHES TO THE MIRROR SYSTEM

One of the earliest modeling study related to the neural mechanisms of imitation has been proposed by Arbib et al. (2000). In this work, the authors primarily proposed a framework in which the results of their modeling studies could be compared to real fMRI data obtained on human subjects (Arbib et al., 1995). This method consists in assigning to each sub-network of a model, a given cortical area. Then by measuring the strength of synaptic activity of each region of the model, which was suggested to correspond to what is effectively measured during brain imaging experiments, they were capable to reproduce neurophysiological data related to imitation in humans as well as to provide experimental predictions.

Next, as mentioned above, this work considered several modeling studies of the authors, which were also published separately. First, a connectionist model which primarily aimed at learning the coordination patterns of arbitrary human movements was proposed by Billard and Mataric (2001). In this work, a hierarchy of time-delayed neural networks capable of learning to reproduce sequences of arbitrary signals was developed (Billard & Hayes, 1999). Each of the elements described in the model was associated with a brain region among those involved in visuo-motor control, including parts of the monkey mirror re-

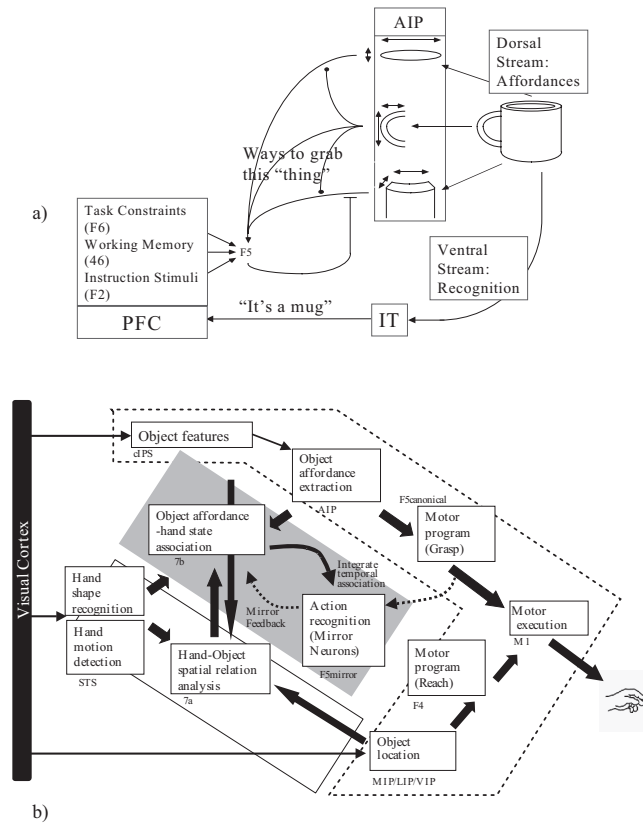


Figure 2.16: FARS and MNS1 models: a) Minimal sketch of the FARS model related to the execution of grasping actions. The observation of a cup activates possible affordances within area AIP which in turn selects the corresponding possible motor programs in area F5. In parallel, another stream flows through IT which informs the prefrontal cortex of the precise nature of the object. PFC finally decides which is the correct affordance for grasping that object. **b)** The Mirror Neuron System 1 extends the FARS model by considering additional brain areas and by distinguishing the role of the F5 canonical and mirror neurons. A specific area processes the hand-state which corresponds to an invariant representation of the relationship between any hand and objects. (Adapted from Fagg and Arbib (1998); Oztop and Arbib (2002))

gions. After learning from human demonstrations, the reproduction of observed movements showed a high qualitative and quantitative agreement with human data. It is important to note here that the mapping of the internal model of movement generation with the visual perception of another individual was realized by means of an invariant representation of arm movements. Indeed, in this model, Billard and Mataric (2001) used the orientation and velocity of each body joint as sensory cues to be learned.

The Mirror Neuron System 1 and 2

The second model that was described by Arbib et al. (2000) consisted of a detailed modeling of the mirror neuron system in monkeys which includes a neural implementation (Oztop & Arbib, 2002). The so-called Mirror Neuron System 1 model (MNS1) primarily focused on understanding the patterns of

activation within monkey area F5 and its interactions with its connected brain regions. This model extends the Fagg-Arbib-Rizzolatti-Sakata model (FARS) which represents the brain circuitry for visually guided grasping of objects (Fagg & Arbib, 1998), by connecting this original model with instances of F5 mirror neurons. Schematics of both FARS and MNS1 models are illustrated in Figure 2.16. In order to be able to map sensory and motor representations into a single mirror area, this system-level model proposed an invariant or goal-centered representation of the possible interactions between the hand and objects: the hand-state hypothesis. Although this hypothesis can be seen as an over simplification of the problem of understanding the process of visual information, it is in accordance with neurophysiological data related to mirror neurons. Indeed, these neurons do fire with respect to a goal-centered reference frame.

Interestingly, this modeling study suggests that the original function of the mirror system is to provide appropriate visual feedback for on-line control of manual object grasping. The mirror property which seems to be related to imitative abilities would have developed later. Further, the important contribution of this work to the field is that Oztop and Arbib (2002) showed that the mirror neurons within F5 can be trained to recognize actions already in the motor repertoire of the observer, assuming that this sensorimotor repertoire was learned by self-observation prior to the recognition of others actions. The ability of the system to further generalize to anyone performing the movements is allowed by the invariant representation, i.e., the hand state.

Recently, the second version of this model, the MNS2, was proposed by Bonaiuto et al. (2007). MNS2 extends MNS1 by means of three major improvements. First, the new model uses a recurrent architecture and a learning paradigm that are biologically more plausible than that of the original model⁸. Moreover, MNS2 is capable of addressing data on audio-visual mirror neurons and on the contextual responses of mirror neurons when the target object was recently visible but is currently hidden (Bonaiuto et al., 2007).

A Dynamical System Approach

Recently, another connectionist modeling study adopted a dynamical system approach to imitation inspired by the mirror system was proposed by Erlhagen, Mukovskiy, et al. (2006); Erlhagen, Mukovsky, and Bicho (2006). The model was primarily built following the theory of Goal-Directed Imitation (GOADI) discussed above, in that the model is biased toward reproducing the outcome of an observed action sequence rather than reproducing the exact action means. The model architecture is composed of building blocks where each of them represents the functionality of neurons belonging to a specific brain area.

⁸MNS1 used back-propagation with some time-dependent preprocessing in order to learn the sensorimotor associations (Oztop & Arbib, 2002), while MNS2 applies the well-known hebbian rule (Bonaiuto et al., 2007).

Similarly to the MNS1 model, the considered brain areas and their connectivity are fundamentally based on current biological evidence of the structure of the mirror system, i.e., the network formed by areas F5, PF and STS (See Figure 2.17). In addition, Erlhagen, Mukovskiy, et al. (2006) also considered prior task knowledge and contextual information in order to allow the model, having a dissimilar embodiment, to reproduce the perceived or inferred end-state of a grasping-placing sequence. Interestingly, the model implementation shows that, in the case where visual information is incomplete or where environmental constraints changes across trials, the motor simulation loop within the mirror network is a powerful mechanism to achieve imitation, and, by extension, to endow robots with the ability to understand the motor intention of other agents. Considering the technical aspects of this model, the Dynamic Field Theory⁹ (Erlhagen & Schöner, 2002; Schöner, 2002) consisting of continuous neural representation was adopted and a hebbian learning rule was used in order to learn sensorimotor associations.

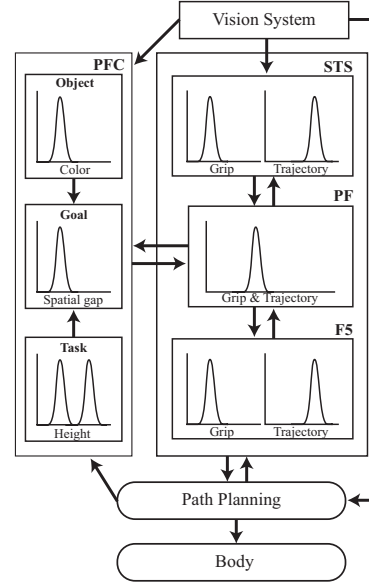


Figure 2.17: Illustration of a connectionist model which has been proposed to account for goal-directed imitative abilities. (Adapted from Erlhagen, Mukovsky, and Bicho (2006))

GROUNDING THE MIRROR SYSTEM ON THE MOTOR PRODUCTION SYSTEM

Next, while the studies presented above mainly focus on a high-level connectionist representation, other approaches consider the mirror system as an integrative part of the motor system and even show how mirror properties may naturally emerge from the modular architecture of the motor control process. Two major streams of models were developed. While one of them is more concerned with imitation at a behavioral level (Demiris & Hayes, 2002), the other is more strictly grounded on motor control mechanisms (Wolpert et al., 2003). Nevertheless, both models are very similar in their architecture. Their key ingredients are the modularity and the distributed cooperation and competition of several internal models. As shown in Figure 2.18a, each of these internal models is composed of a forward model associated with an inverse model, which can also be considered as a composition of a predictor and a controller, respectively. Because of the existence of multiple predictor-controller pairs, the main idea is that each controller competes against the others in order to take the control

⁹Since this approach was also adopted in this thesis, a detailed description of this approach is given in Chapter 3.

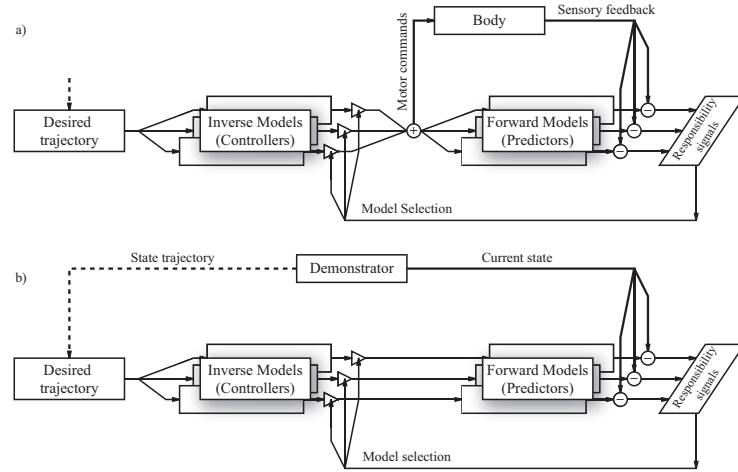


Figure 2.18: The modular architecture for motor control, action recognition and imitation which was proposed by Demiris and Hayes (2002); Wolpert et al. (2003). **a)** In motor control mode, the desired trajectory is fed to all the internal inverse controllers which compute in parallel possible motor commands. These commands are next combined and send to to the body muscles. Simultaneously, each forward model associated with a inverse controller predicts the next state of the system, which is further compared with real sensory feedback. The errors are normalized, and the resulting responsibility signals are finally sent back to the motor selection process to favor the best controller. **b)** In observation mode, the trajectory observed from the demonstrator is fed to the inverse controllers computing what would have been the possible motor commands to reproduce that movement. These commands are then fed to the corresponding forward models and their prediction is then compared with the actual state of the demonstrator. The pair of inverse and forward models having the highest responsibility signal is the one which would reproduce the trajectory best. (Adapted from Demiris and Hayes (2002); Wolpert et al. (2003); Oztop et al. (2006))

over the system. This competition is made possible by assigning a relevance value to each controller proportional to the accuracy of its associated predictor. This value is often described as a responsibility signal (Demiris & Hayes, 2002; Wolpert et al., 2003). Furthermore, it important to note that the controller-predictor pairs are adaptive, what allows the system to learn new and refine known control strategies or movements.

From now on, this type of model is capable of producing actions. However, as will be described next, this model can also be used for imitation and action recognition. The only problem, common to many others computational studies, is that this architecture requires the visual information of the observed movement to be converted into a reference frame compatible with the inputs of the system in normal conditions, i.e., the state variables. First, during movement observation, each controller generates the motor command required to achieve the demonstrated trajectory but none of them is actually executed. In order to close the mental simulation loop, the output of each controller is fed into the predictor it is paired with. As a final step, the parallel estimation of the possible next states of the demonstrator is then compared with the actual observed next state so as to obtain prediction errors. These signals can hence

further indicate which of the controllers would perform best the same movement as that executed by the demonstrator, and hence might be considered analogous to mirror neuron activity (Demiris & Hayes, 2002; Wolpert et al., 2003; Oztop et al., 2006). This additional use of this type of models, illustrated in Figure 2.18b, provides some similarities to the mirror neuron system, which has also been recently suggested to be part of a brain motor control system involving inverse and forward models of the body (Miall, 2003).

As rapidly mentioned above, the models of Demiris and Hayes (2002) and Wolpert et al. (2003) do exhibit one important difference. The model of Demiris and Hayes (2002) considered only kinematic variables in order to encode movements or behaviors, whereas the work of Wolpert et al. (2003), which concentrates on motor variables, was derived from an earlier model developed for understanding how the brain do motor control, namely the **Modular Selection and Identification for Control** model (MOSAIC) (Wolpert & Kawato, 1998).

Extensions of these Models

Later, subsequent studies have extended these models. First, the **Hierarchical Attentive Multiple Models for Execution and Recognition** (HAMMER) was proposed (Demiris & Johnson, 2003; Demiris & Simmons, 2006). A new control module was added on top of the behavioral controllers so that the HAMMER model can learn new actions composed of primitives already present in the motor repertoire. In addition, this hierarchical structure was shown to provide a mechanism for the top-down control of attention during action perception. In agreement with the theory of goal-directed imitation (GOADI), significant performance gains in terms of resource allocation were reported.

In parallel, the **Mental State Inference** model (MSI) was developed (Oztop et al., 2005). This computational model aims at providing a system with the ability to infer the mental state of others which is built upon the circuitry that subserves sensorimotor control. The major hypothesis of this work is that cortical regions, in particular monkey mirror areas, are involved in executive and predictive motor processes (Miall, 2003) such as in the MOSAIC (Wolpert et al., 2003) and the HAMMER model (Demiris & Johnson, 2003). By assuming that these brain areas can produce mental simulation of movements, they can also provide the capacity to understand the movements performed by others. However, in contrast to the preceding models, the MSI model involves only a single arrangement of paired inverse and forward models. Therefore, the mechanisms of mental state inference is based on a gradient search within the behavioral space, which sequentially affects the internal model. As long as the prediction of the mental simulation loop is not good enough, the search continues, until the most appropriate mental state is found. The analysis of the system behavior showed that mental states may be inferred from the observation of more or less complex movements. Thus, Oztop et al. (2005) showed that the computational strategy developed for sensorimotor control is effective in

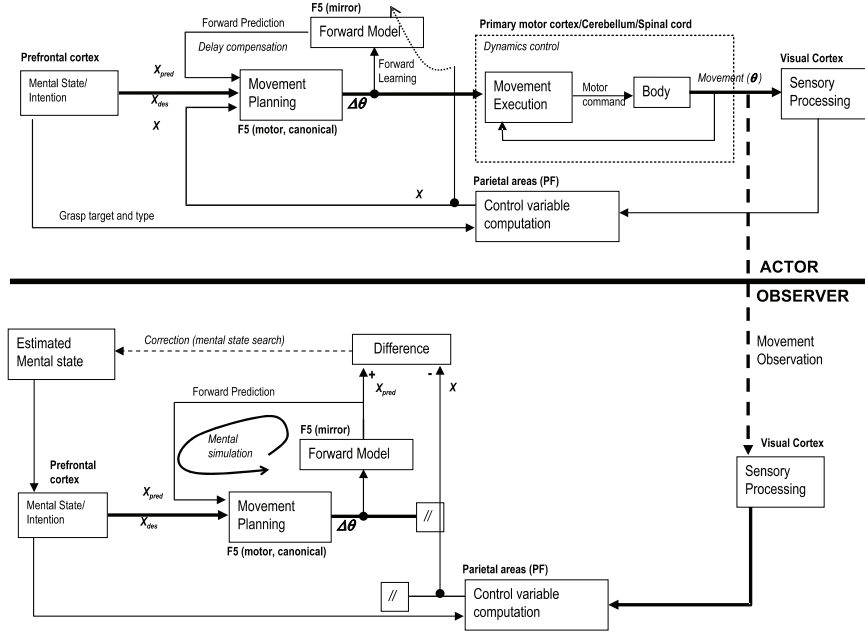


Figure 2.19: The mental state inference model (MSI): Upper panel: the MSI model is based on the illustrated visual feedback control organization. Lower panel: observer’s mental state inference mechanism. Mental simulation of movement is mediated by utilizing the sensory prediction from the forward model and by inhibiting motor output. The difference module computes the difference between the visual control parameters of the simulated movement and the observed movement. The mental state estimate indicates the current guess of the observer about the mental state of the actor. The difference output is used to update the estimate or to select the best mental state. (Adapted from Oztog et al. (2005))

inferring the mental states of others.

DEVELOPMENTAL APPROACHES TO IMITATION

In this section, another class of models are described. The approach which they adopted consists in considering the mirror system as the result of the development of the internal sensorimotor loops. These works closely follow the hypothesis that self-observation is the principle to bootstrap imitation (Piaget, 1978). Initially, an internal multisensory and motor representation is build from self-observation, allowing an agent to execute and monitor its actual behaviors. By generalizing the visual inputs entering this representation, representing the actions of others becomes possible, which is further used to trigger imitation.

A Hebbian View of the Mirror System

First of all, a high-level theoretical study was proposed by Keysers and Perrett (2004) who described an extremely simple system-level hebbian model in order to explain the emergence of the mirror system. This model only addresses the network composed of the monkey mirror areas, i.e., F5, PF and STS. Bas-

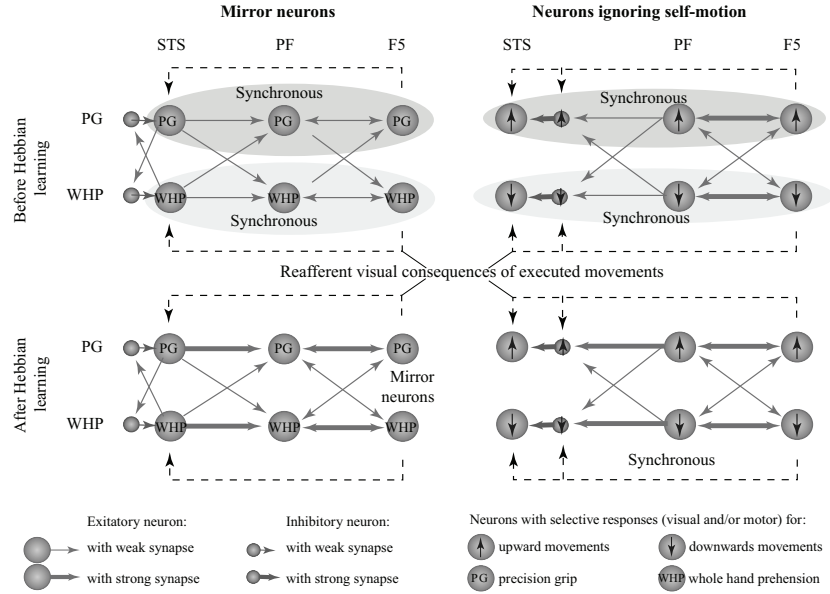


Figure 2.20: Hebbian Model of the Monkey Mirror System: **a)** On top, a chain of interconnected STS, PF and F5 neurons before learning is shown. PG and WHP stand for neurons selective for precision grip and whole hand prehension, respectively. During self-observation of PG or WHP, F5 activity leads to motor output, and thereby to visual re-afferent activation of the corresponding STS-PG neurons (dashed arrow). Hebbian learning then builds strong connections between neurons sharing similar properties, from which mirror sensitivity emerge (bottom). **b)** In this example, neurons sensitive to upwards and downwards movements are considered. During the execution of upwards movements, the visual consequences of the action activate upwards STS neurons. At the same time, a PF/STS corollary discharge conveys the motor and kinesthetic information regarding the action to the inhibitory interneurons controlling STS neurons sensitive to both upwards and downwards motion. Synchrony with the visually evoked activity occurs only in the upwards neurons, as the monkey is currently viewing his upward movement, resulting in hebbian enhancement of only the matching synapses. In addition, the inhibitory interneurons receive also visual input, and hebbian learning occurs between PF and STS. As a consequence, STS upward neuron gets slowly inhibited by self-motion. (*Adapted from Keyzers and Perrett (2004)*)

ing their work on current neurophysiological data, the authors basically describe that the mirror properties of the neurons found in these areas emerge, from a pure motor or visual sensitivity to a multisensory representation through development. They also mention an hypothesis as to how these neurons specialize, by refining their internal representation for specific actions. based on known anatomical connections across these areas Keyzers and Perrett (2004) proposed an associative hebbian network given in Figure 2.20, through which the monkey brain may learn to understand the actions of others by associating them with self-produced actions. An additional interesting issue was also addressed. This network may simultaneously learn to discriminate its own actions from those of others. The important hypothesis raised by this study is that the mirror system, despite being central to imitation and social communication, probably evolved originally for different purposes such as the monitoring of the visual feedback

to improve motor control and the reduction of redundant visual information (Keysers & Perrett, 2004; Oztop et al., 2006).

Embodiment and Imitation

Further, a developmental approach was followed by Andry et al. (2004) in order to provide robots with an embodied and autonomous ability to learn their internal sensorimotor mapping. The core of this model is a neural network which associates, by hebbian learning, visual, proprioceptive and motor signals related to the movements produced by the robot. This mainly allows to endow the system with a common multisensory and motor representation which also reduces the high dimensionality of the inputs. Similarly to the other neural models applying the hebbian paradigm within associative network, when applying partial sensory inputs, missing data can be completed and hence motor actions can still be driven accurately. Then, by means of the internal dynamics of this mapping, various sensorimotor behaviors such as tracking, pointing, spontaneous imitating, and sequences learning can then be obtained on their robotic platform. Indeed, the important idea that the work of Andry et al. (2004) comforts, is that imitative and sensorimotor capabilities may co-develop, what allows the acquisition and the building of increasingly complex behavioral abilities (Piaget, 1978).

Next comes the computational study of Weber et al. (2006), who described a hybrid generative and predictive neural model of the motor cortex where imitation emerge from the internally learned sensorimotor loops. The generative part of their model consists of building a topographically organized hidden representation through hierarchically directed cortico-cortical connections from the sensory inputs trained by an unsupervised learning method. Similarly to the work of Andry et al. (2004), a common multisensory and motor representation is produced. Complementarily, predictive abilities were added to the model by training lateral intra-area and inter-area cortical connections to predict the future state of the network using asymmetric hebbian rule. By closing the loop and avoiding overt motor execution, this predictive part of the model can also mentally simulate longer perception and action sequences. Finally, by considering the hidden layer of the model as a shared representation in which visual information related to other agents is also fed into, imitation abilities have shown to develop (Weber et al., 2006).

Imitation in the Cortical Hemispheres

A quite different approach was adopted by Petreska and Billard (2006), who investigated the neural mechanisms of visuo-motor imitation in humans by considering imitation deficits following callosal brain lesions. The neural network developed in this study also hold an internal multisensory representation which is learned through motor babbling. Importantly, when impaired with respect to callosal lesions reported in human patients (Goldenberg, Laimgruber, & Herms-

dörfer, 2001), this model was shown to account for the scores found in a clinical examination of imitation. Interestingly, the work by Petreska and Billard (2006) is one of the rare study which considers inter-hemispherical processes. In accordance with recent brain imaging studies, the strong proposal of this model is that the processes involved in active imitation are primarily localized in the left hemisphere.

HIGH-LEVEL IMITATION

In this last section, attempts to model higher cognitive functions related to imitation are presented. Higher cognitive functions are meant here to correspond to those requiring forms of reasoning which integrate for instance task knowledge, prior information, mental inferences and abilities for collaborative work. Although some of the previous models do integrate some of these skills, the studies presented here differ from the fact that they try to consider them as a whole within an integrated framework. However, since modeling together these high-level cognitive functions with neural networks may be difficult, the approach adopted by the models that will be described below is more theoretical in the sense that a probabilistic method was adopted. Nevertheless, probabilistic models have already provided interesting insights for explaining a large number of neurophysiological and behavioral phenomena (Van Beers et al., 2002; Pouget et al., 2003; Ma, Beck, Latham, & Pouget, 2006; Yuille & Kersten, 2006). For instance, several studies suggest that neural representations may actually encode probability distribution and implement the Bayes' theorem. Therefore, this would allow the nervous system to solve a variety of problems by combining prior knowledge and contextual information with its current state. Moreover, such a probabilistic approach can also help remove the noise present in the sensory and neural processing structures.

One of the earliest modeling studies which addressed imitative behaviors with a Bayesian methodology was proposed by Rao, Shon, and Meltzoff (2004). The authors of this work emphasized the importance of probabilistic reasoning along the development of imitative abilities. Indeed, they suggested how the different developmental stages of imitation can be associated with a Bayesian approach of increasing complexity. Furthermore, they also stressed the advantage of this probabilistic approach to imitation learning. Indeed, unlike supervised learning methods usually adopted in such tasks, a system endowed with a Bayesian reasoning process can deduce how to resolve tasks rather than simply learn to reproduce movements by demonstration.

In a later study, Cuijpers et al. (2006) developed a similar approach, but which aimed at understanding computational mechanisms responsible for goal-directed imitation and inference in tasks requiring collaboration between two agents. Similarly to the work of Rao et al. (2004), this model integrates possible actions and goals within a probabilistic framework which allows inferring the

actions of others. However, Cuijpers et al. (2006) also considered explicitly actions goals, task knowledge as well as personal preference, which provided to their model interesting properties useful for accomplishing collaborative tasks.

Summary

The modeling work in this thesis is largely inspired by the studies presented above. Although it primarily follows the connectionist approach, which models the brain as a collection of interconnected and interacting neural networks, this thesis is distinct from some of these models in several ways. For instance, the general purpose of the MNS1 and 2 models is to reproduce the activation patterns of the mirror neurons, but in this thesis this is not considered an end in itself. Instead, this thesis focuses on the properties of these groups of neurons, with the goal of deciphering the cortical pathways that mediate imitative behaviors.

Furthermore, most of the approaches reviewed here - while inspired by neurophysiological evidence - address imitation in a functional way in order to achieve other goals (for instance, to endow robots with imitative abilities). One of their primary aims is to deal with the issue of how imitative skill could be learned, either from scratch or from a predefined cognitive system; these studies also try to show how neural models can produce imitative behaviors. Because of this application-based approach, most of these studies do not provide predictions that can be directly tested experimentally. In contrast, the present thesis focuses explicitly on the brain mechanisms responsible for imitative behaviors, including neural processes at a single network level as well as at the level of cortical pathways. This work goes beyond current knowledge by proposing several alternative cognitive routes that might mediate imitation, in combination with sound experimental paradigms that can be used to validate or refute the proposed hypotheses. Another major difference of the approach adopted here is that it does not focus on the issue of learning. Indeed, the experimental results that the present work attempts to reproduce were often obtained in tasks which do not explicitly require any form of learning - that is, this work primarily addresses the automatic side of imitative behaviors.

As mentioned previously, two possible representations in the models were adopted in order to trigger imitative actions: one in which action representations were invariant, and one in which the system variables for both the demonstrator and the imitator were encoded within the same reference frame. It remains to be explained how the brain can compute such representations from its sensory inputs. Furthermore, the question of how discrimination between the self and others occurs has not been addressed here. Most of these models basically consider a two-phase processing scheme. While one of them drives the system in an action-recognition mode, the other triggers the reproduction of actions. Switching between modes is performed artificially. Self-related signals, as well as those generated by others, are usually considered identical. However, the nervous system is capable of disentangling these signals, which allows it to differentially recruit the shared representations, depending on the ownership of

the processed signals.

By evaluating the details of the computational tools adopted here, this work is more in line with that of Erlhagen, Mukovsky, and Bicho (2006) who used the dynamic neural field approach to model the components of their network. The models described in this thesis, which are used to investigate the neural processing pathways responsible for imitative behaviors, are also built using such neural network models. The next chapter therefore reviews the state of current knowledge about this class of models.

ARTIFICIAL NEURAL NETWORKS

THE general methodology adopted by computational modelers consists of taking inspiration from experimental data in order to produce models of the phenomenon in which they are interested. Since the principal aim of this thesis is to develop sound models of the neural mechanisms that underlie human imitation, artificial neural networks were naturally chosen as ideal computational tools for investigating the brain processes involved.

The present chapter describes the fundamental modeling tools adopted in this thesis. After a brief introduction related to computational neuroscience, a rapid description of the artificial neural network approach is provided - including models of single neurons as well as of some neural network models that were proposed during the past few decades. The core of this review defines and describes the class of neural models (neural fields) on which the work presented here is primarily based, with a focus on some of their important properties as well as their biological plausibility.

3.1 COMPUTATIONAL NEUROSCIENCE

Computational neuroscience is an multidisciplinary approach that links together diverse research fields such as neuroscience, cognitive science, computer science, physics and applied mathematics. It investigates and aims at understanding the mechanisms and functions of the nervous system using theoretical and computational methods. The models developed along this discipline primarily focus on replicating existing data by trying to stick as close as possible to biology. Consequently, they may provide predictions related to hypotheses that have not been experimentally tested yet, or that could not be verified in real biological systems. Computational neuroscience encompasses various subdisciplines whose respective research focus lies within different abstraction levels, depending on the primary aims of each study. Indeed, the modeling of the complex system which is the brain, has to be done through multiple levels, *the art of scientific explanations being to find the right level of abstraction rather than the detailed reproduction of nature* (Trappenberg, 2005). Given a phenomenon to explain, considering a too high-level approach may not allow to get the relevant mechanisms underlying it, but a too low-level approach may also be inappropriate.

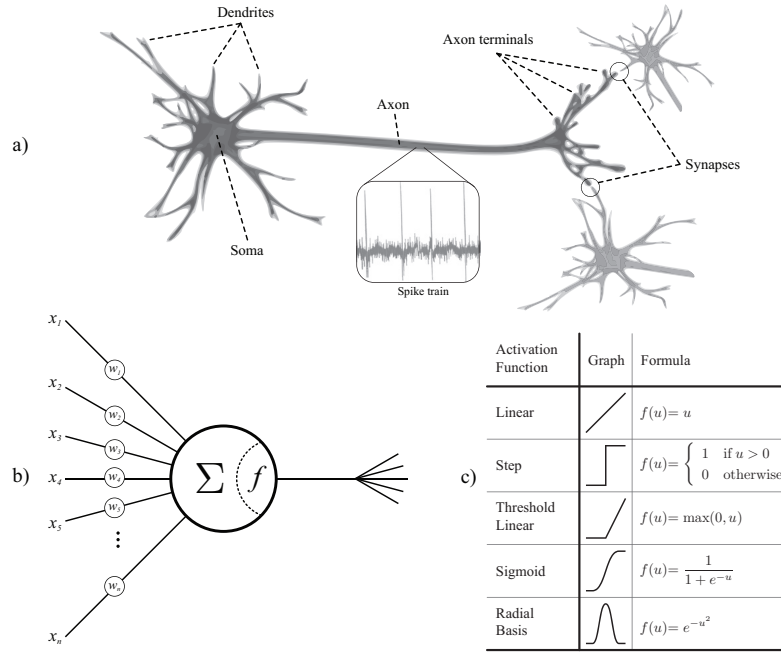


Figure 3.1: **a)** Illustration of a biological neuron and its subparts. **b)** An artificial neuron, as proposed by McCulloch and Pitts (1943), shares several structural and computational similarities with real neurons. **c)** Activation functions transforming the membrane potential of the artificial neuron into a firing rate as usually found in the literature (*Adapted from Trappenberg (2005)*).

ate since it may provide too much details, hence obscuring the important ones. For instance, in order to model the motion of an apple falling from a tree, it is unnecessary to apply the general rule of gravitation and the theory of relativity. The Newton's law is clearly more appropriate in this particular case.

As stressed above, computational neuroscience is deeply grounded on current neural evidence . This discipline tries to links the huge amount of experimental data together in order to provide a coherent view as to how the nervous system works. One particular trend in this field is to understand the dynamics of single neurons, and also how they communicate with their neighbors. Research studies attempt to shed light on the mechanisms underlying neural processing through realistic descriptions of the physiology and dynamics of neurons. These works try to capture the essential features of this biological system which require to consider membrane currents, protein and chemical coupling and the mechanisms of synaptic modification.

However, when coupling many neurons together, the amount of system parameters becomes huge. Therefore, simplified models of neurons have also been proposed in order to still be able to obtain a clear understanding of the neural dynamics at the neural ensemble level. Although the knowledge of the precise neural mechanisms occurring at the cellular level is fundamental, the present review of the artificial neural network literature will not address these low-level

processes. Instead, since the aim of this thesis is to understand how large cortical networks may be at the origin of high-level behaviors, this review will rather concentrate on artificial neural networks composed of simplified models of neurons which are suggested to still capture the essential dynamics of large neural ensembles. In the following, some important neural models will be described as well as the class of networks that was adopted in this thesis, namely, the Continuous Attractor Neural Networks (CANN) or Dynamic Neural Fields (DNF).

3.1.1 ARTIFICIAL NEURAL NETWORKS

The fundamental processing unit of the nervous system is the neuron. Neurons communicate with each other using action potentials or spikes, which are pulsed electrical signals. As illustrated in Figure 3.1a, a biological neuron is composed of several parts. The soma, or the cell body, is the core of the neuron. It integrates the electrical signals incoming from the other neurons through its dendrites. When a sufficient amount of spikes enters the neuron within a short period of time, the membrane potential of the neuron may reach its firing threshold. At this moment, the stored energy is transformed into a spike which is emitted by the neuron through its axon. When the spike reaches an axon terminal connected to the dendrite of another neuron, i.e., a synapse, the spike is transformed into a chemical signal which is in turn re-transformed into an electrical signal within the dendrite of the target cell.

The biochemical mechanisms of information transduction of neurons were incorporated in several detailed models of neurons (Hodgkin & Huxley, 1952; Gerstner & Kistler, 2002). These models of spiking neurons are currently studied with great attention since they may provide a better understanding of several fundamental mechanisms of the brain such as those mediating learning (Abbott & Nelson, 2000; Melamed, Gerstner, Maass, Tsodyks, & Markram, 2004). However, when considering large groups of neurons or cortical networks, computational modelers sometimes prefer to study the dynamics of neural networks with simpler models of neurons which consider the average firing rates of these cells rather than their individual spikes. Indeed, under certain conditions, a rate code (McCulloch & Pitts, 1943) has been shown to adequately represent the average response of an ensemble of neurons (Gerstner & Kistler, 2002). In this thesis, since the principal aim consists of understanding the mechanisms of a high-level cognitive function that is imitation, the simplification provided by the rate code model of the neurons was adopted.

The schematic of such an artificial neuron is shown in Figure 3.1b. It can receive several inputs denoted by x_1, x_2, \dots, x_n which are respectively modulated by synaptic weights w_1, w_2, \dots, w_n . The artificial neuron then sums all these weighted inputs into its membrane potential u . Finally, the output firing rate of the neuron depends on a function f of membrane potential, namely the

activation function. Formally, the output y of a neuron is given by

$$y = f(u) \quad \text{where} \quad u = \sum_{i=1}^n w_i x_i \quad (3.1)$$

or in a more compact vectorial form

$$y = f(\vec{w}^T \vec{x}) \quad (3.2)$$

where $\vec{w} = (w_1, w_2, \dots, w_n)^T$ and $\vec{x} = (x_1, x_2, \dots, x_n)^T$ correspond to the synaptic weights and input vectors. The activation function f has been defined in different ways. The most commonly used activation functions are given in Figure 3.1c. The choice of either function usually depends on problem to be solved and on the mathematical analysis to be performed. As can be seen from Equ. (3.1), a neuron alone can not perform any really complex computations. However, by combining and connecting neurons together, much more interesting properties can emerge from their interactions. Indeed, the computational outcome of neural networks can be greater than the sum of their parts.

Examples of Artificial Neural Networks

During the past decades, several neural network architectures which could be separated in several classes were developed. For instance, the well known multi-layer perceptrons belong to the class of feed-forward neural networks. Interestingly, these networks were shown to be capable to learn almost any non-linear mappings of their inputs. They could be seen as universal function approximators (Rumelhart, Hinton, & Williams, 1986; Hornik, Stinchcombe, & White, 1989). This important property has nowadays led multi-layer perceptrons to become a standard tool in computer science and engineering.

However, when trying to mimic brain neural processes, this type of neural networks were often criticized as not being biologically plausible. Indeed, real cortical networks are intensively recurrent. Furthermore, because of the existence of feedback loops within this class of network, different approaches, such as the dynamical systems theory, has mainly been adopted in order to understand them. Among this class of networks, Winner-Take-All were proposed, which, by means of recurrent inhibitory connections, were shown to exhibit several interesting properties. Of major interest, principal and independent component analysis of input data, (PCA) and (ICA) respectively, may be obtained by this network structure (Oja, 1992; Dumitru, King, Nandedkar, & Oja, 1997).

A similar network model also based on a winner-take-all architecture was proposed by Kohonen (1990), better known as the self-organizing map (SOM). Motivated by how sensory information such as visual information is represented and processed in the brain, SOMs are intended to produce a low dimensional representation of high-dimensional training data while preserving the topological properties of the input space. The interesting property of this model is that it

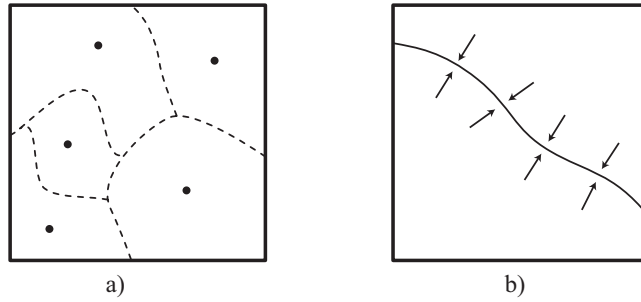


Figure 3.2: a) Point attractors and associated basins of attraction as may be found in associative neural networks. Each point would then correspond to a memorized input pattern toward which the network may converge according to its initial state. b) Continuous line attractor which could be produced by increasing the number of neurons in the network, as well as the number of point attractors. This form of neural dynamics is usually found in neural fields.

has been shown to reproduce patterns of neural sensitivity found in several areas of the brain such as the visual and motor cortices (Swindale & Bauer, 1998; Aflalo & Graziano, 2006b).

Another class of competitive recurrent networks are associative networks such as that proposed by Hopfield (1982). Associative networks are meant to store neural activation patterns in their recurrent connectivity by means of the Hebb's rule for synaptic modification (Hebb, 1949). The interesting feature in associative networks is that, from partial or noisy information, the system can dynamically retrieve the whole stored pattern corresponding best to the actual inputs.

From Associative Networks to Continuous Attractor Networks

Recurrent networks can be seen as dynamical system which may exhibit various behaviors such as attractor dynamics, limit cycles, and chaotic behavior. As illustrated in Figure 3.2a, in an associative network, patterns are ideally stored in the system as point attractors. A partial or noisy input pattern corresponds then to a different point in the state space, located in the basin of attraction of a given pattern. When the network evolves, its state simply moves toward the closest attractor. Interestingly, as the number of neuron in a Hopfield network increases, the number of possible attractor may increases. Therefore, it would be possible to build a continuous attractor such as that shown in 3.2b. This is exactly the approach followed by another class of models, namely the Continuous Attractor Neural Networks (CANN) or Dynamic Neural Fields (DNF). While point attractors may be useful for building discrete associative memories, continuous attractor network may be more useful for building short-term memories (Trappenberg, 2005). Indeed, neural fields can dynamically represent a continuum of states, and hence any variable value of a given space. By this attractor dynamics, the state of the actual input can thus be remembered as long as no external perturbation affects the system. This property of neural

fields is only one among many others. As will be mentioned in the next section, neural fields are powerful tools endowed with many fundamental properties for exploring the neural mechanisms and dynamics of the nervous system.

3.2 NEURAL FIELDS

Neural fields or continuous attractor neural networks (Wilson & Cowan, 1973; Amari, 1977; Erlhagen & Schöner, 2002; Trappenberg, 2005) share many similarities with all the three classes of recurrent networks mentioned above. As will be described further in more details, neural fields are an implementation of a winner-take-all mechanism, are topographically organized by means of a reciprocal connectivity between its elements, and, by their attractor dynamics, can be used as a form of memory. A key assumption made when modeling cognitive functions using neural fields is that a variable is encoded in the network in a distributed manner. All neurons in the field are preferentially tuned to a different instance of the variable represented in the network. Moreover, the recurrent connectivity of these networks have primarily a center-surround excitatory-inhibitory characteristic, where neurons sharing similar firing properties cooperate by exciting themselves and where neurons distant in their preferential tuning inhibit each other. This inter-neuron relationship is at the basis of their fundamental ability to allow one single localized activity packet, also known as bump or pulse solutions, to emerge from this neural implementation of a winner-takes-all operation (Wilson & Cowan, 1973; Amari, 1977; Erlhagen & Schöner, 2002; Trappenberg, 2005).

In the past decades, these mathematical tools were increasingly used in order to model and implement brain cognitive functions. Indeed, in addition to their biologically plausible structural relationship with real cortical neural ensembles, their numerous computational properties have made them very attractive to the computational neuroscience community. For instance, these models, originally proposed to explain the formation patterns of cortical representations, were then applied to other research topics such as visual processing (Ben-Yishai et al., 1995; Mineiro & Zipser, 1998; Giese, 2000), visual attention (Rougier, 2006; Vitay & Rougier, 2006), spatial navigation (Zhang, 1996; Xie et al., 2002; Redish, 1999; Stringer et al., 2004), memory (Compte et al., 2000; Rolls et al., 2002), motor control (Kopcz & Schöner, 1995; Lukashin et al., 1996), decision making (Kopcz & Schöner, 1995; Erlhagen & Schöner, 2002; Schöner, 2002), sensorimotor transformations (Salinas & Thier, 2000; Deneve et al., 2001; Meñard & Frezza-Buet, 2005), stimulus binding (Wersing et al., 2001), parameter estimation (Deneve et al., 1999; Pouget et al., 2003), and even higher-level cognitive functions such as imitation (Andry et al., 2004; Erlhagen, Mukovskiy, et al., 2006; Erlhagen, Mukovsky, & Bicho, 2006).

In the following, after a rapid mathematical formulation of the structure and

dynamics of neural fields, a review of their principal properties will be given. The aim of this section is to provide the reader with a flavor of the powerful tools for cognitive modeling that are the continuous attractor networks. Rather than being too formal, this review is intended to describe qualitatively the numerous computational properties embedded in this class of neural networks. The reader interested in the proofs of the mathematical properties presented here is instead suggested to refer to the associated literature.

3.2.1 DEFINITION

Neural fields are recurrent neural networks developed in order to address neural computations in the continuous domain. Although neurons are discrete elements, when an important number of cells is considered, a mathematical approximation in the continuous domain has been shown to be an interesting simplification for understanding the neural dynamics of large networks (Amari, 1977; Gerstner & Kistler, 2002; Wilson & Cowan, 1973). In addition, neural fields were primarily designed using a simplified model of neuron, similar to that presented earlier, which considers the average firing rate as a means to transmit information (McCulloch & Pitts, 1943). However, since this class of model also assumes a continuous-time integration of information rather than a discrete time one, the model of neuron used in this class of models is slightly different. Instead of integrating instantaneously their external inputs, the cells do it progressively through time according to their membrane time constant. As a consequence, when a constant input is applied to a neuron, it takes a non-negligible amount of time for the neuron to integrate its input. This type of artificial neuron was named a *leaky-integrator* neuron, which should not be confused with the leaky-integrate and fire model of spiking neurons. Basically, the dynamics of a leaky-integrator neuron can be written as follow:

$$\tau \dot{u}(t) = -u(t) + x(t) \quad (3.3)$$

where $u(t)$ corresponds to the state of the membrane potential of the neuron, τ to its time constant, and $x(t)$ to the global external inputs entering the cell.

As mentioned earlier, a dynamic neural field is a topographically organized network. By topographically organized, it means that the cells composing the neural field share a preferential tuning to a set of parameters. Moreover, each neuron is preferentially tuned to a specific instance of these input parameters, often called the preferred direction of the neuron. Importantly, the preferred directions of all the cells are uniformly distributed in the input space. A continuous neural population is defined by an ensemble of neurons where each cell is preferentially tuned to a preferred direction \vec{r} uniformly distributed along Γ , which corresponds to the parameter space such that $\vec{r} \in \Gamma$. Specific examples of possible parameter spaces Γ will be given further in the text.

A continuous attractor neural network is also characterized by the strong recurrent interactions among its elements. These interactions follow a center-surround excitation-inhibition pattern where neurons sharing a similar preferential tuning tend to excite each others, whereas inhibitory interactions take place between neurons having a large difference in their preferential tuning. In addition, this interaction kernel must be invariant across all neurons. Indeed, this condition ensures that the attractor formed by the network is continuous. Non-uniformity of the recurrent usually produces discrete attractors such as those found in Hopfield networks. Several interaction kernels have been used in the literature. Importantly, gaussian and mexican-hat shaped kernal functions were considered. These two weight functions are illustrated in Figure 3.3 and can respectively be written as

$$W^R(\vec{r}', \vec{r}) = \begin{cases} \alpha \left[\exp\left(-\frac{\|\vec{r}-\vec{r}'\|^2}{2\sigma^2}\right) - \delta \right] & \text{for a gaussian kernel} \\ \alpha \left[\exp\left(-\frac{\|\vec{r}-\vec{r}'\|^2}{2\sigma_E^2}\right) - \alpha_I \exp\left(-\frac{\|\vec{r}-\vec{r}'\|^2}{2\sigma_I^2}\right) - \delta \right] & \text{for a mexican hat kernel} \end{cases} \quad (3.4)$$

$W^R(\vec{r}', \vec{r})$ denotes the synaptic projection from the neuron with preferred direction \vec{r}' to that sensitive to \vec{r} . The index R stands for recurrent. α and α_I are amplification factors and δ an offset, allowing the weights to have an inhibitory component. In the gaussian kernel, σ corresponds to the breadth of the interaction weights, whereas σ_E and σ_I in the mexican-hat kernel are, respectively, the breadth of the excitatory and inhibitory gaussians forming the hat.

Practically, however, the use of preferred direction distributed on an infinite space is problematic since computer systems have limited resources. Moreover, boundary effects also appear if the space of the parameters has limits. Indeed, the recurrent interaction must be invariant so that the possible attractors of the network form a continuum. Imposing limits on the neural space breaks this symmetry, which usually results in stronger attractors at the boundaries. An usual solution to these problems is to adopt periodical parameter spaces¹. The commonly used limited and periodical manifolds may be circles, hyperspheres and N -dimensional torus. Therefore the equation of the weights can be rewritten by using a periodical gaussian function, better known as the Von Mises distribution, so that the weight function is always continuous and differentiable. In addition, by assuming that the parameter space consists of unitary vectors, i.e., $\forall \vec{r} \in \Gamma, \|\vec{r}\| = 1$, the dot product can be used as a distance function rather than the typical euclidian distance. Before rewriting the synaptic weights, let define the parameter space Γ more precisely. The simplest form of periodical

¹It is important to note that any kind of information can be encoded in a periodical space Γ , since a unidimensional variable λ defined in an arbitrary domain can be transformed into the considered space using a mapping function $m(\lambda) \rightarrow \vec{r}$ such that $\vec{r} \in \Gamma$ and $\|\vec{r}\| = 1$.

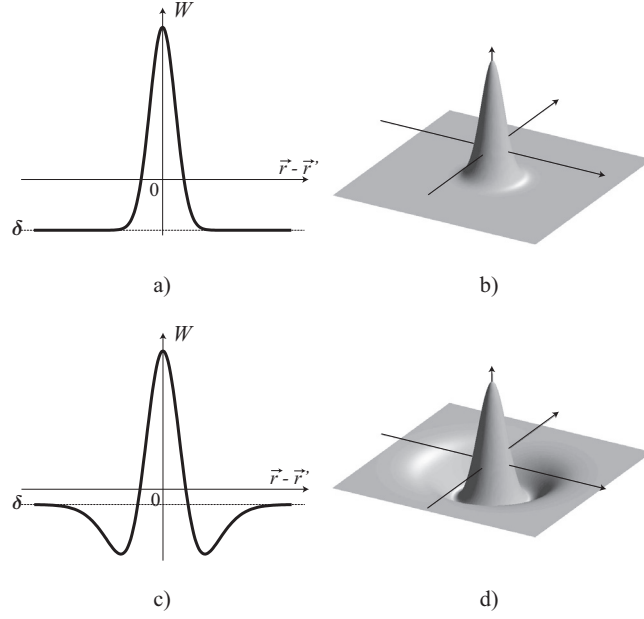


Figure 3.3: The interaction kernels encoded in the recurrent synaptic weights are symmetric and of center-surround type. **a)** A gaussian-like weight kernel in one and **b)** two dimensions. **c)** A mexican-hat like weight kernel in one and **d)** two dimensions.

closed manifold consists of the unitary hyper-sphere of dimension N . In this case the preferred directions are distributed within

$$\Gamma = \{ \vec{r} \in \mathbb{R}^{(N+1)} \mid \|\vec{r}\| = 1 \} \quad (3.5)$$

When $N = 1$, this manifold is simply a circle. Next, in order to follow the same notation to define a toroidal space of dimension N , the preferred directions \vec{r} have to be decomposed into N sub-vectors such that $\vec{r} = (\vec{r}_1, \dots, \vec{r}_N)$. Then, the N -dimensional toroidal space can be defined as

$$\Gamma = \{ \vec{r} = (\vec{r}_1, \dots, \vec{r}_N) \mid \vec{r}_i \in \mathbb{R}^2 \text{ and } \|\vec{r}_i\| = 1 \} \quad (3.6)$$

As can be noticed, when $N = 1$, this manifold is also equivalent to a circle. Next, the periodical gaussian function $\mathcal{G}(\vec{r}', \vec{r}, \sigma)$ which defines the interaction kernel having breadth σ , between neurons preferentially tuned to \vec{r}' and \vec{r} , respectively, is given by

$$\mathcal{G}(\vec{r}', \vec{r}, \sigma) = \begin{cases} \left[\exp\left(\frac{\vec{r} \cdot \vec{r}' - 1}{2\sigma^2}\right) - e^{-1/\sigma^2} \right] / (1 - e^{-1/\sigma^2}) & \text{if } \Gamma \text{ is an hyper-sphere} \\ \prod_{i=1}^N \left[\exp\left(\frac{\vec{r}_i \cdot \vec{r}'_i - 1}{2\sigma^2}\right) - e^{-1/\sigma^2} \right] / (1 - e^{-1/\sigma^2}) & \text{if } \Gamma \text{ is an } N\text{-dimensional torus} \end{cases} \quad (3.7)$$

Moreover, this function is normalized such that $\mathcal{G}(\vec{r}', \vec{r}, \sigma) \in [0, 1]$. The recurrent

weights can here be expressed as a function of

$$W^R(\vec{r}', \vec{r}) = \begin{cases} \alpha [\mathcal{G}(\vec{r}', \vec{r}, \sigma) - \delta] & \text{for a periodical gaussian kernel} \\ \alpha [\mathcal{G}(\vec{r}', \vec{r}, \sigma_E) - \alpha_I \mathcal{G}(\vec{r}', \vec{r}, \sigma_I) - \delta] & \text{for a periodical mexican hat kernel} \end{cases} \quad (3.8)$$

Now that the recurrent interaction weights were defined, let consider the differential equation governing the dynamics of neural fields. It is given by

$$\begin{aligned} \tau \dot{u}(\vec{r}, t) = & -u(\vec{r}, t) + \oint_{\Gamma} W^R(\vec{r}', \vec{r}) f(u(\vec{r}', t)) d\vec{r}' + \\ & + x(\vec{r}, t) + h(t) \end{aligned} \quad (3.9)$$

where $u(\vec{r}, t)$ is the membrane potential of the neuron with preferred direction \vec{r} at time t . The non-linear activation function f can be any of those mentioned in Figure 3.1c except the radial basis function. $x(\vec{r}, t)$ corresponds to the external input and $h(t)$ to a global homogeneous input or resting potential. This non-linear dynamical equation has many interesting properties for modeling cortical cognitive functions which are described next.

3.2.2 PROPERTIES OF NEURAL FIELDS

The development of the neural field approach started from the evidence that sensory and motor information is represented in a distributed fashion within populations of neurons influenced by an important amount of recurrent connections (Douglas, Koch, Mahowald, Martin, & Suarez, 1995). This computational paradigm has been shown to be shared by several areas of the nervous system, including proprioceptive receptors, such as muscle spindles (Ribot-Ciscar et al., 2003), the motor cortex (Schwartz et al., 1988; Kakei et al., 1999; Cisek & Kalaska, 2005), the visual cortex (Bienenstock, Cooper, & Munro, 1982), the posterior parietal cortex (Batista et al., 1999; Scherberger & Andresen, 2003) and the superior temporal sulcus (Ashbridge et al., 2000). This widespread network structure has then be suggested to be a general architecture providing high-level computational abilities to cortical neural ensembles. In the following paragraphs, the most relevant properties of this class of network as well as their implication for neurobiology are reviewed. But before starting, although most of the properties, if not all, will be illustrated using a one-dimensional parameter space Γ , it is worth noting that they are nevertheless still valid in higher dimensions (Taylor, 1999; Laing & Troy, 2003).

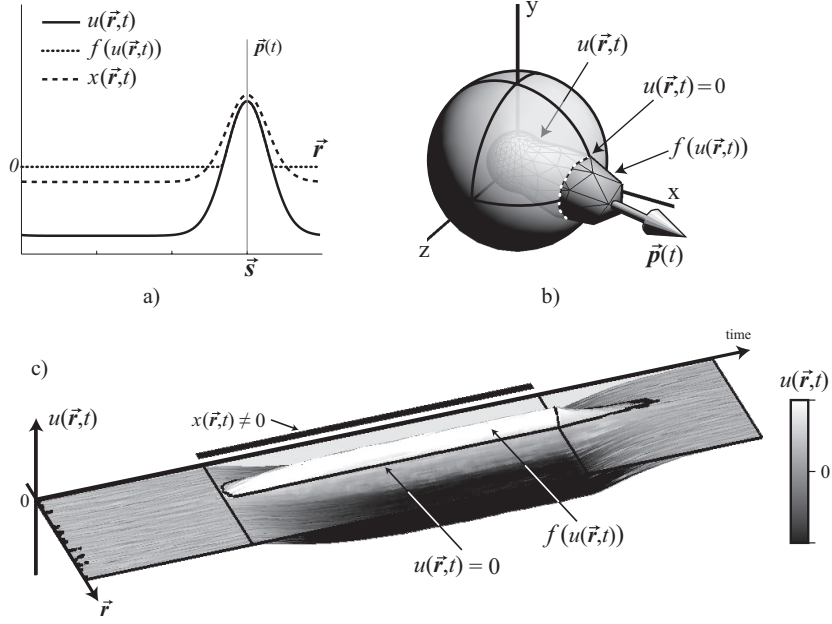


Figure 3.4: **a)** Illustration of the resulting activity of a neural field which receives an external input spatially localized at \vec{s} . In this example, the neural space Γ corresponds to a circular space. For convenience, the visualization of this space is shown on a line where both extremities should be seen as interconnected. Superimposed on this figure, the population vector of the neural activity profile is also shown. **b)** The same information than that shown in **a)** is displayed, but where the space in which the network is defined corresponds to the surface of a sphere. This spherical representation may be seen as a two dimensional surface which has been folded around a sphere. **c)** Temporal evolution of the membrane potential of a circular neural field when a spatially localized external input is applied. Because the activation function f was chosen to be the linear threshold function, the region which is over zero corresponds to the actual firing rate of the neurons. In order to help reading the figure, black contours were drawn around the regions of positive activation, i.e., along the lines where $u(\vec{r}, t) = 0$. Finally, the black strip indicates the period of time where the external input was presented.

SINGLE CUE INTEGRATION

As described in Equ. (3.9), a neural field can receive an external input $x(\vec{r}, t)$. Usually, the considered external inputs encode a variable $\vec{s} \in \Gamma$ by means of a representation having a shape similar to the recurrent connectivity profile. In the present review, since periodic representation are assumed, an external input is written as a periodic gaussian function of \vec{r} similar to that given in Equ. (3.8). It gives

$$x(\vec{r}, t) = \beta(t) \left[\mathcal{G}(\vec{s}, \vec{r}, \sigma) - \delta \right] \quad (3.10)$$

where $\beta(t)$ is the input amplitude, and δ is a constant usually set so that $x(\vec{r}, t)$ is a zero-mass distribution², i.e., $\oint_{\Gamma} x(\vec{r}, t) d\vec{r} = 0$. As illustrated in Figure 3.4,

²The external input $x(\vec{r}, t)$ is considered to have a zero-mass distribution in order to clearly

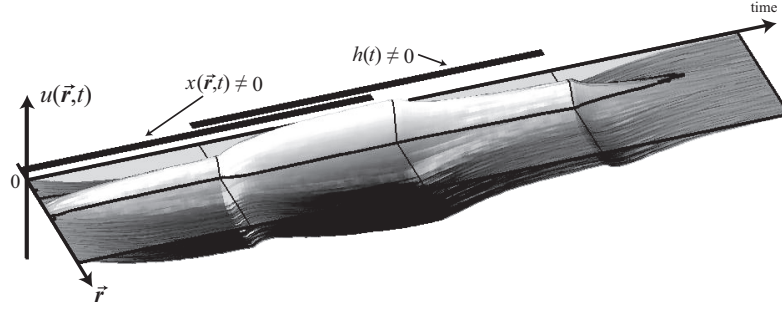


Figure 3.5: Temporal evolution of a neural field activity profile when the background input is increased above the threshold which allows the network to produce a self-sustained memory of the last seen input. The time during which the external input $x(\vec{r}, t)$ and the background input $h(t)$ were presented to the network are represented by the two black strips on the left of the time axis. The same graphical representation as in Figure 3.4c was used here.

an external input produces a unimodal increase of the population potential. Since its shape is similar to the recurrent weights profile, it matches an element of the continuous memory pattern of the continuous attractor neural network. Note here that in almost all examples provided in this chapter, the activation function f which was used consists of the linear threshold function defined as $f(y) = \max(y, 0)$.

Decoding Population Codes: The Population Vector

As shown in Figure 3.4, the neural dynamics of neural fields is known to form an attractor bump on the surface of the neural field through which this class of networks is suggested to convey information. A typical read-out mechanism, the *population vector* $\vec{p}(t)$, has been suggested to measure the macroscopic effect of the joint activities of large sets of neurons on ensemble of real neurons (Georgopoulos, 1996). It consists of a weighted summation of the firing activity of each neuron with its preferred direction. Through the similarity of the representation of neural fields with groups of biological neurons, the same decoding mechanism is often applied to this class of models. The estimate $\vec{p}(t) \in \Gamma$ of the variable value encoded in the neural field can be written as

$$\vec{p}(t) = \oint_{\Gamma} f(u(\vec{r}, t)) \vec{r} d\vec{r} / \left\| \oint_{\Gamma} f(u(\vec{r}, t)) \vec{r} d\vec{r} \right\| \quad (3.11)$$

Since the distribution of the preferred directions is assumed to be uniform, it is not necessary to consider more complex decoding techniques such as those using optimal linear estimators as proposed by Salinas and Abbott (1994); Deneve et al. (1999). In Figures 3.4a and b, where the activity of neural fields receiving an unimodal external input is illustrated, the resulting population vector is also

distinguish the effects produced by the two types of input found in the equation describing the network's dynamics (Equ. (3.9)), i.e., $x(\vec{r}, t)$, and $h(t)$ which is constant along the whole network. This is more a modeling constraint rather than a biological one.

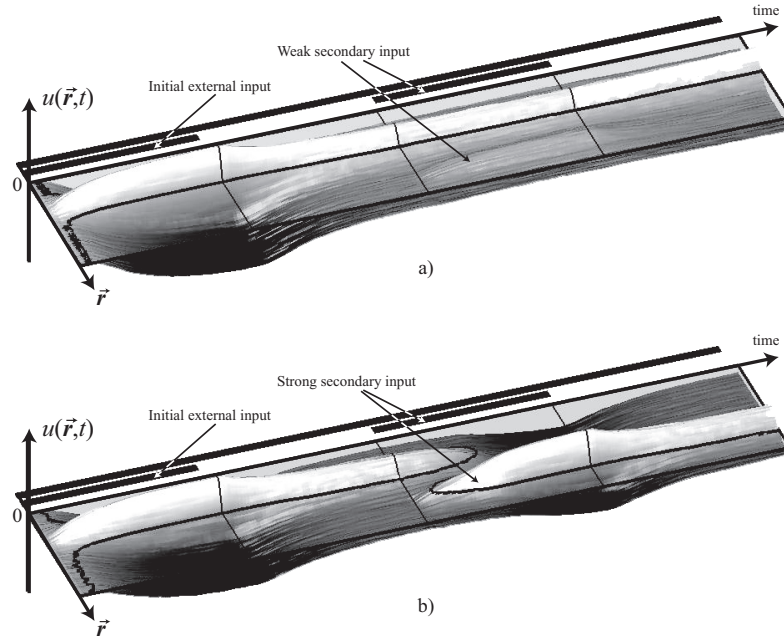


Figure 3.6: Temporal evolution of the network’s activity profile when an input encoding a value different from that currently self-sustained is applied. **a)** The external input is not strong enough to modify the internal representation. **b)** A stronger external input can however update the internal memory of the network. The strength of the input which is necessary to update the memory usually depends on the parameters of the network. In this case, it simply needs to be capable of inducing a positive activation packet on the surface of the field. The same graphical representation as in Figure 3.4c was used here.

shown.

Self-Sustained Memory

Following from the internal dynamics of the neurons and their recurrent connectivity, an input imposed to the field is not instantaneously integrated by the cells. Similarly, when removing an external input, it takes time for the network to relax to its resting activity. While this sort of phasic memory effect can be observed in Figure 3.4c, an active memory maintenance can also be realized by continuous attractor neural networks. As may have been noticed, the effect of the second external input $h(t)$ has not been described, yet. By setting this external input, which will be called the homogeneous or the background input, above a certain threshold depending on the model parameters, a self-sustained memory of the last presented input variable can be observed (Salinas & Abbott, 1996; Compte et al., 2000). This pattern of neural activation is shown in Figure 3.5, where an external input is temporarily applied to the network. By increasing the background activity, the neurons maximally firing for the given input keep firing even in its absence. Interestingly, according to neurophysiological evidence, this mechanism mediating short-term memory has been suggested to be actually found in the brain (Compte et al., 2000)

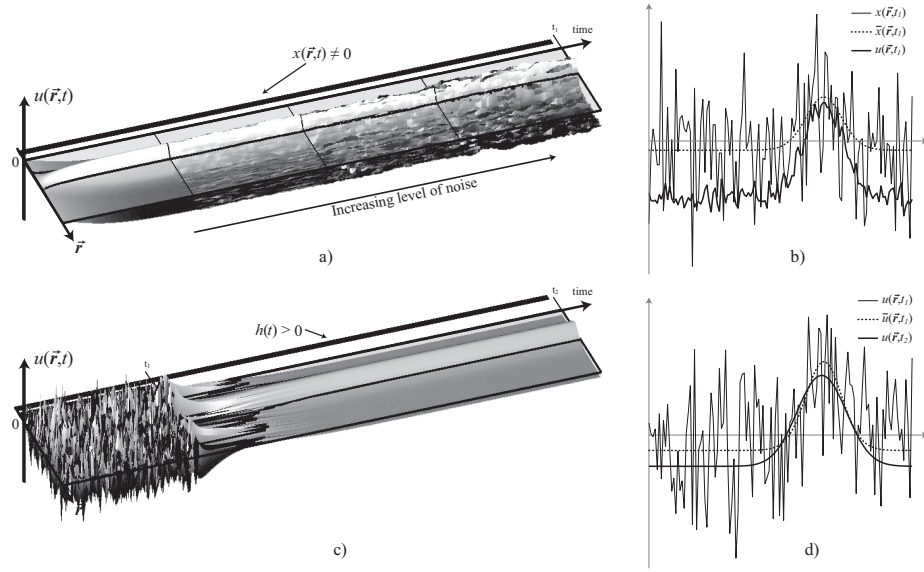


Figure 3.7: Neural fields have an intrinsic ability to reduce noise conveyed by their inputs. The same graphical representation as in Figure 3.4c was used here. **a)** Although an increasing amount of noise is applied to the network’s input, the neural field can still focus on the relevant information transmitted by its input. **b)** A snapshot at time t_1 shows that the noise influencing the internal representation $u(\vec{r}, t_1)$ is low as compared to that present in the input $x(\vec{r}, t_1)$. Moreover, the network faithfully keeps locked on the noise-free original signal $\bar{x}(\vec{r}, t_1)$. **c)** When the membrane potential of the neurons initially representing a bell-shaped signal corrupted by an important amount of noise relaxes, the network converges almost toward the original signal. **d)** As described by Deneve et al. (1999), the neural field computes an nearly optimal maximum likelihood estimation of the clean signal. Indeed, as can be seen, the membrane potential at time t_2 is almost equal to that at time t_1 without noise.

Updating the Internal Representation

Next, although an active memory mechanism is useful for guiding behaviors going beyond a simple input-output information processing scheme, one has to consider how this memory could be updated with new information and how it may resist to eventual distractors. The intrinsic dynamics of continuous attractor networks provide an extremely powerful embedded solution to these issues. When the external input is removed and the neural field sustains the information conveyed by that input, an hysteresis-like dynamics can be observed. While an extremely small input (and even noise) can activate the memory, in order to modify the memorized internal representation, an input stronger than a threshold determined by the parameters of the system must be applied to the network (Compte et al., 2000; Erlhagen & Bicho, 2006). As mentioned above, the self-sustained memory mechanism is mediated by the background homogeneous input $h(t)$. The minimal strength of the input which is necessary for updating the information encoded by the network, is approximatively proportional to that of the background homogeneous input. Figure 3.6 shows two examples where a spatially localized input is applied to a neural field which is al-

ready self-sustaining a previously memorized input. The first illustration shows that a weak input is not sufficient to affect the memory, whereas the second example depicts a successful update of the representation of the network.

Noise Reduction

The recurrent cortical architecture observed within neuronal ensembles is believed to amplify and stabilize noisy or corrupted feed-forward synaptic input. Computational studies addressing this hypothesis using neural fields have shown that this property can effectively be reproduced in simulation (Salinas & Abbott, 1996; Deneve et al., 1999). When a noisy spatially localized input enters the neural field, the dynamics produced by the recurrent connectivity drives the network toward one of its attractors. Since these attractors are not or less noisy, they compete with the external input and hence reduce its noise. An illustration of this form of filtering is shown in Figures 3.7a and b.

Following from this intrinsic network property, Deneve et al. (1999) showed that the noise reduction ability of this class of neural structure can be near-equivalent to a maximum likelihood operation. An illustration of their finding can be observed in Figures 3.7c and d. A noisy gaussian-shaped input is fed to the network by means of a direct assignation of the membrane potential of the neurons. Then, the input is removed and the network relaxes to one of its predefined attractors. As shown in Figure 3.7d, the final profile of the neurons' membrane potential corresponds quasi-perfectly to the original input without noise. Consequently, this form of recurrent neural dynamics was suggested to exist throughout the cortex so that neural ensembles could estimate in a near optimal way the value of the variables encoded in the noisy activity which they receive from other brain areas³ (Deneve et al., 1999; Wu, Amari, & Nakahara, 2002). Although the relaxation phase described above is not really plausible, Deneve et al. (1999) pointed out that the recurrent process could be implemented by a series of networks which are connected in a feed-forward manner, hence capable of mimicking recurrent dynamics.

Non-Uniform or Noisy Distribution of Preferred Directions

In the previous paragraphs, noisy inputs were considered and neural field were described to be capable to reduce this noise by means of their recurrent connectivity. However, this connectivity was assumed to be free of noise. Indeed, a continuum of attractors in a recurrent network only exists if the weight interaction kernel is shift invariant. It has been shown that a small deviation from this hard constraint produces drifts of the population response, especially when the network is used as a self-sustaining memory (Zhang, 1996; Compte et al., 2000; Stringer, Trappenberg, Rolls, & Araujo, 2002). As shown in Figure 3.8, noise in the synaptic weights creates point attractors instead of the continuous attractor manifold. In this example, the network, built on noisy

³In the work of Deneve et al. (1999), the noise was drawn from a gaussian distribution.

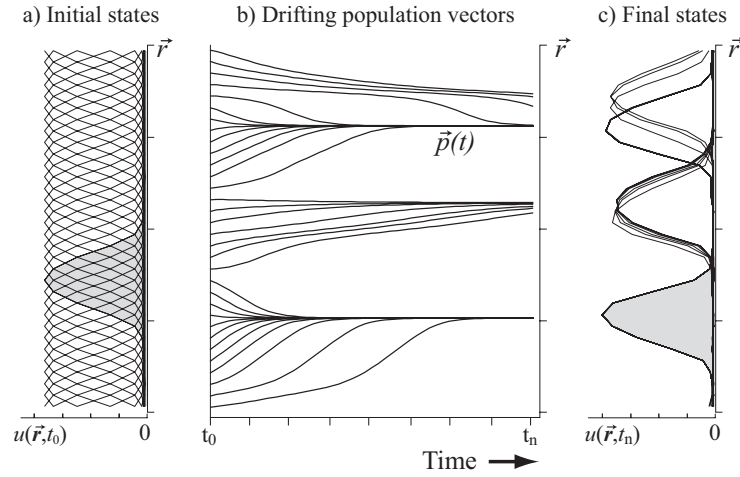


Figure 3.8: Effects of noisy recurrent weights on the attractor dynamics. **a)** Initial states which were imposed to the network. **b)** The temporal evolution of the population vector corresponding to each initial states shows that noisy weights break the continuum of attractors by producing discrete attractors shown in **c)**. (*Adapted from Zhang (1996)*)

recurrent weights, was initialized with a self-sustained activity bump and then evolved without any external influences. For several trials, the temporal evolution of the population vector was recorded and then displayed as lines shown in Figure 3.8b. As can be seen, rather than keeping representing the original information, the network activity packet drifts and converges toward one of the point attractor produced by the noisy weights. Interestingly, such dynamical phenomena has been reported in the behavioral literature when addressing the cognitive processes of working memory (Schutte, Spencer, & Schöner, 2003). Nevertheless, this drifting is still a problematic issue, which needs be resolved. Biologically plausible solutions to this problem of learning an uniform mapping of sensory receptors to a neural representation have nevertheless been proposed by Compte et al. (2000) and Stringer et al. (2002). The developed methods mainly consist either of modifying the recurrent connectivity by an appropriate learning rule, or by upgrading the basic model of the artificial neuron with an adaptation mechanism.

INTEGRATION OF MULTIPLE CUES

Until now, only unimodal external inputs were considered. It is however interesting to address the case of several distinct inputs. Indeed, neurophysiological studies have reported that bimodal activity profiles can be observed within populations of neurons (Amirikian & Georgopoulos, 2000; Cisek & Kalaska, 2005). This evidence suggests that a neural ensemble may process simultaneously different inputs such as those arising from different modalities or multiple sensory cues (Pouget, Deneve, & Duhamel, 2002; Cisek & Kalaska, 2005).

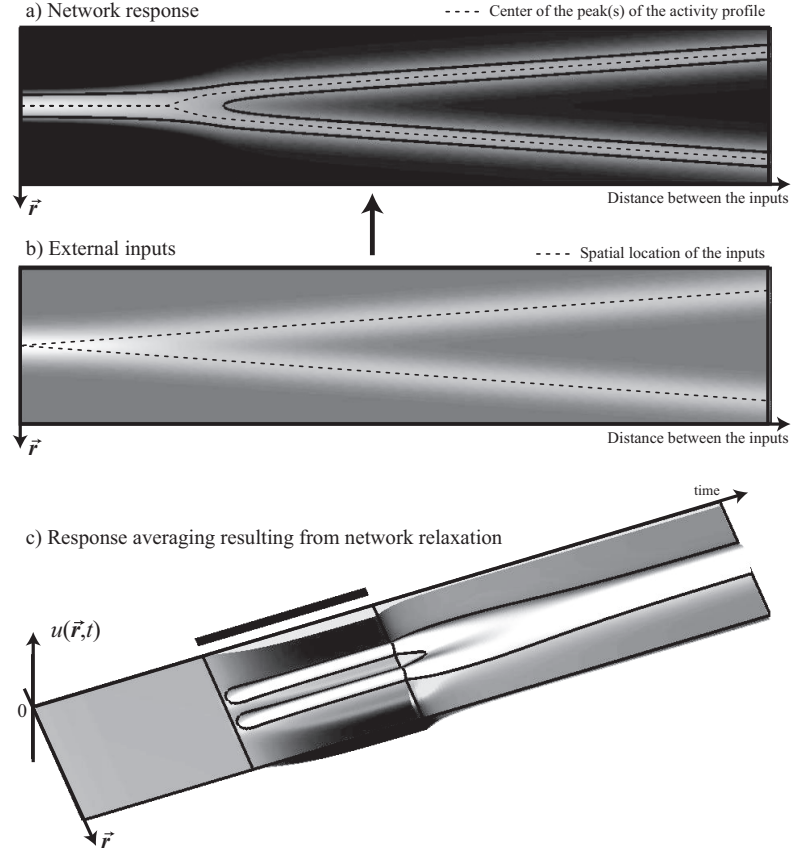


Figure 3.9: Fusion of two distinct input signals located at various distances in the neural space. **a)** A typical response of a continuous attractor network to the inputs pattern shown in **b)**. As can be noticed, when the two inputs get sufficiently close, the neural response indicates that the network can not discriminate them anymore. The inputs are fused. **c)** A neural field can also estimate the signal averaging those conveyed by two inputs. This estimation is performed by a relaxation process where the inputs are first presented to the network and then removed. The final self-sustaining activity profile represents their average, or, as suggested by Pouget et al. (2003), corresponds to an optimal bayesian estimation of the variable encoded by the two signals, given that they do represent the same but noisy information. This mechanism is however valid only for sufficiently close inputs.

Representing Multiple Cues: Separation and Fusion

In order to represent distinct external inputs within a neural field, Equ. (3.10) has to be rewritten so that several bell-shaped inputs corresponding to different variables $\vec{s}_1, \dots, \vec{s}_n \in \Gamma$ can be represented. It gives

$$x(\vec{r}, t) = \sum_{i=1}^n \beta_i(t) \left[\mathcal{G}(\vec{s}_i, \vec{r}, \sigma) - \delta \right] \quad (3.12)$$

where n denotes the number of distinct inputs. Because of the gaussian-like distributed encoding of a single variable, the distance in the neural space Γ between two of them as well as the breadth σ of their representation, play an important

role regarding how these inputs are integrated in the network. Figure 3.9b first shows that when two inputs, equivalent in strength, get sufficiently close, their respective representation starts to overlap and may even fuse (Schöner, 2002). In addition, when these inputs are projected into the neural field, the recurrent interactions, which tend to find the internal attractor matching best the network's inputs, may even merge the inputs for larger distances (See Fig. 3.9a). Indeed, the closeness between two inputs depends on the breath of their respective representation as well as that of the interaction weights. For a given distance in the neural space, when the breadth of the input representation gets larger, the neural field dynamics considers them as getting closer. Therefore, the number of inputs which can simultaneously be presented to a neural field, while being represented distinctively, clearly depends on the breadth of the inputs.

Furthermore, when two close inputs are fed to the network and then removed, the neural fields' relaxation dynamics tend to average the spatial location of these inputs. This effect, illustrated in Figure 3.9c, can also be applied to non-identical inputs having a different profiles in breath and amplitude. Under certain conditions, the activity profile of the relaxed network can be seen as an optimal bayesian estimator of a variable given different probability distributions provided by each external inputs (Pouget et al., 2003).

Competition, Selection and Decision

As already well described in the literature (Amari, 1977; Salinas & Abbott, 1996; Erlhagen & Schöner, 2002; Erlhagen & Bicho, 2006), given sufficiently strong gaussian-shaped recurrent interactions, distinct inputs may compete within the neural surface of the network. This competition usually results in a selection process where only one input wins against the others for being represented on the active surface of the neural field. This property may be seen as a form of decisional process implemented neurally (Salinas & Abbott, 1996; Erlhagen & Schöner, 2002). Figure 3.10a illustrates the dynamics of a neural field in which two distant inputs compete for being selected. This example shows the unstability of the network when two symmetric inputs are applied. Indeed, a small amount of noise added to the membrane potential of the neurons can break this symmetry, hence letting the network choose randomly one of the two inputs.

In general, this form of selection process is assumed to be successful or completed when the activity of either the whole network, or the neural region surrounding the winning bump exceeds a given threshold E_0 . This quantity is defined here as the energy $E(t)$ of the network response. Formally, the selection is completed when

$$E(t) = \oint_{\Gamma} f(u(\vec{r}, t)) \, d\vec{r} > E_0 \quad \text{global network activity} \quad (3.13)$$

$$E(t) = \oint_{\Gamma} f(u(\vec{r}, t)) \vec{r} \cdot \vec{s} \, d\vec{r} > E_0 \quad \text{local network activity} \quad (3.14)$$

In the second case, $\vec{s} \in \Gamma$ denotes the location of the winning input.

Use of Prior Information for Biasing Selection

A random selection process may be disadvantageous in many behavioral conditions. Animals normally take advantage of prior information in order to bias and accelerate their decisions. Commonly applied methods for biasing the selection process consist of either increasing the strength of the input to be chosen, or pre-activating the neural region, i.e., the variable instance in Γ , having the highest probability to be selected. This pre-activation is done by applying a sub-threshold input at that precise location. Figures 3.10b to d show how this biasing method may accelerate the selection process toward the pre-cued external input.

Neural correlates of this computational mechanism where potential choices pre-activate the ensemble of neurons preferentially tuned to these choices have been recently reported (Cisek & Kalaska, 2005). The use of multi-electrode recoding techniques allowed to monitor simultaneously an important number of neurons simultaneously in the premotor and motor cortex, while a monkey was performing a delay-reaching task. As illustrated in Figure 3.10e, by sorting the neurons according to their preferred direction of movement, Cisek and Kalaska (2005) were able to illustrate the spatiotemporal neural activation patterns in a way similar to that used in the neural field literature. In addition, a model composed of continuous attractor neural networks neural fields has been shown to be capable of reproducing such activation patterns (Cisek, 2006).

Effect of the Representation Metric

The importance of the metric of the neural representation is now considered, especially in the case of a competition between two different inputs. When two competing inputs encoding respectively two variables \vec{s}_1 and $\vec{s}_2 \in \Gamma$, their distance in the neural space⁴ is a fundamental parameter influencing the time needed for the selection (Erlhagen & Schöner, 2002). As shown in Figures 3.11a and b, during the decisional process, the time course of the peak of the population activity profile highly depends on the distance between the competing inputs. It rises faster for close inputs and slows down according to the distance between these inputs. As already mentioned above, another aspect of the input signals also influences the speed of the selection, namely their amplitude or strength. Indeed, in order to be sure that a given input will be chosen by the network, it has to be the strongest. As could be easily deduced, the larger is the strength difference between the competing inputs, the faster is the selection. Therefore, both the distance and the difference in amplitude are important factors. Figure 3.11c illustrates how the time course of the selection process is influenced by these two factors, i.e., the distance metric and the amplitude of the

⁴Recall that the distance metric is primarily defined by the breadth of the interaction kernel, as well as the breadth of the input representations.

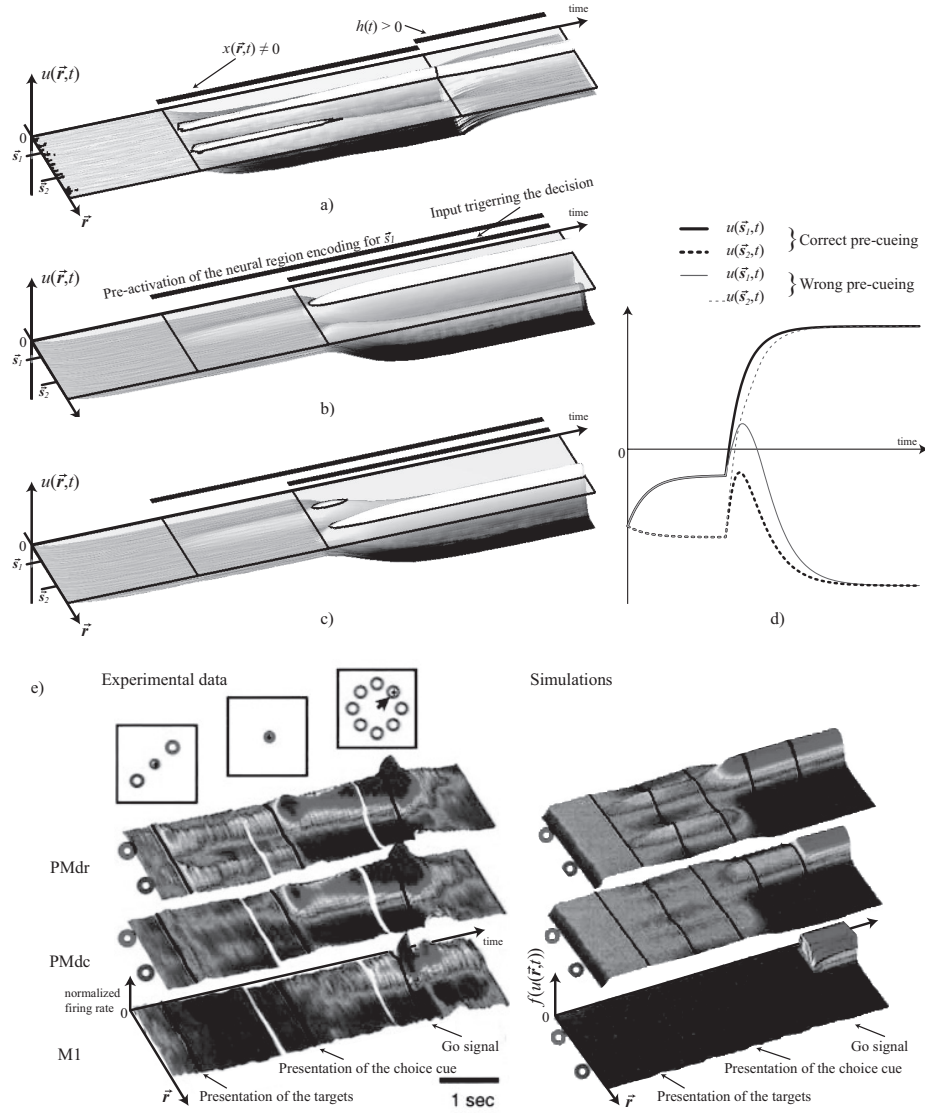


Figure 3.10: Selection of an input among two of them. **a)** Two symmetric inputs are presented to the network. Since some white noise was introduced in the membrane potential of the neurons, the symmetry is rapidly broken, which results in a random selection. At the end of the trial, the inputs are removed and the signal conveyed by the winning input is sustained. **b)** The same symmetric input signals are also fed to the network, but a bias was introduced in order to facilitate the selection of the one representing the variable \bar{s}_1 . This bias takes the form of a pre-activation of the neural region corresponding to that variable. **c)** In this example, the external signals are asymmetrically balanced, so that the input representing \bar{s}_2 is selected by the neural field. The final asymmetry in amplitude combining both the external inputs and the pre-activation is equivalent to that observed in **a)**. **d)** This plot summarizes the selection process observed in **b)** and **c)**. The time course of the membrane potential of the neurons respectively tuned to the value represented by the input variables \bar{s}_1 and \bar{s}_2 is shown. As can be noticed, a correct pre-activation of the further selected input accelerates the network response. **e)** (Left) Neural activity recorded in three motor brain areas of a monkey during a two-target choice task. On top, the stimuli viewed by the monkeys are shown. The two circles drawn on the activity plots indicate the location of the target stimuli in the neural space. (Right) Simulation results reproducing the activity patterns of the experimental data. As can be seen, experimental and simulation data are fairly equivalent (*Adapted from Cisek (2006)*).

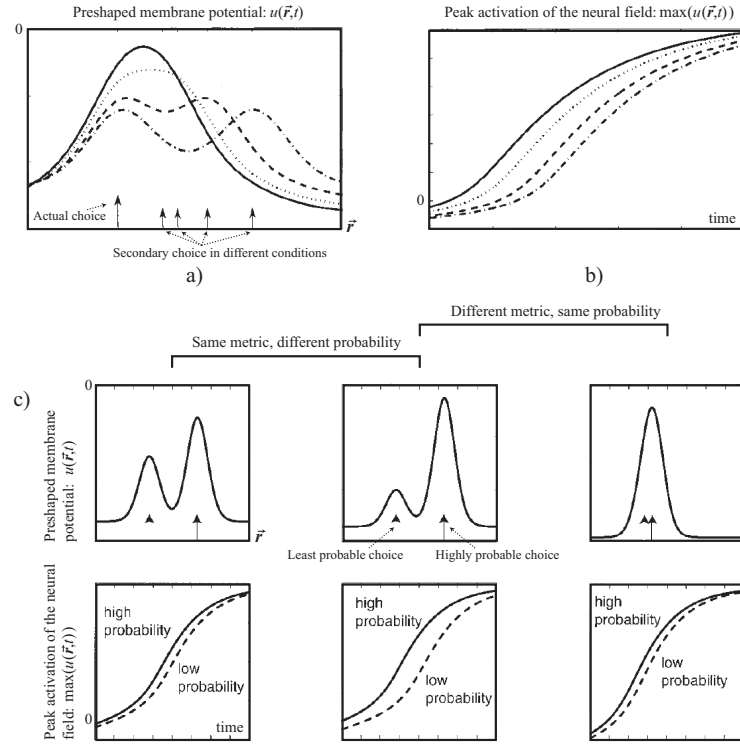


Figure 3.11: Effects of the metric of the internal neural representation on the selection process. **a)** Several examples of pre-activation patterns corresponding to different distances between two symmetric inputs are shown. **b)** According to this distance, the presentation of two asymmetrically balanced inputs produces distinct increases of the activity of the most active neuron which corresponds to the peak of the winning region. The more distant are the inputs, the longer is the network response. **c)** An illustration of the combined effect of the distance metric and the relative amplitude between both inputs on the network response time is shown. The amplitude of the pre-activation pattern may be seen as a probability of selection. The more distant and asymmetric are the input signals, the larger is the difference in response time between the correctly and the incorrectly pre-cued network. (*Adapted from Erlhagen and Schöner (2002)*)

signals, the latter of which could also be understood as a probability measure.

GAIN MODULATION

The next property of neural fields that is considered here is that the information they convey can be gain modulated. Gain modulation is a change in the response amplitude of a neuron that is independent of its preferential selectivity, sometimes also referred to its receptive field (Salinas & Thier, 2000). Fundamentally, when considering distributed neural representations such as those implemented in neural fields, the modulation of the overall activity of the network corresponds to a multiplicative interactions between two separate sources of inputs. The networks endowed with such a property were then designated as gain fields. As will be described later, gain modulation has been suggested to

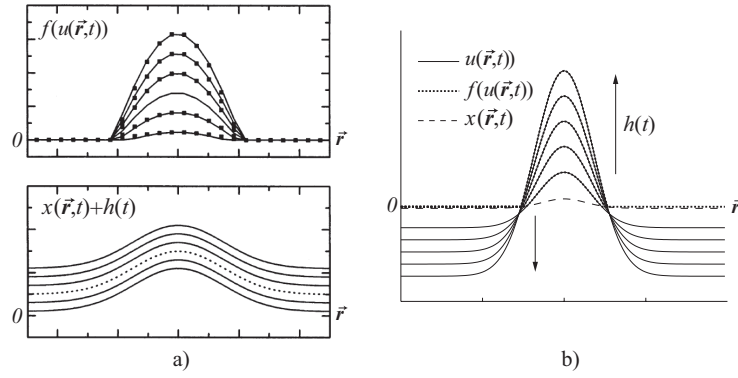


Figure 3.12: Gain modulation of the information conveyed by the neural field is produced by varying the background input $h(t)$. This effect can be produced by a network having **a)** a mexican-hat-like or **b)** gaussian-like recurrent synaptic interactions. (Figure **a)** is adapted from Salinas and Abbott (1996))

be a mechanism by which the brain may combine or integrate information from different sensory, motor, and cognitive modalities in a non-linear way (Salinas & Thier, 2000; Pouget et al., 2002).

In continuous attractor neural network, gain modulation has been shown to be implemented by changing the background homogeneous input $h(t)$ (Salinas & Abbott, 1996). As a consequence, when the background input is increased, the effect of the recurrent interaction kernel on the network is amplified almost proportionally, hence modulating the information currently represented in the network. Another interesting property following from this mechanism, is that the network may be *switched off* by decreasing this input to a very low level, corresponding to a strong global inhibition of the network (Salinas, 2003a). Some illustrations of the activity profile of neural fields which homogeneous input was varied are given in Figure 3.12.

Sensorimotor Transformations

According to several theoretical considerations, gain fields have been suggested to be ideally configured to facilitate certain kinds of computations such as coordinate transformations (Salinas & Thier, 2000; Pouget & Snyder, 2000) and the processing of invariant representations (Salinas & Abbott, 1997; Deneve & Pouget, 2003). Indeed, as reported by neurophysiological studies, gain fields seems to be implicated in eye and reaching movements (Bruneo, Jarvis, Batista, & Andersen, 2002; Scherberger & Andresen, 2003), spatial perception (Avillac et al., 2005), attention (Connor, Preddie, Gallant, & VanEssen, 1997), and object recognition (Ashbridge et al., 2000).

The principle mechanism of transformation across frame of reference can be described as follow. In Figure 3.13b, a target reaching example is considered. In order to be able to reach the target, its location x in eye-centered coordinates has to be transferred sequentially into a head-, body-, shoulder- until a hand-

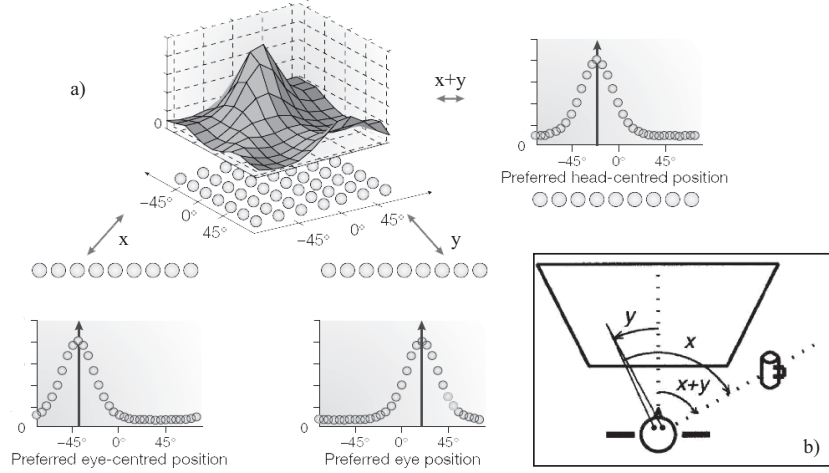


Figure 3.13: Sensorimotor transformations are suggested to be performed through gain fields. **a)** The gain field, composed of neurons arranged on a two-dimensional torus is capable to perform the coordinate transformation illustrated in **b)**. A target is located at an angle x in eye-centered coordinates and the eyes are looking at an angle y in head-centered coordinates. In order to compute the target location $x + y$ in a head-centered frame of reference, the network combine multiplicatively the activity profile of the two input populations representing respectively x and y . From such non-linear and distributed combination of input variables, almost any transformations such as their sum can be computed. (Adapted from Pouget et al. (2002); Salinas and Thier (2000))

centered frame of reference so that the necessary information for planning a reach becomes available. By considering only the first step here, the use of the information related to the eye position y in their orbit, the location $x + y$ of the target in a head-centered coordinate system can be computed. In this example, a neural field, which internal space Γ corresponds to a two-dimensional torus, receives projections from the neural populations encoding both variables x and y along each of its dimension, respectively. Despite of the additive nature of the synaptic projections, the recurrent interactions are capable to produce a non-linear combination of these inputs. Consequently, the sensitivity of each neuron of the gain field corresponds more or less to a radial basis function centered on the a specific value in a space combining those in which x and y are encoded. As illustrated in Figure 3.13a, a third neural population, by appropriate projections from the central field, is finally encoding the target in head-centered coordinates, i.e., $x + y$. The fundamental idea behind the gain field approach, is that, out of a non-linear combination of input variables represented by means of basis functions, any transformation can be calculated, its precision depending on the breadth of the tuning curves as well as the number of neurons in each population (Salinas & Abbott, 1995; Baraduc & Guigon, 2002; Pouget et al., 2002).

Self-sustained patterns of excitation have been described as a means to memorize the information conveyed by the last seen input. However, the question as to how should a neural field behave in the case of a moving input may arise. A theoretical answer has been proposed following from the observation of traveling waves of excitation within neural ensembles which may be either spontaneous (Chagnac-Amitai & Connors, 1989), or resulting from epileptic seizure (Wadman & Gutnick, 1993).

Traveling waves

Neural field models of traveling waves have been developed in the past decades (Amari, 1977; Zhang, 1996; Ben-Yishai, Hansel, & Sompolinsky, 1997; Xie et al., 2002; Erlhagen, 2003). Two major approaches have been proposed to reproduce this phenomenon. While one of them introduces an additional inhibitory layer on top of the classical continuous attractor model (Amari, 1977; Ben-Yishai et al., 1997; Erlhagen & Schöner, 2002), the other considers asymmetries in the recurrent interaction kernel (Zhang, 1996; Xie et al., 2002).

In the first approach, by coupling two layers, the first-order differential equations driving the dynamics of each layer can be combined which results in a differential equation of the second order. Consequently, supplementary behaviors can emerge from the network dynamics such as oscillations. In addition, when appropriately tuned, this coupled neural field may damp these oscillations, and hence, the network activity may stabilize. In a perfectly symmetric and noise-free world, the resulting stable activity bump would be stationary. However, this particular case is, dynamically speaking, unstable. When the symmetry is broken, a traveling activity packet develops. Along the direction in which the bump moves, the tail of the bump mostly gets inhibited by the previous activation of that precise region, while its front side gets more easily excited since no activity was present before at this location. Two examples of reported shapes of traveling wave in such a neural medium are shown in Figures 3.14a-c.

In the second approach, an asymmetry in the recurrent interaction kernel has been proposed. Zhang (1996) demonstrated that if the additional asymmetric component of the synaptic weights corresponds to the directional derivative of the original part along a direction in the neural space, a symmetric-shaped traveling activity bump emerges from the network. As such, the network's dynamics, originally producing a continuum of marginally stable attractors, exhibits now a periodic attractor dynamics or a limit cycle (See Figure 3.14e). Formally, by renaming the original recurrent weights W^R with W_0^R , the recurrent connectivity producing a traveling wave can be written as

$$W^R(\vec{r}, \vec{r}') = W_0^R(\vec{r}, \vec{r}') - \lambda \nabla W_0^R(\vec{r}, \vec{r}') \cdot \vec{d} \quad (3.15)$$

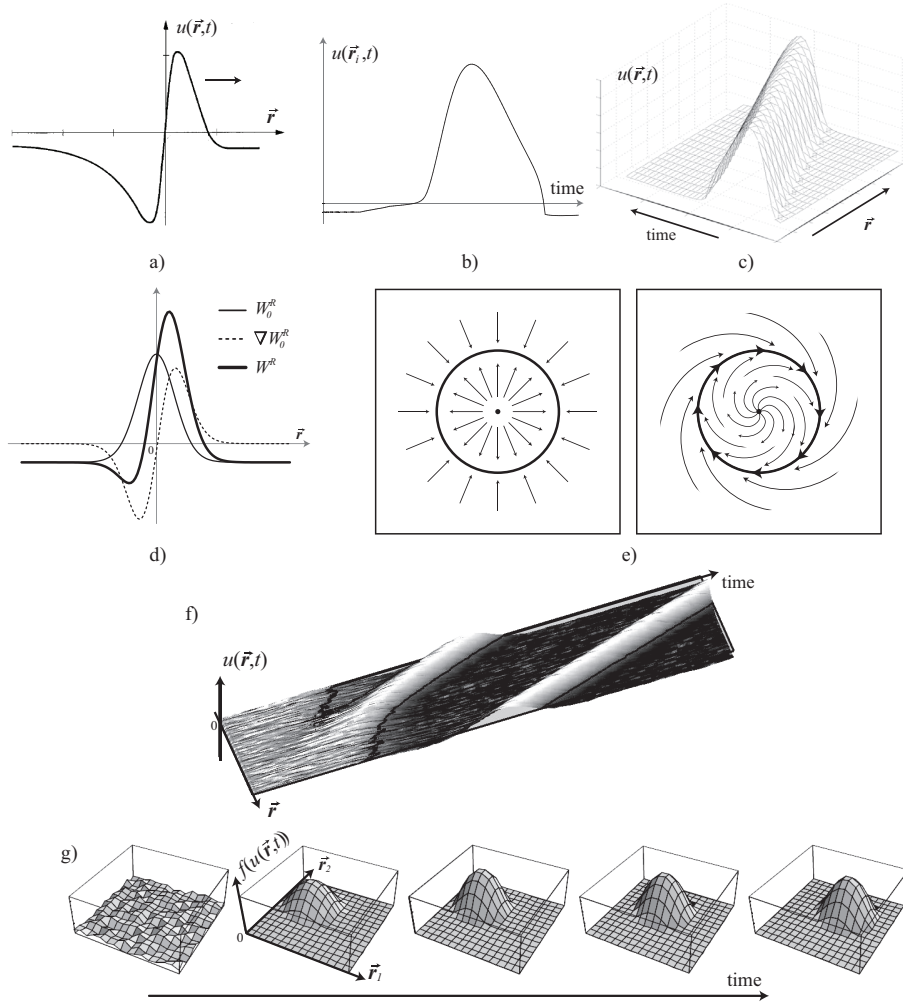


Figure 3.14: Traveling waves can be generated on the surface of neural fields according to two computational methods. **a)-c)** By adding a supplementary layer on top of the neural field, where each neuron of that layer inhibits preferentially the neurons of the original layer sharing a similar preferential tuning, traveling wave may develop on the field. As can be noticed, along the direction of motion, the tail of the bump is more inhibited than its head, which allows the activity packet to move. (*Adapted from Amari (1977); Erhlagen (2003)*). **d)** The second method consists of adding an asymmetric component to the recurrent interaction kernel. **e)** This modification results in a change of the network intrinsic dynamics. As illustrated by the state spaces, the original marginally stable attractors of the network (left) become unstable in the sense that they now form a directed limit cycle. **f)** This simulation example shows the effect of the asymmetric shape of the recurrent weights. From a homogeneous initial state, the noise breaks this unstable state and an activity bump emerges. Because of the biased connectivity, this bump moves in one direction at a given constant speed. **g)** (*Adapted from Zhang (1996)*).

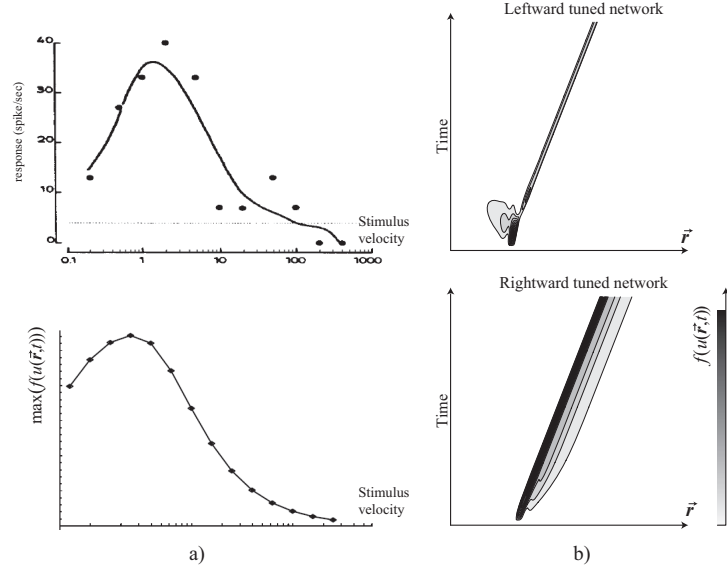


Figure 3.15: Network sensitivity to moving inputs: **a)** (Top) Results of an experiment where a neuron of the visual cortex was recorded while a moving visual stimulus was presented at different speeds. This plots shows that this particular cell is preferentially tuned to a given stimulus velocity (*Adapted from Orban et al. (1986)*). (Bottom) Simulation results where a moving input stimulus is presented to a neural field with an asymmetric connectivity. The profile of the maximal activity of an arbitrary neuron of the network shows that this model can faithfully reproduce the response pattern of a real biological neuron. **b)** The spatiotemporal response of two neural fields preferentially tuned to opposite directions to a input stimulus moving in the right direction shows that when the input direction of motion is compatible with that of the network, the field response is greater than in the opposite case. (*Adapted from Mineiro and Zipser (1998)*)

where ∇ corresponds to the differential operator and \vec{d} to the direction along which the activity bump will move. $\lambda > 0$ is a scaling factor which has been shown modulate the speed of the traveling wave (Zhang, 1996). This speed will further also called the intrinsic velocity of the network. An illustration of the ideal profile of the recurrent weights as well as a traveling activity bump generated by such an interaction kernel are respectively shown in Figures 3.14d, f and g.

Velocity Sensitivity

As shown in previous sections, a stationary input, which is fed into a neural field, develops a stationary bump on its neural surface. However, when the continuous attractor network is tuned to produce a traveling excitation pattern, the effect of such a stationary input is modified. Indeed, instead of entering in resonance with a marginally stable attractor, it now compete with the network's tendency to move the current bump of activation along its directed limit cycle. As a consequence of this competition, the strength of the network response to this input is reduced as compared to that observed in a standard neural

field. Next, following from this observation, one could intuitively imagine that an input moving in the same direction as the intrinsic velocity of the network would produce a stronger response. As shown by Mineiro and Zipser (1998), this directional selectivity effect can effectively be observed. Moreover, this sensitivity is not only directional, but depends also on the intrinsic velocity of the network. An input stimulus moving at a speed more or less equivalent to that of the network appears to produce a greater activation than when moving at other velocities. Interestingly, biological neurons exhibiting such directional selective responses were found in the visual cortex (Orban et al., 1986). Simulation results reported by Mineiro and Zipser (1998) as well as the results of the mentioned neural recordings are given in Figure 3.15.

Velocity Integration and Update of the Internal Representation

Self-sustained patterns of excitation have been described as a means to memorize the information conveyed by the last seen input. However, sometimes, it may be important for the brain to be capable of updating this represented information by means of other input signals, different by nature. For instance, when observing an object passing behind an occluder at a constant speed, one can predict at which precise moment the object will reappear. In this example, an internal representation of the object location must integrate its velocity through time in order to keep that representation up to date.

In the rat hippocampal formation, it has been reported that a group of cells fire according to the current heading direction of the animal (Taube & Bassett, 2003). Interestingly, these so-called head-direction cells have a firing pattern which can be modeled by means of neural fields coding for a direction in space. Moreover, these neurons also have the characteristic that their activation can be kept updated based on self-motion cues even in complete darkness. Consequently, the animal can hence always be aware of which allocentric direction it is looking at. Technically, it is believed that an angular head velocity signal shifts the localized activity pattern by integration (in the mathematical sense) (Zhang, 1996; Redish, 1999). As such this integration mechanism has further been proposed to be also involved in predictive processes such as that described above (Erlhagen, 2003). Indeed, several groups of neurons widespread in various areas of the cortex exhibits predictive responses associated with motion perception (Unema & Goldberg, 1997; Eskandar & Assad, 1999; Umiltà et al., 2001; Nakamura & Colby, 2002; Jellema et al., 2004)

Neural implementations derived from the asymmetric recurrent synaptic connectivity producing traveling waves have been proposed by several computational studies (Redish, Elga, & Touretzky, 1996; Zhang, 1996; Goodridge & Touretzky, 2000; Stringer et al., 2002; Xie et al., 2002). First, a category of models assume that the recurrent asymmetry can be modulated by a third-part neuron (Redish et al., 1996; Zhang, 1996; Stringer et al., 2002). As mentioned earlier, the amplitude λ of the asymmetry given in Equ. (3.15) directly controls

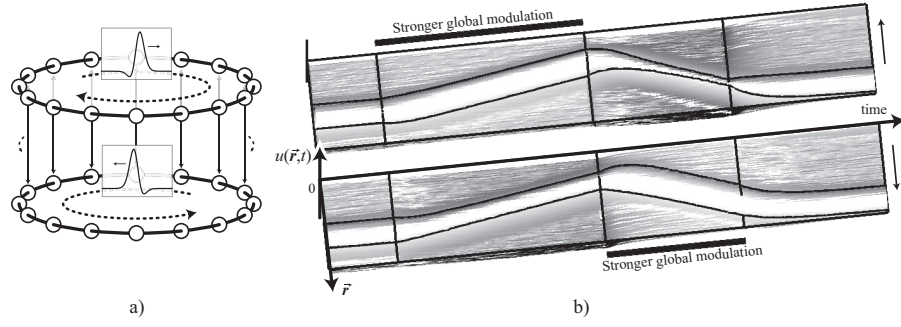


Figure 3.16: Neural mechanism of velocity integration. **a)** The structure of a double-ring network as that proposed by Xie et al. (2002) is shown. The two sub-networks are respectively tuned to drive the activity bump in an opposite direction. Their reciprocal strong coupling forces them to represent the same information. By balancing asymmetrically their respective global excitation level, a velocity-controlled traveling wave is produced. **b)** In this simulation, the global activation of both sub-networks is initially balanced. Then, the leftward population becomes more excited and hence, a traveling motion of the bump is observed in that direction. This tendency is further reversed and finally, the balance of excitation is restored.

the intrinsic speed of the network, and hence the amount of velocity integration. From this, a third-part neuron representing the velocity signal may modulate the network integration level through this variable. However, the use of such so-called sigma-pi neurons has often been considered as controversial in terms of biological plausibility. Indeed, keeping the additive property of neural integration of synaptic inputs has been suggested to be fundamental to remain in line with a biological account of the neural response (Xie et al., 2002). Therefore, neural field models with additional layers and strictly additive neurons were suggested in order to reflect more accurately the neural processes occurring in the brain (Goodridge & Touretzky, 2000; Xie et al., 2002). The fundamental principle mechanism developed by Xie et al. (2002) consists of two neural populations respectively endowed with an intrinsic tendency to move an activity bump in opposite directions. Next, a competitive push-pull interaction dynamics between these two layers is introduced through a tight coupling between them. Indeed, by balancing the respective amount of excitation fed to these oppositely tuned populations, the one which is the most excited tends to drive the activity packet in its associated direction. An illustration of the structure of the two-layer model as well as simulation results of velocity integration are given in Figure 3.16.

This ends the present section addressing some of the numerous computational properties of neural fields that make this approach an ideal framework for analyzing the cognitive processes present in the brain. Until now, continuous attractor neural networks were mostly considered alone, although they do receive some external inputs. However, in order to model higher brain functions such as those with which the present thesis is concerned, networks of such networks

have to be developed. Indeed, the cortex is composed of many interconnected regions processing different types of information as well as being involved in different cognitive functionalities. Similarly to the gain in computational power provided by an appropriate combination of neurons in large networks, more complex behaviors are expected to emerge from networks composed of neural fields connected together.

3.2.3 MODELING CORTICAL PATHWAYS: NETWORK OF NETWORKS

In order to represent the brain neural pathways involved in various cognitive functions, networks of neural fields have been proposed, where each subnetwork is often associated with a specific brain region. Recent works have developed large networks in order to model several cognitive abilities such as visual processing (Itti & Koch, 2001), selection of target for movement execution (Cisek, 2006), sensorimotor learning (Andry et al., 2004; Meñard & Frezza-Buet, 2005), sensorimotor transformations (Pouget et al., 2003), invariant representations (Salinas & Abbott, 1997; Deneve & Pouget, 2003), spatial navigation (Redish et al., 1996), joint-action task involving several agents (Erlhagen & Bicho, 2006), and goal-directed imitation (Erlhagen, Mukovsky, & Bicho, 2006).

In this thesis, the same approach will be followed when the neural processes of imitation will be addressed. Since several neural fields will be interconnected within large neural networks, the notation as well as the fundamental mechanisms for linking neural neural fields together are described in the following. In order to identify a neural network representing a brain region or a sub-ensemble of neurons within a region, an index i , corresponding to the name of the network, is added to its internal variables. For example, the membrane potential of a neuron belonging to a population i , preferentially tuned to a direction \vec{r} at time t will be written as $u^i(\vec{r}, t)$.

Next, in addition to the direct external inputs that a neural field may receive (see Equ. (3.10)), the network can also be subjected to synaptic projections arising from other populations. The projections between two neural fields can basically be of two types. First, topology preserving projections ensure that a localized peak of activity in the source population also produces a localized input in the target population. Secondly, homogeneous projections uniformly modulate the target population proportionally to the global activity of the source field.

Topology preserving projections are mediated by synaptic weights W^{ji} from a population j to a population i . Since the preferential sensitivity of each neuron is important in this case, these synaptic projections affect the input $x^i(\vec{r}, t)$ of the network, defined in the equation describing the field dynamics (Equ. (3.9)).

An additional term is thus added to Equ. (3.10) which gives

$$\begin{aligned}
x^i(\vec{r}, t) &= \sum_{j \in N^i} \beta_j^i(t) \left[\mathcal{G}(\vec{s}_j^i, \vec{r}, \sigma) - \delta \right] \\
&\quad + \sum_{j \in M^i} \oint_{\Gamma} W^{ji}(\vec{r}', \vec{r}) f(u^j(\vec{r}', t)) \, d\vec{r}' \quad (3.16)
\end{aligned}$$

where M^i and N^i correspond, respectively, to the ensemble of external inputs and the ensemble of the population projecting to network i . β_j^i is the strength of the representation of the sensory variable $\vec{s}_j^i(t)$. The synaptic projections W^{ji} are usually defined similarly to the recurrent weights given in Equ. (3.8), but with different parameters⁵. In contrast to the topology preserving projections, homogeneous projections consist of an uniform modulation of the target population that could either be excitatory or inhibitory. By their homogeneous nature, they affect the background input $h^i(t)$ of neural field i such that

$$h^i(t) = h_0^i(t) + \sum_{j \in K^i} W^{ji} \oint_{\Gamma} f(u^j(\vec{r}, t)) \, d\vec{r} \quad (3.17)$$

where K^i corresponds to the ensemble of population projecting homogeneously to the network i . Similarly to the first term of the external inputs given in Equ. (3.16), $h_0^i(t)$ is the background modulation being directly applied to the network. In this case W^{ji} is constant and belongs to \mathbb{R} .

⁵If the weights W^{ji} need to be defined differently, their generating equation will be explicitly mentioned in the text when appropriate.

Summary

This chapter reviewed the modeling approach adopted in this thesis. The artificial neural networks employed here, known as neural fields, possess a range of computational abilities for processing information, almost all of which have strong neurophysiological correlates. They include the ability to represent information, memorize it, reduce noise, select an input between several competing inputs, perform sensorimotor transformations and achieve motion integration. In addition, interconnecting neural fields allows for the development of large neural networks that may, to a certain extent, model the connectivity patterns of the brain areas involved in imitation. It is well accepted that the processes underlying imitation deal with sensorimotor transformations, sensory predictions, decision-making and memory: all computations that neural fields are capable of. Models based on neural fields, therefore, might be expected to successfully reproduce the behavioral expressions of imitation in humans and monkeys.

The following chapter begins to describe the modeling work accomplished in this thesis. An investigation of the cortical pathways involved in automatic imitation is presented. Two plausible neural field architectures accurately reproduce the results of a behavioral experiment. A novel experimental paradigm for differentiating between these two models is proposed. Performing such an experiment would enable identification of the most probable cortical information pathways involved in such automatic imitative behaviors.

BRAIN PATHWAYS OF IMITATION AND THE IDEOMOTOR PRINCIPLE

*The study presented in this chapter was significantly
adapted from the work published in:*

Sauser, E. L. and Billard, A. G.. Parallel and distributed neural models of the ideomotor principle: An investigation of imitative cortical pathways. *Neural Networks*. 19(3):285-298, 2006.

THIS chapter begins to describe the modeling work undertaken in this thesis. It presents an investigation of the neural mechanisms and cortical pathways responsible for automatic imitative behaviors, especially those involved in the behavioral expression of the principle of ideomotor compatibility. This work raises a question formulated with respect to current knowledge of human neurophysiology, one suggested by neural evidence of the existence of shared representations between one's own movements and those performed by others: is there really a direct sensorimotor route responsible for triggering imitative behaviors, or does the information encoded within these shared representations influence only the higher cognitive centers involved in the selection of motor responses? In order to investigate this issue, two neural models are developed, each following one of these two hypotheses. Both models are capable of replicating the results of the behavioral study described in the next section. In order to test which of the two models best describes human behavior, a novel behavioral experiment is proposed.

4.1 RELATED EXPERIMENTAL STUDY

Brass et al. (2000) conducted a behavioral study which was based on a stimulus-response (SR) experimental paradigm. It was used in order to verify two hypotheses of the ideomotor theory. These two hypotheses are based on the neural correlate that the human brain appears to possess highly specialized neural circuits devoted to the recognition of movements performed by others and that these circuits are likely to be shared by the motor preparation circuits (Iacoboni et al., 1999; Decety & Sommerville, 2003). The first of the hypotheses

states that if a subject is requested to respond to a movement of a demonstrator, then (s)he would experience a motor facilitation resulting in faster reaction times as compared to the case where the subject is asked to respond to an more abstract spatial cue. The second hypotheses states that the facilitatory effect would be greater if the movements of the demonstrator and subject are similar (ideomotor compatible) than if they are of a different type (ideomotor incompatible).

Their experimental setup is comprised of three independent binary variables, which leads to eight conditions plus four baseline conditions. The experimental stimuli consisted of a combination of a finger-lifting movement (either index or middle finger) and of a spatial cue consisting of a cross painted on the corresponding or opposite fingernail (see Fig. 4.1a). The subjects reaction times (RTs) were measured while they were asked to respond to the various stimuli by moving the finger that was the closest to either cue (e.g. by moving their index finger in response to a demonstration of a movement of the index finger or to the presentation of a cross on the demonstrator's index fingernail). These instructions determined the first experimental variable, i.e., the relevant stimulus dimension. Furthermore, an interference paradigm was used in order to examine the effect of the presentation of congruent or incongruent¹ stimuli against a baseline condition in which only the relevant stimulus was presented to the subjects. Finally the experiment was varied in order to examine the effect of ideomotor similarity between observed and executed movements. In one case the subjects were asked to lift their finger (ideomotor compatible condition), and in the other case, they were asked to produce a finger-tapping movement (ideomotor incompatible condition).

The results of this experiment, shown in Figure 4.1b, were in agreement with the hypotheses. Indeed, responses to finger movements were faster than responses to spatial cues, and ideomotor compatible pairs of observed-executed movements generally produced better RTs. Moreover, typical facilitatory and interference effects were observed between congruent and incongruent conditions respectively. Next, two neural models which account for these results are presented.

¹Congruent condition: a left (right) finger movement with a cross on the left (right) fingernail. Incongruent condition: a left (right) finger movement with a cross on the right (left) fingernail.

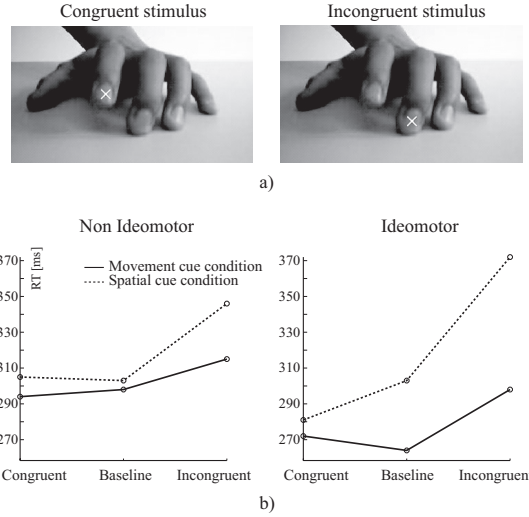


Figure 4.1: **a)** Examples of congruent and incongruent stimuli used by Brass et al. (2000) in their experiment. **b)** The reaction times observed in the original experiment are redrawn with a small modification. As the ideomotor variable was tested among two distinct groups of subjects, the reaction times were shifted in order to make the baseline condition in the spatial cue task coincide in both experiments. Indeed, this is the only case in which the experimental conditions are identical.

4.2 NEURAL MODELS

The modeling approach of this work is primarily based on the well-accepted hypothesis that the brain uses parallel pathways to process information. This parallel and distributed processing principle has been successfully applied in explaining a variety of effects observed during stimulus-response experiments as well as a wide range of human behaviors (Hasbroucq & Guiard, 1991; Kornblum, 1994; Zhang, Zhang, & Kornblum, 1999; Erlhagen & Schöner, 2002). These models usually assume a feed-forward layered network organization, where each layer corresponds to a specific sensorimotor processing stage, such as sensory perception, multisensory integration or motor preparation. Generally, several streams of information coexist and are then combined together within further layers until an output response is provided by the last part of the model.

In this chapter, a similar type of neural architecture is developed in order to account for the visuo-motor flow at the basis of the imitation task described above. In particular, this work is interested in determining whether automatic imitative behaviors, as described at length in Section 2.2.1, are effectively mediated by a direct route between action observation and motor execution, or whether they are mandatorily gated by the higher-cognitive processes of motor selection as reported in usual sensorimotor stimulus-response tasks. Therefore, two models are developed, and each of them follows one of the two hypotheses. As will be described in the next section, the processing components of the two models are identical, but the connectivity among them is different in order

to account for either a separate or a unique route for imitative behaviors and stimulus-response mapping. Finally, the implementation of these models follows the dynamic neural field approach presented in Chapter 3. Consequently, rather than describing again the equations ruling this type of network, references to them will be provided along the description of the architecture of the models.

4.2.1 ARCHITECTURE OF THE MODELS

As shown in Figure 4.2, the two neural models can be split into three major parts. The visual perception, the decisional and the motor preparatory layers have for respective tasks to process visual information, to determine the right response to external stimuli, and to prepare and trigger motor execution. Each model is basically a network of interconnected neural fields, where the processing is mainly feed-forward. The external inputs thus drive the network activity along one direction of propagation. These models are input-driven. Next, The processing of the task instructions is not explicitly addressed here. Frontal cortices are simply assumed to drive the models by modulation of the neural activity of their components. In the next sections, a technical description of the architecture of the two models is first provided, followed by their relationships to real cortical networks.

VISUAL PERCEPTION

First of all, as described in Sections 2.3.2 and 2.3.3, movement observation is suggested to recruit two separate cortical pathways. The route primarily involved in decision-making flows ventrally, from the temporal cortex to the prefrontal lobe. In contrast, the other, which is more concerned with the control of movements in sensorimotor terms, passes dorsally along the parietal and frontal cortices. Therefore, the description of the visual inputs to these two pathways has been separated.

Visual Perception for Decision-Making: Neural Encoding

In this modeling study, visual information is assumed to have already been processed by highly specialized circuits and to be represented in a manner relevant for the task. According to the characteristics of the stimuli used in this experiment, two major types of visual inputs are considered. They correspond to the abstract or spatial cue, i.e., the cross on a fingernail, and the movement cue. These two components of the stimuli are encoded in two distinct neural fields which represent their spatial location in the visual field. The parameter space in which this information is encoded is given by $\vec{r} \in \Gamma_{N=1}$ (See Equ. (3.5)). Here, since Γ is unidimensional, the ensemble of preferred directions \vec{r} of this network can be associated to a periodic variable $\theta \in [-\pi, \pi[$ such that

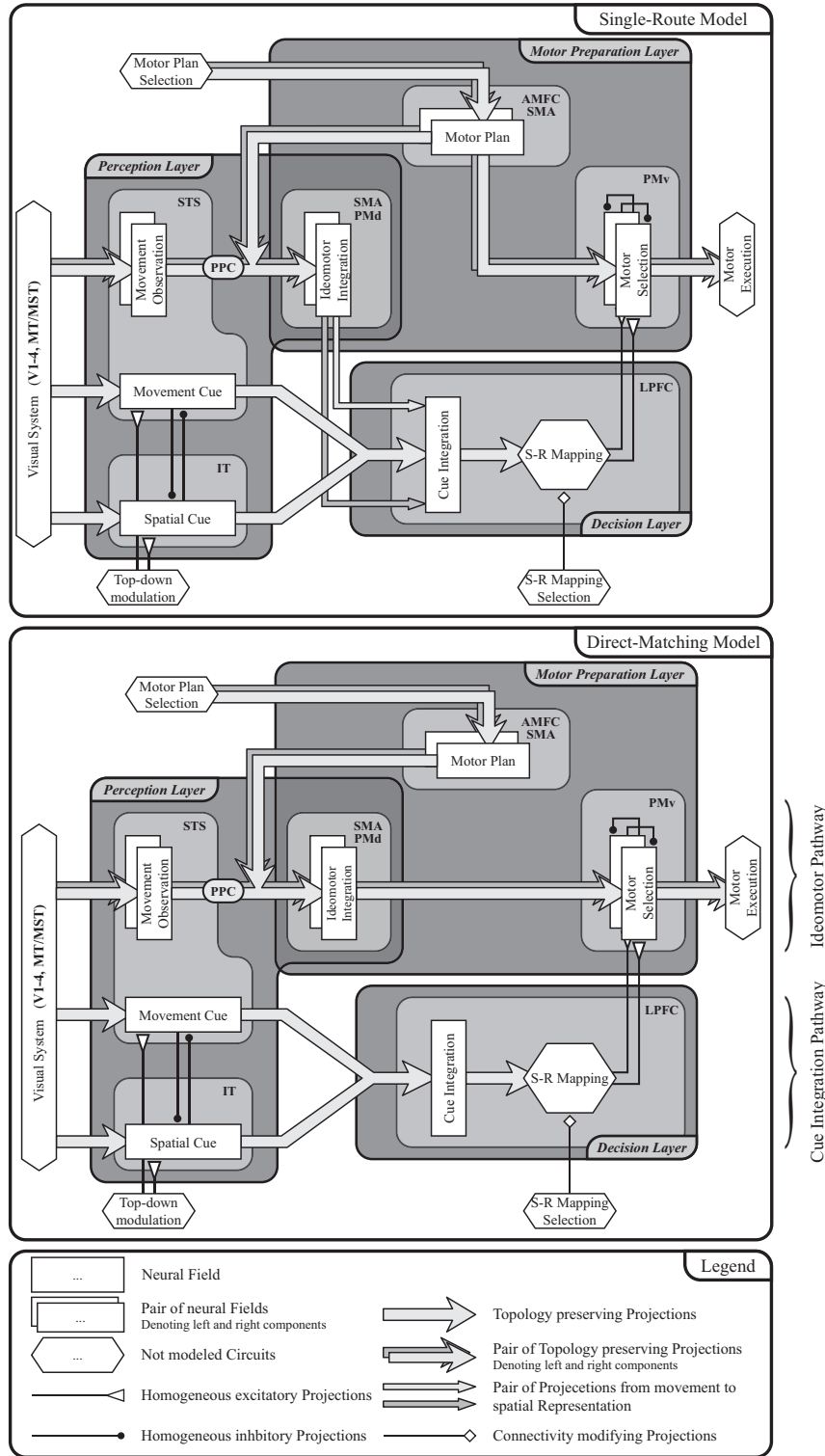


Figure 4.2: Schematic of the two architectures proposed to model the behavioral effect related to the ideomotor principle. On top, the single-route architecture assumes that all the processed stimuli interact within the same integration layer. At the bottom, two separate pathways are involved in the computation. The cue integration pathway accounts for the response selection, whereas the ideomotor pathway influences the motor selection mechanism by means of a direct connectivity.

$\vec{r} = (\cos \theta, \sin \theta)$, where θ represents the neural space where stimuli are localized. Importantly, θ does not necessarily correspond to an angular position. Further, the dynamics of the neural fields is governed by Equ. (3.9). In Equ. (3.10) which describes the profile of external inputs, the location of one input is denoted by \vec{s}^i and its amplitude by β^i , where i indicates the network which receives that input (See Section 3.2.3). Finally, note that \vec{s}^i is replaced here by a variable φ in a way similar to how the preferred direction \vec{r} were exchanged with θ .

Effect of Task Instructions

Since the task instructions have to favor responses to one of the two types of cues, the neural representation of the relevant stimulus should be positively enhanced. Technically, a top-down modulatory input h^T is fed into the background homogeneous input of each neural population such that

$$h^i(t) = \delta(i, j) h^T \quad \text{and} \quad \delta(i, j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad (4.1)$$

where i, j denote either the network corresponding to the spatial or movement cue. In particular, j indicates the neural field encoding the task-relevant stimulus (See Figure 4.2).

Perceptual Biases

Studies in experimental psychology reported several human behavioral characteristics indicating the existence of perceptual biases which affect the processing of visual inputs. Two of them are considered here. First, it has been shown that moving stimuli receive more attention than static ones, which usually result in faster responses to the former class of stimuli (Franconeri & Simons, 2005). In this study, a bias toward movement cues has thus to be introduced, i.e., the amplitude β of the external input corresponding to the movement cue is artificially set to a higher value than that of the abstract or spatial cue.

Secondly, a phenomenon usually referred as the Hick's law, which basically states that the more information is perceived, the longer is the reaction time², may be at play in this experiment (Hick, 1952). Indeed, the baseline condition of the task involves only one stimulus component, whereas the other conditions require two of them. A modeling study by Ernhagen and Schöner (2002) already showed how the Hick's law could be reproduced from the competitive interactions between several external inputs simultaneously fed into a neural field. Here, the same technique is considered but by means of reciprocal inhibitory connections between the neural representations of both cues. This inhibition is applied in the models using homogeneous projections as described in Equ.

²More precisely, this law describes that when the brain has to process N different stimuli simultaneously, the reaction time approximatively increases proportionally with the binary logarithm of N .

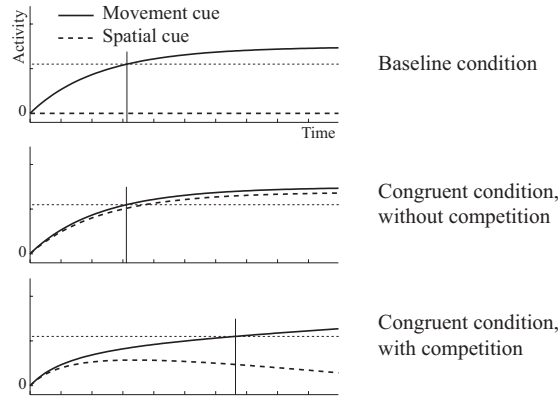


Figure 4.3: Illustration of the mechanism influencing the integration time when more than one type of stimulus is applied to the network. Each plot shows the evolution of the activity of the neurons with maximal activation which represent either the movement or the spatial cue. The light dotted line indicates an arbitrary threshold level, and the light filled line gives the time at which the maximal activation reaches that threshold. As can be seen, when competitive interactions are introduced between the stimulus representations, the integration time is longer.

(3.17). As illustrated in Figure 4.3, when a stimulus input is fed to each cue representation, their respective amplitude increases more slowly as compared to during the baseline condition, where only a single input is applied.

The Sensorimotor Route of Movement Representation

As mentioned above, there is a second processing route in which the observation of movements has an influence. Because this route is more concerned with the representation of movements in sensorimotor terms, the neural encoding along this pathway should be similar to that used for the direct control of movements. Therefore, along this processing pathway, to each finger is associated a neural population coding for its movement direction (Georgopoulos et al., 1988). According to the experimental task, fingers could move either up or down. The external input of these neural fields thus indicates a movement in one of the two opposite directions (See Appendix B.1 for the exact values used in this model). Importantly, these internal representations of the movements of others are assumed to be encoded in motor coordinates, i.e., within the same frame of reference as self-planned movements.

MOTOR PREPARATION

As shown in Figure 4.2, the motor preparation layer consists of three areas, which code respectively for the motor plan of each finger, for the motor commands to be executed, and for the shared representation between movement observation and motor execution designated as the ideomotor integration area. By definition, the ideomotor region is where information related to both move-

ment observation and movement preparation overlaps, and it is thus the area responsible for ideomotor effects to occur (Greenwald, 1970; Brass et al., 2000). Possible movements, represented in all the regions of the motor preparation layer are thereby represented in the same way as observed movements, i.e., two distinct neural fields encode the direction of movements of the index finger and the middle one.

Considering now the ideomotor region and its connectivity patterns, it receives projections from the representation of observed movements as well as that of planned ones by means of synaptic connections defined in Equ. (3.16). Because of its visuomotor property, this area will induce the ideomotor effect, i.e., high ideomotor compatibility happens when the demonstrator and the imitator move the **same** finger in the **same** direction, whereas low ideomotor compatibility is reached when demonstrator and imitator produce movements with different fingers in the opposite direction.

Response Selection

The output of the models consists of a network which receives the possible motor plans to be executed. As will be described in more details in the paragraphs presenting the decisional layer, this area may either receive the original motor plans directly, or the result of the visuomotor integration preformed in the ideomotor region. Next, in order to trigger the correct motor response, the motor selection area also waits for an execution signal coming from the decisional layer. Since this final motor selection process has to trigger the movement of one of the two fingers only, homogeneous inhibitory connections were introduced across the networks representing the direction of movements of each finger. A motor response is then executed as soon as the global activity of its corresponding population, i.e., its energy $E(t)$ as defined in Equ. (3.14), reaches the threshold E_0 such that

$$E^i(t) > E_0 \quad \Rightarrow \quad \text{The motor plan of network } i \text{ is executed.} \quad (4.2)$$

where i denotes the motor selection network associated with either the left or right finger. Finally, actual motor execution processes were not considered in this work. Indeed, since the execution time of finger movements is assumed to be constant under all conditions, modeling these processes or not is assumed to result in no significant relative difference on the reaction times produced by the models.

DECISION AND RESPONSE SELECTION

The principal task of the proposed models is to select a response according to either the spatial cue or the movement cue. Basically, this computation is associated with the neural network located in the decisional layer of the models shown

in Figure 4.2. This neural field receives projections from the representations of spatial and movement cues. Because the strength of these representations are asymmetrically balanced with respect to the task instructions, the cue integration network can thus always select the correct response. However, the time required for performing the decision will depend on the congruency between the inputs (See Figure 3.10 of Chapter 3 for an illustration of this mechanism).

Differing Processing Pathways between the Models

As can be seen in Figure 4.2, the difference between the two proposed models consists primarily of a different connectivity to the cue integration network. While the direct-matching model assumes direct projections from the ideomotor region to the motor selection centers, the single-route model processes the second stream mediating movement observation within the decisional layer. In this second model, projections from the ideomotor area to the cue integration area have to be introduced. Moreover, since the representation of both areas differs, a transformation has to be defined in order to transfer the motor representation of the former into the spatial representation of the latter. This is realized by means of two additional localized inputs feeding the cue integration area, which encodes the location of the two fingers. Their respective amplitude $\beta_i^j(t)$, where i denotes the ideomotor representation of either the left or the right finger and j the cue integration network, is set proportionally to the energy $E^i(t)$ (See Equ. (3.14)) of its corresponding network, i.e., $\beta_i^j(t) = \gamma^i E^i(t)$, where γ^i is a scaling factor.

Stimulus-Response Mapping

Finally, as soon as the selection process ends, the decision which has been made in stimulus space, has to be transformed into motor space. In this work, although a plausible neural mechanism for performing such stimulus-response associations has been proposed by Wilimzig and Schöner (2005), it was simply assumed that this transformation is carried out by hard-wired connections from the decisional process to the motor selection stage. In order to produce these associations, the local energy $E_j^i(t)$ of the network around the location of the finger φ_j as defined in Equ. (3.14) is fed into the homogeneous input of the motor selection network, i.e., $h^j(t) = W^{ij} E_j^i(t)$, where i applies for the cue integration network, and j for the neural field of the motor selection region corresponding either to the left or right finger. W^{ij} corresponds the strength of the projections.

In addition, with respect to the task instructions, this connectivity may be modified so that different stimulus-response associations can be obtained. However, associations that are different from spatially compatible responses, i.e. a left/right movement in response to a left/right stimulus, have been shown by behavioral studies to produce longer reaction times (Hedge & Marsh, 1975; Hasbroucq & Guiard, 1991; Proctor & Pick, 2003). Therefore, an artificial

processing time Δ is introduced in the system such that the equation described above becomes $h^j(t) = W^{ij} E^i(t - \Delta)$. When the mapping is natural, $\Delta = 0$, but otherwise, it is set according to behavioral data (See Proctor and Pick (2003) and Appendix B.1).

RELATIONSHIPS OF THE MODELS WITH CORTICAL PATHWAYS

The models presented above describe several cortical pathways which are responsible for the behavioral expression of automatic imitative tendencies. The correlates of each part of the models with neurophysiological data are given next.

As shown in Figure 4.2, the visual perception of the abstract and movement cues is assumed to be first processed by early visual areas, i.e. V1 to V5 as well as the medial temporal cortex (MT/MST) (Hubel & Wiesel, 1977; De Yoe & Van Essen, 1988). Then, the cortical route differs for both cues. While the identification of abstract cues may involve the inferotemporal cortex (IT) (Booth & Rolls, 1998; Ashbridge et al., 2000), that of movement cues may rather recruit the superior temporal sulcus (STS) (Perrett, Harries, Bevan, et al., 1989; Oram & Perrett, 1996; Jellema et al., 2004). Importantly, these areas are known to project high-level descriptions of the visual scene to the prefrontal areas responsible for decision making, namely, the lateral prefrontal cortex (LPFC) (Lauwereyns et al., 2000; Sakagami et al., 2006).

In parallel, the cortical pathway involved in the processing of observed movements in sensorimotor terms is suggested to flow more dorsally along the parieto-frontal network (Rizzolatti et al., 2002). The representation of observed movements, processed in STS which is specifically activated by the sight of movements of the body and limbs (Perrett, Harries, Mistlin, & Chitty, 1989; Jellema et al., 2004), is then projected, through the posterior parietal cortex (PPC), to the cortical network in charge of preparing and executing movements. This network is known to be partly formed by the anterior medial frontal cortex (AMFC), the supplementary motor area (SMA), and the premotor cortex (PM) (Alexander & Crutcher, 1990; Rizzolatti & Luppino, 2001; Decety et al., 2002; Brass et al., 2005). Importantly, the human cortical areas which exhibit both motor preparatory and mirror activity, i.e. PM and SMA (Fogassi & Gallese, 2002; Rizzolatti et al., 2002), are here believed to correspond to the network which is responsible for the manifestation of the ideomotor principle.

As rapidly mentioned above, LPFC is believed to integrate a high-level description of the environment in order to produce deliberate decisions in response to visual stimuli (Watanabe, 1986; Sakagami et al., 2001). Interestingly, the role of this region is not to produce motor command per se, but rather to produce go or no-go signals for motor commands prepared within the motor cortices (Lauwereyns et al., 2000; Sakagami et al., 2006). Importantly, this work asks whether the decisional process within LPFC is influenced by imitative cues relayed by the dorsal processing route, or whether these cues only interfere at the

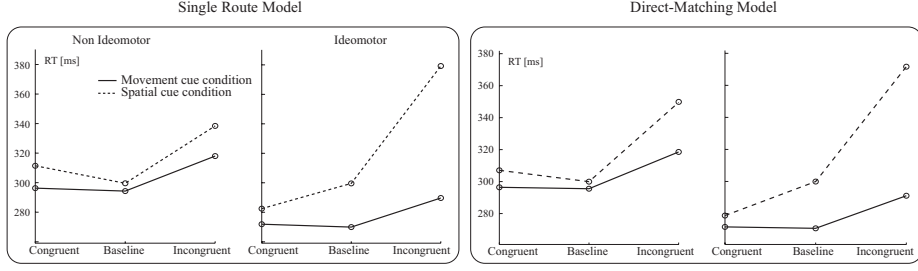


Figure 4.4: Simulation results of the two models under the same conditions as in the experiment by Brass et al. (2000). The time scale on the y-axis is not shown as the models do not simulate real cortical processing time. Both models are in good agreement with behavioral data reported in Figure 4.1.

level of the motor selection in the ventral premotor cortex (PMv). As will be presented in Section 4.3.2 a method to clarify this issue could be to perform an experimental study which implies an increase of the computational load in LPFC.

4.3 RESULTS

4.3.1 REPLICATION OF THE RELATED EXPERIMENTAL STUDY

The two models were simulated under the same conditions as those used by Brass et al. (2000) in their behavioral experiments. The system parameters, i.e., the amplitude of the inputs, the recurrent profile of the neural fields as well as the strength connectivity patterns of the networks, were initially tuned according to the hypotheses of this work, which were presented in Section 4.2.1. Then, they were fine-tuned using a gradient descent method in order to minimize the error between the behavioral data and the simulation results. Moreover, since it is beyond the scope of the present study to account for the precise time course of neural processes, simulated RTs were also fitted to the original data using a first order least squares error regression method. The thereby obtained simulation parameters are summarized in Appendix B.1.

The measured reaction times performed by the models are shown in Figure 4.4. As can be seen, the two models are in good agreement with the original data and hence exhibit a similar behavior. Indeed, despite their conceptually different architectures, all the necessary components which determine the interactions between the perceptual parts of the stimuli are identical. All the processing stages are treated similarly and the differing connectivity only forces stimuli interactions to be processed in different stages.

More details concerning the dynamics of the second network are illustrated in Figure 4.5. Only the spatial cue task condition is shown, as it best represents the model’s interesting characteristics. Firstly, it can be seen in the cue

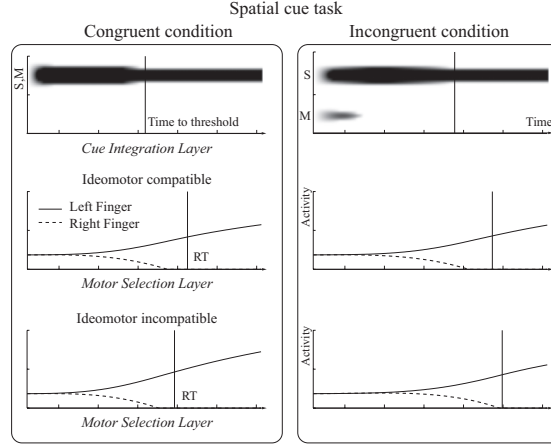


Figure 4.5: Dynamics of the direct-matching model. The activity profiles of the cue integration population are shown on top. Dark areas correspond to regions of neural activity and the vertical lines denote the time at which the population activity reached a certain threshold. Labels S and M on the y-axis indicate the spatial location of spatial and movement cues respectively. At the bottom, the time profile of the neuron with maximal activity of each motor selection area is shown. The vertical bar indicates the time at which the execution threshold E_0 was reached.

integration layer that the selection process takes longer in the incongruent condition than in the congruent condition. Indeed, the selection mechanism must inhibit the incompatible movement cue. Similarly, ideomotor compatibility has also an influence on RTs. In the congruent conditions, the slope of the motor selection activity profile is sharper in the ideomotor case as compared to the non-ideomotor case. The selection process is thus faster in the former condition, where the ideomotor region is facilitating the selection process. Conversely, in the incongruent condition, the ideomotor region favors the response in an opposite fashion to the one given by the cue integration layer. This interference effect slows down the final decision process, which causes the increase of the RTs.

4.3.2 STIMULUS-RESPONSE INCOMPATIBLE MAPPING

Since the previously reported simulation results of the two models are barely distinguishable (see Figure 4.4), this shows that the used experimental paradigm cannot clearly discriminate between these two plausible architectures. Therefore, it is important to devise a method for determining which model best reflects the information pathways of the brain.

To achieve this, inspiration was taken from the large amount of literature on stimulus-response compatibility. The experimental paradigm was modified in order to instruct the models to perform an incompatible stimulus-response mapping. The task is to respond to a left cue with a motion of the right finger (middle finger), and conversely to respond to a right cue with a motion of the left

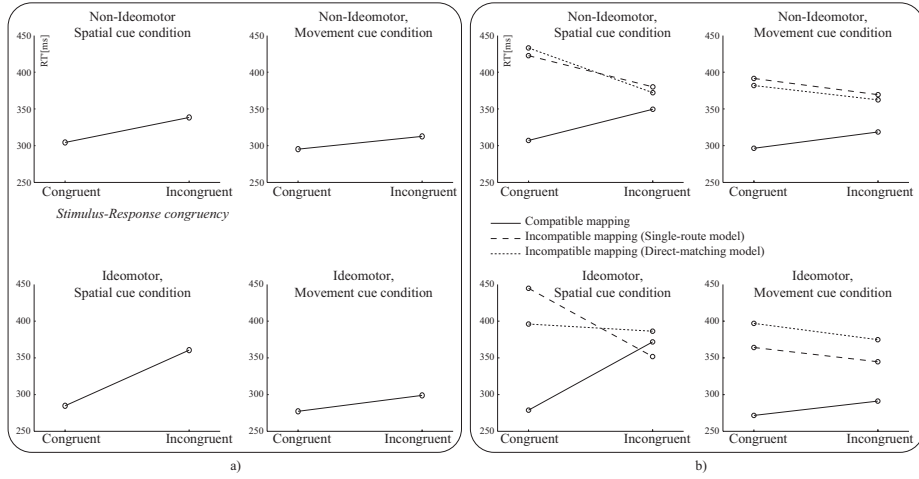


Figure 4.6: **a)** Results of Brass et al. (2000) redrawn, in order to account for the congruency of the irrelevant stimulus location with the response, as in a typical Hedge and Marsh experiment (Hedge & Marsh, 1975). The baseline conditions were omitted to conform to that notation. The labels on the x-axis correspond to compatible and incompatible relationships of the irrelevant stimulus location with the response. **b)** Predictive results of the models while confronted to an SR incompatible mapping task.

finger (index finger). This switching paradigm has already been developed in experimental psychology in order to give more insights on a well-known behavioral effect: the so-called Simon effect³ (Simon et al., 1981). Indeed, switching task instructions from a compatible to an incompatible stimulus-response mapping has been shown to result in a reversal of the classic Simon effect Hedge and Marsh (1975). Several explanations of this phenomenon have been proposed (Simon & Berbaum, 1990; Hasbroucq & Guiard, 1991; Proctor & Pick, 2003). The most relevant argument to the present modeling is the stimulus-stimulus congruency hypothesis. It stresses that facilitatory and interference effects are mainly caused by the integration of spatial cues occurring in an intermediate processing level, rather than in late motor preparation stages. This hypothesis suggests that the reversal effect may be produced by a higher cognitive mechanism, involved in the incompatibility inversion process and that it should occur during the stimulus-response mapping process. As suggested in Section 4.2.1, this inversion may lead to an increase of the computational load of the brain area in charge of this process. Therefore, this mechanism was implemented by switching the wiring of the stimulus-response association module and by applying an additional processing time.

In Figure 4.6a, the results of Brass et al. (2000) were redrawn, so as to account for the compatibility of the irrelevant stimulus location with the response, as in a typical Hedge and Marsh experiment (Hedge & Marsh, 1975). Figure 4.6b shows the predictions of the two models, which display now different be-

³The Simon effect relates to the observation that even if the stimulus location is an irrelevant dimension, a spatial congruency between that irrelevant stimulus dimension and the response significantly facilitates the initiation of the response.

haviors. The first model reproduces the classic reversal effect, i.e. the relative reaction times between the two conditions is reversed in contrast to the original data. In other words, the RTs are faster when the irrelevant stimulus is located at the opposite side as that of the motor response.

The second model exhibits qualitative differences. In the ideomotor and spatial cue condition the reversal effect is reduced. This observation was expected as the direct route between action observation and action execution permanently activates the motor preparation centers in an ideomotor compatible way. This effect can be seen in Figure 4.7, where a comparison of the neural dynamics of the second model between the SR compatible and incompatible mapping conditions. It shows that even in the incompatible mapping task an ideomotor compatible movement observation still strongly favors the corresponding movement execution, as the finger movement stimulus was unaffected by the inversion process required by the task. Moreover, in the ideomotor and movement cue condition, the overall RT is increased. A close look at Figure 4.7 provides the explanation. Firstly, the movement observation is used to determine the correct response, which is always on the opposite side. However, at the same time, the ideomotor system enhances the spatially matching finger movement, and as these two parallel processes always favor opposite responses, an interference effect is constantly present and results in an overall RT increase.

The reverse phenomena, although less significant, can be observed during the non-ideomotor conditions, see Figure 4.6. Indeed, in the spatial cue condition, the Hedge and Marsh reversal effect increases slightly, whereas in the movement cue condition the overall RT decreases slightly. This can be explained by the fact that, in the models, when the observed/planned movement-pair is ideomotor incompatible, the network tends to favor, but to a lesser extent, the execution of the finger opposite the observed one. This opposite facilitatory effect therefore reverses the interference observed in the ideomotor condition.

4.3.3 METRIC OF SPATIAL REPRESENTATION

Predictions of the models concerning a variation along the metric of the representation of spatial cue are presented. This was made possible by the continuous, rather than discrete, representation of spatial and motion cues used in the models. This modeling hypothesis, which follows the Dynamic Field approach, predicts modifications in reaction times when stimuli are displaced according to their representation metric. A new experimental paradigm is now proposed in order to measure this effect.

It consists of the same paradigm as employed by Brass et al. (2000), with the difference that the spatial cue, i.e. the cross, is no longer placed on a specific fingernail, but in a variable position between that fingernail and the midline of both fingers (see Fig. 4.8a). Simulation results showing the mean RT and

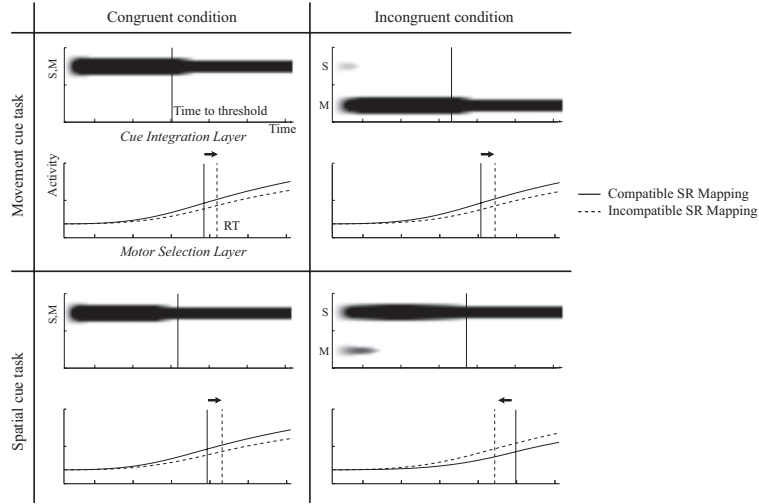


Figure 4.7: Illustration of the dynamics of the direct-matching model as in Figure 4.5. For clarity reasons, in the plots showing the maximal neural activities of the motor selection areas the profile of the less active population (i.e left or right finger) was omitted. The SR compatible and incompatible mapping are compared for ideomotor compatible observation-execution movement pairs. Moreover, both spatial and finger cue conditions are shown. The additional delay of the inversion process was omitted here, in order to allow an easier comparison of the results. Finally, the small black arrow shows the effect, in RT, of switching from a compatible to an incompatible mapping task. For comments on the figure, see text.

the RT difference between incongruent and congruent conditions as a function of the relative location of the spatial cue are presented in Figure 4.8b. The horizontal axis represents the normalized spatial cue location so that a value of 0 corresponds to the fingernail position and 1 to the midline.

Firstly, it can be seen in both ideomotor and non-ideomotor conditions, that the mean RTs of the conditions requiring a response to the movement cue and to the spatial cue change in an opposite fashion according to the spatial cue location. Indeed, as the location of the spatial cue moves toward the midline, its neural representation moves away from the finger location it corresponds to. Consequently, its interfering effect in the movement cue condition, and triggering effect in the spatial cue condition is reduced, producing respectively a decrease and an increase of the mean RT. Moreover, since the effect of the spatial cue is smaller in the ideomotor case, the decrease of the mean RT in the movement cue condition is less significant than in the non-ideomotor case. Second, the variation of the RT difference between incongruent and congruent trials is described. During tasks requiring a response to a movement cue, a similar decrease in RT difference in both ideomotor and non-ideomotor conditions is observed. As mentioned above, a spatial cue getting closer to the midline reduces the interfering effect of its internal representation on the decisional process. Therefore, the original RT difference between congruent and incongruent conditions gradually disappears and tend to zero. The reported effect in tasks

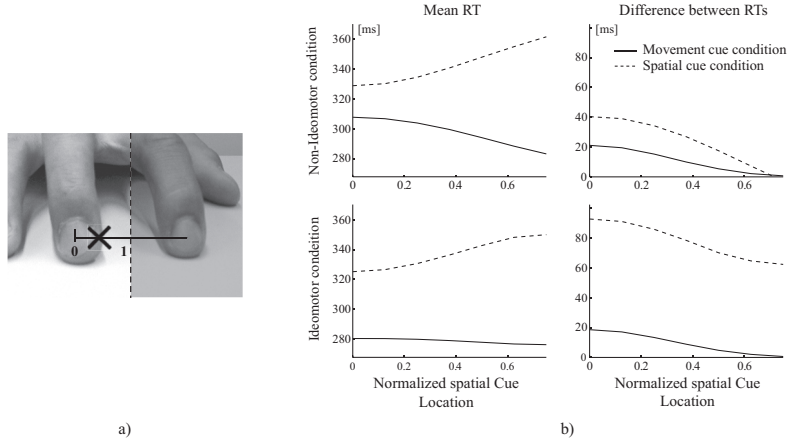


Figure 4.8: **a)** The modification of the original experimental paradigm as used by Brass et al. (2000) is illustrated. **b)** The predictions on the mean RTs and their difference is shown while varying the location of the spatial cue.

involving a response to the spatial cue is similar. In both ideomotor and non-ideomotor conditions, the RT difference does also decrease due to the reduced compatibility between the location of the cue and that of the response. When the spatial cue is not located on a fingernail, it takes more time to the system to select the correct response. Thus, the interference effect of produced by the movement cue is less pronounced in terms of reaction times, even if it is still present internally. As can be noticed in the non-ideomotor condition, the interference in terms of the RT difference even disappears. Indeed, the time needed by the system to trigger the response is greater than that needed to suppress the interference due to to movement cue. In the ideomotor condition, since the interference effect is greater than in the non-ideomotor one, it is still observable.

4.4 DISCUSSION

This work addressed the cortical pathways subserving the neural mechanism underlying human imitation. A computational neuroscience approach was applied on paradigms developed in experimental psychology. Two biologically inspired computational models capable of reproducing the experimental results obtained by Brass et al. (2000) were presented. These models are in line with other computational models addressing imitation mechanisms in both humans and monkeys (Arbib et al., 2000; Demiris & Hayes, 2002; Wolpert et al., 2003), in that they all assume a shared representation between movement observation and action execution. In contrast, this modeling work is also largely inspired by human behavioral phenomena reported in experimental psychology such as the Hick's law, the Simon effect and the Hedge and Marsh reversal effect (Hick, 1952; Hedge & Marsh, 1975; Simon et al., 1981). Furthermore, although one may cast some doubt on the usefulness of the complexity of the models to reproduce

the original data, it is important to understand that each computational stage of the models is necessary in order to obtain a good fit of the behavioral data. Indeed, based on the well-known dynamical properties of the neural field model (see Section 3.2.2), the investigation of the possible network connectivity has led to the conclusion that the architecture of the networks presented here is as minimal as possible to account for the data. Nevertheless, according to this, one may then question the generality of the results provided by the original experiment on which this work is grounded.

Next, although the architecture of the two models differed in their information pathways, the ideomotor compatibility principle was successfully displayed by both models. The fundamental mechanism subserving the ideomotor principle is based on a comparison of observed action with internally planned actions. The difference between the models is that the direct-matching model relies on the hypothesis that a direct route between movement observation and movement execution exists (Rizzolatti et al., 2001; Decety & Sommerville, 2003), whereas the single-route model forces movement observation to be merged with the other cues. In other words, the former model hypothesizes that the spatial relationship between relevant and irrelevant stimuli is processed first at an intermediate level, while the comparison between planned and observed motions is conducted separately but simultaneously through an ideomotor pathway. Both pathways then get merged in a final stage. In contrast, the latter model merges all the information simultaneously. Since both models are biologically plausible and do reproduce behavioral results, this raises the question of which model might correspond best to the real cortical network.

As suggested in the presentation of the architecture of the models, to each sub-network of the complete system was associated a specific brain area with respect to the neurophysiological literature. Each experimental condition⁴ involves the activation of all parts of the models, which respects the brain activation patterns as reported in several studies of imitation using fMRI. However, this brain imaging technique only reveals which brain areas are active at a given time. Consequently, the exact pathways followed by information across simultaneously activated processes can not be uncovered by such a technique. An alternative form of study, which could provide an answer to the question raised by this modeling work, may use the analysis of the impairments following from specific brain lesions. Here, in order to discriminate between the proposed models, a novel stimulus-response experiment is proposed. The experimental protocol of this new experiment is similar to that conducted by Brass et al. (2000), except that the subjects are asked to respond to any of the spatial cues with an incompatible response, i.e. they should respond to a left (right) cue with a right (left) finger movement. If that experiment was to be conducted and a strong Hedge and Marsh reversal effect was to be measured, this would refute the direct-matching model and let the single-route model appear to be

⁴except the baseline conditions

the most plausible. However, if one would observe the Hedge and Marsh reversal effect, this may not necessarily refute the plausibility of the direct-matching model. Indeed, the automatic imitative tendency can be easily overridden after only a brief training (Heyes et al., 2005; Bertenthal et al., 2006). Therefore, it would be essential that the experiment be performed with subjects who remain totally unfamiliar with the task. Subsequently, it would be interesting to see if such SR incompatible training may result in an effective suppression of the ideomotor effect.

The results of the simulations also showed that the observation of a non-ideomotor compatible finger movement results in a slight facilitation of the initiation of the opposite finger. This anti-facilitatory effect seems questionable, as, to current knowledge, such a phenomenon has never been reported. This may be a weakness of the models. The main reason for such effect to occur is that, in the motor selection area, the process of ideomotor facilitation and interference acts within the same metric as that of the decisional layer, i.e. by considering the global amplitude of the population activity. Then, as the competition among the motor plans is performed in that metric, the decrease in amplitude caused by an ideomotor incompatible observed movement favors the execution of the opposite finger's motor plan. This raises the question of if, in an effector selection task, the observation of a non-ideomotor movement with a given effector, does negatively influence its selection, as shown by this computational study. Indeed, it was assumed that both observed and planned finger movements, such as tapping and lifting, are encoded by means of directional information within the same neural layer, a layer which also produces internal competitive interactions. The principal arguments in favor of this hypothesis are the facts that parts of the premotor cortex respond equally to observation and execution of finger movements (Iacoboni et al., 1999), that movements in this cortical area are partly encoded by means of directional information (Schwartz et al., 1988), and finally that neural correlates of decisional processes have also been reported in this brain region (Cisek & Kalaska, 2005).

At the first glance, these arguments may seem to contradict with the non-observation of such an anti-facilitatory effect in the behavioral literature. Nevertheless, an alternative hypothesis for which the present models do not account for, could be suggested. The neural processing of the ideomotor compatibility may be driven by an additional mechanism on top of that proposed here. Instead of only considering the dynamic aspect of the movements, another mechanism may favor the use of the same effector as that observed. This process would then counteract and hence suppress the anti-facilitatory effect which was reported by the present study. This hypothesis remains yet to be verified.

This work also assumed a continuous representation of visual stimuli. This hypothesis is mainly based on neurophysiological evidence indicating that large sets of neurons encode visual information in a distributed fashion, as reported in brain areas such as the visual cortices and the superior temporal sulcus (Perrett,

Harries, Mistlin, & Chitty, 1989; Andersen et al., 1997; Schadlen & Newsome, 2001; Jellema et al., 2004). Similarly to motor representations, competitive interactions influence dynamically the internal representations of the visual inputs. In order to investigate whether such a continuous metric of stimulus representation may faithfully corresponds to real brain mechanisms, a variation of the original experiment was proposed. The models predict that a spatial cue which location gets away from that of the fingers reduces its interfering and triggering effect. As a consequence, it accelerates the initiation of the response to movement cues and increase the reaction times to spatial cues. Moreover, it also reduces the RT difference between congruent and incongruent conditions for similar reasons. The present results also suggest that a maximum operation-like effect might be involved between ideomotor and spatial compatibility processes, the observed reaction times reflecting the time needed by the slower process to either select the correct response or to inhibit the incongruent stimulus.

Finally, this modeling study assumed that the movements performed by others are represented and encoded within the same frame of reference as self-generated movements, which is in accordance with the direct-matching hypothesis and the firing patterns of mirror neurons (Meltzoff & Moore, 1997; Rizzolatti et al., 2001). Consequently, the spatial relationships at stake in these experiments were assumed to be encoded in a limb-centered frame of reference. The neural mechanisms underlying the transformation of visual information related to the movements of others into a representation in ego-centric coordinates was not considered here. However, since in this experiment the observed hand is facing like in a mirror, the effect of this transformation was assumed to be negligible. It is nevertheless an important issue in other situations where the transformation is not constant or when several types of transformations may be involved. This issue is addressed in the next chapter, which describes plausible neural models allowing the transfer of information across different frames of reference, and their implications on cortical sensorimotor processes.

Summary

This chapter presented an investigation of the neural pathways and mechanisms responsible for automatic imitative behaviors. The goal of this study was to illuminate the cortical processes underlying the behavioral expression of the ideomotor principle by evaluating how the observation of others influences the quality of one's own performance. How does the observation of movements performed by others interfere with or facilitate one's own motor execution? One hypothesis states that movement observation follows almost the same cortical pathways as those mediating classical stimulus-response processes, and that a high-level cognitive process is responsible for processing all these information together in order to make decision. A second hypothesis suggests that a direct route - independent from the route that mediates the decisional process - links action observation processes with motor execution centers. Both models are capable of reproducing the behavioral results reported by Brass et al. (2000) in a stimulus-response task comprising abstract and imitative stimuli. Both models also integrate the principle of ideomotor compatibility, since they were built based on the existence of a shared representation between movement observation and action execution. In order to discriminate between these two biologically plausible models, a novel stimulus-response experiment, derived from that proposed by Brass et al. (2000), was suggested. The idea behind this experiment consists of requesting that the models modify their original stimulus-response associations, and hence respond to the instruction stimuli in an inverted manner. Since this reversal process is believed to recruit a network that is independent from the network mediating automatic imitative behaviors, one may thus see at which level imitative responses and those associated to abstract stimuli are combined to produce the final motor response. If the reversal equally affects the responses to both types of cues this would suggest that a single selection process is at play. However, if it only affects responses to abstract cues, this would support a direct sensorimotor route between action observation and motor execution - a route that bypasses motor selection processes.

In this study observed movements were assumed to be encoded within the same frame of reference as self-generated movements. Despite the fact that this modeling hypothesis accords well with the hypothesis of a common shared representation between self and others' actions, it may call for an explanation as to how the visual information related to others' movements can be transformed into a representation compatible with that of one's own movements. Although the effect of the transformation in this experiment could be assumed to be negligible because of the configuration of the experimental setup, in other situations it may nevertheless have a critical impact on human behavior. Consequently, this issue is addressed in the next chapter, which describes plausible neural models that allow the transfer of information across different frames of reference.

FRAMES OF REFERENCE TRANSFORMATIONS

The modeling study presented in this chapter consists of an extension of the work which was published in:

Sauser, E. L. and Billard, A. G.. Three dimensional frames of references transformations using recurrent populations of neurons. *Neurocomputing*. 64:5-24, 2005.

Sauser, E. L. and Billard, A. G.. View sensitive cells as a neural basis for the representation of others in a self-centered frame of reference. *In Proceedings of the Third International Symposium on Imitation in Animals and Artifacts, Hatfield, UK*. 119-127, 2005.

THE modeling study described in the previous chapter made an important assumption concerning the representations on which the neural information processes take place. It was hypothesized that the frame of reference used by the brain for imitation is common to most of the cortical areas described. Indeed, the neurophysiological evidence of the existence of shared neural structures implies that they should communicate within the same representation metric. However, since visual information related to gestures or actions performed by others enters the brain in a retinocentric frame of reference, it must first be converted into a representation in a reference frame compatible with the encoding of motor representations. Since this frame of reference has been suggested to be primarily goal- or body-part- centered, the generated motor commands must also be transformed into a coordinate system consistent with that of the muscles.

In order to fill this gap, the problem of the transfer of information across different reference frames must be addressed. This is precisely the aim of the work described in this chapter, which is mainly concerned with how cortical networks may produce these fundamental transformations. After an introduction to the general cognitive issues related to this problem, including their neurophysiological foundations and its formal description, the results of this modeling work are presented. There are several models, each built by combining the building blocks underlying such transformations. The models mainly consist of variations around the same theme, but also may be involved in different situations.

5.1 INTRODUCTION

The transfer of information across multiple frames of reference is a problem that the nervous system has to face in many cases. For instance, visuomotor tasks, such as visually-guided movements, require that visual information gathered in a retina-based frame of reference be transferred into a head-centered, a body-centered, and a hand-centered reference frame. The manipulation of objects also requires retinal information to be transferred into an object-centered representation. Moreover, when tightly linked together, proprioceptive, vestibular, visual, and sensorimotor representations are suggested to be the basis of the body-schema, a meta representation of the body in the environment (Gallagher, 2000). Since these representations are encoded in various reference frames, transformations are necessary to build a coherent view of the self. Other cognitive skills such as motor imagery and mental rotation, are also dependent on the brain faculty to perform such transformations.

Evidence that the brain encodes part of the visual and motor information in different frames of reference is corroborated by a number of neurophysiological experiments (Batista et al., 1999; Kakei et al., 1999; Ashbridge et al., 2000). Links between these representations were hence supposed to be grounded on the combination of various sensory inputs which further allows the transfer of information across different reference frames. For example, it has been suggested that along the visuo-motor pathway the series of transformations, required for transferring retina-based information into a body-centered representation, are computed gradually by different groups of cells following a sensory gradient of increasing complexity (Perrett, Harries, Mistlin, & Chitty, 1989; Burnod et al., 1999). Considering now the ventral visual pathway, information has been shown to flow from the primary visual cortex (V1) to the temporal lobes, including the inferior temporal area (IT) and the superior temporal sulcus (STS). While the primary visual areas encode visual information in a retina-based coordinate frame, neural sensitivity gradually moves from this purely sensory-centered representations toward more complex ones. For example, IT contains populations of neurons that separately exhibit a sensitivity to a variety of objects. Some of these neural ensembles are sensitive to the size and orientation relative to an viewer-centered frame of reference, whereas others react more in an object-centered fashion (Booth & Rolls, 1998). Similarly, neurons in macaques' STS, have been found to respond to specific human body parts and correlate with various quantities such as the position, rotation and translation of limbs, hands, faces and eyes, in multiple reference frames. Specific neural responses to complex motions such as walking have also been reported (Perrett, Harries, Mistlin, & Chitty, 1989).

In imitation, in addition to strict motor control issues, abilities for transformations across different frames of reference are also necessary for transfer-

ring the representation of the perception of others movements into self-motor representation, which in turn would allow imitation. Different forms of transformations may be useful. First, goal-centered representations are prerequisites for goal-directed imitation. They are a high-level abstract substrate which can be readily be obtained by projecting the information related for instance to the position and posture of the hand into the coordinate frame of the target object. The result of this process is suggested to be responsible for the invariant response of mirror neurons which sensitivity was reported to be goal-centered (Fogassi & Gallese, 2002). However, the neural correlates of imitation are often reduced to the sole goal-centered activity patterns of the mirror neurons, which, for recall, were recorded in the monkey brain. In human, the analysis of the activity patterns of specific cortical areas involved during imitation, revealed that the human mirror system may also show a sensitivity to body parts-centered representations (Iacoboni et al., 2001; Koski et al., 2003). Thus, it has become necessary to address the mechanisms underlying the imitation of intransitive actions or movements. This however raises several important computational problems which can be avoided while considering goal-directed imitation. Importantly, in order to be able to imitate body movements, the visual information representing the configuration of the limbs of another individual, which are perceived in a viewer-centered frame of reference, must be converted into a joint-centered representation.

To summarize, the ability to perform arbitrary transformations across frames of reference is fundamental and necessary for more complex forms of imitation. However, its underlying mechanisms remains still ill-understood and hence, its understanding should be considered as important in order to provide the stages for modeling the leap from simple to complex forms of imitation in animals. In addition, it may also provide fundamental mechanisms on which higher cognitive skills such as mind-reading may develop. Indeed, transformations across different reference frames, are necessary steps to be able to put oneself in the feet of others. Next, before addressing precisely the neural mechanisms underlying the brain ability to perform transformations across reference frames, this problem, along with its solutions, is first stated in mathematical terms.

5.1.1 THE PROBLEM

The transformation across reference frames that is considered here basically consists of transferring a vector representation from an orthonormal basis to another. An illustration is given in Figure 5.1. In mathematical terms, the question is how a vector \vec{v} given in a referential \mathcal{R} can be transformed into \vec{v}' in \mathcal{R}' , knowing the vector \vec{v}_0 across the origins of the two reference frames, and the axes of the referential \mathcal{R}' expressed in \mathcal{R} . Let assume \mathcal{R} and \mathcal{R}' to be given

by

$$\begin{aligned}\mathcal{R} &= \{O, \vec{e}_x, \vec{e}_y, \vec{e}_z\} \\ \mathcal{R}' &= \{O', \vec{e}'_x, \vec{e}'_y, \vec{e}'_z\}\end{aligned}\quad (5.1)$$

where $\overline{OO'} = \vec{v}_0$, and $\vec{e}'_i, \vec{e}_i, i \in \{x, y, z\}$ correspond to the principal axes of \mathcal{R}' and \mathcal{R} , respectively. The orientation of \mathcal{R}' with respect to \mathcal{R} is then given by the transformation matrix $\mathbf{M}^{\mathcal{R}'}$, where

$$\mathbf{M}^{\mathcal{R}'} = \left(\begin{array}{c|c|c} \vec{e}'_x & \vec{e}'_y & \vec{e}'_z \end{array} \right). \quad (5.2)$$

By writing down the classical transformations across reference frames, the following forward and inverse equations can be obtained.

$$\vec{v} = \mathbf{M}^{\mathcal{R}'} \vec{v}' + \vec{v}_0 \quad (5.3)$$

$$\vec{v}' = \left(\mathbf{M}^{\mathcal{R}'} \right)^{-1} (\vec{v} - \vec{v}_0). \quad (5.4)$$

The projection of a vector \vec{v} from \mathcal{R} to \mathcal{R}' can be thus decomposed into one translation across the origins and a transformation $(\mathbf{M}^{\mathcal{R}'})^{-1}$. Then, depending on the available information, $\mathbf{M}^{\mathcal{R}'}$ can be written, for example, as a series of three rotations, such that

$$\mathbf{M}^{\mathcal{R}'} = \mathbf{R}_x(\phi_x) \mathbf{R}_y(\phi_y) \mathbf{R}_z(\phi_z) \quad (5.5)$$

where $\mathbf{R}_i(\phi_i), i \in \{x, y, z\}$ correspond to the rotation matrices around axis i with angle ϕ_i . Following this, the inverse of $\mathbf{M}^{\mathcal{R}'}$ becomes

$$(\mathbf{M}^{\mathcal{R}'})^{-1} = \mathbf{R}_z(-\phi_z) \mathbf{R}_y(-\phi_y) \mathbf{R}_x(-\phi_x) \quad (5.6)$$

Therefore, the transformation can be performed serially through *a translation followed by three consecutive rotations*. However, an alternative way to perform the transformation can also be proposed. Since $\mathbf{M}^{\mathcal{R}'}$ is considered as orthonormal, $(\mathbf{M}^{\mathcal{R}'})^{-1} = (\mathbf{M}^{\mathcal{R}'})^T$. Equation (5.4) can then be rewritten by using the dot product, which gives:

$$\vec{v}' = \sum_{i \in \{x, y, z\}} (\vec{e}'_i \cdot (\vec{v} - \vec{v}_0)) \vec{e}_i. \quad (5.7)$$

By applying this second solution, the overall transformation is reduced to a *combination of relatively simple vectorial operations, consisting of sums, dot products, and vector scaling*.

By means of these two methods, a vector \vec{v} given in a referential \mathcal{R} can be mapped into \vec{v}' in \mathcal{R}' . In the following, neural mechanisms allowing cortical networks to perform these operations are suggested.

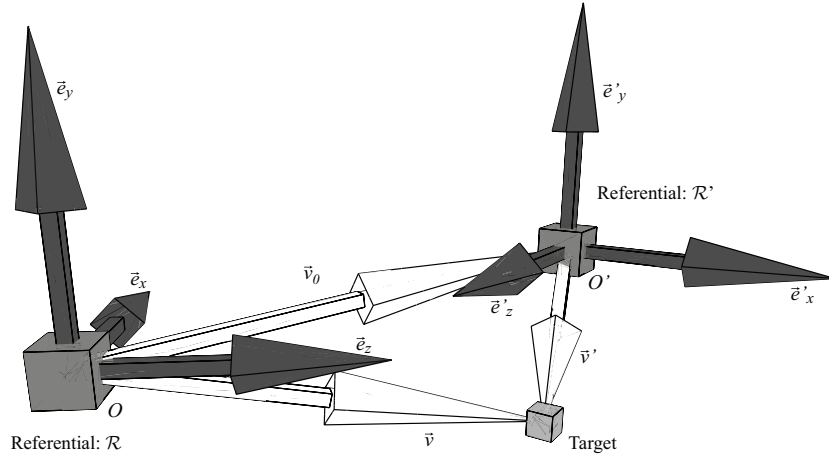


Figure 5.1: A transformation across reference frames consists of transferring a vector \vec{v} given in a referential \mathcal{R} into \vec{v}' in \mathcal{R}' .

5.2 A FUNDAMENTAL NEURAL BUILDING BLOCK

This modeling work started from the idea that sensory and motor information is represented in a distributed fashion within populations of neurons. As already mentioned in previous chapters, this computational paradigm has been shown to be shared by several areas of the nervous system, including proprioceptive receptors, (Ribot-Ciscar et al., 2003), the motor cortex (Schwartz et al., 1988; Kakei et al., 1999), the posterior parietal cortex (Batista et al., 1999; Scherberger & Andersen, 2003) and the superior temporal sulcus (Ashbridge et al., 2000). Moreover, the population vector coding has also been shown to be a plausible way to interpret the macroscopic effect of the joined activities of these large sets of neurons (Georgopoulos et al., 1982; Schwartz et al., 1988). Among the types of information that are encoded in such distributed representations, movement parameters such as the direction of movement, the velocity and the spatial location of the body and body parts were shown to be represented in both motor and visual terms. The neural encoding of objects revealed a sensitivity to similar parameters. Importantly, these representations were shown to be encoded within multiple frames of reference, ranging from viewer-centered cartesian coordinates to muscle-centered reference frames. Thus, distributed representations exhibiting a population vector-like encoding appear to be an important principle of brain information processing.

Considerable attention has already been devoted to this problem and, consequently, several relevant solutions have been proposed. The key point, common to almost all studies, relies on the multiplicative response observed in large sets of neurons which combine non-linearly different information sources such as eyes

position in their orbit and the retinal location of an object. Combining these information sources has been shown to be sufficient to compute the location of the object in head-centered coordinates. Here, the present investigation seeks whether population vector coding can be used as a principle mechanism to accomplish frames of reference transformations. The majority of related works (Salinas & Abbott, 1995; Burnod et al., 1999; Baraduc & Guigon, 2002) model the non-linear multiplicative response of the neurons by either assuming that the activation function of the neurons produces a multiplicative response of the neuron’s input (sigma-pi neurons) or that the synaptic strength between two neurons could be gated by a third part neuron. In contrast, the present approach follows the work of Salinas and Abbott (1996). The multiplicative property of the population output is derived from the concurrent activity of a population of integrative neurons. Indeed, keeping the integrative properties of neurons is fundamental to remain in line with a biological account of the neural response. Prior studies considering frames of reference transformations using population codes with strictly additive synaptic inputs (Salinas & Abbott, 1996; Deneve et al., 2001; Van Rossum & Renart, 2004) have overlooked a major effect: the population vector, resulting from such transformation, exhibits a difference in its amplitude relative to its original value. The need to consider this issue is motivated by the importance of the amplitude of the signal conveyed by distributed representations. For instance, in the motor cortex, the global strength of the population response to a being executed movement represents the instantaneous velocity of the end-effector, as well as its position in spherical coordinates with respect to the center of its workspace (See Figures 2.7, 2.8 and 5.3 for illustrations of the information conveyed by the population response.) (Kettner et al., 1988; Schwartz & Moran, 1999; Aflalo & Graziano, 2007). Consequently, transformations of these signals have to ensure that the part of the information considered here is not affected in an undesired manner. To fill this gap, this work investigates a method by which one can reduce and constrain the error produced by the neural field dynamics within strict and acceptable bounds.

5.2.1 VECTORIAL REPRESENTATIONS AND COSINE-TUNING

In order to transform vectorial information through neural mechanisms, one has first to consider a neural code capable of representing vector. Among the different methods, distributed representations allowing population vector techniques are considered. In contrast to the general form of the tuning which was previously described in Chapter 3, the model of neuron was chosen here to exhibit a cosine tuning to vectorial information. This choice was motivated by two important points. First such a tuning for vectorial representations has been shown to be highly probable since it has been suggested to be the most opti-

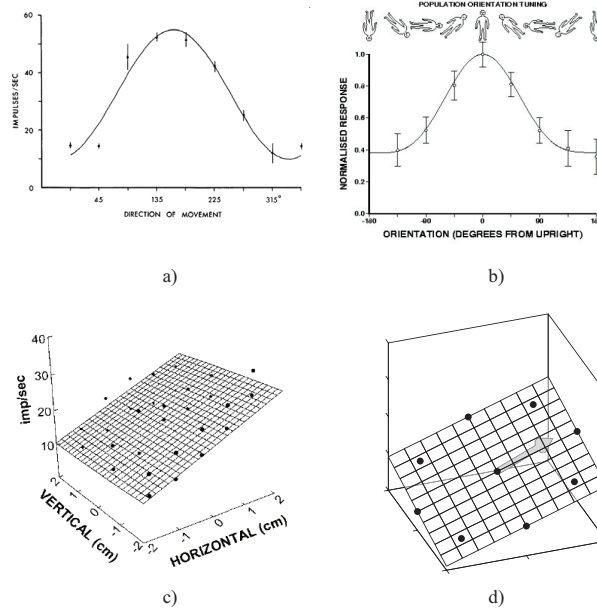


Figure 5.2: Neural correlate of population vector coding. Tuning curves of **a)** a neuron in the primary motor cortex sensitive to the movement direction of the hand, and **b)** a neuron in the superior temporal sulcus responsive to the orientation of an observed body. **c)** Activity of a cerebellar neuron sensitive to the foot position recorded in the rat. A plane can be fitted to the data, which shows a linear relationship between the foot location and the neural activity. **d)** Simulated neuron sensitive to a position in space. It was also encoded using a cosine tuning centered on its preferred direction indicated by the light arrow. (Adapted from Ashbridge et al. (2000); Casabona et al. (2004); Georgopoulos et al. (1988)).

mal way to process visual and motor information (Zhang & Sejnowski, 1999; Todorov, 2002). In Figure 5.2 neural correlates of such a cosine-like encoding are illustrated. The second point is less related to biology. Indeed, cosine tuning curve have shown a better compliance to analytical calculations (Ben-Yishai et al., 1995; Xie et al., 2002). Although the dynamics of neural fields has already been described at length in Chapter 3, a non negligible amount of definition will be rewritten in this chapter. Indeed, since the subclass of neural fields which exhibit a cosine tuning curve is described here along with a formal analysis of their dynamics, it is important to redeclare the fundamental equations which governs their behavior. Moreover, slightly different definitions will be provided here. Nevertheless, the properties detailed in this chapter are still valid for networks having other types of unimodal tuning curves, but their analytical description would have been much more difficult to realize.

The definition of populations composed of neurons exhibiting a cosine tuning is as follow. Let a continuous neural population be defined by an ensemble of neurons where each of which is preferentially tuned to a preferred direction \vec{r} . For a given population, the preferred directions are assumed to be uniformly distributed along Γ_N , a N dimensional subspace defined according to Equ. (3.5),

which corresponds to the surface of a unitary hypersphere of dimension $N+1$. In the following, only the cases where $N \in \{1, 2\}$, i.e., subspaces representing two and three dimensional unitary vector space, are considered. Moreover, unless N needs to be specified, it will be omitted.

Within a neural population, the membrane potential of a neuron is denoted by $u(\vec{r}, t)$, where \vec{r} corresponds to its preferred direction. $f(u(\vec{r}, t))$ is its firing activity, where f is a non-linear function given by $f(x) = \max(0, x)$. The neurons of a population exhibiting an ideal cosine-tuning to a vectorial input $\vec{v}(t) = \beta_s(t) \vec{s}(t)$, where $\beta_s(t) = \|\vec{v}(t)\|$ and $\vec{s}(t) = \vec{v}(t)/\|\vec{v}(t)\| \in \Gamma$, should have a membrane potential equal to

$$u(\vec{r}, t) = \beta_s(t) (\vec{r} \cdot \vec{s}(t)) + \alpha(t) \quad (5.8)$$

up to a scaling factor. $\alpha(t) \in \mathbb{R}$ can be an arbitrary constant. An illustration of the neural tuning and the response at a population level is given in Figure 5.3. Then, in the present case, the *population vector* $\vec{p}(t)$ is redefined in order to consider the amplitude of the representation. It is given by:

$$\vec{p}(t) = \frac{1}{\kappa(\alpha(t), \beta_s(t))} \oint_{\Gamma} f(u(\vec{r}, t)) \vec{r} \, d\vec{r} \quad (5.9)$$

where $\kappa(\alpha, \beta)$ is a normalization factor, such that, when the population encode a vector $\vec{v}(t)$ according to Equ. (5.8), $\vec{p}(t) = \vec{v}(t)$. The exact expression of $\kappa(\alpha, \beta)$ is given in Table 5.1. Note that the population vector is properly defined only when $\beta > -\alpha$. Otherwise, since α is inhibitory and stronger than β , $f(u(\vec{r}, t)) = 0$, i.e. the population is silent.

5.2.2 A CONTINUOUS ATTRACTOR NEURAL NETWORK

Definition

A fully connected continuous attractor neural network is now considered. Its dynamics is similar to that defined in Chapter 3 by Equ. 3.9 but with

Table 5.1: Expression of $\kappa(\alpha, \beta)$ under different conditions.

$\kappa(\alpha, \beta)$	$N = 1$	$N = 2$
$0 < \beta \leq -\alpha$	0	0
$\beta > \alpha $	$\left(\frac{\alpha}{\beta} \sqrt{1 - \left(\frac{\alpha}{\beta} \right)^2} + \arccos \left(-\frac{\alpha}{\beta} \right) \right)$	$\frac{\pi}{3} \left(2 + 3 \frac{\alpha}{\beta} - \left(\frac{\alpha}{\beta} \right)^3 \right)$
$0 < \beta \leq \alpha$	π	$\frac{4\pi}{3}$
$\kappa(0, \beta) = \kappa_0$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$

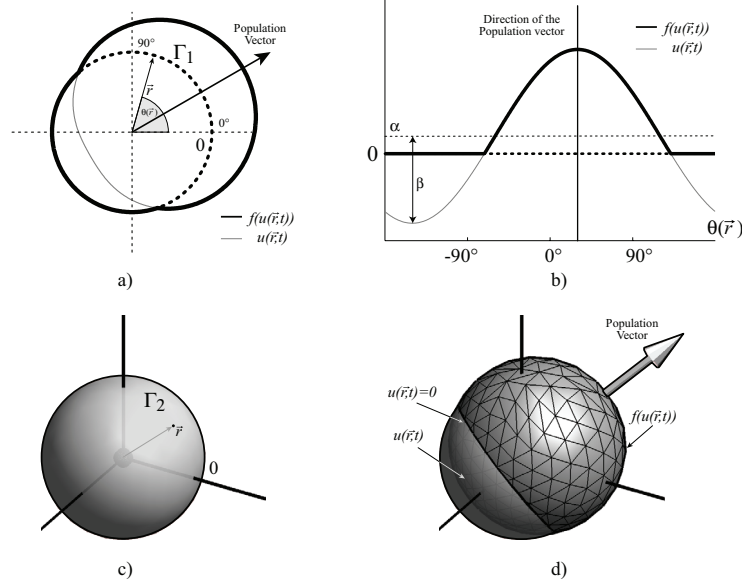


Figure 5.3: Representation of the activity of a neural population exhibiting a cosine tuning. **a)** Circular representation of a population of neurons whose preferred directions are distributed along Γ_1 . The dotted unit circle corresponds to a null activity. **b)** Linear representation of the activity of the same population. **c)** Spherical representation of Γ_2 . **d)** Population activity in that three dimensional space.

cosine-shaped recurrent weights. The neural dynamics then follows

$$\begin{aligned} \tau \dot{u}(\vec{r}, t) &= -u(\vec{r}, t) + \oint_{\Gamma} W^R(\vec{r}', \vec{r}) f(u(\vec{r}', t)) d\vec{r}' + x(\vec{r}, t) + h(t) \quad \text{where} \\ W^R(\vec{r}', \vec{r}) &= \gamma(\eta) (\vec{r}' \cdot \vec{r}) \end{aligned} \quad (5.10)$$

where τ is the time constant of the neurons. W^R are the recurrent weights that exhibit symmetric, rotation invariant, and center surround excitation inhibition characteristics. Note that, according to the definition of the recurrent weights provided in Section 3.2.1 by Equ. (3.8), this weight profile corresponds to a weight kernel where the breadth σ of the recurrent connectivity tends to ∞ . Indeed, in this case, the weights defined by Equ. (3.8) becomes cosine-shaped. $x(\vec{r}, t)$ is the external inputs, and $h(t)$ the constant modulatory input. $\gamma(\eta) = \kappa(\eta, 1)^{-1}$ corresponds to a scaling factor of the recurrent weights, and $\eta \in]0, 1[$ is a network parameter.

Stable Solutions or Attractors

In order to find the stable solution of this dynamical system, the membrane potential of the neurons is supposed to exhibit a cosine-tuning such that

$$u(\vec{r}, t) = \alpha(t) + \beta(t) (\vec{r} \cdot \vec{r}_0(t)) \quad (5.11)$$

where $\alpha(t) \in \mathbb{R}$, $\beta(t) \geq 0$ and $\vec{r}_0(t) \in \Gamma$. Using this assumption, the recurrent

convolution produces an amount of feedback given by

$$\oint_{\Gamma} W^R(\vec{r}', \vec{r}) f(u(\vec{r}', t)) d\vec{r}' = \kappa(\alpha(t), \beta(t)) (\vec{r} \cdot \vec{r}_0(t)) \quad (5.12)$$

where the exact form of the function κ is given in Table 5.1. In addition, the external input is assumed to have the following form

$$x(\vec{r}, t) = \beta_s(t) (\vec{r} \cdot \vec{s}(t)) \quad (5.13)$$

Then, replacing $u(\vec{r}, t)$ in Equ. (5.10) using Equ. (5.11), and by dropping the time variable results in

$$\tau(\dot{\alpha} + \dot{\beta}(\vec{r} \cdot \dot{\vec{r}}_0)) = -\alpha + h - \beta(\vec{r} \cdot \vec{r}_0) (1 - \gamma(\eta) \kappa(\alpha, \beta)) + \beta_s (\vec{r} \cdot \vec{s}) \quad (5.14)$$

where the time variable t has been omitted. Since Equ. (5.14) should be true $\forall \vec{r} \in \Gamma$, this equation can be split into three different equations, which can be solved separately. They are

$$\begin{aligned} \tau \dot{\vec{r}}_0 &= \vec{s} - \vec{r}_0 \\ \tau \dot{\alpha} &= h - \alpha \\ \tau \dot{\beta} &= \beta_s - \beta (1 - \gamma(\eta) \kappa(\alpha, \beta)) = g(\beta) \end{aligned} \quad (5.15)$$

where $g(\beta)$ is the function ruling the dynamical system associated with β . Then, by defining α^* , β^* and \vec{r}_p^* as the values of the corresponding variables at convergence ($t \rightarrow \infty$), it can easily be found that $\vec{r}_0^* = \vec{s}$ and $\alpha^* = h$. However, a general solution for β^* can not be found analytically. An approximation of the value taken by β^* will thus be considered. Nevertheless, in some special cases, analytical solutions for β^* can still be obtained.

First, the case in which the external input $x(\vec{r}, t) = 0$ and the homogeneous modulatory input $h(t) > 0$ is considered. By choosing a value for $\gamma(\eta)$ such that $\gamma(\eta) = \kappa(\eta, 1)^{-1}$, rewriting Equ. (5.15) leads to

$$g(\beta) = -\beta \left(1 - \frac{\kappa(\alpha, \beta)}{\kappa(\eta, 1)} \right) \quad (5.16)$$

The equation $g(\beta) = 0$ has at least one solution, which is given by $\beta^* = 0$. It is stable for $\alpha \leq 0$ and becomes unstable when $\alpha > 0$. Indeed,

$$\frac{\partial}{\partial \beta} g(\beta = 0) = -1 + \kappa(\alpha, 0) \begin{cases} > 0 & \text{if } \alpha > 0 \\ < 0 & \text{otherwise} \end{cases} \quad (5.17)$$

When $\alpha > 0$, the dynamical system bifurcates and produces another solution such that $\kappa(\alpha^*, \beta^*) = \kappa(\eta, 1)$. Since $\kappa(\alpha, \beta) = \kappa(\eta, 1)$ when $\frac{\alpha}{\beta} = \eta$, the second solution of the system is $\beta^* = \frac{1}{\eta} \alpha$. Since $\partial g(\beta) / \partial \beta < 0$, this solution is stable. It corresponds to a response amplitude β^* proportional to the modulatory input h .

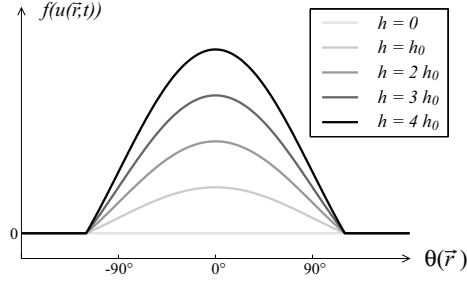


Figure 5.4: Activity profile of a population of neurons, sorted relative to their preferred direction \vec{r} (see Fig. 5.3), which are under the influence of an homogeneous input h . By considering the global population activity, the effect of this modulatory input is a multiplicative response.

The membrane potential of the neurons has thus the following stable attractor state:

$$u^*(\vec{r}) = \begin{cases} h \left(1 + \frac{1}{\eta} (\vec{r} \cdot \vec{r}_0) \right) & h > 0 \\ h & h \leq 0 \end{cases} \quad (5.18)$$

where \vec{r}_0 depends on the initial state of the network. Similarly to Salinas (2003b), when the network receives a constant excitatory global activation, it will converge to an *active state*, in which the amplitude of the population vector is amplified proportionally to the external homogeneous activity (see Figure 5.4). Conversely, when the external activity is inhibitory, the network is turned *off*. Each neuron becomes silent. This mechanism, known as *gain modulation* (Salinas & Abbott, 1996; Salinas, 2003b) was already presented in Section 3.2.2 of this thesis.

Now, another case is considered in which the network receives an external input $x(\vec{r}, t) \neq 0$, and no modulatory input, i.e. $h(t) = 0$. By rewriting Equ. (5.15), the amplitude β^* after convergence can be found. It is given by

$$\beta^* = \frac{1}{1 - \gamma(\eta) \kappa_0} \beta_s \quad (5.19)$$

where $\kappa_0 = \kappa(0, \beta)$ is a constant. $u^*(\vec{r})$ then becomes

$$u^*(\vec{r}) = \frac{1}{1 - \gamma(\eta) \kappa_0} \beta_s (\vec{r} \cdot \vec{s}) \quad (5.20)$$

From this result, the network matches its external input $x(\vec{r}, t)$ up to a scaling factor.

Finally, in order to consider the whole solution of Equ. 5.15 when the two types of inputs are applied, a linear approximation of previous solutions (Eqs. (5.18) and (5.20)) is assumed. Summing up the two solution gives

$$u^*(\vec{r}) \approx h + \left(\frac{1}{\eta} h + \frac{1}{1 - \gamma(\eta) \kappa_0} \beta_s \right) (\vec{r} \cdot \vec{s}). \quad (5.21)$$

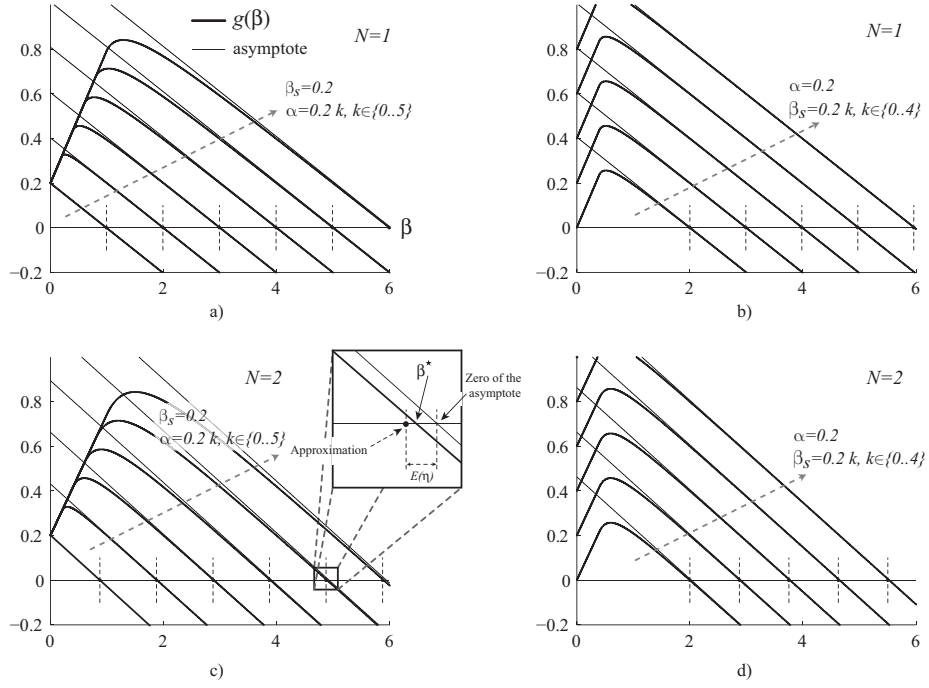


Figure 5.5: This figure shows the phase portrait of the dynamical system governing the amplitude of the field response. **a-d)** The solution at convergence, the proposed approximation and the error boundary are illustrated for different conditions. The dimension of the network as well as the input parameters are given on each subplot. In all these examples, $\eta = 0.3$.

Approximation Errors

Simulations of this system shows that Equ. 5.21 is not exact. Nevertheless, a boundary on the error of this approximation can be found. In order to find it, Equ. (5.15) which describes the dynamics of the system over β is considered. In Figure 5.5, plots of this function for different values of the constants are shown. In addition, the approximation given by Equ. (5.21) is also drawn. As observed, the equation $g(\beta) = 0$ has only one solution for $\beta > 0$ and the function $g(\beta)$ exhibits an asymptotical convergence when $\beta \rightarrow \infty$. The boundary on the error that is considered here, corresponds to the distance between the value of the approximation and the zero of the asymptote of g . The parameters of the linear asymptote, defined as $c_1\beta + c_0$, can be found using the following system of equations.

$$c_1 = \lim_{\beta \rightarrow \infty} \frac{\partial}{\partial \beta} g(\beta) \quad (5.22)$$

$$c_0 = \lim_{\beta \rightarrow \infty} g(\beta) - c_1\beta \quad (5.23)$$

Then, since the derivative of $g(\beta)$ is given by

$$\frac{\partial}{\partial \beta} g(\beta) = -1 + \gamma(\eta)\kappa(\alpha, \beta) - \beta \frac{\partial}{\partial \beta} \kappa(\alpha, \beta) \quad (5.24)$$

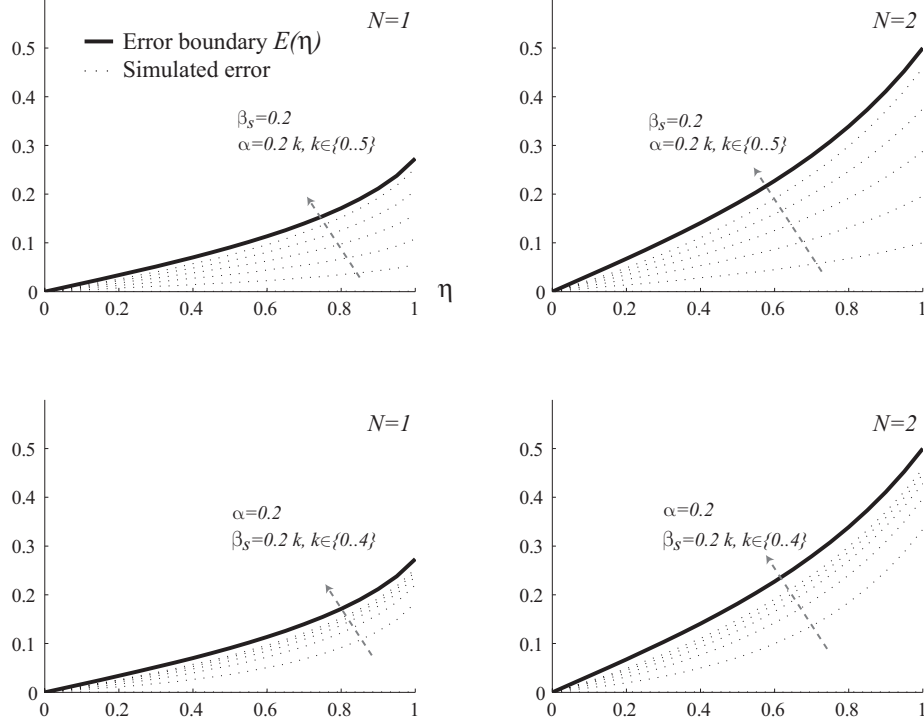


Figure 5.6: a-d) The errors due to the approximation are constrained under the bound shown in thick line for different conditions. The range of the used parameters are given on each subplots.

from Eqs. 5.23 and 5.23, the parameters of the asymptote are, for $N \in \{1, 2\}$,

$$c_1 = \gamma \kappa_0 - 1 \quad (5.25)$$

$$c_0 = \alpha \gamma(\eta) \chi + \beta_0 \quad \text{where} \quad \chi = \begin{cases} 2 & N = 1 \\ \pi & N = 2 \end{cases} \quad (5.26)$$

As mentioned earlier, the error boundary $E(\eta)$ is the difference between the zero of the asymptote given by $-c_0/c_1$ and the approximation given in Equ. (5.21). It gives

$$E(\eta) = \alpha \hat{E}(\eta) \quad \text{where} \quad \hat{E}(\eta) = \frac{\gamma(\eta) \chi}{1 - \gamma(\eta) \kappa_0} - \frac{1}{\eta} \quad (5.27)$$

As can be noticed, the bound is proportional to the modulatory input $h(t)$ to which $\alpha(t)$ converges, and is independent of the input amplitude $\beta_0(t)$. Figure 5.6 shows the value of this bound for difference values of η . Numerical calculations of the exact error in several conditions are also displayed.

Now that the behavior of the recurrent network in response to its inputs is known, the necessary information for performing frames of reference transformations should be extracted. As will be shown in the next section, adding a second layer to the network is sufficient.

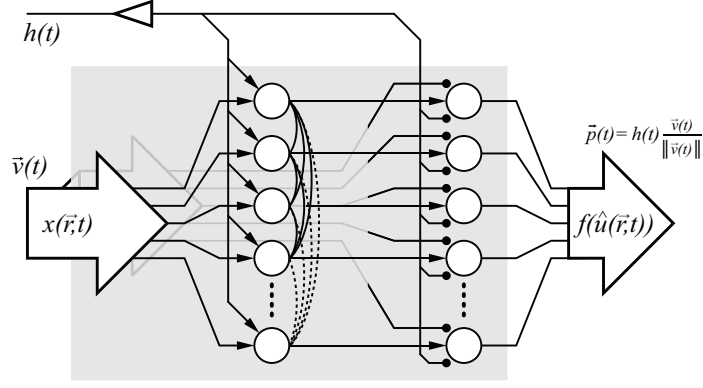


Figure 5.7: Architecture of the fundamental building block. It consist of a two-layer neural network capable of producing a non-linear composition of its inputs.

5.2.3 A TWO-LAYER NEURAL NETWORK

The goal of this section is to show how an extension of the presented neural network can produce a multiplicative combination of its inputs, which will be useful for performing transformations across frames of references (Salinas & Abbott, 1995). Indeed, up to now, the output of network consists of a sum of several terms, two of which reflect the network inputs, i.e., the vectorial and the modulatory inputs, and a third term which is a non-linear multiplicative combination of the inputs. In order to strictly keep that last term, a two-layer architecture is built. It is illustrated in Figure 5.7.

The first layer of this extension consists of the original recurrent attractor network. The second layer is composed of another population of neurons without recurrent connectivity. It receives direct projections from the first population, and inhibitory inputs designed to remove the contribution of the vectorial and constant inputs. The input $\hat{x}(\vec{r}, t)$ of this second network, which variables will further be denoted by a small hat, is given by

$$\hat{x}(\vec{r}, t) = \eta f(u(\vec{r}, t)) - \eta h(t) - \frac{\eta}{1 - \gamma(\eta)\kappa_0} x(\vec{r}, t) \quad (5.28)$$

After substitution of the firing rate of the first layer given by Equ. (5.21) into the previous equation, and by assuming that these neurons support a very fast integration of their inputs, i.e., $\tau \rightarrow 0 \Rightarrow \hat{u}(\vec{r}, t) = \hat{x}(\vec{r}, t)$, the firing rate of the second layer becomes

$$f(\hat{u}(\vec{r})) \approx \begin{cases} h(\vec{r} \cdot \vec{s}) & \vec{r} \cdot \vec{s} > 0, h > 0 \\ 0 & \text{otherwise} \end{cases} \quad (5.29)$$

This shows that the network is capable to encode independently two separate quantities, which are the direction $\vec{s}(t)$ and the amplitude $h(t)$, regardless of

the intensity of the directional input $\beta_s(t)$. In vectorial terms, this means that given a vector $\vec{v}(t) = \beta_s(t)\vec{s}(t)$ and a scalar $h(t)$, the population vector of the network tends toward $h(t)\vec{s}(t)$. This property, as will be explained in the next paragraph, will greatly help design neural models of transformations between frames of reference.

But before going into the details of the mechanisms of transformations, synaptic projections across networks will be described, followed by a description of the architecture of gain fields, which are neural substrates capable of combining different sources of vectorial information non-linearly.

5.2.4 SYNAPTIC PROJECTIONS ACROSS NEURAL POPULATIONS

In addition to the previously described form of the external input to a network, a neural population can also receive synaptic projections from another population. As already described in Section 3.2.3, in such network of networks, in order to deal with several populations simultaneously, a new index is added to the variables of the network. The index corresponds to the name of the network.

Two populations denoted by A and B are now considered. A projects its activity to B through synaptic weights denoted by $W^{A,B}(\vec{r}', \vec{r})$ such that the external input of population B becomes

$$\begin{aligned} x^B(\vec{r}, t) &= \oint_{\Gamma} W^{A,B}(\vec{r}', \vec{r}) f(\hat{u}^A(\vec{r}, t)) d\vec{r}' \quad \text{where} \\ W^{A,B}(\vec{r}', \vec{r}) &= \frac{1}{\kappa_0} (\vec{r} \cdot \vec{r}'). \end{aligned} \quad (5.30)$$

κ_0 is the weights normalization factor given in Table 5.1. By considering the output shape of the two-layer network given by Equ. (5.29), the result of the synaptic projections can be calculated. It gives

$$x^B(\vec{r}, t) = h^A(t) (\vec{r} \cdot \vec{s}^A(t)). \quad (5.31)$$

These synaptic projections to population B thus preserve the vectorial information encoded in population A .

5.2.5 A GAIN FIELD

Following from the neurophysiological findings of neurons sensitive to two or more behaviorally related variables such as the position of the eyes and that of an object, the gain field ¹ hypothesis was proposed (Salinas & Thier, 2000). It defines itself as a neural substrate where different sources of information are com-

¹As often confused in the literature, it is important to mention that the term *gain field* may refer to two different computational principle. Here, it mostly refers to the work by Salinas and Abbott (1996); Pouget and Snyder (2000).

bined. Neural models of gain field were then developed to model the underlying mechanisms of the cortical ability to combine information non-linearly (Pouget & Snyder, 2000; Salinas & Thier, 2000). Here, this work mainly follows the basic architecture proposed by these previous modeling studies, but gain field are deigned here using an assembly of the building blocks described in Section 5.2.3. As already mentioned, this structure will allow the system to preserve the amplitude of the transformed vectorial information. Then, since this assembly produces a new dimension in the neural representation, the neurons of the gain field are now preferentially tuned to a direction $\vec{r} = (\vec{r}_A, \vec{r}_B) \in \Gamma \times \Gamma$, where \vec{r}_A corresponds to preferred direction of the building block network, whereas \vec{r}_B to that along the new dimension. The gain field architecture is depicted in Figure 5.8. The external inputs come from two different neural populations, denoted by A and B , which project separately to each dimension of the gain field. The external inputs of the gain field are then

$$\begin{aligned} x^{\text{GF}}(\vec{r}_A, \vec{r}_B) &= \oint_{\Gamma} W^{A, \text{GF}}(\vec{r}', \vec{r}_A) f(\hat{u}^A(\vec{r}')) d\vec{r}' + \\ &+ \oint_{\Gamma} W^{B, \text{GF}}(\vec{r}', \vec{r}_B) f(\hat{u}^B(\vec{r}')) d\vec{r}' \end{aligned} \quad (5.32)$$

where GF denotes the variables of the gain field. By setting the synaptic weights according to Equ. (5.30), the input of the gain field becomes

$$x^{\text{GF}}(\vec{r}_A, \vec{r}_B, t) \stackrel{(5.31)}{=} h^A(t) (\vec{r} \cdot \vec{s}^A(t)) + h^B(t) (\vec{r} \cdot \vec{s}^B(t)) \quad (5.33)$$

Then, the output activity profile of the gain field can be found by substituting Equ. (5.33) into Equ. (5.29), which gives

$$\begin{aligned} f(\hat{u}^{\text{GF}}(\vec{r}_A, \vec{r}_B, t)) &\approx \\ &\begin{cases} h^B(t) (\vec{r}_B \cdot \vec{s}^B(t)) (\vec{r}_A \cdot \vec{s}^A(t)) & \text{if } h^B(t)(\vec{r}_B \cdot \vec{s}^B(t)) > 0, \vec{r}_A \cdot \vec{s}^A(t) > 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (5.34)$$

The resulting activity profile is symmetric with a peak located at the intersection of the directions encoded by the two input sources. Moreover, the amplitude only reflects that of population B . This property will be useful in solving the problem of transformations across frames of reference, which is described next.

5.3 NEURAL MODELS OF FRAME OF REFERENCE TRANSFORMATIONS

Recall the problem of transformation across frames of reference described

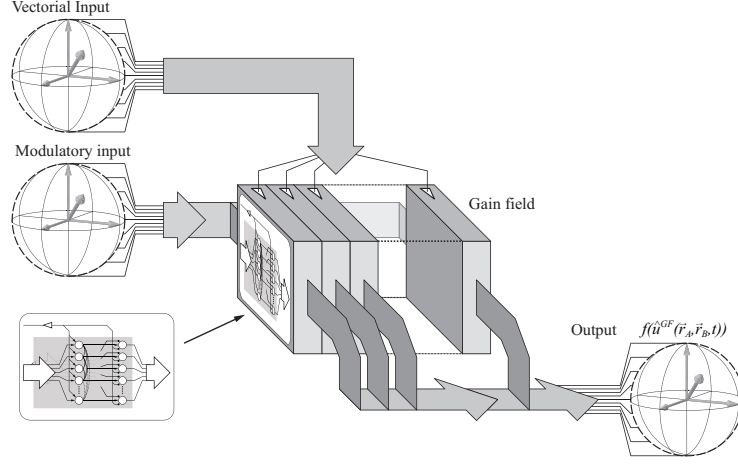


Figure 5.8: Gain field architecture.

in Section 5.1.1. It consists of transferring a vector \vec{v} encoded in a referential \mathcal{R} into a vector \vec{v}' encoded in \mathcal{R}' , knowing the translation vector \vec{v}_0 across the origins of the referentials, and either the series of rotation necessary to transfer \mathcal{R} into \mathcal{R}' , or the principal axes of \mathcal{R}' expressed in \mathcal{R} . The possible use of these inputs in several neural models are described in the following sections. But first, the method for performing translations of vectors is described.

5.3.1 TRANSLATIONS

Let \vec{v} be a vector in referential \mathcal{R} , and \vec{v}' its projection in the referential \mathcal{R}' . Assuming that \vec{v}_0 is the vector across the origins of both reference frames, and that both referentials are aligned, i.e., $\mathbf{M}^{\mathcal{R}'} = \mathbf{I}$ in Equ. (5.2), it gives $\vec{v}' = \vec{v} - \vec{v}_0$. To perform the translation, the neural population encoding the result \vec{v}' is assumed to receive in its input $x^{v'}(\vec{r}, t)$, the difference between the synaptic projections from the populations encoding \vec{v} and \vec{v}_0 , i.e.,

$$\begin{aligned}
 x^{v'}(\vec{r}, t) &= \oint_{\Gamma} W^{v, v'}(\vec{r}', \vec{r}) f(\hat{u}^v(\vec{r}', t)) d\vec{r}' \\
 &\quad - \oint_{\Gamma} W^{v_0, v'}(\vec{r}', \vec{r}) f(\hat{u}^{v_0}(\vec{r}', t)) d\vec{r}' \\
 &\stackrel{(5.31)}{=} (\vec{r} \cdot \vec{v}) - (\vec{r} \cdot \vec{v}_0) = \vec{r} \cdot (\vec{v} - \vec{v}_0) \\
 &= \vec{r} \cdot \vec{v}'.
 \end{aligned} \tag{5.35}$$

where $W^{v, v'}$ and $W^{v_0, v'}$ are defined according to Equ. (5.30). Then, by Equ. (5.9), the output population vector can be seen as representing the result of the translation which is \vec{v}' .

5.3.2 PLANAR ROTATIONS

A mentioned earlier, a method for aligning the referential \mathcal{R}' with \mathcal{R} consists of performing a series of rotations serially. In the following a single rotation model is described. Indeed, a complete transformation in three dimensions corresponds simply in replicating the process three times.

The case of a planar rotation will first be considered. A vector \vec{v} in referential \mathcal{R} is rotated by an angle $-\phi$ to project onto \vec{v}' , where $-\phi$ corresponds to the angle between the two planar reference frames \mathcal{R} and \mathcal{R}' , with superposed origins. \vec{v} , \vec{v}' and the amount of rotation $-\phi$ are separately encoded in three neural populations. The angle $-\phi$ is represented as a vector having direction $-\phi$ with an arbitrary amplitude $\beta^{-\phi} > 0$. The output activity profile of the corresponding population is described by

$$\hat{u}^{-\phi}(\vec{r}, t) = \beta^{-\phi} (\vec{r} \cdot \vec{s}^{-\phi}(t)) \quad (5.36)$$

where $\vec{s}^{-\phi} = (\cos(-\phi), \sin(-\phi))$. In contrast to the translation operation, a rotation is a non-linear transformation. An intermediary gain field, described in Section 5.2.5, will help resolve this non-linearity. In the present case, the inputs of the gain field (Equ. (5.32)) consist of projections from the populations encoding the amount of rotation $-\phi$ and the input vector \vec{v} . They respectively project to the first and second dimensions of the gain field. According to Equ. (5.35), its output response becomes

$$f(\hat{u}^{\text{GF}}(\vec{r}_A, \vec{r}_B, t)) \approx \beta^v(t) (\vec{r}_B \cdot \vec{s}^v(t)) (\vec{r}_A \cdot \vec{s}^{-\phi}(t)) \quad (5.37)$$

Then, similarly to Salinas and Abbott (1995); Baraduc and Guigon (2002), the rotation is encoded in the synaptic projections from the gain field to the output population of the system. Using an analogy to the complex division, the rotation of a unitary vector \vec{r} of an amount $-\phi$, i.e., the angle given by the orientation of the unitary vector $\vec{s}^{-\phi}$, is expressed by $\vec{r}/\vec{s}^{-\phi}$. The synaptic weights between the gain field and the output population are then given by

$$W(\vec{r}_A, \vec{r}_B, \vec{r})^{\text{GF}, v'} = \frac{1}{\kappa_0^2} (\vec{r} \cdot (\vec{r}_B / \vec{r}_A)). \quad (5.38)$$

By Eqs. 5.37 and 5.38, the input $x^{v'}(\vec{r}, t)$ of the output population becomes

$$\begin{aligned} x^{v'}(\vec{r}, t) &= \oint_{\Gamma_1 \times \Gamma_1} W(\vec{r}_A, \vec{r}_B, \vec{r})^{\text{GF}, v'} f(\hat{u}^{\text{GF}}(\vec{r}_A, \vec{r}_B)) \, d\vec{r}_A \, d\vec{r}_B \\ &\approx (\vec{r} \cdot (\vec{v} / \vec{s}^{-\phi})). \end{aligned} \quad (5.39)$$

Finally, by using the population vector (Equ. (5.9)), the input of the output population encodes approximatively the vector \vec{v}' , which corresponds to \vec{v} rotated by $-\phi$. Note that the quantification of the errors produced by this transformation is provided further in Section 5.4.3. Further, this mechanism

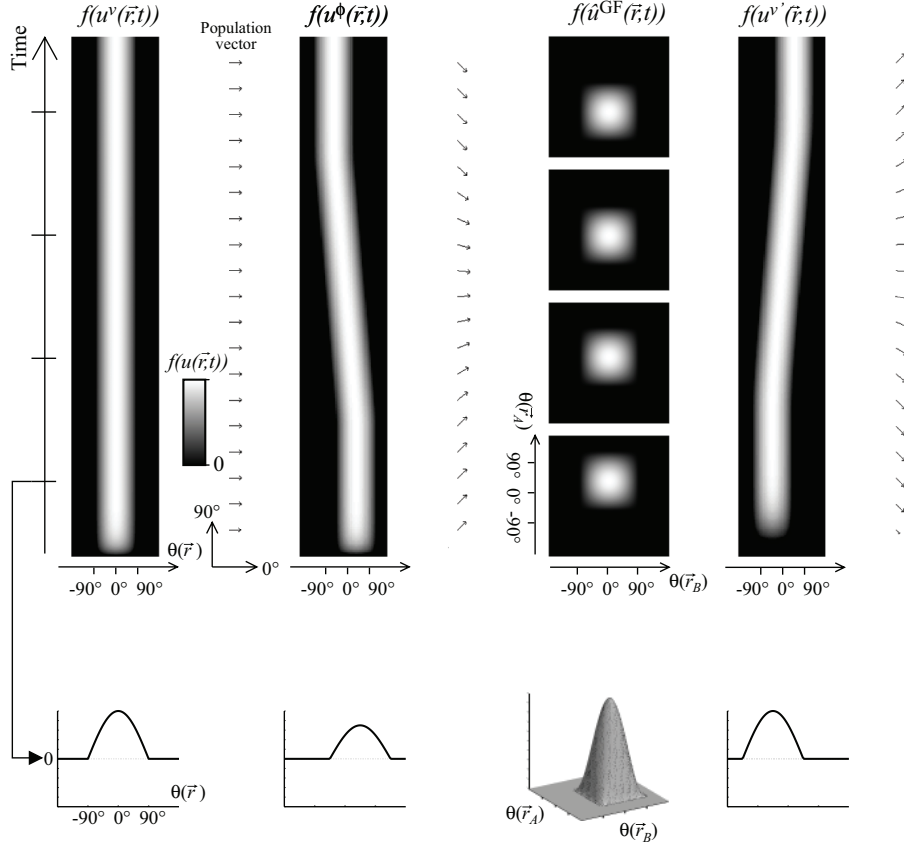


Figure 5.9: (top) Evolution over time of the activity profiles of the involved populations while performing a constantly varying rotation. The activity of the output neurons of the gain field are shown for several time steps denoted by the marks on the time axis. On the right-hand side to each activity plot, the corresponding population vector is shown. (bottom) Snapshot of the activities of the same populations at the time step given by the left black arrow on the time axis.

underlying this transformation is illustrated in Figure 5.9, on top of which, the evolution of the activity of the involved populations is shown while performing a constantly varying rotation. It can be seen that the activity profile of the gain field output is symmetric and that the peak is located at the intersection of the directions currently encoded by the source populations. At the bottom of this figure, a snapshot of the activities at a given time is shown. The amplitude of the cosine shape of the input and output populations are equal, meaning that the amplitude is preserved through the transformation.

5.3.3 EXTENSION TO 3D ROTATIONS

While previous section described a rotation in 2D space, the present paragraph addresses the rotation in 3D space. this transformation can be described by a rotation of angle $-\phi$ around an axis $\vec{d} \in \mathbb{R}^3$, $\|\vec{d}\| = 1$. The present model

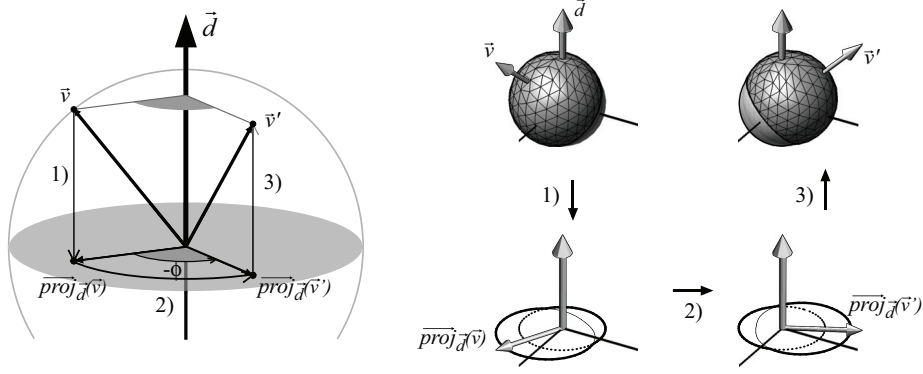


Figure 5.10: (left) A rotation in 3D space can be decomposed into three sub-transformations (see main text). (right) Illustration in spherical coordinates of the populations activity profile during this transformation.

needs the rotation to be decomposed in three sub-transformations, shown in Figure 5.10. Indeed, a rotation of vector \vec{v} in 3D space can be seen as a succession of three steps: 1) a projection of \vec{v} on the plane perpendicular to the rotation axis \vec{d} , 2) a rotation of an angle $-\phi$ around \vec{d} and 3) the restoration of the component parallel to \vec{d} lost during the projection. This process results in a vector \vec{v}' that corresponds to \vec{v} rotated by an angle $-\phi$ around axis \vec{d} .

In order to project \vec{v} on the plane perpendicular to \vec{d} , connecting a network encoding $\vec{v} \in \mathbb{R}^3$ with a two dimensional population using synaptic projections as defined in Equ. (5.30) is sufficient. The only constraint is that the preferred directions \vec{r} of the second network have to be defined on the projection plane, i.e., in $\{\vec{r} \in \Gamma_2 | \vec{r} \cdot \vec{d} = 0\}$. After this projection, the projected vector denoted by $\overrightarrow{proj}_{\vec{d}}(\vec{v})$ is rotated as described in Section 5.3.2. Since the resulting vector lies only in the projection plane, it is necessary to add the component of \vec{v} parallel to \vec{d} lost during the projection. This component is given by $(\vec{v} \cdot \vec{d})\vec{d}$. Using direct synaptic links from the source to the destination population defined by

$$W^{v,v'}(\vec{r}', \vec{r}) = \frac{1}{\kappa_0} (\vec{r}' \cdot (\vec{r}' \cdot \vec{d})\vec{d}). \quad (5.40)$$

and according to Eqs. (5.35) and (5.39), the synaptic inputs of the output network becomes

$$\begin{aligned} x^{v'}(\vec{r}, t) &= \oint_{\Gamma \times \Gamma} W^{\text{GF}, v'}(\vec{r}_A, \vec{r}_B, \vec{r}) f(u^{\text{GF}}(\vec{r}_A, \vec{r}_B, t)) d\vec{r}_A d\vec{r}_B + \\ &\quad \oint_{\Gamma} W^{v,v'}(\vec{r}', \vec{r}) f(u^v(\vec{r}, t)) d\vec{r} \\ &\approx (\vec{r} \cdot (\overrightarrow{proj}_{\vec{d}}(\vec{v})/\vec{r}^{-\phi})) + (\vec{r} \cdot (\vec{v} \cdot \vec{d})\vec{d}) \\ &\approx (\vec{r} \cdot (\vec{v}/\vec{r}^{-\phi})) \end{aligned} \quad (5.41)$$

As a result, the input of the output population encodes effectively \vec{v}' , which corresponds to \vec{v} rotated by an amount $-\phi$ around axis \vec{d} .

5.3.4 3D ROTATIONS

In the previous section, a three dimensional rotation around a fixed axis required the transformation to be split into two streams. While one of them processes the rotation of a vector projected on the plane perpendicular to the rotation axis, the other computes the constant vectorial component lost during the projection. Since such a split process could seem questionable in terms of biological plausibility, this section presents a variation of that model which processes the rotation directly. This new model involves a gain field of a higher dimension. As will be shown, this additional cost allows the computation of the transformation through projections from the gain field only.

At first glance, one could think that the rotation in three dimensions could be directly derived from the model given in two dimensions (see Section 5.3.2. However, the use of the synaptic weights given by Equ. (5.38) when the gain field is defined in $\Gamma_1 \times \Gamma_2$ rather than $\Gamma_1 \times \Gamma_1$, the network response obtained from the output synaptic projections, exhibits a bias toward the poles, i.e., the points on the sphere intersecting with the rotation axis. Indeed, rewriting Equ. (5.39) using $W^{\text{GF},v'}(\vec{r}_A, \vec{r}_B, \vec{r}) = (\vec{r} \cdot (\vec{r}_B/\vec{r}_A))$, in this three dimensional case gives

$$\begin{aligned} x^{v'}(\vec{r}, t) &= \oint_{\Gamma_1 \times \Gamma_2} W(\vec{r}_A, \vec{r}_B, \vec{r})^{\text{GF},v'} f(\hat{u}^{\text{GF}}(\vec{r}_A, \vec{r}_B)) d\vec{r}_A d\vec{r}_B \\ &= \frac{\pi}{3} \left(\pi \vec{a}^\perp(\vec{v}') + 4 \vec{a}''(\vec{v}') \right) \cdot \vec{r} \end{aligned} \quad (5.42)$$

where $\vec{a}^\perp(\vec{v}')$ and $\vec{a}''(\vec{v}')$ correspond respectively to the components perpendicular and parallel to the rotation axis \vec{d} of the rotated vector \vec{v}' . As shown by this equation, the two components do not have the same scaling factor. Since the factor corresponding to the parallel component is bigger, this explains the bias toward the poles that was observed during preliminary simulations. In order to balance equally the influence of each component, and hence to compute the rotation accurately, a new term in the synaptic weights used in Equ. (5.42) was added. The synaptic weights originating from the gain field are now given by

$$W^{\text{GF},v'}(\vec{r}_A, \vec{r}_B, \vec{r}) = c_0 (\vec{r} \cdot (\vec{r}_B/\vec{r}_A)) [1 - c_1 (1 - \vec{r} \cdot \vec{d})^2] \quad (5.43)$$

where c_0 and c_1 are constants which value will be given next. Recalculating Equ. (5.42) using the synaptic weights defined in Equ. (5.43) gives,

$$x^{v'}(\vec{r}, t) = c_0 \frac{\pi}{15} [\pi(5 - 4c_1) \vec{a}^\perp(\vec{v}') + 4(5 - 2c_1) \vec{a}''(\vec{v}')] \quad (5.44)$$

Since the scaling factors of a^\perp and a'' must be equal, c_1 can be determined.

$$\pi(5 - 4c_1) = 4(5 - 2c_1) \quad \Rightarrow \quad c_1 = \frac{5}{4} \frac{4 - \pi}{2 - \pi} \quad (5.45)$$

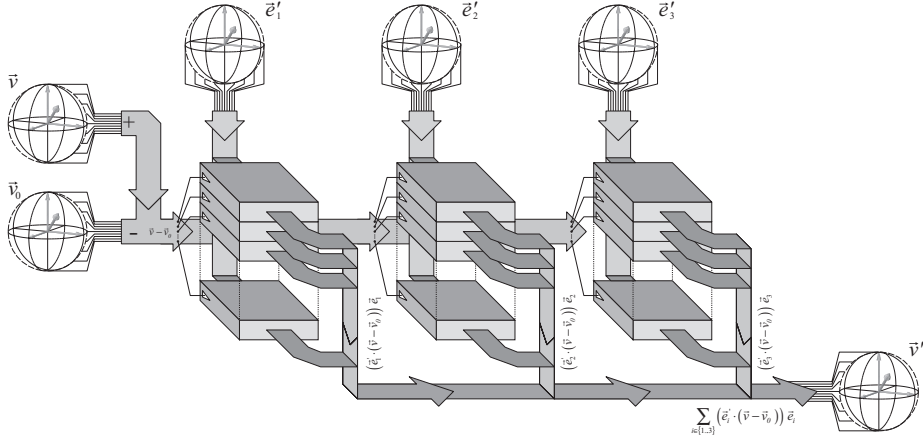


Figure 5.11: Architecture and connectivity of the model proposed to perform a complete transformations of frames of reference through projections on the principal axes.

Then, replacing c_1 in Equ. (5.44), gives,

$$x^{v'}(\vec{r}, t) = c_0 \frac{\pi^2}{3} \left(1 - \frac{4 - \pi}{2 - \pi} \right) \underbrace{[\vec{a}^\perp(\vec{v}') + \vec{a}^{\prime\prime}(\vec{v}')]_{\vec{r} \cdot (\vec{v}/s - \phi) = \vec{r} \cdot \vec{v}'}}_{\vec{r} \cdot (\vec{v}/s - \phi) = \vec{r} \cdot \vec{v}'} \quad (5.46)$$

Finally, by setting c_0 to the inverse of the global amplification factor, i.e.,

$$c_0 = \frac{3}{2} \frac{\pi - 2}{\pi^2} \quad (5.47)$$

the input amplitude across the transformation is preserved. Following from this development, by using the synaptic weights described in Equ. (5.43), a network model built with a gain field having dimension $\Gamma_1 \times \Gamma_2$ will conserve the amplitude of the input vector in all possible directions.

5.3.5 PROJECTIONS ON THE PRINCIPAL AXES

The models presented up to here assume that the mechanisms of transformations across frames of references are based on a series of rotations, computed one after the others. Although this technique may allow to compute all possible transformations, it requires as inputs the exact sequence of rotation angles, which, in certain cases, may be difficult to obtain. An alternative method has already been suggested in Section 5.1.1. Rather than performing three rotations serially, transformations across reference frames can be performed by using projections on the principal axes forming a referential.

The neural model applying this principle and hence Equ. (5.7) is shown in Figure 5.11. Five sources of information coming from five neural populations are considered. They respectively encode the principal axes of \mathcal{R}' expressed in \mathcal{R} , i.e., \vec{e}'_i , $i \in \{1..3\}$, the input vector \vec{v} , and the vector \vec{v}_0 across the origin of

both reference frames. Three gain fields perform the three dot products. Their modulatory inputs are connected to the populations coding for the principal axis $\vec{e}'_{i \in \{1..3\}}$ and their vectorial inputs are linked to the two populations coding for \vec{v} and \vec{v}_0 . These gain fields then project to the output population using the following synaptic weights, $\forall i \in \{1..3\}$

$$W^{\text{GF}_{i,v'}}(\vec{r}_A, \vec{r}_B, \vec{r}) = \frac{1}{\kappa_0^2} (\vec{r}_A \cdot \vec{r}_B) (\vec{r} \cdot \vec{e}_i). \quad (5.48)$$

Then, according to the output activity profile of the gain fields given by Equ. (5.37), each neuron of the final population receives a net synaptic input equal to

$$\begin{aligned} x^{v'}(\vec{r}, t) &= \sum_{i \in \{1..3\}} \oint_{\Gamma \times \Gamma} f(\hat{u}^{\text{GF}_i}(\vec{r}_A, \vec{r}_B)) W^{\text{GF}_{i,v'}}(\vec{r}_A, \vec{r}_B, \vec{r}) d\vec{r}_A d\vec{r}_B \\ &= \sum_{i \in \{1..3\}} ((\vec{v} - \vec{v}_0) \cdot \vec{e}'_i) (\vec{r} \cdot \vec{e}_i) \\ &= \vec{r} \cdot \vec{v}' \end{aligned} \quad (5.49)$$

which shows that these synaptic projections convey the vector \vec{v}' , corresponding to \vec{v} expressed in \mathcal{R}' .

5.4 SIMULATION RESULTS

The results presented here consist first of simulation illustrations of the transformation across reference frames as performed by the different models which were proposed. Secondly, the accuracy of the methods is also quantitatively reported. An important point must be mentioned concerning the simulations involving neural populations representing three dimensional vectors. Since the present modeling approach assumes continuous and uniform distributed representations, and since in the three dimensional case an ideal uniform distribution of direction on a unit sphere is not feasible (Marsaglia, 1972), an approximation had to be made in order to perform the simulations. Thus, an iterative algorithm was applied to generate a quasi-uniform distribution of preferred directions. For more details on this algorithm, the interested reader is suggested to refer to Appendix A.1.

5.4.1 VISUOMOTOR TRANSFORMATIONS FOR REACHING

In order to illustrate the rotation mechanism as performed by the model described in Section 5.3.2, a simple target reaching task is considered. The target location \vec{v} perceived and encoded by the visual system in head-centered coordinates has to be transferred in a body-centered frame of reference in order

to plan the reaching commands. The eyes are assumed to be fixed in their orbit. The input population of the model is encoding both the direction and the distance to the target. The angle between the body and the head is then assumed to be encoded in a population receiving proprioceptive information transmitted by the corresponding muscles receptors. These data are injected into the model which computes the vector \vec{v}' encoded with respect to a body-centered frame of reference. Figure 5.12 shows the firing activity over time of these three populations during the following three scenarios. In the first scenario, the head rotates to follow closely the motion of a target that moves from left to right (from -45° to 45°). In the second scenario, the head remains fixed with respect to the body, and only the target moves. Finally, in the last scenario, the head rotates with respect to the body, while the body and the target remain static. The simulations show that \vec{v} is correctly rotated according to ϕ . There is, however, a delay in the representation of the target in body centered coordinates, which is due to the time required for the network to propagate its information.

5.4.2 TRANSFORMATIONS OF MOVING TARGETS

The simulations described in this section illustrate the ability of the model which applies projections on the principal axes of the target referential, to perform transformations across reference frames. Simulations were conducted, in which the target \vec{v} is following eight different trajectories. Figure 5.13 shows superimposed the trajectories produced by the model and by classical mathematical manipulation (Equ. (5.4)). Similarly to previous section, the trajectories were injected in the model in a viewer-centered reference frame. The task was however to transform the location of the moving target with respect to a referential centered on another part of the space and oriented differently to that of the viewer, i.e., the transformation required both a translation and a 3D rotation.

Both trajectories show a high qualitative resemblance. However, one can observe a systematic shift in space and a slight deformation of the drawn figure. This is an artifact resulting from the non-uniformity of the distribution of preferred directions in the neural populations. In other words, the neural populations produce a non-uniform map of their inputs, resulting in a slight deformation of the three dimensional representation of the target vectors. This naturally depends on a given distribution of preferred directions. Moreover, it is worth mentioning that the duration of the trajectories were deliberately slow with respect to the system time constant. Indeed, if the movement to follow is faster than the time for the network to converge, which depends on the time constant τ of the network, the latter will not be able to follow accurately the observed trajectory. It will only tend to match the current position of the target.

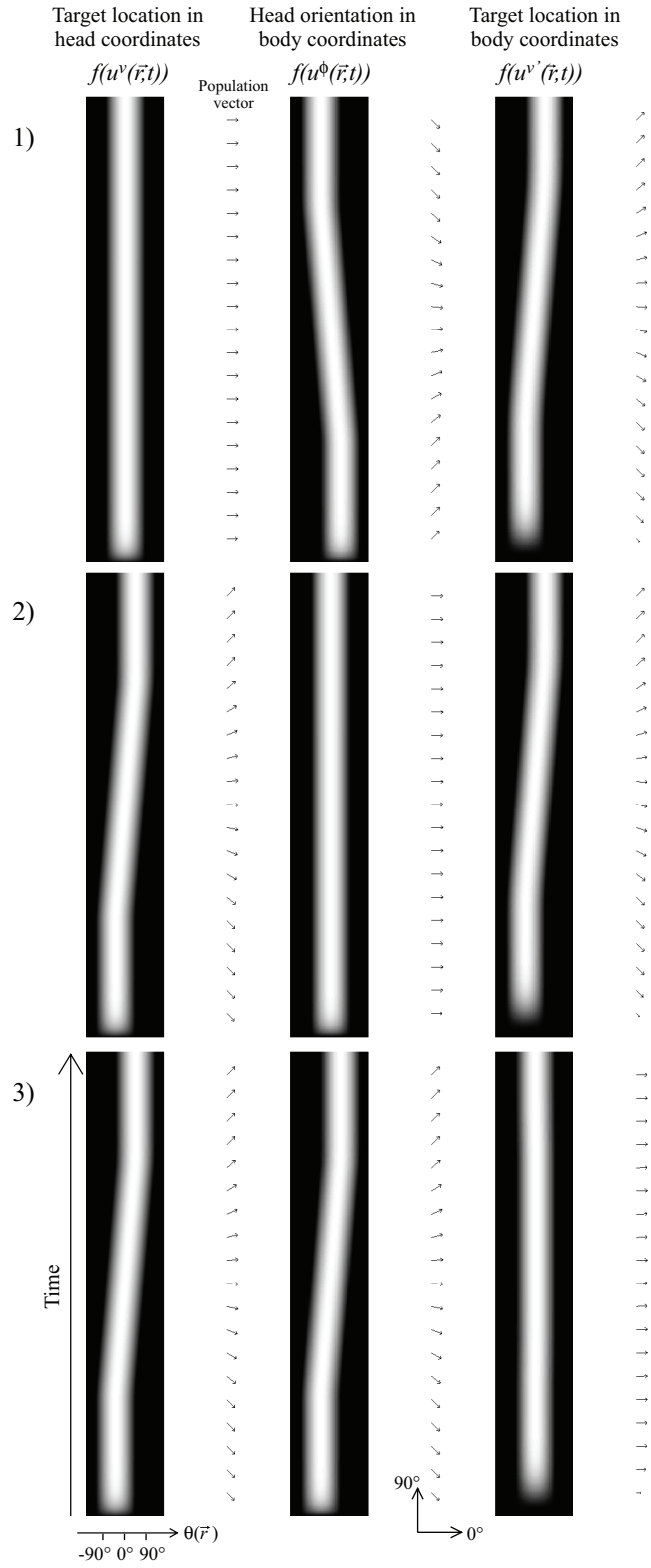


Figure 5.12: Results of simulations of the network that realizes a 2D rotation, and which is illustrated in Figure 5.8. Each row corresponds to a different scenario (see the main text for explanations). Each column represents the activity of the involved neural populations over time. On the right-hand side to each activity plot, the corresponding population vector is drawn.

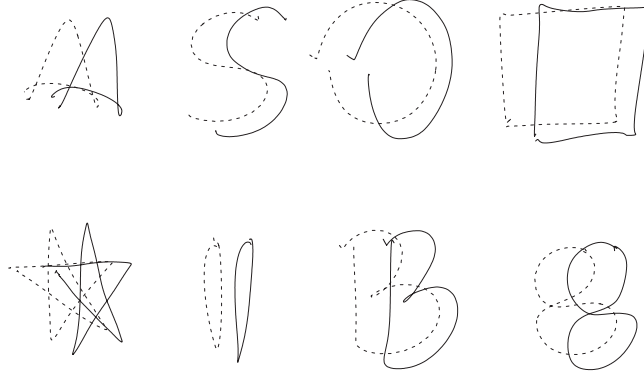


Figure 5.13: This figure shows eight target trajectories that were followed by the model which had to project the visually perceived location of the moving target on the principal axes of a static frame of reference arbitrarily localized and oriented in space. Dotted and plain lines correspond to the movement of the target movement respectively transformed using classical mathematical techniques, and the model.

5.4.3 APPROXIMATION ERRORS

In addition to the errors that appear by discretizing continuous equations, the approximation made in the mathematical development (see Equ. (5.21)) is also a source of systematical errors between the theoretical result vector \vec{v}'^* , computed with classical mathematical methods, and the output \vec{v}' produced by the model. In order to quantify them, error measures were defined. E_β is the normalized error between the ideal amplitude of the transformed vector and that given by the models, and $E_{\vec{r}}$, the angular error on the direction. Both measures are defined by

$$E_\beta(\vec{v}', \vec{v}'^*) = \frac{|\|\vec{v}'\| - \|\vec{v}'^*\||}{\|\vec{v}'^*\|} \quad (5.50)$$

$$E_{\vec{r}}(\vec{v}', \vec{v}'^*) = \text{acos} \left(\frac{\vec{v}' \cdot \vec{v}'^*}{\|\vec{v}'\| \|\vec{v}'^*\|} \right) \quad (5.51)$$

Errors in the Representation of Input Vectors

First of all, the ability of neural populations representing three dimensional vectors to faithfully reflect their inputs was tested. Indeed, as mentioned earlier, an ideal uniform distribution of preferred directions in three dimensions is not possible, which may have some influence on the model accuracy. Several networks with different sets of preferred directions were simulated. Their population vector response was then compared with different vectorial inputs they received. The error measures resulting from these simulations are summarized in Figure 5.14. They first show that the size of the network considerably influences its precision, as well as the system parameter η . The influence of η explained by looking at Equ. (5.10) describing the network dynamics. Since

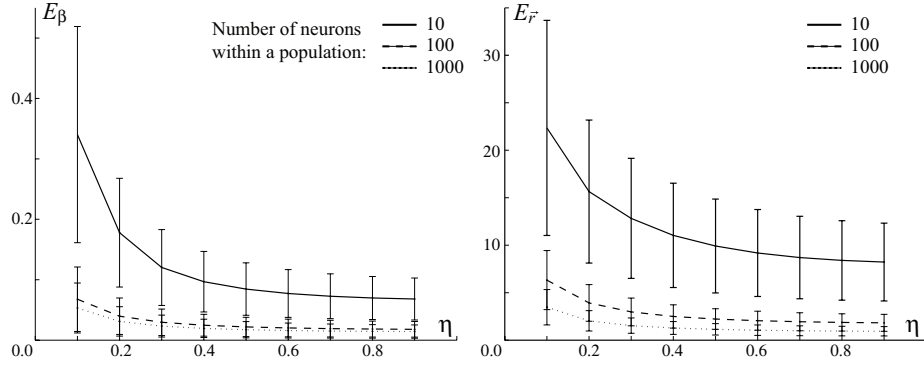


Figure 5.14: Error measurements for different population sizes, expressing the ability of an arbitrary three dimensional population with a quasi uniform distribution of preferred directions to represent its inputs. (left) Amplitude error E_β , (right) directional error E_r .

the factor $\gamma(\eta)$, which defines the strength of the imperfect recurrent feedback, decreases as η grows, the amount of errors closely follows this tendency.

In order to illustrate that this drift is mainly caused by the lateral weights, Figure 5.15a shows the difference between the population vector recorded out of a neural population, and the direction encoded by the input. As can be seen, in almost all cases, there is a natural tendency for deviating from the direction indicated by the external input. Figure 5.15b shows that, due to this imperfection, it may not be possible for an activity bump to sustain while keeping the same direction. In this simulation, a constant modulatory input was applied to the recurrent network while keeping its vectorial contribution null. In addition, the membrane potential of each neuron was initialized so that the network is encoding a random initial direction. Then, the evolution of the direction coded by the population vector until convergence was recorded. The resulting trajectories are drawn on a unit sphere given in Figure 5.15b. This indicates that, for a non uniform distribution of preferred directions, the lateral weights define specific directional attractors to which, in absence of vectorial inputs, the population vector slowly converges. Nevertheless, since the distribution of preferred directions is quasi-uniform, these attractors can be considered as relatively weak if the strength of the external inputs is set to be strong enough.

Errors in the Transformations across Frames of Reference

Considering now the errors in the transformations across frames of reference, Figure 5.16 shows the errors produced by the model which performs rotations for different network parameters. Error measures of the models performing planar and 3D rotations as described in Sections 5.3.2 and 5.3.3 are presented in this figure. As expected by the theoretical boundary on the error (Equ. (5.27)), the bigger η , the bigger is the error. In addition, an increase in the amplitude β^ϕ of the population expressing the amount of rotation, also generates an augmentation of the error. This secondary source of errors, which is not

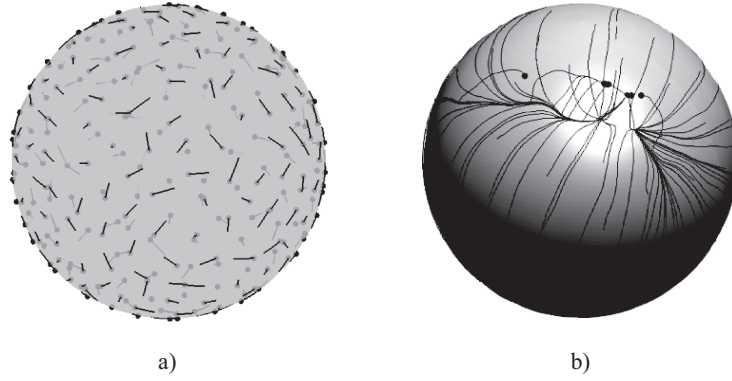


Figure 5.15: **a)** The difference between the location of the external input and the resulting population vector is shown for a neural population encoding a three dimensional vector. **b)** Illustration of the directional attractors generated by an imperfect uniform distribution of preferred directions, while the population receives only an homogeneous input. Each line terminated by a dot on the unit sphere corresponds to a trajectory followed by the population vector of the network.

accounted in the expression of the boundary, does however not exceed that limit. The directional error $E_{\vec{r}}$ is not shown here. In the 2D case, since the uniformity of the preferred direction is ideal, this error is negligible. In the 3D case, since the rotation is performed within a neural population evolving in two dimension, the resulting errors in direction is equivalent to that presented in Figure 5.14(right), which corresponds to errors in the internal representation of the vectorial input. Indeed, by comparing the error of the networks with $N = 1$ and $N = 2$, besides the fact that the variance is bigger in three dimensions ($N = 2$), one can observe a combined effect on the system parameter η . The difference between the 3D and the 2D case is that, in the former case, the mean error E_{β} is smaller for high values of η , and bigger for small values. This property is the result of the mixture of the effects of η on the ability of a population to faithfully represent its input and of generating errors due to the mathematical approximation.

Next, Figure 5.17 shows the errors produced by the model based on projections on principal axes as described in Section 5.3.5. Similarly to the previous models, the bigger is a population, the smaller are the errors. Moreover, the parameter η has also an ambivalent influence on the network errors. On the one hand, small values of η increase the strength of the recurrent connections, and consequently increase the errors due to an imperfect distribution of preferred directions. On the other hand, big values induce more errors as reported in the Equ. (5.27) describing the boundary of the error resulting from the mathematical approximation given by Equ. (5.21). These two properties explain why, in Figure 5.17a, an optimum can be observed. Further, this model may seem to produces bigger errors than the previous ones. However it should be mentioned that it performs a complete transformation, including translation and three projections, whereas the previous model has to be replicated three times in order

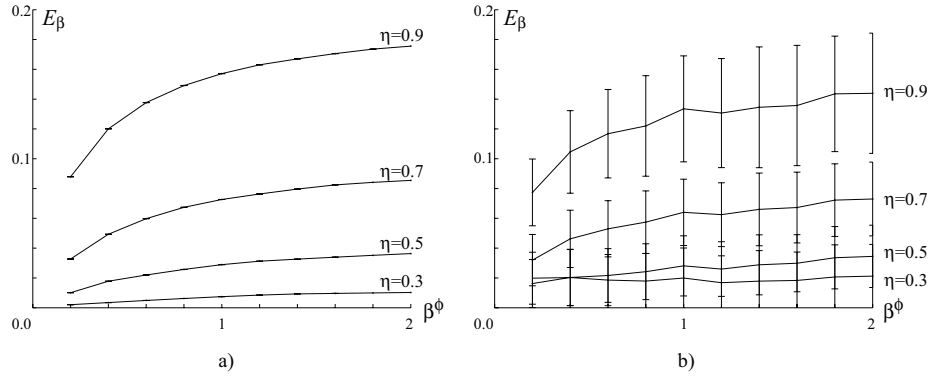


Figure 5.16: **a)** Each line corresponds to the mean error E_β against β_ϕ for different values of parameter η , while simulating planar rotation as described in Section 5.3.2. The standard error is not shown as it is too small. **b)** Same error plot, but for a three dimensional rotation as was described in Section 5.3.2.

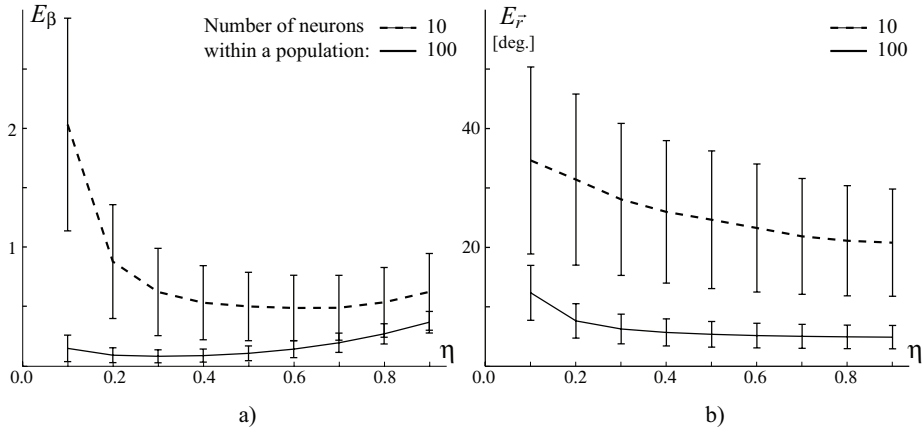


Figure 5.17: Error recorded during a simulation batch of the model described in Section 5.3.5. Different populations with different parameters were given random input vectors and frame of references. **a)** Errors in the conservation of the amplitude of the input vector. **b)** Errors in the accuracy of the direction of the output vector, resulting from the transformation.

to realize the same computation.

5.5 DISCUSSION

Biologically plausible mechanisms of transformations across frames of reference have been presented. They are based on neural computations of vectorial operations such as translations, rotations and projections on principal axes. Two main strategies were described. One assumes that three dimensional transformations are realized through a serial composition of rotations, whereas the other considers vectorial projections on principal axes. In order to combine multiple information sources, both models are based on a neural substrate following from neurophysiological findings: the gain field (Salinas & Abbott, 1995; Salinas &

Thier, 2000; Xing & Andersen, 2000; Deneve et al., 2001; Deneve & Pouget, 2003; Van Rossum & Renart, 2004), which has been observed in several sensorimotor areas (Kakei et al., 1999; Scherberger & Andersen, 2003). In contrast to common uses of gain fields, this modeling approach considers a neural architecture composed of two layers, which allows to compensate for an effect resulting from the attractor dynamics of a population of neurons. Indeed, as shown by this work, a purely multiplicative interaction between inputs within a gain field is not straightforwardly produced. Therefore, if one assumes that the amplitude of the neural response is a relevant factor used by the nervous system, which is indeed the case (Kettner et al., 1988; Casabona et al., 2004), a mechanism suppressing the undesired effects produced by a gain field has to be considered. In addition, the present investigation assumed a mathematical approximation in the network response, but the resulting error was shown to lie within strict and acceptable bounds.

The simulations results suggest that the transformation accuracy could be correlated with the uniformity of the distribution of the preferred directions along populations of neurons. This results raise the hypothesis that learning to be precise may consist of recruiting more neurons within a population and then of uniformizing their preferred directions. However, the model's assumption that a reference frame is represented by a population of neurons, having a uniform distribution of preferred direction and exhibiting a cosine tuning curve that depends on the coding direction, is not representative of all neurophysiological data (Amirikian & Georgopoulos, 2000; Scott, Gribble, Graham, & Cabel, 2001). Nevertheless, several solutions to this problem have been proposed. For instance, Scott et al. (2001) suggested that non-uniform distributions do not prevent specific brain areas to pick up uniform sub-populations or to ponderate the weights of each neuron in an inversely proportional manner relative to the distribution of preferred directions. In parallel, mechanisms of synaptic regulation and of neural adaptation have also been shown to partly solve this problematic uniformity issue in a model of working memory (Compte et al., 2000; Stringer et al., 2002; Turrigiano & Nelson, 2004).

Two different classes of models were presented in this chapter. One class assumes that the transformations across frames of reference in three dimensional space is done through a series of rotations performed serially, whereas the other class assumes parallel projections of vectorial information onto the principal axes of the corresponding reference frame. An important point which should be emphasized here is that both mechanisms may be used by the brain, but for different purposes, depending on which information is available in a given stage of the sensorimotor processes. Indeed, the hypothesis that transformations across reference frames are performed serially is in line with the gradient hypothesis for sensorimotor transformations which was suggested to follow separately parieto-medial and fronto-medial pathways, respectively for the computation of visuomotor spatial relationships and for the control of movements (Burnod et

al., 1999; Battaglia-Mayer, Caminiti, Lacquaniti, & Zago, 2003). On one hand, performing a series of rotation may be more efficient when considering sensorimotor transformations for reaching. Since reaching-plans have been suggested to be encoded in head-centered coordinates (Batista et al., 1999), the information related to the position of the hand has thus to be transferred into this frame of reference for an accurate planning. Then, due to the mechanical constraints of the configuration of the limbs, which are arranged in a serial fashion, it seems more natural for the brain to compute the forward kinematics of hand position in visual space through a series of rotations which parameters are provided by proprioception that is partly encoding the orientation of the limbs.

On the other hand, the use of a mechanism of projections on the principal axes may be more adapted to transformations within cortical areas concerned with visual processing such as STS or IT, which are brain regions known to be associated with the recognition of various body and object features in multiple reference frames (Perrett, Harries, Mistlin, & Chitty, 1989; Perrett et al., 1990). For instance, let consider a simple imitation task requiring the mimicry of the location of the hand of a demonstrator with respect to his/her body. This task mainly needs the transfer of the visual representation of that hand in a self-centered frame of reference. By observation only, one has access to all the necessary information in a viewer-centered reference frame, which are the location of the demonstrator's body and hand, as well as his/her principal axes determining his/her three dimensional orientation in space. Therefore, in such a situation, the second method should be more efficient, because obtaining the corresponding rotation vectors would require more computations.

The main hypothesis, raised by the part of this modeling study concerned with projections on principal axes, is that it is based on the assumption that orientation sensitive cells in the visual areas STS and IT, to state an example, may be grouped in populations that encode the principal axes of the observed body or object. If this was shown to be true, this would imply that such groups of neurons are used in the visual processing pathway as a vectorial basis for body- or object-centered representations. Unfortunately, there is no conclusive evidence of such an encoding, yet. Indeed, no systematic experiment have been found to show a complete description of single cell sensitivity to all possible orientations. Moreover, it may seem dubious that these cells would encode the three principal axis, since such a representation would be redundant. Nevertheless, this hypothesis deserves to be more extensively investigated, as this could provide further evidence on the mechanisms of transformations across frames of reference in the cortical visual streams.

Further, by construction of that model, the time required for the model to perform a frames of reference transformation is independent on the orientation of the frames of reference. Such a result would seem to be in contradiction with the observation that humans produce longer reaction times, when required to perform mental rotations in an "unusual" orientation, such as shifting an image

upside-down. One could, however, imagine that another mechanism is at play. In the absence of visual input, this mechanism may set the representation of the principal axes to a default state which may express an expectation of the system to find a person or object in a given orientation, e.g., in a vertical and standing posture for a person. In this case, the network state in the model will take more time to match an observed orientation different from that predicted. Several findings from experimental psychology may argue in this direction. For instance, the time required to perform motor imagery tasks has been shown to be highly dependent on the actual posture of the person executing such tasks (Jeannerod & Decety, 1995; Lange, Helmich, & Toni, 2006). In this case, the default state would be provided by proprioceptive signals. In addition, an alternative hypothesis related to this timing issue is that this delay could also be the result of a longer processing phase during the recognition of the features of bodies and objects, which are then used to set the landmarks for determining the axes.

Finally, concerning the issues of learning, it is clearly not in the scope of this work to describe mechanisms allowing the learning of sensorimotor transformations. Indeed, several powerful techniques have already been proposed by other research teams (Zhang, 1996; Stringer et al., 2002; Meñard & Frezza-Buet, 2005). Moreover, since local synaptic modification rules may be tuned to drive the synaptic weights to converge toward a given desired state, the aim of this work was rather to determine that final stage, that may be reached after learning. Importantly, by determining this target state, the approach adopted in this work has nevertheless made possible a mathematical analysis of the behavior of the network.

Summary

This chapter described two biologically plausible neural mechanisms allowing the transformation of vectorial information across different frames of reference. One class of models assumes that three dimensional transformations are realized through a series of rotations, whereas the other considers a direct mapping into the target reference frame through vectorial projections on principal axes. Both mechanisms may be involved in different parts of the overall cortical processing scheme: the former in the processing of proprioceptive information, the latter in purely visual information processing. In contrast to related modeling studies on sensorimotor transformations, this work paid special attention to the amplitude of the signal to be transformed, i.e., the neural response of the source population. The amplitude of the population vector conveys important information that must be preserved across transformations.

These neural mechanisms only make up a small building block of the complex cortical network involved in imitation. They also do not provide any clue as to how transformations across reference frames may be understood in terms of cognitive structures and pathways. For instance, since behavioral studies have shown that different imitative strategies involving different transformations can be in charge of imitative behaviors, one might suggest the existence of separate pathways processing separate strategies. How might such a hypothesis be tested? In order to address this question, this work develops a neural model that uses the mechanisms of transformations across frames of reference proposed in this chapter. The next chapter describes this model and an experiment designed to test it. In conjunction with experimental results, the behavioral predictions obtained from the simulations of this model help to clarify the cortical mechanisms of imitation.

INTERFERENCES IN THE TRANSFORMATION OF FRAMES OF REFERENCE

The study presented in this chapter consists of an extension of the work which was published in:

Sauser, E. L. and Billard, A. G.. Interferences in the transformation of reference frames during a posture imitation task. *In Proceedings of the International Conference on Artificial Neural Networks, ICANN 07, Porto, Portugal. 2007.*

THIS chapter describes a modeling study which illustrates how the mechanisms of transformations across reference frames may be applied in order to model a specific part of the neural processes underlying imitation. Like the work described in Chapter 4, this study aims to propose a neural model along with predictions related to its behavior. In particular, a biologically-inspired neural model addressing the problem of transformations across frames of reference in a posture imitation task is presented. This work is based on the hypothesis that imitation is mediated by two concurrent transformations that are selectively sensitive to spatial and anatomical cues. In contrast to classical approaches, separate instances of this pair of transformations are assumed to be responsible for the control of each side of the body. In addition, an experimental paradigm was also devised; this paradigm allows modeling of the interference patterns caused by the interaction between the anatomical and the spatial imitative strategy.

6.1 INTRODUCTION

This work focuses on the processes of transformations across frames of references required for imitation of arbitrary gestures, and particularly on two different strategies that are anatomical and spatial imitation. In psychology, anatomical and spatial imitation are usually considered distinct (Brass et al., 2000; Koski et al., 2003; Heyes & Ray, 2004; Chiavarino, Apperly, & Humphreys, 2007). On one hand, anatomical imitation considers observed movements with respect to the observed person's body. It respects the joint angles between the

body and limbs, and hence facilitates the mapping between the left and the right side of the demonstrator and those of the imitator. On the other hand, spatial imitation considers the demonstrator movements, but only according to the spatial location of the limbs with respect to the observer, regardless of the orientation of the demonstrator. When the imitator and the demonstrator are facing each other, this form of imitation is usually denoted as specular or mirror (Brass et al., 2000; Koski et al., 2003; Bertenthal et al., 2006; Chiavarino et al., 2007).

The main hypothesis adopted by this study is that the computations associated with these two forms of imitation are simultaneously computed in the brain. According to a developmental point of view, these strategies may have evolved within two different stages. While spatial imitation may originate from early visuomotor development (Piaget, 1978), anatomical imitation may have evolved later along with the development of the body schema which respects strictly body configurations (Gallagher, 2000). Next, a competitive process is needed to select, with respect to the task constraints, the correct response among those provided by the two imitative strategies (Brass et al., 2005; Bertenthal et al., 2006; Chiavarino et al., 2007). As already mentioned in Chapter 2 of this thesis, the competition between multiple sources of information usually produces measurable interferences on reaction times and movement variability (Simon et al., 1981; Brass et al., 2000; Kilner et al., 2003). This is precisely how this work will attempt to validate or refute its modeling hypotheses. Indeed, in addition to the previous hypothesis, anatomical imitation is suggested to be considered less strictly than in psychology. In fact, this work proposes that an anatomical mapping between ipsilateral limbs, which mirrors the relationship between the limb joint, also exists. Therefore, distinct pairs of spatial and anatomical transformations are assumed to be responsible for the control of each side of the body. As a consequence, when an arm posture is presented with either the left or the right arm, an imitative response is hypothesized to be computed in parallel for both arms of the imitator.

Attempts to understand the neural mechanisms and brain pathways responsible for such automatic behaviors have been reported by many researchers. Similarly to the modeling study addressed in Chapter 4 of this thesis, the neural field approach, has been applied to the problem of conflicting transformations across frames of reference. Indeed, among other properties, and in contrast to classical binary models of stimulus-response compatibility, this framework allows the modeling of continuous variables (Erlhagen & Schöner, 2002; Schöner, 2002), which are more common in imitative tasks. Next, an experimental paradigm is first presented, which will help determine the possible interferences between different imitative strategies during a task requiring the imitation of meaningless body postures. Then, a neural model, capable of computing both anatomical and spatial imitative transformations concurrently is described. And finally, the particular interference patterns predicted by the model and their implications

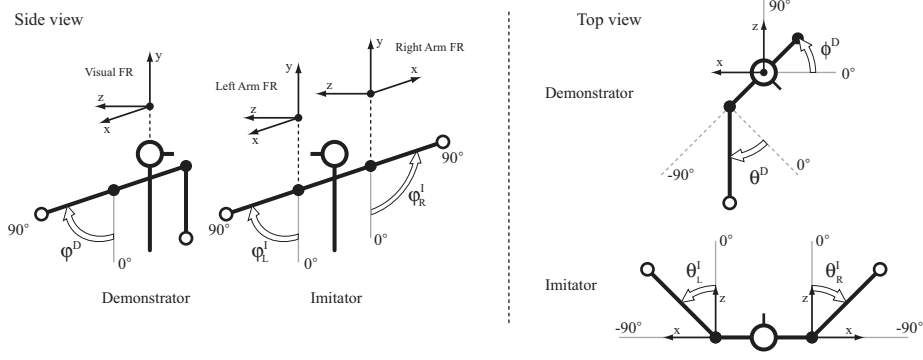


Figure 6.1: Stimulus and response variables and frames of reference (FR) relative to the observer's point of view, and to his/her left and right arm.

for future research are discussed.

6.2 EXPERIMENTAL SETUP

This work considers the imitation of body postures where the orientation of the right upper arm can be varied. The visual perspective of the demonstrator's body can also vary from side to front view. The task instructions require either a spatial or an anatomical imitative response with either the right arm (the corresponding one) or the left arm (the opposite one). The stimulus variables, shown in Figure 6.1, are: φ^D , the demonstrator arm elevation, θ^D , its orientation relative to the body in the horizontal plane, and ϕ^D , the orientation of the body with respect to the observer. The response variables are: φ_L^I and φ_R^I , the elevation of the left and right arm of the imitator, and θ_L^I and θ_R^I , their orientations on the horizontal plane. The desired responses are:

$$\begin{aligned} \theta_L^{I,A} &= \theta_R^{I,A} = \theta^D \\ \theta_L^{I,S} &= -\theta_R^{I,S} = \begin{cases} -180 - (\theta^D + \phi^D) & \theta^D + \phi^D < -90 \\ \theta^D + \phi^D & |\theta^D + \phi^D| \leq 90 \\ 180 - (\theta^D + \phi^D) & \theta^D + \phi^D > 90 \end{cases} \end{aligned} \quad (6.1)$$

where the additional index, A or S, denotes the anatomical or spatial imitative strategy, respectively. An illustration of these transformations are shown in Figure 6.2. Since both transformations are hypothesized to be processed in parallel, both are expected to have an influence on the resulting imitative behavior. So let define the discrepancy D between the response of both strategies, which is given by the difference between the response of the instructed strategy and that of the other. Spatial and anatomical transformations are said to be perfectly congruent when the discrepancy $D = 0$. Note that ideal congruency conditions are not equivalent for both arms.

An experimental trial consists first of the presentation of a starting posture

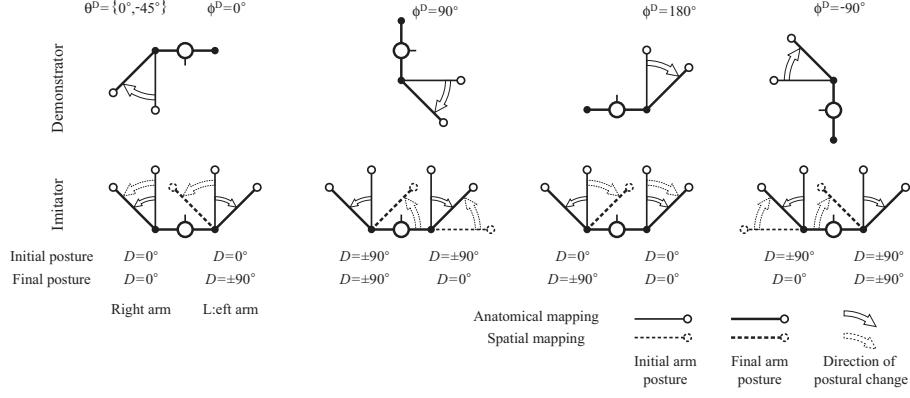


Figure 6.2: Examples of anatomical and spatial imitative strategies in various conditions. The discrepancy between the response of the transformations are given for each arm.

which the subject (or the model) is requested to imitate according to the task instructions (either spatially or anatomically, and with either the left or the right arm). Then, the arm posture is abruptly changed, and the subject has to keep imitating as fast as possible. During a single trial, only the arm posture is modified, whereas the body orientation is left unchanged. Experiment 1 investigates the interferences produced by both imitative strategies when their initial responses are congruent and the amplitude of the change of arm posture is kept constant across the trials. The pair of initial and target postures consists of the arm raising from a neutral down position ($\varphi^D = 0^\circ$), where the responses of both spatial and anatomical transformations are always congruent, to a position on the horizontal plane ($\varphi^D = 90^\circ$). The arm elevation is thus the only degree of freedom which changes during a trial. Complementarily, Experiment 2 investigates the influence of a horizontal postural change, in which amount is denoted by $\Delta\theta^D$. Indeed, in such conditions, depending on the stimuli, the discrepancy between the responses of the transformations may vary.

6.3 NEURAL MODEL

This section first briefly reminds the reader about the present modeling approach which was already described at length in Chapter 3 and applied in the modeling of an experimental study in Chapter 4. It consists of building networks composed of neural fields for representing cortical functions. To recall, a neural field is composed of a continuous set of neurons $u(\vec{r}, t)$, where each of them fires maximally for a specific value \vec{r} uniformly distributed in the parameter space Γ_2 . Indeed, since the modeled variables consist of arm and body orientations, the considered parameter space is the ensemble of directions in the three dimensional

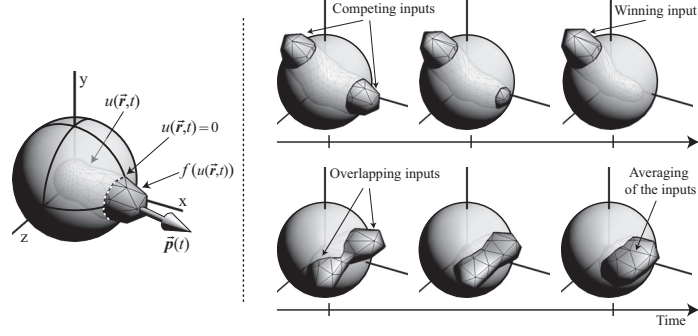


Figure 6.3: (left) Graphical representation of a neural field activity. (right) Evolution of the neural activity during a selection process involving two competing inputs (top), and between two partially overlapping inputs (bottom).

space. As already defined in Chapter 3, the neural field dynamics follows

$$\begin{aligned} \tau \dot{u}(\vec{r}, t) &= -u(\vec{r}, t) + \oint_{\Gamma} W^R(\vec{r}', \vec{r}) f(u(\vec{r}', t)) d\vec{r}' + \\ &+ x(\vec{r}, t) + h(t) \end{aligned} \quad (6.2)$$

where τ is the membrane time constant of the neurons, W^R the recurrent synaptic weights and f the activation function. $x(\vec{r}, t)$ and $h(t)$ are, respectively, the external input and the background, task-related, input. There is, however, a small difference with the model definition provided in Chapter 3. The gaussian function \mathcal{G} (Equ. (3.7)) which is used to generate the synaptic weights had to be redefined in order to include the ability to perform linear transformations. It is given by

$$\mathcal{G}(\vec{r}', \vec{r}, \sigma) = \left[\exp\left(\frac{(\mathbf{M}\vec{r}')^T \vec{r} - 1}{2\sigma^2}\right) - e^{-1/\sigma^2} \right] / (1 - e^{-1/\sigma^2}) \quad (6.3)$$

where \mathbf{M} refers to a transformation or mapping matrix. According to definition introduced in Equ. (3.8) of Chapter 3, the recurrent weights, as well as the projections across neural populations are generated following

$$W(\vec{r}', \vec{r}) = \alpha \left[\mathcal{G}(\vec{r}', \vec{r}, \sigma) - \delta \right] \quad (6.4)$$

where α denotes the strength of the weights and δ is a normalization term. In the case of the recurrent connectivity, $\mathbf{M} = \mathbf{I}$, i.e., the identity matrix, but different mappings will be described later in the text. This type of neural dynamics is known to form an attractor bump on the surface of the neural field (see Fig. 6.3), through which this class of networks is suggested to convey information. As a read-out mechanism, the population vector $\vec{p}(t) \in \Gamma_2$ defined in Equ. (3.11). Finally, the formalism used to consider the connectivity and the inputs of neural fields within larger networks corresponds to that described in Section 3.2.3.

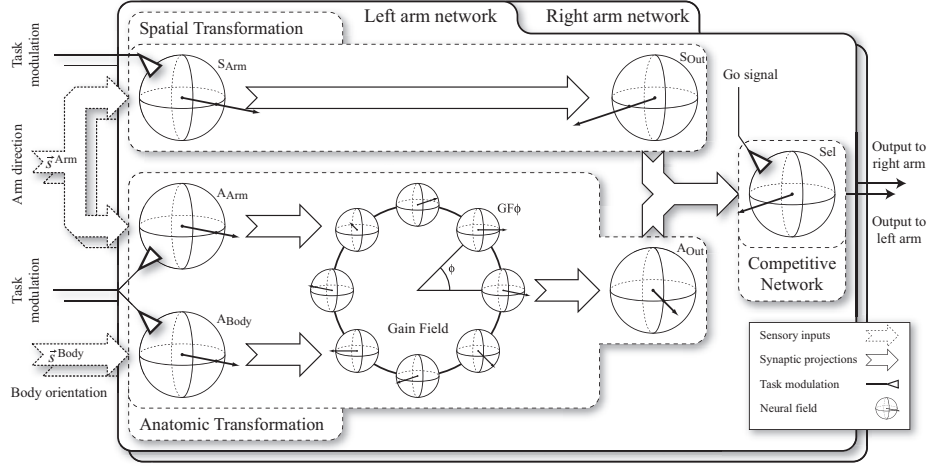


Figure 6.4: Architecture of the model where two networks process both anatomical and spatial transformations, and separately the for each arm. Within each network, each sphere represents a neural field which dynamics is driven by Equ. (3.9). The complete network is finally governed by the set of equations defined in Section 3.2.3, whose parameters are summarized in Appendix B.2.

6.3.1 NETWORK ARCHITECTURE

The model architecture, depicted in Figure 6.4, consists of two networks, each of which is associated with a specific arm. In addition, within a single network there are two main streams, each of which separately processes the spatial and the anatomical transformation. Their respective output is then projected to a competitive network, performing the selection of the appropriate response. Since the considered task instructions clearly specify which arm should be used, an effector selection process was not modeled.

EXTERNAL INPUTS

As external inputs, the two streams receive the visually perceived arm and body orientation vectors \vec{s}^{Arm} and $\vec{s}^{\text{Body}} \in \Gamma_3$. According to the visual reference frame of the observer (shown in Fig. 6.1), they can be written as

$$\vec{s}^{\text{Arm}} = \begin{pmatrix} \sin(\varphi^D) \sin(\theta^D + \phi^D) \\ \cos(\varphi^D) \\ -\sin(\varphi^D) \cos(\theta^D + \phi^D) \end{pmatrix} \quad \text{and} \quad \vec{s}^{\text{Body}} = \begin{pmatrix} \sin(\phi^D) \\ 0 \\ -\cos(\phi^D) \end{pmatrix} \quad (6.5)$$

These inputs are fed into the input populations of each transformation using Equ. (3.16). Note that the spatial transformation does not need information relative to the orientation of the demonstrator's body. In addition, the input populations also receive an external modulation applied asymmetrically to each stream, so that the global activity of the network performing the relevant trans-

formation (given by the task instructions) is stronger. While the inputs of the relevant network receive a positive modulation, those of the other receive inhibition, i.e., h^{Task} and $-(h^{\text{Task}} + \Delta h)$, respectively, where $h^{\text{Task}} > 0$ is a constant modulatory term, and $\Delta h > 0$ is an additional term which allows us to vary the strength of the competition. Indeed, as Δh increases, the network corresponding to the irrelevant transformation becomes more inhibited. Above a certain threshold, it is completely inactive. This case, where only the relevant transformation is active, will further be considered the baseline condition.

SPATIAL TRANSFORMATION

The spatial transformation consists of a mapping between the visually perceived orientation of the demonstrator’s arm and the imitator’s left and right arm, regardless of the demonstrator’s body. For a given arm, two neural populations are required. The former receives the visual input and is connected to the latter through synaptic projections. Since the spatial mapping does not require any other information, it can be performed by a direct projection. Using Eqs. (6.1) and (3.16), the correct mapping functions for the left and the right arm are given by $\mathbf{M} = \mathbf{M}^{\text{L,Sp}}$ and $\mathbf{M} = \mathbf{M}^{\text{R,Sp}}$, respectively, where

$$\begin{aligned} \mathbf{M}^{\text{L,Sp}} &= \begin{cases} \mathbf{diag}(1, 1, -1) & \vec{s}_z^{\text{Arm}} < 0 \\ \mathbf{I} & \text{otherwise} \end{cases} \quad \text{and} \\ \mathbf{M}^{\text{R,Sp}} &= \mathbf{diag}(-1, 1, 1) \mathbf{M}^{\text{L,Sp}} \end{aligned} \quad (6.6)$$

Since the frames of references of each arm are symmetric, so are the mapping matrices.

ANATOMICAL TRANSFORMATION

The anatomical transformation requires the combination of two variables, i.e., the orientation of the demonstrator’s arm and that of his/her body. Neurophysiological data suggest that such a transformation is performed through gain fields, which are neural populations combining inputs from several external sources (Scherberger & Andresen, 2003; Deneve & Pouget, 2003). Similarly to a neural model of gain field for transformations across reference frames described in Section 5.3.4 of this thesis, a gain field is defined as a continuous set of neural fields denoted by $\text{GF}\phi$, where each of them is preferentially tuned to a specific body orientation ϕ . The population encoding the demonstrator’s arm orientation projects to each of these sub-networks using Equ. (3.16) with $\mathbf{M} = \mathbf{R}_y(-\phi)$ as the mapping function. $\mathbf{R}_y(-\phi)$ corresponds to the rotation matrix around axis Y with angle $-\phi$. An additional term in the synaptic weights, as described in Section 5.3.4, was also applied in order to avoid a bias toward the pole in the transformation and hence to perform an accurate rotation. Next, the body

orientation is fed to the subfields through their modulatory input $h^{\text{GF}\phi}(t)$ using Equ. (3.17), with $\mathbf{M} = \mathbf{I}$ and $\tilde{\mathbf{r}}^{\text{GF}\phi} = (\sin \phi, 0, -\cos \phi)$. Finally, the gain field projects to the output population of network by synaptic projections with $\mathbf{M} = \mathbf{I}$.

RESPONSE SELECTION

The response selection is performed by a neural field receiving projections from the output population of both transformations. As illustrated in Figure 6.3, and already described at length in Chapter 3 and applied in Chapter 4, the competition is known to arise naturally from recurrent network connectivity, producing a sort of winner-take-all operation (Erlhagen & Schöner, 2002). When two inputs of a neural field are separated in neural space, they compete, and the input with the highest intensity wins the selection process. The cost of this competition is a longer convergence time with respect to the case where two close inputs cooperate. In this model, since the output strength of both streams are asymmetrically balanced, the correct response is always selected by the network. The network also receives a go signal by means of its modulatory input. Prior to the presentation of the target posture, $h^{\text{Sel}}(t) = -h^{\text{Go}} \ll 0$ so that the neural field is completely inactive. When the target posture is presented, the go signal goes to zero, i.e., $h^{\text{Sel}}(t) = 0$, and the selection process begins. The network response is read-out using the population vector (Equ. (3.11)), which directly represents the arm posture in the frame of reference of the imitator.

6.4 RESULTS

The two experiments described in Section 6.2 were simulated. The demonstrator's arm and body postures, were systematically varied across each trial during both experiments. The simulation parameters can be found in Appendix B.2. Moreover, in experimental conditions involving the use of the left arm, the subnetwork corresponding to the right arm was not considered, and vice versa.

6.4.1 REACTION TIMES AND ACCURACY

The mean reaction times and the errors resulting from the transformations were measured for both experiments described in Section 6.2. Reaction times were defined as the time when the response energy $E(t)$ given by Equ. (3.14) of the selection network reached a given threshold, whereas the transformation errors was defined as the angular distance between the population vector response $\vec{p}(t)$ after network convergence, and the correct target position. Moreover, since the dynamics of arm movements is not modeled, reaction times should

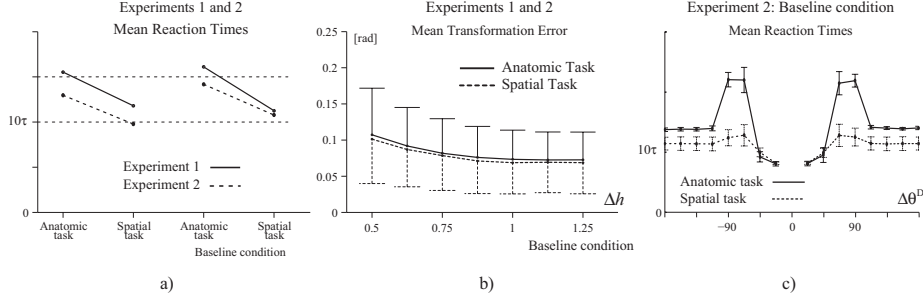


Figure 6.5: **a)** Mean reaction times and **b)** transformation errors observed in both experiments. **c)** Reaction times during Experiment 2 in the baseline conditions shown according to the amount of postural change $\Delta\theta^D$.

be considered to be times of movement initiation rather than times of movement completion.

As reported in Figure 6.5a, in both experiments, the average reaction time of the system in anatomical conditions was systematically longer than during the spatial task. Indeed, the former transformation requires more computations. In addition, increasing the inhibition Δh of the irrelevant transformation produced a slight increase in average reaction times, which was not really significant. Nevertheless, this modulation had a clear effect on the transformation errors (see Fig. 6.5b). Indeed, a weak inhibition produces a stronger competition between the parallel transformations, which results in larger errors. In Experiment 2, the amplitude of the postural change $\Delta\theta^D$ was different across trials. The reaction times dependency on this experimental variable in the baseline conditions is shown in Figure 6.5c. For small postural changes, reaction times were longer, but then decreased for larger $\Delta\theta^D$. This effect is caused by the center-surround recurrence in the neural dynamics, resulting in longer convergence times when moving from one attractor state to another, which is sufficiently close.

6.4.2 INTERFERENCE PATTERNS

The interference patterns resulting from the competition between the two transformations are now described. Only the conditions expected to show maximum interferences are considered, i.e., where the inhibitory modulation Δh is minimal. The influence of the amount of this inhibition will be described later in Section 6.4.3. The reaction times and transformation errors were considered relative to the baseline conditions. They are respectively defined by $\Delta RT_{\Delta h} = RT_{\Delta h} - RT_{\Delta h_0}$, and $\Delta Err_{\Delta h} = Err_{\Delta h} - Err_{\Delta h_0}$, where $RT_{\Delta h}$ and $Err_{\Delta h}$ correspond respectively to the reaction times and errors measured for a given Δh . Δh_0 corresponds to the inhibitory term in the baseline conditions.

In Figure 6.6, data from Experiment 1 are given according to the discrepancy D between the responses of the anatomical on one hand, and the spatial transformation on the other hand. First, since the processing time of the spatial

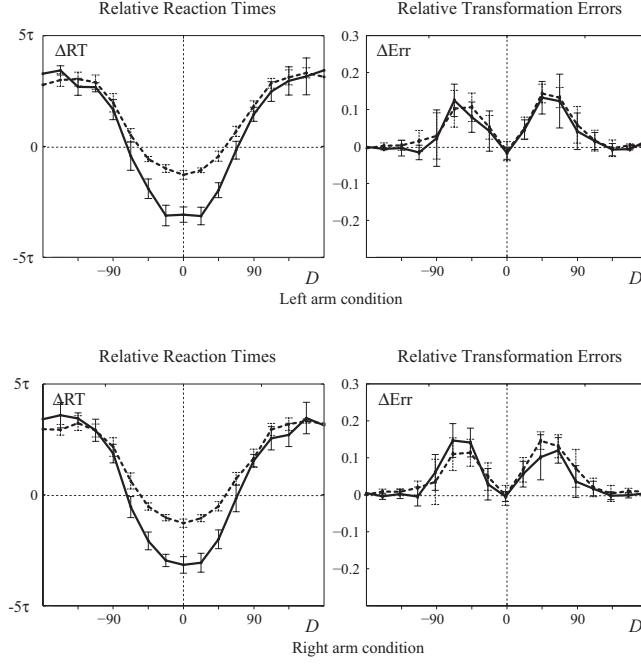


Figure 6.6: Results of Experiment 1: Reaction times and transformation errors relative to the baseline condition are shown.

transformation is shorter, it interferes earlier with the anatomical transformation, and conversely. As an effect, the strength of the interferences on reaction times were globally higher in anatomical conditions. Next, the reaction times increased with the discrepancy between the responses, whereas transformation errors behaved slightly differently. The errors did also increase with the discrepancy, but only within a small range. For outermost distances, they decreased until approximately zero. This effect is the result of the averaging of close responses on the neural field.

Similar effects were observed in Experiment 2 (see Figure 6.7), i.e., the interference patterns were globally more important under anatomical conditions and the error patterns also depended on the discrepancy between the responses. Further, the interference patterns on reaction times exhibited a combination of the effects of both the discrepancy D between the responses and the amount $\Delta\theta^D$ of arm postural change that were shown earlier in Figures 6.5c and 6.6. In conditions close to ideal congruency between the transformations, a general facilitatory effect was primarily produced which was even stronger for mid-range distances. In addition, an interaction between both variables on reaction times was observed. It produced a small shift of the interference pattern relative to the discrepancy D , which depended on $\Delta\theta^D$. In anatomical conditions, when the response of the spatial transformation is in the course of the anatomical transformation, the facilitatory effect is strengthened, whereas when the former is located at a distance, it is weakened. Since this dependency between the responses is primarily caused by the difference in processing times, its effect is

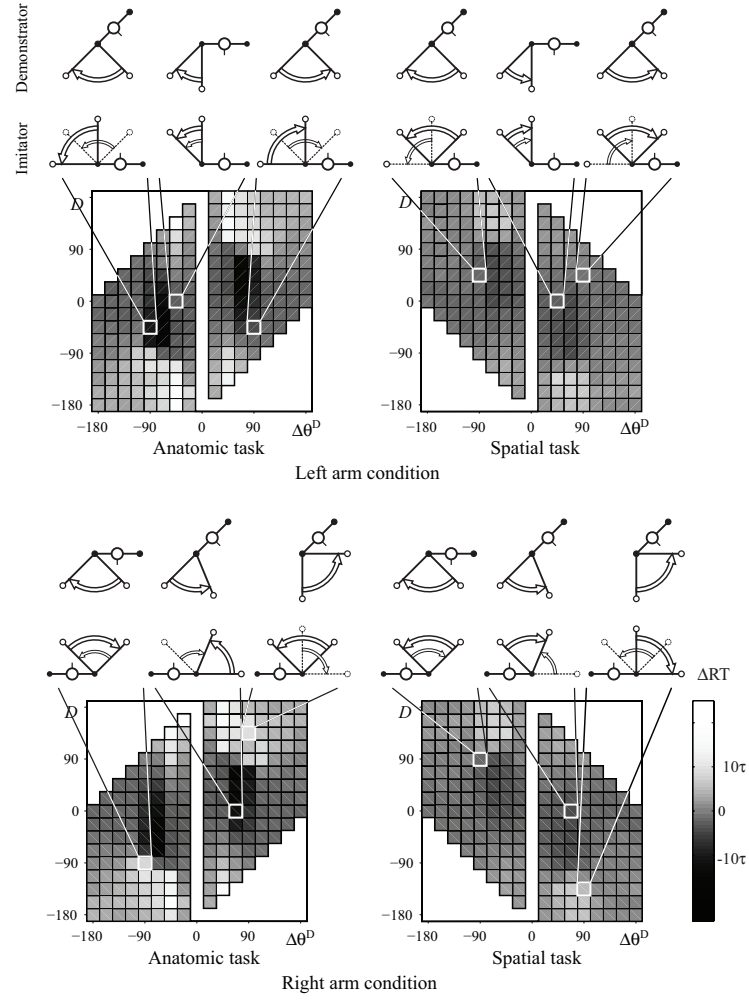


Figure 6.7: Results of Experiment 2: Reaction times relative to the baseline condition. On top of each plot, examples of experimental conditions are shown. In each case, the largest arrow corresponds to the response of the relevant transformation.

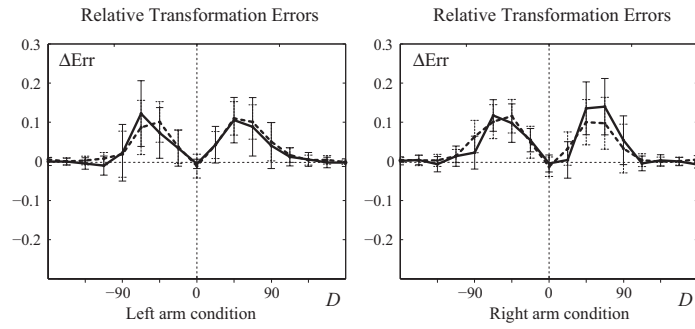


Figure 6.8: Results of Experiment 2: Transformation errors relative to the baseline condition.

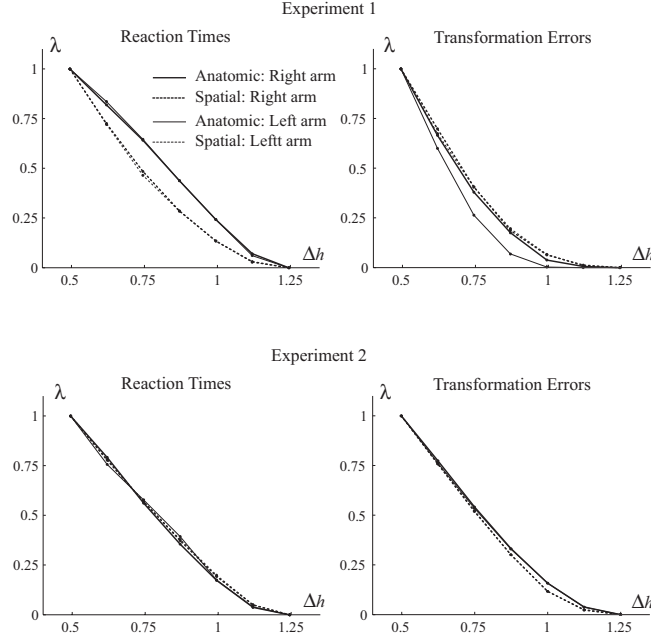


Figure 6.9: Effect of the inhibitory modulation of the irrelevant transformation on the amplitude of the interference patterns.

reversed in spatial conditions. Finally, because the errors were measured after network convergence, they were not different from those reported in Experiment 1 (see Figure 6.8).

6.4.3 EFFECTS OF INHIBITORY MODULATION

As mentioned in Section 6.4.2, the amount of the inhibition Δh is expected to have a modulatory effect on the interference patterns. Thus, the interference patterns of the condition where the least inhibition was applied, was used as a basis for fitting the patterns observed in the other cases. For each amount of inhibition Δh , the factor $\lambda_{\Delta h}$ which minimizes

$$\lambda_{\Delta h} = \min_{\lambda} (\lambda \Delta RT_{\Delta h_{\min}} - \Delta RT_{\Delta h})^2 \quad (6.7)$$

was calculated. $\Delta RT_{\Delta h}$, which was already defined in Section 6.4.2, corresponds to the relative reaction times when the modulatory term has value Δh . The same fitting procedure was also calculated for the network response errors. The results, depicted in Figure 6.9, show that the strength of the interference patterns decreased almost linearly with increasing Δh . As have been expected, the more irrelevant transformation is inhibited, the less interferences (either positive or negative) are observed.

6.5 DISCUSSION

In this chapter, a biologically-inspired neural model addressing the problem of transformations across frames of reference in a posture imitation task has been presented. This modeling work is based on the hypothesis that such an imitation process is mediated by two concurrent transformations, corresponding to the spatial and the anatomical imitative strategies (Brass et al., 2000; Bertenthal et al., 2006; Chiavarino et al., 2007). These strategies may have evolved separately for different specific uses. The former may result from early visuo-motor development where the mapping to own hand would have been extended to be responsive to any hand (Piaget, 1978). Since the observation of one's own hands is strictly experienced in viewer-centered coordinates, the mapping between the executed movements and the observed consequences corresponds to pure spatial associations. Differently from the spatial strategy, the anatomical strategy may be an after effect of the mental development of a body schema which respects strictly body configurations (Gallagher, 2000). By extension, this system would strengthen the anatomical mapping between the sides of observed people and of self.

Further, an experimental paradigm has also been designed, which allows to measure the interference patterns produced by the interaction between the anatomical on one hand, and the spatial imitative strategy on the other hand. In addition, separate instances of the pair of transformations are assumed to be responsible for the control of each side of the body. Since the described experiments did not involve the use of both arms simultaneously, this latter hypothesis does not rule out the fact that the processes of each arm may be coupled and located within a single brain region (Arbib et al., 2000; Koski et al., 2003). As such, the present results provide predictions of real behavioral responses.

Similar to other works which applied the Dynamic Field approach (Erlhagen & Schöner, 2002), this work goes beyond usual binary models often proposed in experimental psychology (Zhang et al., 1999), by modeling continuous stimulus variables and responses. In addition, this approach is of high biological significance. Neurophysiological studies have shown that, in the superior temporal sulcus, body and arm postures are encoded into neural populations where each neuron exhibits tuning to a specific posture (Ashbridge et al., 2000). Similarly distributed representations encoding different modalities have also been reported in many other brain areas such as the motor, the premotor and the parietal cortex and the cerebellum (Georgopoulos, 1996; Cisek & Kalaska, 2005; Roitman, Pasalar, Johnson, & Ebner, 2005; Aflalo & Graziano, 2006a). Moreover, neural correlates of decisional processes within such representations have also been reported in the premotor cortex (Cisek & Kalaska, 2005). Together, these findings strengthen this approach by grounding it on a strong biological

basis.

Behavioral studies on imitation report greater interferences during tasks where the spatial transformation is irrelevant, as compared to tasks where anatomical imitation has to be avoided (Brass et al., 2000; Heyes & Ray, 2004; Bertenthal et al., 2006; Chiavarino et al., 2007). This modeling work supports this observation, but explains it in terms of the longer processing time required by the anatomical imitative strategy, which needs to process an additional variable. Usual accounts for the greater influence of the spatial transformation consider primarily a stronger linkage with the decisional process (Zhang et al., 1999; Brass et al., 2000; Koski et al., 2003; Bertenthal et al., 2006). Although both hypotheses are compatible, one may be interested in determining their respective influence, which would need more investigations.

This work also showed that combining transformations can produce interferences. One may wonder why the nervous system would use a combination of two strategies for solving imitation tasks since they produce interferences. An answer may be given with respect to the reported simulations which show that, in specific conditions, their interaction result in positive effects. For instance, when the imitator and the demonstrator are face to face, mirror imitation is faster, whereas anatomical imitation is more effective when the imitator looks at the back of the demonstrator. From this, an alternative hypothesis can be proposed, which may explain that, in unconstrained conditions and when people are facing each other, mirror imitation is the most usual strategy for copying meaningless gestures (Bekkering et al., 2000). Rather than simply assuming that mirror imitation has a stronger influence on the selection process (Koski et al., 2003; Heyes & Ray, 2004; Bertenthal et al., 2006), this work suggests that this strategy is the one which exhibits the maximal congruency between the concurrent transformations. Additional neurophysiological evidence supporting this hypothesis can be found in an fMRI study showing that some of the brain areas activated during the imitation of finger movements are more active during specular than during anatomical imitation (Koski et al., 2003). In this experiment, the authors did not consider the hypothesis that an anatomical mapping could exist between ipsilateral hands. The mirror condition which they showed to produce higher brain activation, corresponds in this study to a condition where the responses of the parallel strategies are perfectly congruent. Since this case is effectively the one in which the model produces responses with the highest energy, the nervous system may hence be naturally biased toward this strategy.

Behavioral experiments on imitation often involve a binary selection process, which consists of responding to a stimulus with either the index or middle finger, or with either the left or right arm, to state two examples (Brass et al., 2000; Heyes & Ray, 2004; Bertenthal et al., 2006). Here, the suggested experimental paradigm did not requested such an effector selection process. It rather asked for a specific response in a continuous space. This difference could explain why the

simulation results are not completely in concordance with the data of an experimental study addressing the effects of spatial compatibility in imitation, which used similar body postures as stimuli (Heyes & Ray, 2004). Thus, future developments of the model should address the problem of effector selection. Finally, this work focused on the imitation of static postures, whereas in real conditions, imitative tasks are often performed in a dynamic environment. Since the model does not take movement velocity into account, future work should also aim at integrating the sensitivity to this property into neural models. Nevertheless, as suggested by another behavioral study by Stürmer et al. (2000), imitation of postures and movements may also be processed separately, strengthening the predictions of this model concerned with the imitation of body postures.

Summary

This chapter presented a neural model which evaluates two possible forms of transformations across frames of reference in the context of an interference paradigm. An experimental setup has been proposed in order to test the computational hypotheses developed here on real human subjects. Like the neural models proposed in Chapter 4, interference produced by competitive interactions between contradictory inputs affect the reaction times of movement execution. In contrast to those models, the model presented here neither explicitly mentioned nor addressed the issues related to shared representations. This model considers only an input-output sensorimotor mapping of the observed movements to a motor representation; doing otherwise would have required a much more complex network capable of simultaneously integrating the representation of others' movements with those of the imitator, while still being aware of the ownership of each sensory input. As discussed in Chapter 2, humans are capable of discriminating between self and others' actions; this capability requires additional neural mechanisms aside from those presented here. Nevertheless, the absence of such mechanisms in this model would certainly have only small repercussions on the presented behavioral results. Indeed, since imitation was the aim of the task, the movements to be executed are, by definition, always congruent with those observed. Stronger interference would have been reported only if the movements were to be different or incongruent (Jeannerod, 2003; Kilner et al., 2003).

In order to address the issue of discrimination within shared representations, a neural model based on current knowledge related to the monitoring of self versus others movements will be presented in Chapter 8. But before addressing the cortical pathways involved in this cognitive function, the next chapter presents a neural field model capable of discriminating between self-generated and externally produced inputs. The model consists of an extension of the continuous attractor neural network models described in Section 3.2.2. The combination of the capacity to integrate velocity signals and the sensitivity to the speed of the external inputs allows a neural field to select the input related to self-generated movements.

MOTION INTEGRATION, MOTION SENSITIVITY, SENSORY PREDICTION AND SENSORY DISCRIMINATION

*The study presented in this chapter was adapted from
the work which has been published in:*

Sauser, E. L. and Billard, A. G.. Dynamic updating of distributed neural representations using forward models. *Biological Cybernetics*. 95(6):567-588.

SHARED representations are often thought to be responsible for representing motor execution patterns associated with both one's own will as well as observed actions performed by other individuals. This substrate, while providing a fundamental neural basis for the study of the processes mediating imitation, raises several questions. One concerns the issue of what happens in these shared representations when the execution of a movement is done simultaneously with the observation of a movement performed by another individual. In particular, what happens when these two actions are different?

Chapter 4 presented evidence demonstrating interference resulting from conflicts within shared representations when the movement to be executed was different from that indicated by the triggering movement stimulus. However, in order to succeed in this task, the strength of the internal representation of one's own choice was explicitly set to be greater than that of the interfering external cue. This implicitly implied that the model must know which input, among those entering the shared representations, corresponds to its own choice. This worked because there was no ambiguity as to which visual and motor signals belonged to the model. Nevertheless, in ambiguous situations, such as those illustrated in Figure 2.3 of Chapter 2, one must be capable of distinguishing which of the visual cues is controlled by one's own will. Humans are capable of discriminating between actions that are self-generated and those performed by other individuals. The key processing components which have been suggested to be involved in this ability are a forward model of movement execution and a comparator (Decety & Sommerville, 2003; Jeannerod, 2003). The forward model uses motor efference copies in order to predict the consequence of motor

acts, while the latter compares this prediction with actual sensory consequences. A more or less perfect matching then indicates that the perceived movements may effectively be self-generated.

In order to provide a plausible neural mechanism underlying this human ability, this chapter presents an extension of neural field models capable of realizing this operation of discrimination. This model is capable of performing different but related computations, which may allow velocity integration, velocity discrimination and motor imagery. All of these processes imply complex interactions between sensory inputs and the network involved in their cortical representation. In contrast to the previous modeling studies presented here, these interactions require continuously changing sensory inputs which need to be dynamically integrated into the internal neural representation so that the representation can continuously, quickly and accurately be kept up to date. These two issues are reviewed in this chapter, followed by the description of the extension as applied to the classical architecture of neural fields. Several experiments are reported; these demonstrate how the model can be involved in the cognitive processes responsible for velocity integration, velocity discrimination and motor imagery. The neurophysiological and behavioral implications raised by this model are discussed in the light of the simulation results.

7.1 INTRODUCTION

Two important issues are developed in this modeling study. The first point concerns the dynamics of the interactions between external stimuli and a neural field, and more precisely in the case where stimuli are moving along the neural representation. Indeed, apart from purely abstract theoretical works and a few applied to the modeling of biological systems (Zhang, 1996; Mineiro & Zipser, 1998; Giese, 2000; Xie et al., 2002), the use of such networks in practical neurobiological modeling comes along with the assumption of quasi-static neural dynamics. It means that the time scale of the motion of the external inputs is much larger than that of the network. Since the dynamics of neural fields are, by definition, strongly influenced by the recurrent connectivity, the reaction time to a changing stimulus is thus higher than that of a single neuron (Panzeri, Rolls, Battaglia, & Lavis, 2001). Therefore, when a quasi-static network dynamics is considered, the time scale at which the inputs are updated is set to a very large value. Nevertheless, the comparison of these network implementations with real biological systems does not usually lead to biased interpretations since their performance is often considered in relatively slow tasks. However, when dealing, for example, with precise and fast movements like catching a ball or smooth eye pursuit, the internal representations should be updated very quickly, and even in advance, so that the outcome of self-generated movements could be predicted accurately (Miall & Wolpert, 1996; Wolpert & Kawato, 1998). An influential

theory related to motor control suggests that the brain may use forward models in order to better control movements. The key argument in this direction is that, in humans, a closed-loop control system alone would be relatively inaccurate since the time needed for sending motor commands and then receiving the resulting sensory feedback is quite long (Miall & Wolpert, 1996; Wolpert & Kawato, 1998; Miall & Reckess, 2002). For instance, during eye-tracking of a self-moved target experiment, it has been shown that when a subject actively moves a target, the presence of the movement sensory feedback is not necessary to achieve almost zero latency, whereas motor efference copies are fundamental to perform this task (Vercher et al., 1996). Furthermore, neurophysiological data show that predictive neural responses can be found in the monkey visual, parietal and frontal cortices, as well as in the cerebellum (Unema & Goldberg, 1997; Nakamura & Colby, 2002; Roitman et al., 2005), which suggests the presence of neural processes involved in the forward control of movements.

The second motivation concerns the neural mechanisms of self-awareness and recognition. A current theory related to this problem considers two forms of cues which may be at the origin of this human ability (Decety & Sommerville, 2003; Haggard & Clarke, 2003; Jeannerod, 2003). When one has to recognize one's own limb, body cues, such as the spatial and visual attributes of the limb are primarily used. More interestingly, in the case of ambiguous visual attributes, it has been shown that one relies more on the so-called action or movement cues such as the time course of the movement which includes its velocity and acceleration (Jeannerod, 2003). A plausible mechanism derived from motor control theories has been suggested whereby an internal prediction of the consequences of a motor act is compared with the real sensory outcome. Then, depending on their similarity or discrepancy, the brain becomes capable of determining the ownership of the observed movement (Decety & Sommerville, 2003). In addition, another related issue may be brought to light by considering the neural substrates concerned with the perception of one's own actions and those of others. Current body of evidence suggests that a common neural substrate devoted to both the recognition and production of movements exists in both humans and monkeys (Iacoboni et al., 1999; Rizzolatti et al., 2001). A behavioral correlate of this discovery, reported by several psychophysics experiments, is that observing movements of others influences the quality of one's own performance (Kilner et al., 2003; Chaminade et al., 2005). The observation of such an interference effect, while supporting the view of a common pathway for the transfer of visuo-motor information, calls for an explanation as to how the same neural substrate can both integrate multisensory information and, at the same time, determine the ownership of the observed movements.

Despite the apparent differences between these two introduced topics, they both share a common and fundamental requirement. They need a **predictive process**, which allows, respectively, *a*) an almost instantaneous update of the internal representations of the actual sensory state and *b*) the computation of

sensory predictions to be compared with movement outcome. Indeed, the timing of neural updating of the internal sensory information is crucial in motor control for the generation of accurate movements (Vercher & Gattthier, 1988; Miall & Reckess, 2002). Similarly, a short time between movement execution and the perception of its sensory feedback is also crucial for being able to feel the agency of actions (Haggard & Clarke, 2003). The main interest that has driven the work presented in this chapter, is effectively to show how a neural field can integrate the efferent commands incoming from a forward model in order to update its internal representation. As internal representations, the present modeling is restricted to neural ensembles which encode simultaneously a variable value and its first order time derivative. For instance, the position and the velocity of either a visual stimulus in retinal space, or the hand location in cartesian or joint space may be considered. Note that real populations of neurons exhibiting this type of sensitivity have been reported in many brain areas, such as the motor, parietal, visual and temporal cortices and the cerebellum (Hubel & Wiesel, 1977; Kettner et al., 1988; Ben Hamed, Duffy, & Pouget, 2003; Jellema et al., 2004; Roitman et al., 2005; Aflalo & Graziano, 2007). The contributions of this modeling study are threefold. First, a generalized framework for the dynamic integration of velocity commands within continuous attractor neural networks is developed. Secondly, a detailed stimulus encoding function, which compensates for the internal dynamics of both neurons and the network is considered. And thirdly, the properties and several applications resulting from the proposed neural integration mechanism are described and analyzed. They include *a)* a dynamically adjustable sensitivity to the velocity external inputs, *b)* the ability to transfer information between cortical networks in an almost instantaneous and predictive-like fashion, and *c)* the capacity to discriminate stimuli according to their dynamical properties.

The remaining of this chapter is organized as follows. First, Section 7.2 introduces the model core architecture and then describes the mechanisms underlying the integration of velocity commands. Next, the precise profiles of stimulus encoding and the synaptic projections for transferring information across neural populations are defined. Note that detailed mathematical calculations were left in Appendix A.2. Further, in Section 7.3, an analysis of the network properties is provided. Finally, in Section 7.4, the relationships between the properties of the proposed model and biological data is considered in more details. In addition, its implications related to several important cognitive processes in terms of neural mechanisms are provided.

7.2 NEURAL MODEL

7.2.1 ARCHITECTURE

First of all, similarly to Chapter 5, since the work described here presents an extension of classical neural field models, the majority of the equations governing their dynamics and their inputs will be rewritten so that they could account for the proposed technical modifications. First of all, instead of being sensitive to a single neural variable $\vec{r} \in \Gamma$, the continuous attractor neural network developed here is composed of a continuum of neurons preferentially tuned to a primary variable $\vec{r}_{\mathcal{A}}$ and a secondary variable $\vec{r}_{\mathcal{B}}$ following a uniform distribution such that $\vec{r}_{\mathcal{A}} \in \Gamma^{\mathcal{A}}$ and $\vec{r}_{\mathcal{B}} \in \Gamma^{\mathcal{B}}$, respectively. As mentioned in the introduction, these two neural parameter spaces are assumed to represent, respectively, a variable value and its variation in time. For example, if a stimulus location is encoded in the space $\Gamma^{\mathcal{A}}$, a neuron tuned to a specific $\vec{r}_{\mathcal{B}}$ should fire preferentially when the stimulus is moving in the direction given by that $\vec{r}_{\mathcal{B}}$. This hypothesis related to a combined preferential tuning to both a variable value and its direction of variation is primarily motivated by several neurophysiological studies showing neurons exhibiting similar firing properties (Kettner et al., 1988; Fu, Flament, Coltz, & Ebner, 1997; Jellema et al., 2004; Roitman et al., 2005). In addition to the computational power that this type of combined neural representations may provide to the brain (Salinas & Abbott, 1995; Pouget & Snyder, 2000), the following analysis will show that it can also facilitate the updating of its internal representation. For example, a neural network encoding both the position and the instantaneous velocity of a moving target at a given time, should be capable of predicting where the target should be in the near future.

The parameter spaces of the primary variable $\vec{r}_{\mathcal{A}}$ which are considered here, consist of a unidimensional and a two dimensional neural space (see Table 7.1). Moreover, in order to avoid boundary effects, periodic spaces are assumed. Therefore, the considered domains are, respectively, a ring and its two dimensional analog, a torus. Despite of the discrete nature of $\vec{r}_{\mathcal{B}}$ in the case of the ring attractor, the neural ensemble follows a continuous attractor network dynamics where the time evolution of the neurons' membrane potential $u(\vec{r}_{\mathcal{A}}, \vec{r}_{\mathcal{B}}, t)$

Table 7.1: Definition domains of neural preferential tuning considered in this chapter. Note that except in the case where it is mandatory, the index indicating the dimension of the parameter space is omitted.

	Ring, $N = 1$	Torus, $N = 2$
$\vec{r}_{\mathcal{A}} \in$	$\Gamma_1^{\mathcal{A}} = \{\vec{r}_{\mathcal{A}} \in \mathbb{R}^2 \mid \ \vec{r}_{\mathcal{A}}\ = 1\}$	$\Gamma_2^{\mathcal{A}} = \Gamma_1^{\mathcal{A}} \times \Gamma_1^{\mathcal{A}}$
$\vec{r}_{\mathcal{B}} \in$	$\Gamma_1^{\mathcal{B}} = \{-1, 1\}$	$\Gamma_2^{\mathcal{B}} = \{\vec{r}_{\mathcal{B}} \in \mathbb{R}^2 \mid \ \vec{r}_{\mathcal{B}}\ = 1\}$

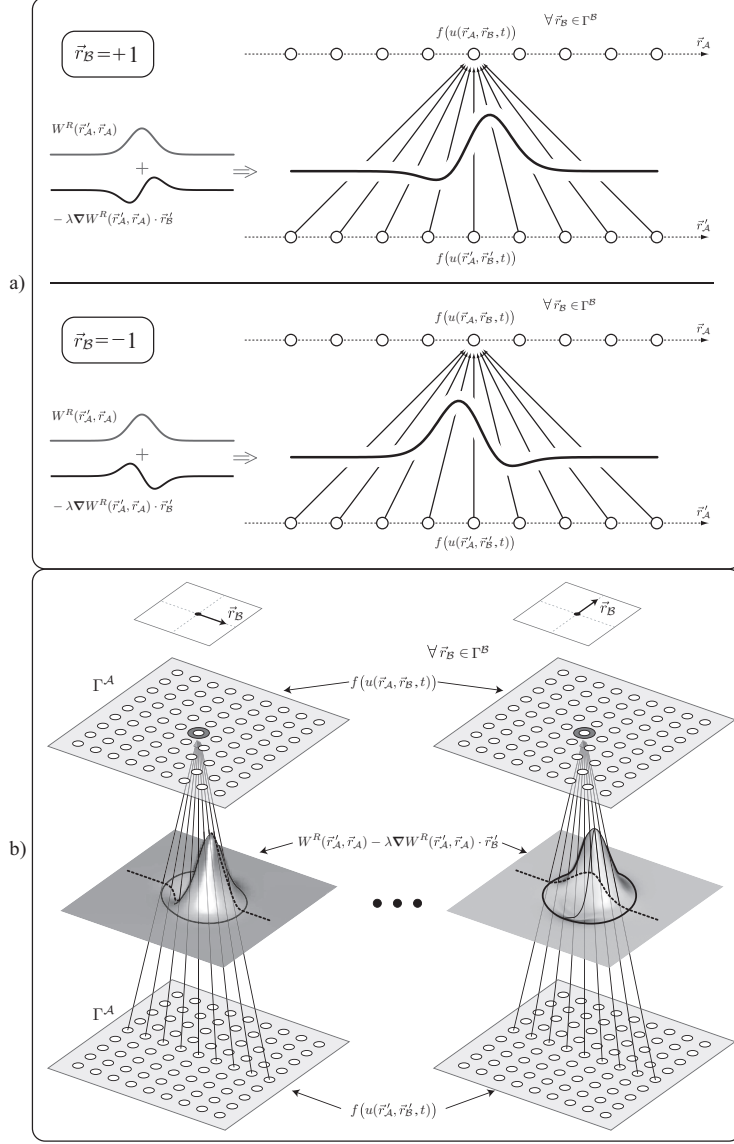


Figure 7.1: Illustration of the model architecture and weight kernels for both the ring (top) and the torus attractor spaces (bottom). Each sub-layer encoding a variable in the neural space defined on \vec{r}_A , has an preferred movement direction \vec{r}_B which is the result of the asymmetric self-connectivity. This weight kernel is shown for different values of \vec{r}_B . It is superimposed on the arrows denoting the synaptic projections across the sub-layers.

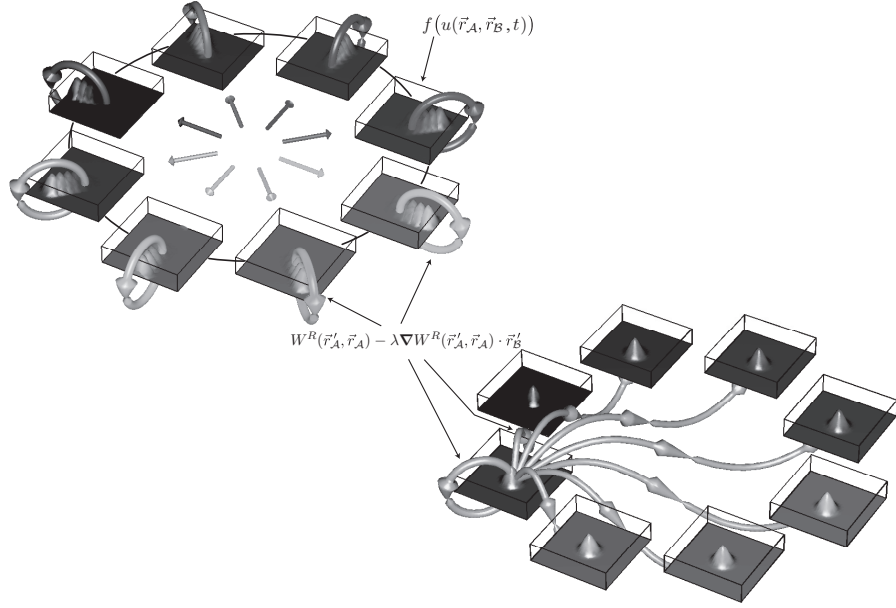


Figure 7.2: Illustration of the model architecture. Each sub-layer encoding a variable in the neural space defined on $\vec{r}_A \in \Gamma^A$, has a preferred movement direction $\vec{r}_B \in \Gamma^B$ which is the result of the asymmetric self-connectivity. On top, the center arrows indicate the preferred movement direction of the sub-layers. If considered alone, each of them would display a traveling activity blob in its preferred movement direction, as shown by the trace of the activity blobs on the neural surface. At the bottom, it is illustrated how each sub-layer projects its activity onto all sub-layers using the same weight profile.

satisfies

$$\begin{aligned} \tau \dot{u}(\vec{r}_A, \vec{r}_B, t) = & -u(\vec{r}_A, \vec{r}_B, t) + h(\vec{r}_B, t) + x(\vec{r}_A, \vec{r}_B, t) + \\ & \oint [W^R(\vec{r}'_A, \vec{r}_A) - \lambda \nabla W^R(\vec{r}'_A, \vec{r}_A) \cdot \vec{r}'_B] \\ & f(u(\vec{r}'_A, \vec{r}'_B, t)) d\vec{r}'_A d\vec{r}'_B \end{aligned} \quad (7.1)$$

The transfer function $f(u)$ is the linear threshold function $\max(0, u)$ and $\tau \in \mathbb{R}_+^*$ is the membrane time constant. The network receives external inputs, which were separated into two distinct forms. $x(\vec{r}_A, \vec{r}_B, t)$ is defined as the **stimulus input** whereas $h(\vec{r}_B, t)$ as the **background input**. As can be noticed, $h(\vec{r}_B, t)$ is, by definition, homogeneous across the group of neurons sharing the same preferential tuning to the variable \vec{r}_B . From now on, these groups of neurons will be designated as sub-layers.

The network is fully and recurrently connected by means of synaptic weights which can be decomposed into two parts: a center-surround Gaussian-like¹, translation-invariant and symmetric component $W^R(\vec{r}'_A, \vec{r}_A)$ and an asymmet-

¹The exact shape of W^R is not crucial as long as it can allow the network to sustain an activity packet on its neural surface (Amari, 1977). Nevertheless, in the simulations that will be reported further in this chapter, the profile of these weights complies with that already defined in Chapter 3 by Equ. (3.8).

ric term $-\lambda \nabla W(\vec{r}'_{\mathcal{A}}, \vec{r}_{\mathcal{A}}) \cdot \vec{r}'_{\mathcal{B}}$ where ∇ corresponds to the gradient operator along $\vec{r}_{\mathcal{A}}$. $\lambda \in \mathbb{R}_+^*$ is a constant scaling factor. The first term consists of a convolution kernel which links the neurons along the neural fundamental variable $\vec{r}_{\mathcal{A}}$, while the second term links them along $\vec{r}_{\mathcal{B}}$. These convolution weights and the network architecture are illustrated in Figures 7.1 and 7.2. As will be described in more details in Section 7.2.2, this connectivity is the core of the system ability to integrate velocity commands and to update its internal representation. The idea behind this recurrent synaptic profile comes from the asymmetric connectivity which allows traveling waves to emerge on the surface of neural fields (See Section 3.2.2). Then by connecting together neural fields having different intrinsic directions of motion, and by breaking the balance of excitation between each field, a traveling pulse solution may then develop in the intrinsic direction given by the strongest field.

Examples of activity packets or attractor bumps produced by this type of neural dynamics are illustrated in Figure 7.2. Since this network possesses two internal variables, the population vector $\vec{p}(t)$ has been modified with respect to that already defined in Equ. (3.11). In the present situation, the estimate $\vec{p}(t) \in \Gamma^{\mathcal{A}}$ of the variable value encoded along $\vec{r}_{\mathcal{A}}$ is given by

$$\vec{p}(t) = \frac{\oint f(u(\vec{r}_{\mathcal{A}}, \vec{r}_{\mathcal{B}}, t)) \vec{r}_{\mathcal{A}} d\vec{r}_{\mathcal{A}} d\vec{r}_{\mathcal{B}}}{\oint f(u(\vec{r}_{\mathcal{A}}, \vec{r}_{\mathcal{B}}, t)) d\vec{r}_{\mathcal{A}} d\vec{r}_{\mathcal{B}}} \quad (7.2)$$

An estimate of the variable encoded in $\Gamma^{\mathcal{B}}$ may be envisaged similarly. However, since this space is an extension of the primary space $\Gamma^{\mathcal{A}}$ and since it will be shown to be primarily used for integration purposes, its explicit definition is not necessary. In the next section, the description of the role of that dimension in the neural mechanism which can update the network internal representation is given.

7.2.2 CONTROL OF THE INTRINSIC DYNAMICS

This section first describes how external commands, such as those which can be produced by forward models, may drive the network internal dynamics so that the neural representation could be updated accordingly. Afterwards, the boundaries in which this velocity integration remains valid will be determined.

First of all, this work starts with the hypothesis that the neural field is already representing and sustaining information and that no stimulus input is presented, i.e., $x(\vec{r}_{\mathcal{A}}, \vec{r}_{\mathcal{B}}, t) = 0$. As already suggested by Xie et al. (2002) who proposed a neural architecture similar to that presented here, the key element for updating the network representation consists of exciting asymmetrically each sub-layers. By activating most that which is preferentially tuned to the desired direction of variation of the internal variable, the network should drive its internal representation correctly. Thus, let the background input $h(\vec{r}_{\mathcal{B}}, t)$ be defined

as

$$h(\vec{r}_B, t) = h_0 [1 + \vec{h}_0^d(t) \cdot \vec{r}_B] \quad (7.3)$$

where $h_0 > 0$ corresponds to a constant excitation level, and $\vec{h}_0^d(t)$ is a directional component which allows to break the network symmetry when different from zero (see Figure 7.3a). Indeed, when referring to Equ. (7.1), it can be noticed that a background input which is homogeneous over all sub-layers, i.e., $\vec{h}_0^d = \vec{0}$, leads the integral along \vec{r}_B of the asymmetric component of the recurrent weights to disappear, i.e.,

$$\oint \lambda \nabla W^R(\vec{r}'_A, \vec{r}_A) \cdot \vec{r}'_B f(u(\vec{r}'_A, \vec{r}'_B, t)) d\vec{r}'_A d\vec{r}'_B = 0 \quad (7.4)$$

In this situation, the system settles by symmetry into a constant state along \vec{r}_B , and produces a marginally stable bump solution. Thus, this network becomes equivalent to a typical continuous attractor neural network which does not possess any asymmetric connectivity. Consequently, in order to break this symmetry, an input $\vec{h}_0^d \neq \vec{0}$ has to be applied. It will favor higher excitation levels for the sub-layers having a preferred direction \vec{r}_B close to that indicated by \vec{h}_0^d .

Now, if a single sub-layer is first considered, it has been shown to develop a traveling activity bump symmetric in shape having a constant velocity. Moreover, this constant velocity has been proved to depend only on the ratio between the strength of the asymmetric component of the weights and the time constant of the neuron, i.e., λ/τ (Zhang, 1996). By introducing a strong coupling between the sub-layers having opposite direction preferences, each of them will try to drive the global network response toward its preferred direction of motion. This form of **push-pull** mechanism is then regulated by the respective balance of excitation across the sub-layers (Xie et al., 2002). In the present model, this effect is controlled by the asymmetric component \vec{h}_0^d of background input.

Further, when the internal representation is updated accurately according to a desired velocity $\vec{v}^*(t)$, the population vector $\vec{p}(t)$ of the network has to respect the following equation:

$$\dot{\vec{p}}(t) = \vec{v}^*(t) \quad (7.5)$$

Detailed calculations, provided in Appendix A.2 to lighten the text, show that near the equilibrium state, i.e., $\vec{h}_0^d = \vec{0}$, the relationship between the variation of the population vector and the asymmetric background input can be approximated linearly such that

$$\dot{\vec{p}}(t) \approx \frac{\lambda}{\tau} \gamma \vec{h}_0^d(t) \quad (7.6)$$

where $\vec{h}_0^d(t)$ is small. γ corresponds to the slope of the linear approximation

which uniquely depends on the parameters of the recurrent weights². In addition, since the update of the neural representation is made possible by modulating the influence of the sub-layers, the maximal velocity that the model may integrate is bounded by that of a single sub-layer, i.e λ/τ . Thus, the estimate of the internal velocity of the network may be rewritten as

$$\dot{\vec{p}}(t) \approx \frac{\lambda}{\tau} \min \left(\gamma \vec{h}_0^d(t), \frac{\vec{h}_0^d(t)}{\|\vec{h}_0^d(t)\|} \right) \quad (7.7)$$

Next, in order to find the correct expression of the background input to be applied to the network given a desired velocity $\vec{v}^*(t)$, Eqs. (7.5) and (7.7) can be substituted into Equ. (7.3), which gives

$$h(\vec{r}_B, t) = h_0 \left[1 + \frac{\tau}{\lambda\gamma} \vec{v}^*(t) \cdot \vec{r}_B \right] \quad \|\vec{v}^*(t)\| \ll \frac{\lambda}{\tau} \quad (7.8)$$

As a consequence, the internal representation of the neural field can be driven by an external command which indicates in which direction as well as at which velocity its update should be performed. This dynamical process thus transforms the continuous and marginal attractor state of the neural field defined on the neural space Γ^A , into a limit cycle having a constant velocity. This network is said to possess an **intrinsic velocity** that corresponds to the velocity at which a self-sustained activation pattern would move in absence of a stimulus input. Indeed, when a stimulus input is fed to the network, its response is affected and may not exhibit a traveling motion characteristic anymore. Before addressing this issue experimentally in Section 7.3.3, stimulus inputs are first defined in order to drive the network dynamics toward their own. That is, if an input is moving at given speed and direction, the dynamics of the neural field should not produce interferences but it should rather resonate with its input.

7.2.3 STIMULUS ENCODING

This section describes the profile a stimulus input should take in order to drive the network dynamics toward its own dynamics. Two possible forms are proposed, a non-linear and a linear one. The non-linear input profile is described first, since its derivation from the results obtained above is the most straightforward. Next, by a linear approximation, a linear form is developed. The interest in this second type of input is that it makes possible to derive the profile that synaptic projections between two neural populations should have in order to transmit both the positional and velocity-dependent information conveyed by the neural fields described here.

A stimulus located at $\vec{s}(t)$ in the neural space Γ^A is considered. Moreover, it is assumed to move in phase with the network intrinsic velocity, i.e., $\dot{\vec{s}}(t) = \vec{v}^*(t)$.

²Except in rare specific cases, γ has to be found numerically.

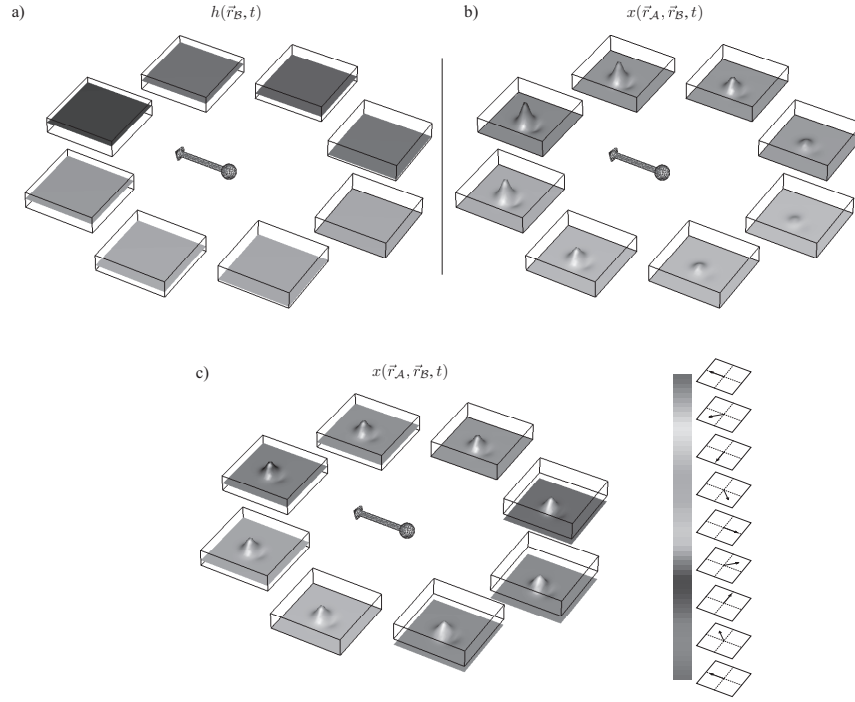


Figure 7.3: Examples of external inputs where the same representation as in Figure 7.2 is used. The color code on the bottom right of the figure indicates the corresponding preferred movement direction of each sub-layer. **a)** The background input $h(\vec{r}_B, t)$ (Equ. (7.3)), **b)** the stimulus input $x(\vec{r}_A, \vec{r}_B, t)$ (Equ. (7.9)) and **c)** the linear form of that stimulus input (Equ. (7.12)) are shown. The arrow in the middle of the figures denotes the intrinsic velocity of the network resulting when the corresponding input is applied.

The input profile $x(\vec{r}_A, \vec{r}_B, t)$ associated to this stimulus is given by

$$x(\vec{r}_A, \vec{r}_B, t) = h_1 \left[\mathcal{G}(\vec{r}_A, \vec{s}(t), \sigma_s) - \tau \dot{\vec{s}}(t) \cdot \nabla \mathcal{G}(\vec{r}_A, \vec{s}(t), \sigma_s) \right] \left[1 + \frac{\tau}{\lambda \gamma} \dot{\vec{s}}(t) \cdot \vec{r}_B \right] \quad \|\dot{\vec{s}}(t)\| \ll \frac{\lambda}{\tau} \quad (7.9)$$

where $h_1 > 0$ corresponds to the stimulus amplitude and \mathcal{G} to the function generating a gaussian profiles as defined in Equ. (3.7). σ_s indicates the breadth of the input gaussian which is centered on the stimulus location $\vec{s}(t)$. An illustration of the resulting shape of such a stimulus input is shown in Figure 7.3b. The second term of the first factor is meant to compensate for the integration time of the neurons. As a consequence, the activity profile of the network response to this input is unaffected by the motion of the stimulus. It stays symmetric, as if it was static in the neural medium.

Finally, the second factor is responsible for driving the push-pull mechanism of the neural field. Indeed, by comparing Equ. (7.9) with Equ. (7.8), it can be noticed that this second factor is similar to the input necessary to set the intrinsic dynamics to the neural field to a given desired velocity. Therefore, by means of this constant modulation along \vec{r}_B , the asymmetric recurrent connec-

tivity will help the neural field track the moving stimulus. Further, since the asymmetry of the stimulus input is proportional to its speed, a directional input factor \vec{h}_1^d is introduced. It has a similar role to that of the background input and is given by

$$\vec{h}_1^d(t) = \frac{\tau}{\lambda\gamma} \dot{\vec{s}}(t) \quad (7.10)$$

This relationship indicates how much the strength of the input asymmetry should be so that the network can follow its speed accurately. Then, rewriting Equ. (7.9) gives

$$x(\vec{r}_A, \vec{r}_B, t) = h_1 \left[\mathcal{G}(\vec{r}_A, \vec{s}(t), \sigma_s) - \tau \dot{\vec{s}}(t) \cdot \nabla \mathcal{G}(\vec{r}_A, \vec{s}(t), \sigma_s) \right] \left[1 + \vec{h}_1^d(t) \cdot \vec{r}_B \right] \quad (7.11)$$

Since the asymmetric factor of the stimulus input is proportional to its velocity, the use of either notation will be considered as equivalent in the further analysis of the experiments. This will be applied to the variable defining the background input as well. Again, the reader interested in a more detailed description of this mathematical development is encouraged to refer to Appendix A.2.2.

LINEAR ENCODING

As mentioned earlier, an alternative form of stimulus input can be defined in order to avoid the non-linear multiplicative factor found in Equ. (7.9). This multiplicative factor has been shown to be needed to drive the network dynamics toward that of the stimulus input. In order to get rid of it, the idea is simply to replace it with a linear term. It has to be constant along the stimulus space Γ^A , and asymmetric along Γ^B . Detailed calculations given in Appendix A.2.2 show that the following input form approximates Equ. (7.9). It is given by

$$x(\vec{r}_A, \vec{r}_B, t) = h_1 \left[\mathcal{G}(\vec{r}_A, \vec{s}(t), \sigma_s) - \tau \dot{\vec{s}}(t) \cdot \nabla \mathcal{G}(\vec{r}_A, \vec{s}(t), \sigma_s) + \eta \vec{h}_1^d \cdot \vec{r}_B \right] \quad (7.12)$$

where η is a constant which depends on both the network recurrent weights W^R and on the profile \mathcal{G} of the input. This form of linear encoding is particularly interesting for transmitting information across populations. Indeed, as will be described next, the positional and velocity-related information conveyed by a neural field can also be transferred to another field by means of strictly linear synaptic projections appropriately chosen.

7.2.4 INFORMATION TRANSMISSION ACROSS NEURAL FIELDS

In large scale neural networks, such as those trying to model the neural pathways responsible for cortical functions, several instances of the neural field model presented here have to be interconnected. The synaptic projections W^{AB} from a neural field A to another one B are defined so that the input $x^B(\vec{r}_A, \vec{r}_B, t)$ of population B is given by

$$x_B(\vec{r}_A, \vec{r}_B, t) = \iint W^{AB}(\vec{r}'_A, \vec{r}'_B, \vec{r}_A, \vec{r}_B) f(u^A(\vec{r}'_A, \vec{r}'_B, t)) d\vec{r}'_A d\vec{r}'_B \quad (7.13)$$

The aim of this section is to provide the profile of these synaptic weights so that the result of their projections are equivalent to the stimulus input form given in Equ. 7.12. As a consequence, the information conveyed by the source population A would be completely transferred into B such that $\vec{p}^B(t) = \vec{p}^A(t)$ and $\dot{\vec{p}}^B(t) = \dot{\vec{p}}^A(t)$.

In classical neural field implementation, information is usually transmitted using a weight kernel having a symmetric and center-surround shape. However, as shown previously, the ability of a neural field for velocity integration is the result of an asymmetric coupling across the network sub-layers. Therefore, a similar technique to that described in the context of the network dynamics (Equ. (7.1)) is applied here. It can be shown (See Appendix A.2.2) that the following projection weights

$$W^{AB}(\vec{r}'_A, \vec{r}'_B, \vec{r}_A, \vec{r}_B) = \left[W^T(\vec{r}'_A, \vec{r}_A) - \lambda \nabla W^T(\vec{r}'_A, \vec{r}_A) \cdot \vec{r}'_B + \mu^{AB} \vec{r}_B \cdot \vec{r}'_B \right] \quad (7.14)$$

produce the desired effect. W^T could be any symmetric and center-surround convolution kernel strictly defined on the neural space Γ^A . The last term $\mu^{AB} \vec{r}_B \cdot \vec{r}'_B$, where μ^{AB} is a constant depending on the recurrent weights and on the profile of the stimulus input, is the analogue of the last term found in Equ. (7.12). Consequently, population A can drive the dynamics of population B according to its own dynamics.

7.3 EXPERIMENTS AND RESULTS

In the following section, several experiments performed using the previously described model are presented. Sections 7.3.1 and 7.3.2 show numerical simulations aiming at illustrating the relevance of the mathematical developments. The ability of the model to integrate velocity commands and its responses to stimulus inputs and synaptic projections across multiple instances of the networks are also described. Further, in Section 7.3.3, the model is shown to be capable of reproducing neurophysiological data related to the preferential tuning

of neurons in the visual cortex to stimulus velocity (Orban et al., 1986). Finally, two additional simulation results are reported. They address the mechanisms that the present approach suggests to be related to the human abilities for motor imagery and sensory discrimination, respectively. The first experiment describes how the excitation level of the network may unbind it from its sensory influences. The second one considers a neural field receiving two contradictory and ambiguous sensory inputs which has to select the input corresponding best to its own internal dynamics.

7.3.1 DYNAMIC VELOCITY INTEGRATION

In the hippocampal formation of the rat, the so-called head-direction cells have been shown to code for the current heading direction of the animal. In addition, these neurons display a rapid updating of their representation, which has been suggested to be a response to motor efference copies or vestibular inputs (Sharp, Blair, & Cho, 2001; Taube & Bassett, 2003). Furthermore, in other brain regions, a similar rapid and predictive update of the neural representations encoding for different sensory states have also been reported (Schwartz & Moran, 1999; Roitman et al., 2005).

Here, the proposed model is shown to be capable of updating its internal representation according to the external commands $\vec{v}^*(t)$ fed to the network by means of its background input $h(\vec{r}_B, t)$. The simulation parameters that were used in each of the following experiments are summarized in Appendix B.3. Figure 7.4 shows the measured velocity response $\dot{\vec{p}}(t)$ of the neural field as a function of $\|\vec{h}_0^d\|$. Each point on the graph was obtained in a different trial with a different value for $\|\vec{h}_0^d\|$. The approximation, given by Equ. (7.7), is also displayed on the same graph. In the one dimensional case, it can be noticed that for a sufficiently strong asymmetric input, the network can reach the maximum velocity given by λ/τ . However, in the two dimensional case, it can be seen that the system only tends toward that maximum asymptotically. Indeed, in the former case, the stronger sub-layer is capable of completely inhibiting its opposite and hence, it drives the whole network alone. In contrast, the continuous nature of the direction preference in the latter case forbids a single sub-layer to win against all the others. Nevertheless, for a relatively small asymmetric component \vec{h}_0^d , the theoretical approximation fits well the simulation results. In Figures 7.5 and 7.6, some trajectories followed by the command \vec{v}^* are compared to the velocity response $\dot{\vec{p}}(t)$ of the network.

7.3.2 TRANSFER OF INFORMATION ACROSS NEURAL FIELDS

This section shows how the network can integrate the information provided by a stimulus input as well as by projections from another network. In addi-

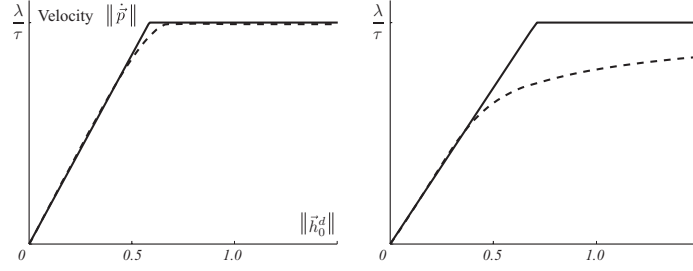


Figure 7.4: Network velocity response (dotted line) to different amplitudes of the asymmetric background input drive $\|\vec{h}_0^d\|$, as displayed by a ring (left) and torus attractor (right). The straight line corresponds to the approximation of the velocity response (Equ. (7.3)).

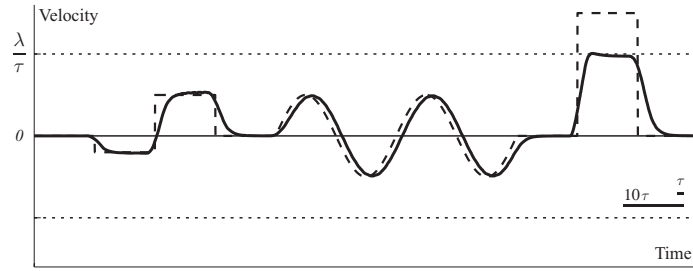


Figure 7.5: Velocity response $\dot{p}(t)$ of a ring network to external commands $\vec{v}^*(t)$ provided through background input (Equ. (7.8)). Dotted and straight lines correspond, respectively, to the desired velocity command $\vec{v}^*(t)$ and to the network velocity response $\dot{p}(t)$. As can be seen when a high velocity command is given, the network saturates at its maximum velocity integration boundary given by λ/τ .

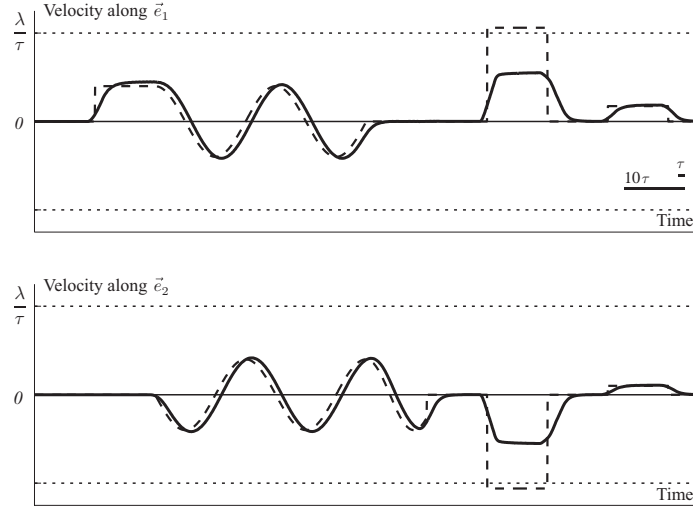


Figure 7.6: Velocity response $\dot{p}(t)$ of a torus network to external commands $\vec{v}^*(t)$. Data are shown relative to the two principal axes of the network representation space Γ_2^A given by the canonical base $\{\vec{e}_1, \vec{e}_2\} \in \Gamma_1^A \times \Gamma_1^A$. The same notation as in Figure 7.5 is used.

tion, its performance will also be compared with that of an usual single-layered continuous attractor neural network. For recall, its neural dynamics is governed by

$$\begin{aligned} \tau \dot{u}(\vec{r}, t) = & -u(\vec{r}, t) + x(\vec{r}, t) + h(t) \\ & \oint W(\vec{r}', \vec{r}) f(u(\vec{r}', t)) d\vec{r}' \end{aligned} \quad (7.15)$$

In comparison to Equ. (7.1), this equation corresponds to a network with a reduced dimensionality. In this case, the stimulus and the background inputs are, respectively, given by $x(\vec{r}, t) = h_1 \mathcal{G}(\vec{r}, \vec{s}(t))$ and $h(t) = h_0$.

Transfer from Stimulus Input to Network Representation

In order to illustrate the validity of the mathematical development addressing the integration of stimulus inputs, the non-linear and the linear form of the stimulus input (Eqs. (7.9) and (7.12), respectively) were separately applied to the network. The background input amplitude h_0 was set to zero. For different stimulus speeds, the spatial lag between the effective stimulus spatial location $\vec{s}(t_i)$ and the neural population vector $\vec{p}(t_i)$ at a given time t_i was measured. The results are plotted in Figure 7.7a and 7.7b. They show that, under both conditions, the model outperforms the classical model. Indeed, below the maximum integration velocity given by λ/τ , the lag stays close to zero, whereas an almost linear velocity-dependent lag is observed for the other model.

Transfer across Networks

Next, similar simulations were performed while the transfer of information across the representations of two interconnected neural fields were considered (See Eqs. (7.13) and (7.14)). Figure 7.7c shows the resulting lag measured between the two neural populations. As expected, the lag effectively stays close to zero as long as the velocity stays below the integration boundary λ/τ .

7.3.3 DYNAMIC VELOCITY TUNING

In the visual cortex, groups of neurons have been reported to be preferentially tuned to the position and velocity of visual stimuli. This sensitivity has further been suggested to take part in the mechanisms underlying human ability for velocity discrimination (Goodwin & Henry, 1975; Cheng, Hasegawa, Kadharbatcha, & Tanaka, 1994; Chey, Grossberg, & Mingolla, 1998; Mineiro & Zipser, 1998). In this section, the behavior of the model tested against this experimental paradigm is described. Prior to this experiment, the model was expected to replicate this finding. Indeed, it has already been shown that a single-layered neural field having a fixed asymmetric recurrent connectivity, can exhibit a preferential tuning to a specific speed of input stimuli (Mineiro &

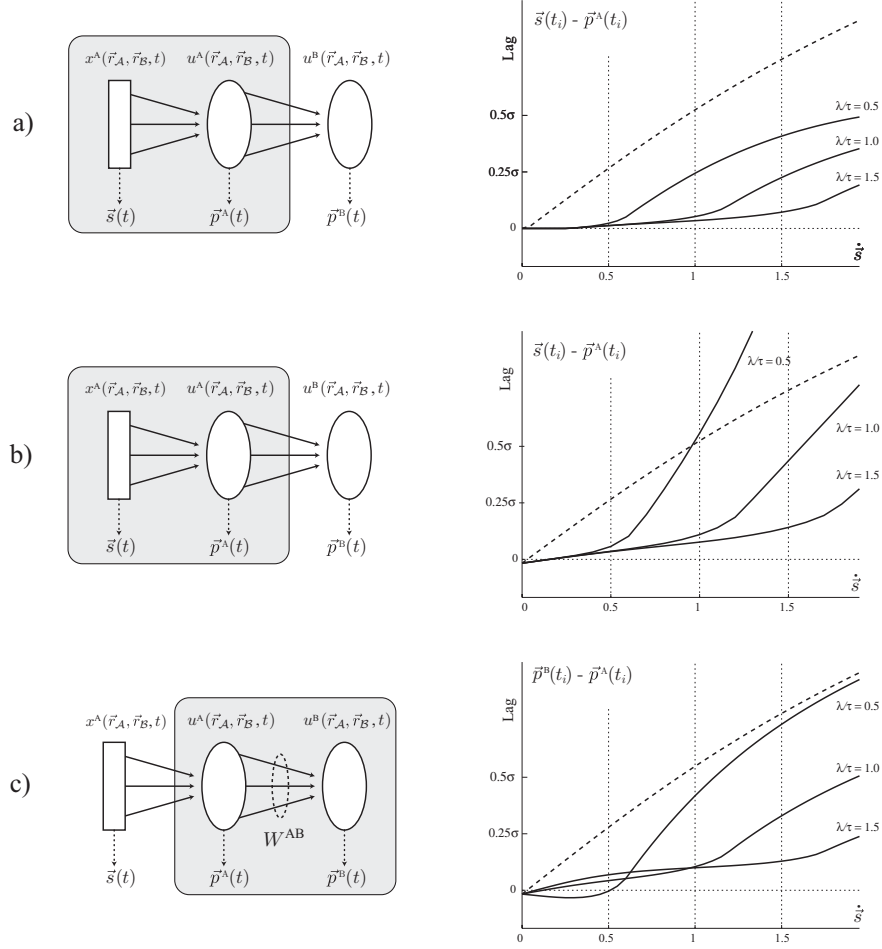


Figure 7.7: This figure illustrates the positional lag at a given time t_i between the network representation $\vec{p}(t_o)$ and **a)** a moving stimulus encoded using the non-linear method (Equ. 7.9), **b)** encoded with the linear method (Equ. 7.12) and **c)** another network which is driven by a moving stimulus (Equ. 7.13). In order to quantify the amount of latency between these representations, the lag is given relative to the breadth σ_s of the input shape \mathcal{G} (See Appendix B.3). On top, a system schematic shows which part of the network is considered in each case. Data is shown for different values of the ratio λ/τ which corresponds to the maximum allowable integration velocity. As can be seen, the lag is very low below this maximal value, but then grows almost linearly with the input speed. Moreover, this figure also compares the developed neural field model with classical continuous attractor network implementation (dotted line). As can be noticed, such a model suffers from a lag which is almost linearly related to the input speed.

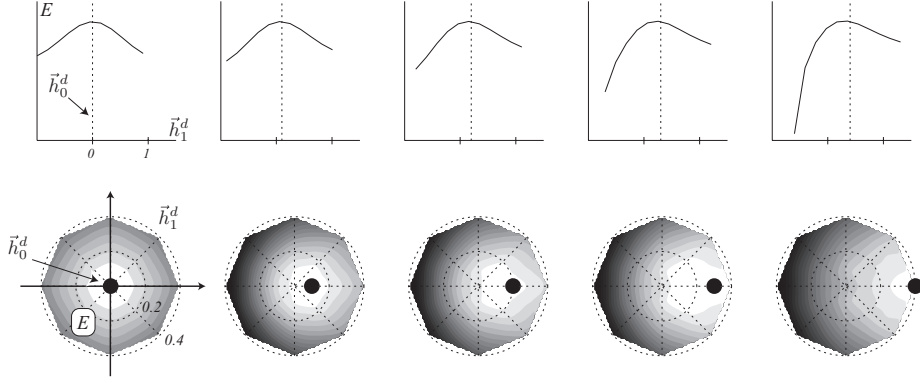


Figure 7.8: On top, the ring network mean energy relative to the speed of the input stimulus is shown. Each plot corresponds to a different intrinsic network speed. Note the shift in the preferential velocity tuning of the network with its intrinsic speed. At the bottom, the same results are shown for a torus attractor. The black dot indicates the network intrinsic speed. The brighter the plot is, the higher the energy.

Zipser, 1998). However an important advantage of the neural model described here, is that the intrinsic velocity to which the neurons are sensitive, can be changed dynamically. Indeed, by considering the changes in the network response to a stimulus moving at different velocities, the model will be shown to be capable of displaying a large range of preferential tuning by only varying the asymmetry of the background input $h(\vec{r}_B, t)$. Indeed, given an intrinsic network velocity \vec{v}^* , a stimulus with a close speed will resonate or cooperate more with the network than divergent ones. Since cooperative interactions in continuous attractor neural network produce higher activation patterns than competitive ones, the mean global firing rate of the neural field was chosen to be a good measure of this resonance effect. The network response energy $E(t)$, already defined in Equ. (3.14), is redefined here in order to account for the second variable. It is given by

$$E(t) = \iint f(u(\vec{r}_A, \vec{r}_B, t)) d\vec{r}_A d\vec{r}_B \quad (7.16)$$

Several simulation trials were performed while the model was given various intrinsic velocities \vec{v}^* . Then, a single stimulus which velocity was varied on a large range of velocities and directions was fed to the network. The field energy measured during each trial is reported in Figure 7.8. It can be seen that the network effectively responds preferentially to stimuli having a similar velocity, and that these tuning curves show a high similarity to those reported in the visual cortex (See Figure 3.15a) (Orban et al., 1986; Cheng et al., 1994).

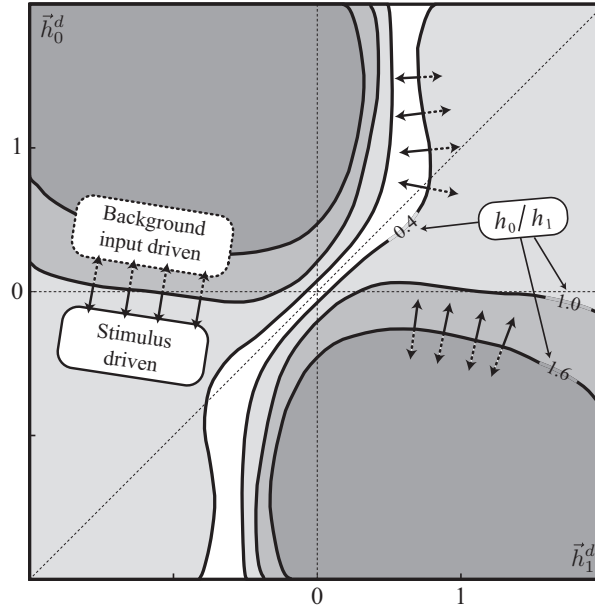


Figure 7.9: Input space in which the network is mostly driven by the stimulus or the background input is shown for different value of the ratio h_0/h_1 . The straight lines indicate the separations of the input velocity space between areas where the network is mostly driven by the stimulus input (straight arrows), and where the background input dominates (dashed arrows).

7.3.4 NEURAL LOCKING AND UNLOCKING TO EXTERNAL STIMULUS

The experiments described above were mostly performed while the mean background input h_0 was kept sufficiently small so that the influence of the recurrent connectivity was weak as compared to the strength of the stimulus input h_1 . Indeed, the ratio h_0/h_1 is of critical importance on the network behavior. It determines which of these two forms of inputs is driving the network. A small value of this ratio corresponds to a predominance of the stimulus input, while a larger one corresponds to a dominance of the network intrinsic dynamics. Interestingly, the ability of this system to exhibit these two distinct operative modes may be correlated to the brain processes of movement execution and motor imagery, which are known to activate similarly several motor cortical areas (Jeannerod & Decety, 1995; Porro et al., 1996; Rizzolatti et al., 2001; Fogassi & Gallese, 2002). On one hand, during motor execution, it is important for the neural representation of movements to keep track of the actual sensory state such as proprioceptive information. The stimulus input must thus predominate. On the other hand, during motor imagery, the brain has to mentally simulate movements which are not overtly produced, and consequently, it has to avoid relying on proprioceptive feedback. In this case, the network intrinsic dynamics has to dominate.

In the neural model proposed here, when $h_1 \gg h_0$, the network stays locked to the input stimulus. However, when $h_0 \gg h_1$, the intrinsic dynamics becomes sufficiently strong to free itself from the influence of the stimulus input. In this case, the network behaves strictly according to its background input. Figure 7.9 shows, for different values of the ratio h_0/h_1 , the space in which the network is mostly driven by either the stimulus or the background input. By mostly driven by the stimulus input, it is meant that the network response $\vec{p}(t)$ is strongly dependent on the spatial location $\vec{s}(t)$ of the stimulus. When the lag between these two values is almost constant through time, it indicates that the network is actually locked to that input. In contrast, an uncorrelated difference between $\vec{p}(t)$ and $\vec{s}(t)$ rather shows that the network is only driven by its intrinsic dynamics. The data shown in Figure 7.9 were obtained by first letting the network and the stimulus evolve for a given and sufficiently long period of time. Then, the lag between the stimulus location and that of the network response was measured. If, for a given trial, the lag was less than twice the breadth of the stimulus input σ_s , the network was considered as stimulus driven. Otherwise, it was considered as driven by its intrinsic dynamics.

7.3.5 SENSORY DISCRIMINATION

The last but not least simulation results concern the mechanisms of sensory discrimination. The principal source of inspiration comes from a behavioral experiment addressing the human ability to recognize one's own hand while observing two hands, identical in terms of their visual attributes (See Figure 2.3) (Van Den Bos & Jeannerod, 2002). One hand is actually that of the human subject, and the other belongs to that of the experimenter. It was shown that even when the two hands are moving in a similar fashion, the subjects can fairly determine which hand is their own (Van Den Bos & Jeannerod, 2002). The cognitive principle which has been suggested to allow this self-recognition ability, is based on a forward model which takes motor efference copies as inputs and then predicts the corresponding sensory consequences (Decety & Somerville, 2003). A comparison between these predictions with actual sensory feedback is then performed. A close match would further indicate that the observed stimulus is controlled by the self, whereas a discrepancy would mean that it is under an external influence. Neurophysiological data clearly suggest that this discrimination process is partly grounded within brain regions containing shared representations for both motor execution and visual recognition of movements (Decety & Somerville, 2003). This implies that some neural populations which receive visual signals irrespective of their owner, i.e., the self or another individual, should be capable of discriminating between self-generated motions and those performed by others, even if these visual signals are ambiguous in terms of their visual attributes.

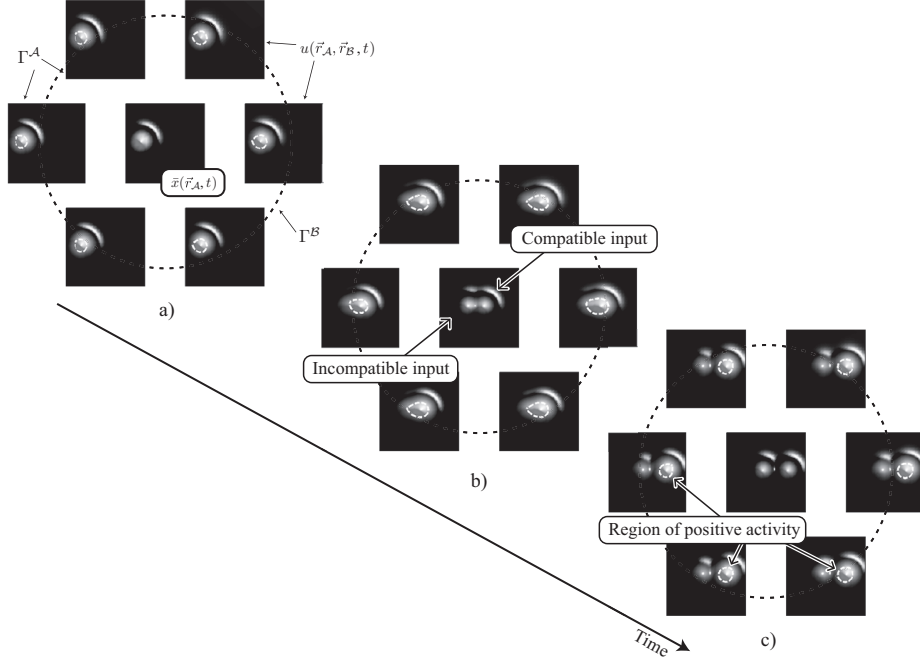


Figure 7.10: Network response during a velocity discrimination task. Each subplot corresponds to a snapshot of the membrane potential $u(\vec{r}_A, \vec{r}_B, t)$ of selected sub-layers sorted according to their preferred movement direction. The surfaces enclosed with the white dotted line indicate regions of neural space in which activity is above zero, i.e., the neurons within these areas are actually firing. In the middle of each sub-figure, the external input $x(\vec{r}_A, \vec{r}_B, t)$ averaged over the preferred directions of movement \vec{r}_B are shown. **a)** Beginning of the discrimination task: Both the compatible and the incompatible stimuli are at the same location in neural space. They are indistinguishable. **b)** The stimuli, moving at different speeds, start to separate but are not distinguishable on the network representation yet. **c)** The stimuli are clearly disjointed and by means of its recurrent interactions, the network naturally selects the stimulus which is the most compatible with its own intrinsic velocity.

This theoretical principle was applied to the model presented here. Two external stimuli located, respectively, at $\vec{s}^c(t)$ and $\vec{s}^i(t)$ are fed to the network. Since they are assumed to be ambiguous, their respective amplitude h_1^c and h_1^i as well as their initial location are considered identical. The indices c and i , designate the stimulus **compatible** with the network intrinsic dynamics, and the **incompatible** stimulus, respectively. The compatible stimulus must thus have a velocity which is equivalent to that of the neural field, i.e., $\vec{h}_1^{d^c} = \vec{h}_0^d$ so that $\dot{\vec{s}}^c(t) \approx \vec{v}^*(t)$. In contrast, that of the incompatible stimulus is different. Simulation results are given in Figures 7.10 and 7.11. Figure 7.10 describes the neural activity of the network at different time steps. It can be seen that when the inputs are separated in the neural space Γ^A , the network naturally selects the input having a velocity corresponding best to its own intrinsic dynamics. Additionally, Figure 7.11 shows the temporal dynamics of stimulus selection in different cases. When the input speed is below the maximum intrinsic network velocity (See Figures 7.11a and b), the selection is successful. However, when

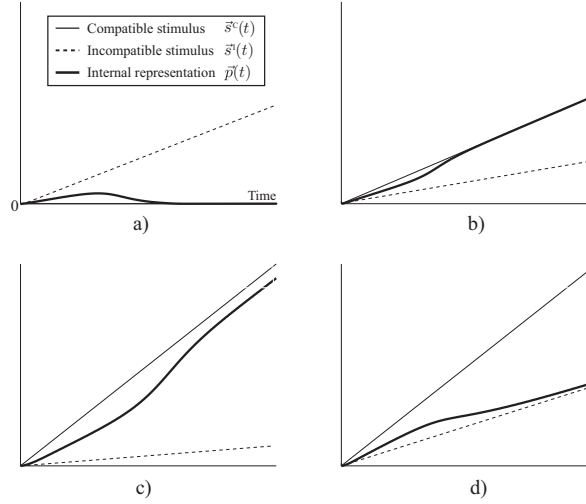


Figure 7.11: Illustration of the network response $\vec{p}(t)$ when a stimulus compatible with the intrinsic network velocity and an incompatible one are applied to the inputs of the network. The positions of the compatible stimulus (light filled line), the incompatible stimulus (dotted line) and the network response (dark filled line) are shown over time in several situations. **a-c)** As soon as the stimuli are sufficiently separated in neural space, the network successfully selects the compatible stimulus. It can be noticed that the trajectory of the internal representation $\vec{p}(t)$ exhibits, initially, a deviation from the compatible stimulus trajectory. As described in Figure 7.10, this is the result of the temporary overlap between the two stimuli representations on neural space; the network thus cannot discriminate between the two. However, as soon as the stimuli are sufficiently separated, the selection is performed. **c)** Since the speed of the compatible stimulus is above the maximal network velocity, a constant lag can be observed. **d)** In a similar out-of-boundaries situation, the network selects the wrong input. Indeed, the input speed is here closer to that of the network.

above, the network may either lag behind the compatible stimulus or even select the incompatible stimulus (See Figures 7.11c and d). Finally, Figure 7.12 summarizes the range of speeds where the network successfully discriminates the correct stimulus input. In the case of a right decision, the spatial lag between the population vector $\vec{p}(t)$ and the compatible stimulus location \vec{s}^c is indicated by a gray level. It can be noticed that the lag is almost zero when the compatible stimulus velocity lies within the allowable range of velocity integration. Indeed, as described in Section 7.3.2, outside that range the lag increases with the stimulus speed. Furthermore, note that the regions of false discrimination are strictly located over the system integration boundary. Since in these cases, the network can not update its internal representation accurately, it naturally selects the input which is the closest to its actual intrinsic velocity. When the velocity of the incompatible input is closer to the maximum integration velocity of the network than that of the compatible input, the incompatible input becomes more compatible with the network than the compatible one, and hence it gets naturally selected.

These results confirm that this model can account for the importance of the precise timing between the prediction of the movement outcome and its

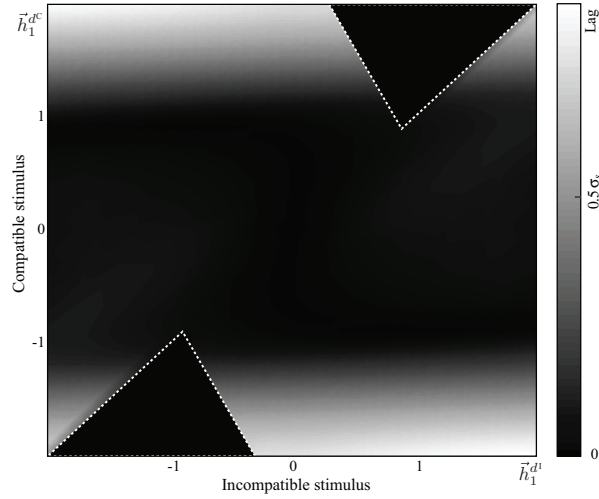


Figure 7.12: This figure shows the range of compatible-incompatible stimulus velocities leading the network to a false decision (triangular black regions). Moreover, in the case of a correct decision, the figure also shows the lag measured between the location $\vec{s}^c(t)$ of the compatible stimulus and the network response $\vec{p}(t)$. The lag is given relative to the breadth σ_s of the stimulus input. As can be seen, when the compatible stimulus speed is over the network integration limit, an increasing lag between the input and the network response is observed.

associated sensory feedback (Haggard & Clarke, 2003). Indeed, this constraint in the timing allows the model to select which of the stimuli is under self-control. Consequently, this model may describe a plausible neural mechanism which can contribute to this cognitive function. It also suggests how, within shared neural representations, neural discrimination may be performed.

7.4 DISCUSSION

This chapter presented a continuous attractor neural network model capable of integrating velocity commands in order to update its internal representation. In addition, adequate profiles of external inputs were defined such that these inputs can drive the network toward their own dynamics. Further, the analysis of the network dynamics revealed various interesting properties which may be shared by several cortical brain regions. By considering current neurophysiological data, the structure and mechanisms of this model may be found in various brain areas, such as the cerebellum and the motor, visual and associative cortices.

Related models addressing the problem of velocity integration have already been described in the literature. The large majority of them considers, as biological background, the rat head direction system, its hippocampal places fields and its abilities for path integration (Redish et al., 1996; Zhang, 1996; Xie et al., 2002; Stringer et al., 2004). Although the biological plausibility of their

implementation is still under debate, a series of models assumes the existence of sigma-pi units, i.e., neurons performing both a sum and a product of their inputs (Redish et al., 1996; Zhang, 1996; Stringer et al., 2004). In contrast, the neural architecture described here avoids the use of such computational units. It is nevertheless largely inspired from the computational principles described by Zhang (1996) and the neural structure proposed by Xie et al. (2002). Despite this similarity, this work first considers explicitly the external inputs to the system and their influences on the network. Moreover, the analysis of the behavior of the model may raise new hypotheses concerning the implications of the model in the neural processes occurring in other brain regions. They are described next.

Neural Sensitivity to Moving Stimuli

Velocity and direction tuning are properties of neurons that were primarily found in the visual cortex. These cells have been shown to fire preferentially during the presentation of visual stimuli moving in a specific direction with a specific velocity and inside their receptive field (Goodwin & Henry, 1975; Orban et al., 1986; Cheng et al., 1994). This neurophysiological property has then been suggested to be the basis of human ability for velocity discrimination (De Bruyn & Orban, 1988). Earlier work by Mineiro and Zipser (1998) has already demonstrated how a neural field endowed with a fixed asymmetric recurrent connectivity can exhibit such a sensitivity. However, their network is constrained to exhibit a preferential tuning to a single velocity. Consequently, in order to allow a neural structure to be sensitive to a broader range of velocities, this model has to be replicated a large number of times with different weight strengths. This approach has been followed by another modeling study addressing the same discrimination problem (Chey et al., 1998). In contrast, the present model proposes a mechanism relying on sub-layers having opposite directional tunings, where the sensitivity to a precise velocity can be tuned dynamically by the background neural activity. However, visual areas such as V1 and V2 possess a wide range of neurons tuned to different velocities (Goodwin & Henry, 1975; Orban et al., 1986), which may argue against the present model. Nevertheless, area MT, which is located further in the visual processing pathways, is endowed with velocity tuned cells which distribution is centered near some extreme values (Cheng et al., 1994). Since MT is directly connected to the parietal cortex (Wise et al., 1997), which is known to be the locus of some decisional processing (Shalden & Newsome, 2001), this may confirm the modeling approach described here. The brain may be capable to adapt the dynamics of its internal representations toward that of visual stimuli for discrimination purposes. Finally, computationally speaking, in order to have large range of velocity sensitivities in the two-dimensional retinal space, the complexity of this model is of order $O(n^3)$, where n is the number of neurons along a single dimension, whereas the previously mentioned approach is of order $O(n^4)$ (Chey

et al., 1998; Mineiro & Zipser, 1998).

Sensorimotor Transformations and Motor Control

The transmission time of information across neural structures is critical for time-dependent tasks such as the control of movements. However, the brain is known to suffer from delays arising from the substrate in which sensorimotor information is encoded, i.e., the neurons. In addition to the integration time constant of a single neuron, the high density of the recurrent connectivity among cortical columns and across brain regions, while providing the brain with high computational power, adds an even stronger inertia to the information flow and consequently increases the overall system time constant. Both neurophysiological and modeling studies have already described the consequences in response latency of intra-layer recurrent and center-surround connectivity during simple stimulus-response tasks (Raiguel, Xiao, Marcar, & Orban, 1999; Panzeri et al., 2001). Moreover, according to hypotheses raised by the neural field approach, this recurrent connectivity has another side effect which was not addressed by these studies. It concerns the difficulty of such neural representations to move from a given attractor state to another.

This problem has several implications regarding recently proposed neural substrates that are suggested to be at the basis of sensorimotor transformations (Burnod et al., 1999; Deneve et al., 1999; Salinas & Thier, 2000; Scherberger & Andresen, 2003) (See also Chapter 5 of this thesis). Indeed, these approaches rely on populations of neurons encoding basis functions and grouped within gain fields, which are neural populations showing a dense recurrent connectivity and which receive changing inputs from several external sources. For example, to compute the location of a visual target in body-centered coordinates, the brain is supposed to merge information related to the target location in retinal space, to the eyes displacement relative to the head and to the heading direction in body-centered frame of reference. Then, in spite of the inertia resulting from the recurrent connectivity, the desired information has to be read out from this combined representation with a latency as short as possible. Nevertheless, the brain successfully executes that operation. For instance, in the visual cortices of the monkey, neurons have been found to fire even before a saccade brings a stimulus into their receptive field (Unema & Goldberg, 1997; Nakamura & Colby, 2002). This neural activity may be the result of a predictive update of the visual representation, that may explain, at least partially, that eye tracking of predictable targets such as one's own hand can be performed with almost no latency (Vercher et al., 1996; Miall & Reckess, 2002). Neurophysiological findings also indicate that large groups of neurons in the cerebellum display a preferential tuning to arm position and movement direction. These neurons also fire with almost no latency when compared to real arm dynamics (Fu et al., 1997; Roitman et al., 2005). Together, these findings suggest that the brain may use forward models in order to predict the consequences of upcoming move-

ments. This work proposed a neural mechanism for updating internal sensory representations by means of velocity integration. Consequently, this study suggests that the integration time of gain fields can be reduced drastically and thus, that sensorimotor transformations can be performed with almost no delay. Therefore, in addition to the suggested role of neurons of the cerebellum and of motor and parietal cortices in the control of movements (Kalaska et al., 1990; Georgopoulos, 1996; Schweigenhofer, Arbib, & Kawato, 1998; Todorov, 2000), their sensitivity to non-linear mixtures of information, such as arm position and velocity, may also reveal that these areas take advantage of an internal integration process. By means of a rapid and predictive-like updating of the information they convey, they could avoid waiting on slow sensory feedbacks.

Shared Representations: Motor Imagery and Imitation

The model developed here has also some implications concerning the neural mechanisms related to motor imagery and imitation. First, motor imagery is the ability to mentally imagine oneself or someone else executing a movement. Recent findings suggest that this mental operation is performed in motor terms, i.e., by activating parts of the motor cortices that would be effectively involved in overt movement execution (Jeannerod & Decety, 1995; Porro et al., 1996; Fogassi & Gallese, 2002). Computationally, this hypothesis implies that some motor areas which receive projections from proprioceptive feedback during normal motor execution, should also be able to process simulated motor commands without being influenced by sensory information. The analysis of the model provides some insight to a potential neural mechanism which may resolve this problem. Indeed, this study suggests that the same neural substrate, by increasing its global excitation level, can detach itself from efferent sensory inputs, and consequently perform an imagery task freely. Then, by lowering this level of excitation, the external inputs can lock the network back to the actual sensory state for the control and the overt execution of movements. In addition, the present model may also, to some extent, contribute to the explanation of the neural processes underlying the illusory perception of one's own body resulting from epileptic seizures, such as the autoscopic phenomena (Blanke et al., 2002; Blanke & Mohr, 2005). Indeed, since epileptic seizures are defined as an abnormal synchronization of the electrical activity of large ensemble of neurons, this would result, in the model, in an abnormally high global excitation of the network. Thus, by producing an undesired unlocking of the network representation from the real sensory state, the representation of self gets disturbed. This could then have for consequence a disintegration of self processing in the brain areas related to self and others (Blanke & Mohr, 2005). Finally, by back-propagating this conflicting information to visual areas, the reported hallucinations may be observed.

Furthermore, shared representations have also been reported by experimental studies which consider the observation of movements as well as their exe-

cution. Indeed, recent neurophysiological studies indicate that the observation and the execution of actions activate both a shared complex of brain areas known as the mirror neuron system (Iacoboni et al., 1999; Rizzolatti et al., 2001). However, this calls for explanations as to what happens when action execution and action observation occur simultaneously, and subsequently as to how the brain minimizes conflicts which may arise from the integration of visual inputs related to the self and to the others. In contrast to motor imagery, sensory feedback is here needed by motor areas for normal performance, but the sensory signals corresponding to others movements should be discarded, or at least not considered. The neural model proposed here provides a computational hypothesis to resolve this issue. As shown in Section 7.3.5, when the network receives a sensory feedback whose dynamics is compatible with its own intrinsic dynamics, it can track and keep locking to that re-afferent input, and this even if a distractor, such as an input signal resulting from movements performed by others, is present. In addition, this model also predicts that errors may increase with the decreasing ability of the network to update its internal representation accurately with speed. At very high movement velocities, the pattern of errors should reflect a bias toward a boundary corresponding to the internal limit for movement integration.

Finally, this modeling study also gives some insights on a plausible neural medium for learning by imitation through motor resonance (Gallese & Goldman, 1998; Wolpert et al., 2003; Oztop et al., 2006). This principle has been proposed to explain the mirror response of the neurons in premotor areas, by suggesting that the perception of others' movements and actions activates in parallel the internal representations of motor plans. Through competition, the plan corresponding best is selected and may further be used for prediction of movement outcomes, action understanding and imitation. Since the presented model can potentially represent any state space, it may implement a motor plan by setting its intrinsic dynamics to code for the movements corresponding to the plan. Then, the state of a demonstrator's movement may be fed into this network. By monitoring the global energy of the neural ensemble which has been shown to be maximal in the case where the internal dynamics match perfectly that of the external input, it may thus become possible for the brain to select the best among activated motor plans. Moreover, as suggested by earlier modeling works (Demiris & Hayes, 2002; Wolpert et al., 2003; Oztop et al., 2006), imitation processes may use such a comparison value to perform a gradient ascent on this energy function. Within this framework, movement imitation may thus consist in maximizing the energy of the shared representation between self and others' movements.

Summary

In this chapter, an extension of the classical neural field modeling approach was presented. Its primary novel property (from which the others are derived) is to allow the network to modify its intrinsic dynamics by acting on its background synaptic activity in a specific manner. As a consequence, the typical marginal and continuous attractor state usually exhibited by neural fields can become a linear trajectory or a limit cycle in the space of the neural representation. For instance, by considering that a forward model is actually driving the network according to specific motor commands, this property allows the neural field to update its internal representation in a predictive fashion without the need to wait for slow sensory feedback. The dynamic properties of the model were analyzed with a special focus on the behavior of the network when one or more external stimuli are influencing the dynamics of the network. Finally, the biological plausibility of the model and its implications concerning several brain computational mechanisms were discussed in light of neurophysiological and behavioral data. The addressed topics include velocity tuning to visual stimuli, sensory discrimination, sensorimotor transformations, motor control, motor imagery and imitation.

Like the models presented in Chapter 5, the model described here proposes a plausible mechanism at the neural level which may contribute to the overall processes underlying imitation. This work does not provide clear answers to the question of which cortical pathways and cortical structures mediate the important cognitive processes related to the human ability to discriminate between the movements produced by the self or by other individuals. In order to investigate this issue, the next chapter proposes a neural model that is in part composed of the model described above. The fundamental idea behind it is to first explore how the model accounts for the results of behavioral experiments that focus on the well-known human ability for multisensory discrimination when visual and proprioceptive information are biased. This work will then attempt to shed some light on the neural mechanisms responsible for the cognitive processes involved in the differentiation between the self and the others. In order to test the hypotheses raised throughout this study, predictions of the model with respect to an experimental interference paradigm are presented; these predictions should be tested on real behavioral data.

MOVEMENT GENERATION, SENSORY INTEGRATION, SENSORY DISCRIMINATION AND INTERFERENCES

CHAPTER 7 developed a neural field model that was designed to provide a neural network with the ability to discriminate sensory inputs which are the consequence of self-generated movements. By integrating external commands, the network was capable of modifying its intrinsic dynamics. Consequently, it exhibited a preferential tuning to sensory inputs which are coherent with its own dynamics. Although the analysis of the properties of the model is already an important step toward a better understanding of the neural mechanisms responsible for the processes of multisensory integration and discrimination, it is also important to provide a framework through which behavioral data may be compared with model predictions. In this chapter several such experimental paradigms are proposed. The topics are primarily related to multisensory integration in situations where biases are imposed on the sensory feedbacks. The main idea is to analyze how the system behaves in such conditions, and subsequently to see whether it reproduces current experimental data and leads to new experimental predictions. Two forms of bias are considered. The first is artificially biased sensory feedback related to one's own movements. The second is interference resulting from observing others' - a sensible possibility, since it is believed that there are neural representations which are shared between self- and others-generated actions.

8.1 INTRODUCTION

In order to better control the execution of movements, the brain may be endowed with both an inverse and a forward model of the body (Miall & Wolpert, 1996; Decety & Sommerville, 2003; Wolpert et al., 2003). While the inverse model computes the motor commands according to a desired target state with respect to the current state of the body, the forward model takes these motor commands, better known as the motor efference copies, and predict the consequences of these commands, i.e., how the current state of the body will change. By monitoring these predictions, the brain may anticipate possible perturbations without the need to wait for slow sensory feedbacks. From an engineering

point of view, such a coupled system is known to allow a more accurate control of movements as compared to a single loop control system. In addition, as already mentioned along this thesis, and more particularly in Chapters 2 and 7, by predicting the outcome of its own motor commands, the brain may compare it with its actual sensory feedback. If both were to be different, it would indicate either that the forward model has failed to predict correctly the movement outcome and has hence to be refined or the commands changed, or that the perceived sensory feedback does not in fact, belong to, or was produced by the body it controls.

This capacity to determine the ownership of sensory feedbacks which may come from different sensory modalities such as vision or audition, has been suggested to be fundamental for the sense of agency to emerge (Decety & Somerville, 2003; Haggard & Clarke, 2003; Jeannerod, 2003). Indeed, the sense of agency, as well as the sense of self appear to be the result from the integration of coherent multisensory signals, where incoherent ones, if present, have to be set apart from that internal representation. Therefore, a neural process capable of determining which signals belong to the self, i.e., those which are the result of self produced actions has to exist. Current body of evidence tends to localize this cognitive skill in a network of cortical regions including the anterior-medial frontal cortex (AMFC) and the tempo-parietal junction (TPJ) (Decety & Somerville, 2003; Decety & Chaminade, 2003). For instance, while lesions or disturbance of the TPJ have been shown to produce hallucinatory perceptions of the self, patients with lesions of the AMFC were often reported as attainted of schizophrenia, which is also a result of an altered representation of the self (Blanke & Mohr, 2005). Furthermore, in well-being people, this discrimination process has been shown to be not perfectly accurate. For example, a delay up to one or two hundreds milliseconds between a movement believed to produce an immediate auditory signal and its actual perception does not affect the feeling of human subjects that the sound was clearly produced by them (Haggard, Clark, & Kalogeras, 2002; Haggard & Clarke, 2003). In addition, when asked to indicate the precise time at which the sound was heard, the subjects tend also to underestimate the delay, as if the execution of the action biases the perception time of its associated response toward that of the action. This behavioral effect was denoted as the temporal binding of action and perception. Several explanations to this phenomenon were provided in terms of behavioral advantages it could provide. It may first compensate for the inherent noise and time delays present in the sensory systems, and second, it may also facilitate the learning of the associations between actions and effects (Haggard & Clarke, 2003; Engbert & Wohlschläger, 2007).

Another set of interesting experiments showing that the process in charge of determining whether a visual signal is controlled by the self or if it is under an external influence has been proposed by Fournieret and Jeannerod (1998) and in a follow-up study by Slachevsky et al. (2001). In an experimental setup illus-

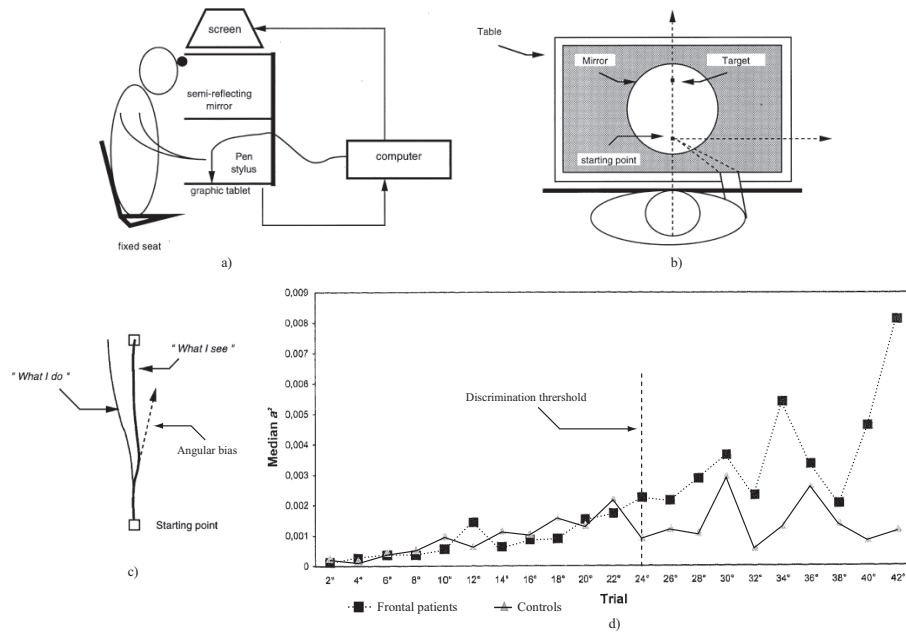


Figure 8.1: **a)** Experimental setup used by Fournieret and Jeannerod (1998), and Slachevsky et al. (2001) in order to test human behavioral responses to biased visual sensory feedback. Subjects hold a pen stylus connected to a computer which may produce an artificial bias in the visual feedback projected on the table in front of the subjects. **b)** On this projection screen, the starting point, as well as the target are also displayed. **c)** When an angular bias (dotted line) is induced in the visual feedback, a typical trajectory followed by the cursor (dark line) does not correspond to that actually performed by the subject's hand (light filled line). **d)** Median values of the errors relative to the bias as produced by normal subjects and patients with lesions of the frontal lobes. The error was measured as the area formed by the ideal straight trajectory and by that followed by the visual feedback. While the error constantly increases with the amount of bias in frontal patients, it stops growing and stays almost constant in normal patients above a certain threshold value, namely, the discrimination threshold. This suggests a conscious monitoring of the integration of conflicting sensory signals. (Adapted from Fournieret and Jeannerod (1998); Slachevsky et al. (2001)).

trated in Figure 8.1, human subjects were asked to reach for a target by tracing a vertical line shown on a computer screen. During each experimental trial, the experimenter introduced bias in the visual feedback associated with the hand location. This bias corresponded to a constant angle of deviation with respect to the starting point and the ideal vertical line. It was then reported that below a certain angle of deviation, the subjects were in general unaware of the bias, while still being capable to reach for the target. In the experiment of Fournieret and Jeannerod (1998), subjects were, after each trial, requested to estimate the direction in which their hand deviated in proprioceptive terms. Two different types of answers were given. One group of subjects indicated that their hand deviated in the approximative direction of the visual signal, i.e., in the opposite direction of the actual movement. In contrast, the answers of the other group reflected the correct direction of deviation. However, importantly, all answers clearly underestimated the actual amount of deviation. Interestingly, in a later

experiment, Slachevsky et al. (2001) used the same paradigm but increased the range of angles of deviation. Similarly to the former experiment, the authors of this study showed that for small deviations, subjects tend to be unaware of the bias. Nevertheless, over a certain deviation threshold, they become aware of that bias, and consequently modify their movement strategy. In addition, by comparing the trajectories performed by the participants which consisted of a control group and a group of patients with frontal lesions, they also pointed-out the involvement of the frontal cortex in the process of discriminating whether a sensory signal believed to be self-controlled behave as internally predicted. Interestingly, the integration of these different sensory inputs, in this case vision and proprioception, which have to be similar on a spatiotemporal metric, reveals that the multisensory binding principle described above is not only temporal but also spatial.

Further, in relation with imitation, the principle of shared representations, where sensorimotor representations related to movements executed by the self and by others overlap has to be considered. Indeed, a monitoring of the ownership of each sensory signal is required, so that the brain can know to whom the information encoded in these shared representations corresponds. A failure of this monitoring process, usually involving frontal and temporal cortices, has for instance been suggested to result in compulsive imitative behaviors such as echopraxia¹ (Lhermite et al., 1986; Shimomura & Mori, 1998). As already mentioned in Section 2.2.1, although neurologically intact people can easily, during everyday experience, discriminate between self- and others-generated movements, interferences in the execution of movements are still reported during the simultaneous observation of actions performed by other individuals (Brass et al., 2000; Kilner et al., 2003; Bertenthal et al., 2006). It is well documented that observing the movements of others may influence one's own performance, in either a differed moment (Paccalin & Jeannerod, 2000) or simultaneously (Kilner et al., 2003). Observed deviations from the originally planned movements may be very subtle and almost imperceptible, but they may also influence the general contextual behavior, or even generate movements which were not initially planned. Indeed, the imitative interference effects were not only observed during the production of movements, but also during the execution of actions involving objects as goals. For instance, if two potential target objects are present at a reach distance, and that one of them has to be selected, the observation of someone grasping one of the two objects can induce a positive bias toward the reaching of that object (Wohlschläger & Bekkering, 2002).

The work presented here investigates how a neural model can exhibit some of the previously mentioned human behavioral characteristics, in particular those related to the multisensory integration of sensory signals associated with the

¹Echopraxia is defined as the involuntary repetition of observed movements performed by other individuals. This syndrome is mostly believed to be the result of neurological disorders.

execution of movements. It will be shown that the proposed model can reproduce the spatiotemporal binding effect, where sensory inputs can be integrated together up to some discrepancy level. Above that threshold, the system keeps only track of the input corresponding best to its internally predicted state. Further, another experiment attempts to describe the possible spatiotemporal profiles of interference patterns when movement execution and movement observation occur simultaneously. Since several hypotheses can be considered as to how the movements of others may influence one's own internal representation, several conditions are considered and the associated movement responses will be described in detail. Consequently, by reproducing the same experiment with human subjects and recording the exact undesired movement response to the observation of movements performed by an experimenter, it would be possible to determine the factors which influence automatic imitative behaviors, and subsequently the neural pathways which mediate them.

8.2 NEURAL MODEL

Similarly to the work presented in Chapter 6, the present computational study considers a neural model based on the neural field approach for controlling kinematically the movement of a single body joint such as the shoulder. Since the tasks addressed here are related to the maintenance of a coherent internal state to dynamically changing sensory inputs, the core building block of this model consists of the continuous attractor neural network described at length in Chapter 7. Indeed, this network has been shown to be capable of integrating velocity commands and also to be able to discriminate between two external inputs, equal in strength, by comparing their dynamical properties.

In this work, a simple neural network based on an inverse and forward model for controlling goal-directed movements along two degrees of freedom is considered. First, the model described in Chapter 7 is slightly modified in order to account for a spherical parameter space. Further, a neural network which can generate goal-directed velocity commands is proposed, followed by a description of the coupling between these two neural fields. Finally, the complete architecture of the model that will be used in the experimental scenarios will be given.

8.2.1 GENERATION OF MOVEMENTS ON A SPHERICAL SURFACE: FORWARD MODEL

In order to represent a variable on a sphere, the definition domains of the neural preferential tuning, as described in Section 7.2, have to be redefined. Here, the primary variable has to be defined in $\vec{r}_A \in \Gamma^A = \{\vec{r}_A \in \mathbb{R}^3 \mid \|\vec{r}_A\| = 1\}$. Next, since a displacement on a sphere is most easily described by a angu-

lar motion around an axis, i.e., a rotation, the possible directions of motion represented by the secondary variable are also defined on a sphere such that $\vec{r}_B \in \Gamma^B = \Gamma^A = \Gamma$. Therefore, in contrast to the original formulation given in Chapter 7, the secondary variable does not anymore represent the direction of movement but rather the rotation axis around which the motion occurs. Considering now the equation governing the dynamics of the system, the definition of the recurrent weight profile has also to be adapted to the previously described change of how movements are encoded. In order to drive an activity bump around a preferential axis \vec{r}_B , the following recurrent weights

$$[W^R(\vec{r}'_A, \vec{r}_A) - \lambda \nabla W^R(\vec{r}'_A, \vec{r}_A) \times \vec{r}'_B] \quad (8.1)$$

have to be used in place of those given in Equ. (7.1). The fundamental weight profile $W^R(\vec{r}'_A, \vec{r}_A)$ is nevertheless still generated by Equ. (3.8), since this equation is general to arbitrary dimensions. As a consequence, imposing a background input $h(\vec{r}_B, t)$ to this network where

$$h(\vec{r}_B, t) = h_0 \left[1 + \frac{\tau}{\lambda \gamma} \vec{\omega}^*(t) \cdot \vec{r}_B \right] \quad (8.2)$$

will drive the intrinsic dynamics of the neural field to develop a tendency to move any activity bump with an angular velocity $\|\vec{\omega}^*(t)\|$ around the axis given by $\vec{\omega}^*(t)/\|\vec{\omega}^*(t)\|$. Note that the system variables γ , λ , τ and h_0 were already introduced in Section 7.2.2.

In Figure 8.2a, a visual representation of the resulting network is shown. Each sphere corresponds to the ensemble of neurons sharing the same preferred angular movement direction indicated by the rotating arrow. In Figures 8.2b, c and d, the intrinsic dynamics of the network is shown when an input given by Equ. (8.2) is fed to the field for different values of $\vec{\omega}^*(t)$. These illustrations are given in the form of vector fields, where each vector corresponds to the instantaneous displacement induced on an activity bump located at the origin of the vector. As can be seen, the recurrent weight profile described in Equ. (8.1) creates vector fields which rotate the activity packet currently encoded in the network.

8.2.2 GENERATION OF VELOCITY COMMANDS: INVERSE MODEL

As described above and in Chapter 7, the proposed neural structure allows a traveling activity pattern to be controlled by an external velocity command. By acting on the background synaptic input of each sub-layer whose neurons share the same preferred movement direction, one can control the direction as well as the velocity of the internal dynamics of the neural field. In order to generate goal-directed motor commands, this section presents a network which,

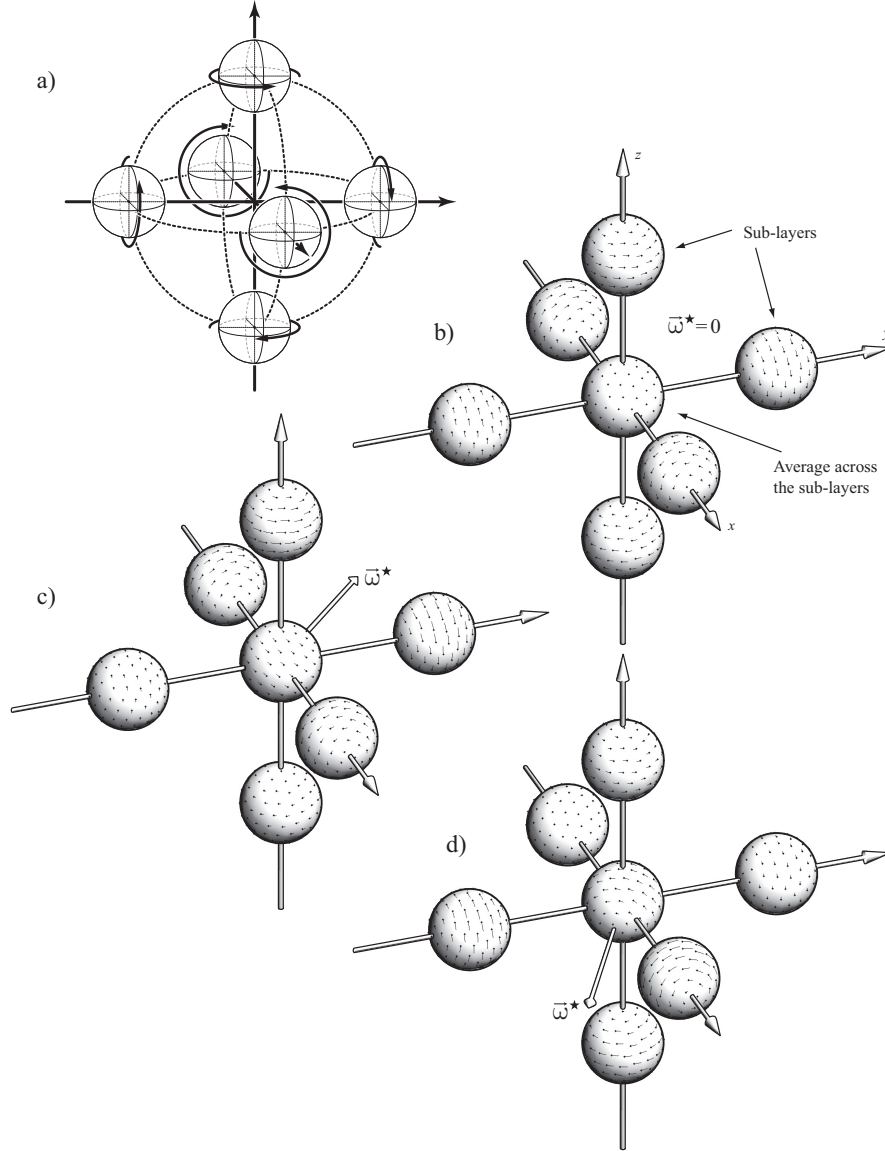


Figure 8.2: Graphical illustration of the structure of a neural field defined on a sphere which can update its internal representation by means of an angular velocity command. **a)** The network structure is composed of sub-layers, denoted by small spheres, where each neuron share the same movement preference. In contrast, each sub-layer is preferentially tuned to a different angular motion indicated by the circular dark arrows. Since all sub-layers are interconnected, the balance among their activation may drive the internal representation of the network in a given direction or not. **b)** On each sphere located in the periphery, a vector field corresponding to the motion that would be induced by its associated sub-layer on an activity bump is shown. Since the background activation is perfectly symmetric in this case, no change in the internal representation is produced, as shown in the sphere located at the center, which consists of the averaging of the vector fields produced by all sub-layers. **c-d)** Two examples of movement induced by an angular velocity input $\vec{\omega}^*$ (Equ. (8.2)) are shown. While the sub-layers having a movement direction preference close to the input vector $\vec{\omega}^*$ are more activated, the background activity of the others is decreased. This generates a rotational vector field around the axis given by $\vec{\omega}^*$.

when given a goal to reach in the primary parameter space, can produce velocity commands to reach it. The general principle behind the proposed mechanism is illustrated in Figure 8.3. A neural field capable of generating a moving activity bump, denoted as the **state** network, is reciprocally connected to another network, namely, the **command** network, which is not necessarily endowed with the same integration ability. In order to keep the modeling as simple as possible, this network is assumed to share the same structure as the state network but without any recurrent connectivity.

By definition, the state network encodes the current state of the system, for example, the current arm position. It projects this information to the command network so that the velocity commands can be dependent on the current state. Next, as shown in Figure 8.3b, the generation of commands is based on the principle that this network receives also another external input which preshapes its subthreshold activation. How this preshaping could be produced by a third part network will be described later in the text. This subthreshold preshaping has to be asymmetric across the sub-layers of the neural field having an opposite direction preference. It will further produce a given command according to the current state of the system. Indeed, as depicted in Figure 8.3c, for a given state, if the subthreshold preshaping of the sub-layer, whose preferential direction corresponds to the desired direction of movement, is set to a higher level than that of its opposite sub-layer, the global firing rate of the former sub-layer becomes stronger than that of the other. Consequently, by projecting the global excitation of each sub-layer on each associated sub-layer of the state network (Figure 8.3d), a movement is initiated. Then, this motion stops as soon as a perfect balance in the excitation of each sub-layer of the command network is reached, i.e., when and where the subthreshold preshaping of each sub-layer is equal. This location in the neural space can thus be seen as the goal or the attractor generated by the command network. As example of evolution over time of the system activity shown in Figure 8.3e. In the next section, the technique as to how a goal-dependent subthreshold preshaping can be produced is described.

GOAL-DIRECTED MOVEMENT GENERATION

Up to now, two networks reciprocally connected were described. The state network represents the state of the system and integrates velocity commands provided by the command network. Then, in order to generate of movements toward a goal, the goal is encoded in a third neural representation, namely, the **goal** network. It consists of a standard neural field defined for simplicity, without the extension proposed in Chapter 7. It receives an external input $\vec{s}^G(t) \in \Gamma$, defined in Equ. (3.10), a variable which indicates the location of the target. Then, in order to produce a goal-dependent subthreshold preshaping of the command network, the goal network projects its activity by means of synaptic

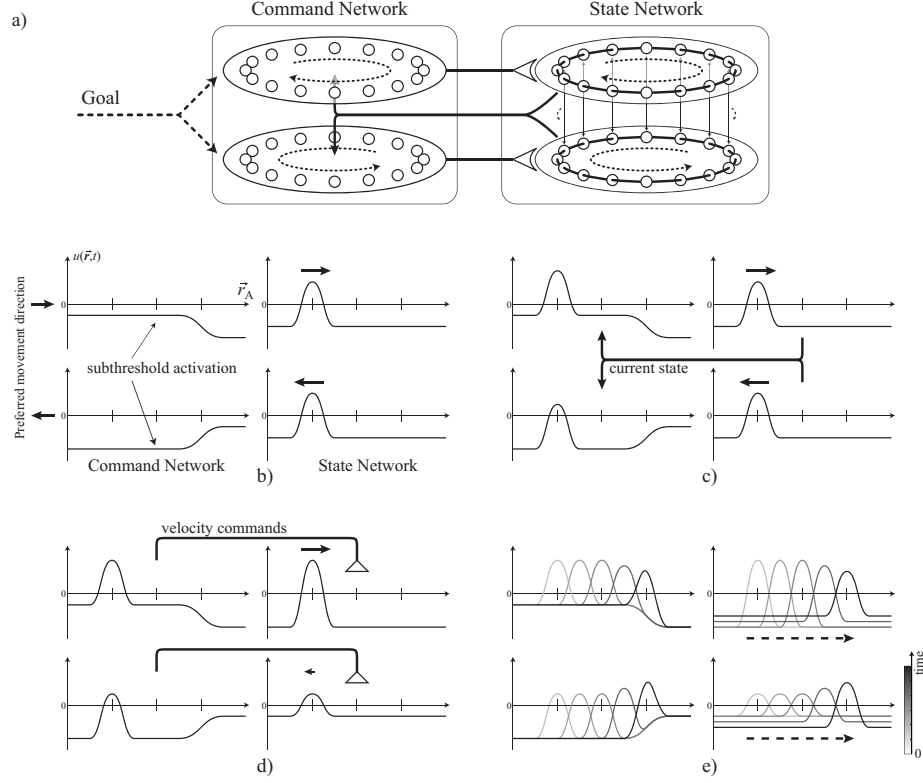


Figure 8.3: The proposed neural mechanism of movement generation is illustrated. **a)** Two coupled neural field-based networks can generate a movement in the internal state representation. The **state** network encodes the current position as well as the current velocity of the system. This state is projected into the **command** network. It computes the velocity commands according to a sensory goal that are sent back to the state network, which in turn, integrates these commands and updates its internal representation accordingly. In order to describe step by step this movement generation process, the neural activation that can be observed within the coupled network is illustrated in the Figures **b-d**). **b)** On the right, the current state of the system is shown. Since the balance of excitation is equal across each sub-layer having different preferences of movement direction, the current state does not change. On the left, the command network is silent, but a goal-dependent subthreshold activation pattern is present. Note that how the goal may preshape the subthreshold activation pattern will be described later in the text. **c)** By projecting the current state into the command network, this network gets activated. Because of this subthreshold activation, the activity across the sub-layers is asymmetrically balanced. **d)** The average excitation level of each sub-layer of the command network is projected back to the state network. This generates an asymmetric balance of excitation in that neural field. Consequently, the strongest sub-layer drives the internal representation into its preferred direction of motion. **e)** The activity packet on the surface of the field starts to move. It stops as soon as the subthreshold activation pattern of all sub-layers of the command network becomes locally equal.

weights $W^{G,C}$ appropriately set. Because a convolution on a gaussian-like input, such as that produced by the synaptic projections from the activation pattern of the goal network, can almost result in any synaptic excitation profile, any desired subthreshold preshaping can almost always be realized by synaptic projections. Its precision only depends on the breadth of the neural representation. The synaptic projections which are considered here are given by

$$W^{G,C}(\vec{r}'_A, \vec{r}_A, \vec{r}_B) = \alpha^{G,C} (\vec{r}_A \cdot (\vec{r}'_A \times \vec{r}_B) - \delta^{G,C}) \quad (8.3)$$

where $\alpha^{G,C}$ corresponds to the amplitude of the weights and $\delta^{G,C}$ to a negative offset. Further, in addition to this projection from the goal network, the command network also receives the state information from the state network. Its external inputs can thus be written as

$$\begin{aligned} x^C(\vec{r}_A, \vec{r}_B, t) = & \oint_{\Gamma} W^{G,C}(\vec{r}'_A, \vec{r}_A, \vec{r}_B) f(u^G(\vec{r}'_A, t)) d\vec{r}'_A + \\ & \iint_{\Gamma \times \Gamma} [W^{S,C}(\vec{r}'_A, \vec{r}_A) - \lambda \nabla W^{S,C}(\vec{r}'_A, \vec{r}_A) \times \vec{r}'_B] \\ & f(u^S(\vec{r}'_A, \vec{r}'_B, t)) d\vec{r}'_A d\vec{r}'_B \end{aligned} \quad (8.4)$$

where $W^{S,C}(\vec{r}'_A, \vec{r}_A)$ are the projection weights from the state network to the command network. They exhibit a center-surround profile as defined in Equ. (3.8). The asymmetric term with the gradient operator allows the state information to be immediately integrated into the command network, i.e., it compensates for the time constant of the neurons (For more details, see Section 7.2). Next, the command network sends velocity commands back to the state network through the background input $h^S(\vec{r}_B, t)$ of the latter. This projection links the mean neural activation of each sub-layer with the sub-layers of the state network according to the similarity of their preferential motion direction. It is given by

$$h^S(\vec{r}_B, t) = h_0^S + \iint_{\Gamma \times \Gamma} W^{C,S} f(u^C(\vec{r}'_A, \vec{r}'_B, t)) (\vec{r}_B \cdot \vec{r}'_B) d\vec{r}'_A d\vec{r}'_B \quad (8.5)$$

where $W^{C,S} > 0$ is the constant strength of the projection weights and h_0^S is a constant background excitation level.

In order to show that this coupling between the state and the command network can effectively drive the current state of the system toward the goal, several mathematical simplifications will be assumed. First, the output activity of the goal network is assumed to be a dirac distribution $\mathcal{D}(\vec{r}_A, \vec{s}^G(t))$ centered on the goal location $\vec{s}^G(t)$, i.e., $f(u^G(\vec{r}_A, t)) \approx \mathcal{D}(\vec{r}_A, \vec{s}^G(t))$. As a consequence, the first integral term of Equ. (8.4) reduces to $W^{G,C}(\vec{s}^G(t), \vec{r}_A, \vec{r}_B)$. Further, the second integral term of Equ. (8.4) produces a periodic gaussian-like input. By the symmetry of weights $W^{S,C}$, and since the command network also en-

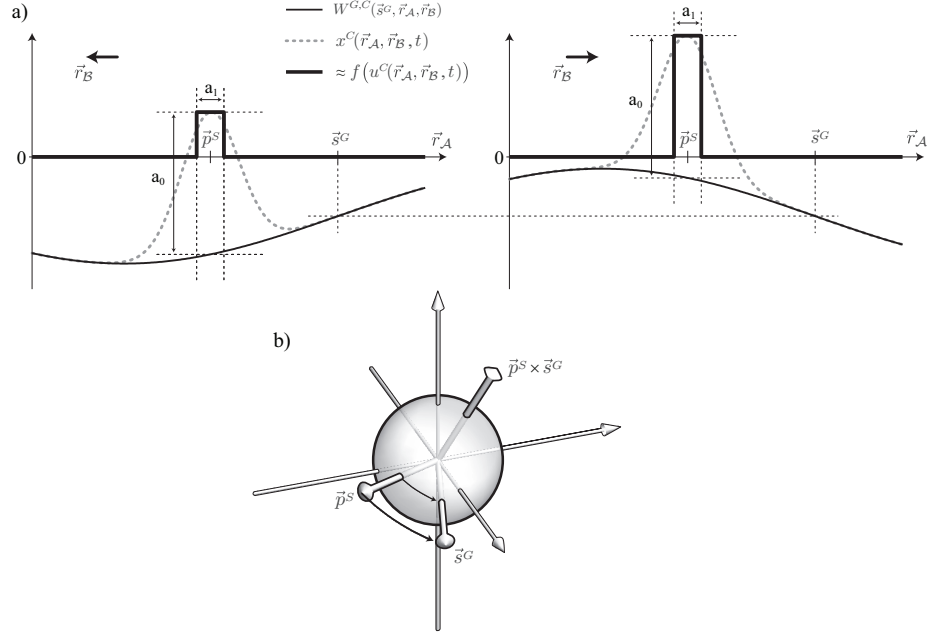


Figure 8.4: **a)** The approximation of the neural activation of the command network is shown for two sub-layers having opposite motion direction preference. This mathematical simplification was further used to calculate the effect of the projection from this network to the state network for generating movements. **b)** A vectorial illustration of the resulting command vector generated by the projections from the command network. Since the angular velocity command corresponds to the cross product of the current state $\vec{p}^S(t)$ with the location \vec{s}^G of the goal, it drives the system's state toward the goal.

codes the actual state of the system, the result of the convolution is centered on the population vector $\vec{p}^S(t)$ of the state network (Equ. (7.2)). Moreover, because of the non-linearity of the activation function f and of the difficulty to integrate gaussian-shaped functions, an approximation of the neural activity $f(u^C(\vec{r}_A, \vec{r}_B, t))$ of the command network is also assumed here. It is shown graphically on Figure 8.4a, and is written as

$$f(u^C(\vec{r}_A, \vec{r}_B, t)) \approx \begin{cases} a_0 + W^{G,C}(\vec{s}^G(t), \vec{p}^S(t), \vec{r}_B) & \vec{r}_A \cdot \vec{p}^S(t) \geq \cos(\frac{a_1}{2}) \\ 0 & \text{otherwise} \end{cases} \quad (8.6)$$

where a_0 and a_1 are constants. The main constraint that must be respected is that $a_0 > W^{G,C}(\vec{r}'_A, \vec{r}_A, \vec{r}_B)$, $\forall \vec{r}'_A, \vec{r}_A, \vec{r}_B \in \Gamma$. This implies that the amplitude of the projection from the state network to the command network has to be higher than the subthreshold activation of the latter network at any location in the neural parameter space. From this, the background input $h^S(\vec{r}_B, t)$ of the

state network given in Equ. 8.5 can be approximated, which gives

$$\begin{aligned}
h^S(\vec{r}_B, t) &\approx h_0^S + W^{C,S} \oint_{\Gamma} k_1 [a_0 + W^{G,C}(\vec{s}^G(t), \vec{p}^S(t), \vec{r}_B)] (\vec{r}_B \cdot \vec{r}'_B) d\vec{r}'_B \\
&= h_0^S + W^{C,S} \oint_{\Gamma} k_2 [\vec{p}^S(t) \cdot (\vec{s}^G(t) \times \vec{r}'_B)] (\vec{r}_B \cdot \vec{r}'_B) d\vec{r}'_B \\
&= h_0^S + W^{C,S} k_3 [\vec{p}^S(t) \cdot (\vec{s}^G(t) \times \vec{r}_B)] \\
&= h_0^S + W^{C,S} k_3 [(\vec{p}^S(t) \times \vec{s}^G(t)) \cdot \vec{r}_B]
\end{aligned} \tag{8.7}$$

where $k_i, i \in \{1..3\}$ are positive constants resulting from the integration. Since the aim of this section is to provide a qualitative description of the behavior of the system, exact values for k_i will not be given. Finally, by rewriting this result in a form similar to that given in Equ. (8.2), it becomes

$$h^S(\vec{r}_B, t) \approx h_0^S [1 + \epsilon (\vec{p}^S(t) \times \vec{s}^G(t)) \cdot \vec{r}_B] \tag{8.8}$$

where $\epsilon > 0$ is a constant. Therefore, as illustrated in Figure 8.4b, this input generates an angular velocity command that drive the internal state of the system toward the goal. In addition, while Figures 8.5a and b show vector fields generated by this mechanism for two different goal locations, Figure 8.5c depicts the neural activation of the command network for a specific state $\vec{p}^S(t)$ of the system.

8.2.3 NEURAL ARCHITECTURE

Figure 8.6 describes the neural model that is proposed here in order to reproduce and explain several behavioral data related to multisensory integration, to the sense of agency and to the interferences on self movements reported during the observation of movements performed by others. The model mainly consists of the paired inverse and forward neural fields described above, on top of which several supplementary external inputs and an output network were added. First of all, the output network, receives the same velocity commands as the state network. The purpose of this additional network is to integrate the received commands in order to simulate their overt execution. The information encoded by this output network is further considered as being the real and overt state of the whole system, whereas the information encoded in the state network corresponds to the perceived state which consists of the integration of multisensory signals. By multisensory inputs, proprioceptive as well as visual inputs are considered. On one hand, the proprioceptive input is an external feedback signal which corresponds to the real state of the system, i.e., the information encoded by the output network retrieved by its population vector $\vec{p}^O(t)$. This input is fed into the state network by means of Equ. (7.11) where the input stimulus location $\vec{s}(t)$ used in that equation is set such that $\vec{s}(t) = \vec{p}^O(t)$. It is impor-

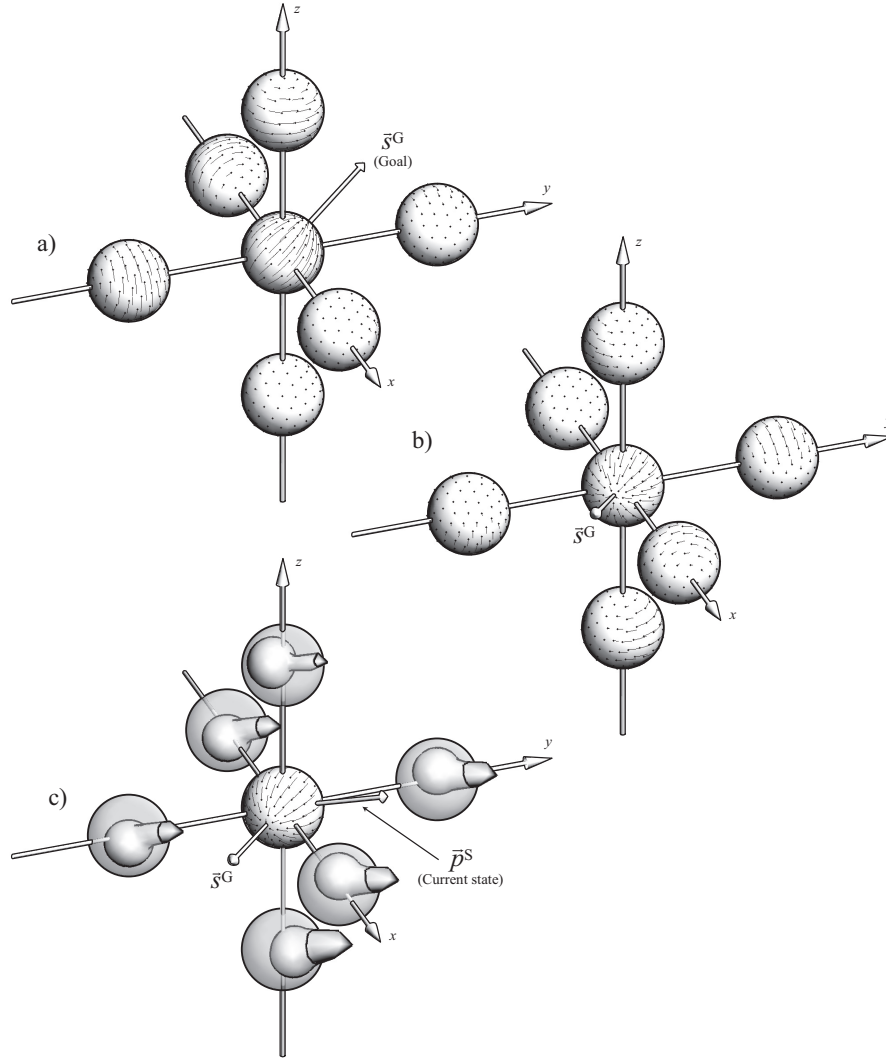


Figure 8.5: Goal-directed movement generation: **a-b)** When a goal input is projected into the command network, the latter generates a vector field which attracts the current network state toward that goal. **c)** Illustration of the membrane activity of the neurons of the command network for a given goal and a given state of the system. These activation patterns are the result of the combination of the subthreshold activation with the state representation given by the state network. As can be seen, since the distance between the current state and the goal is the longest around the z -axis, the difference of excitation of the two sub-layers oppositely tuned to that axis is the largest.

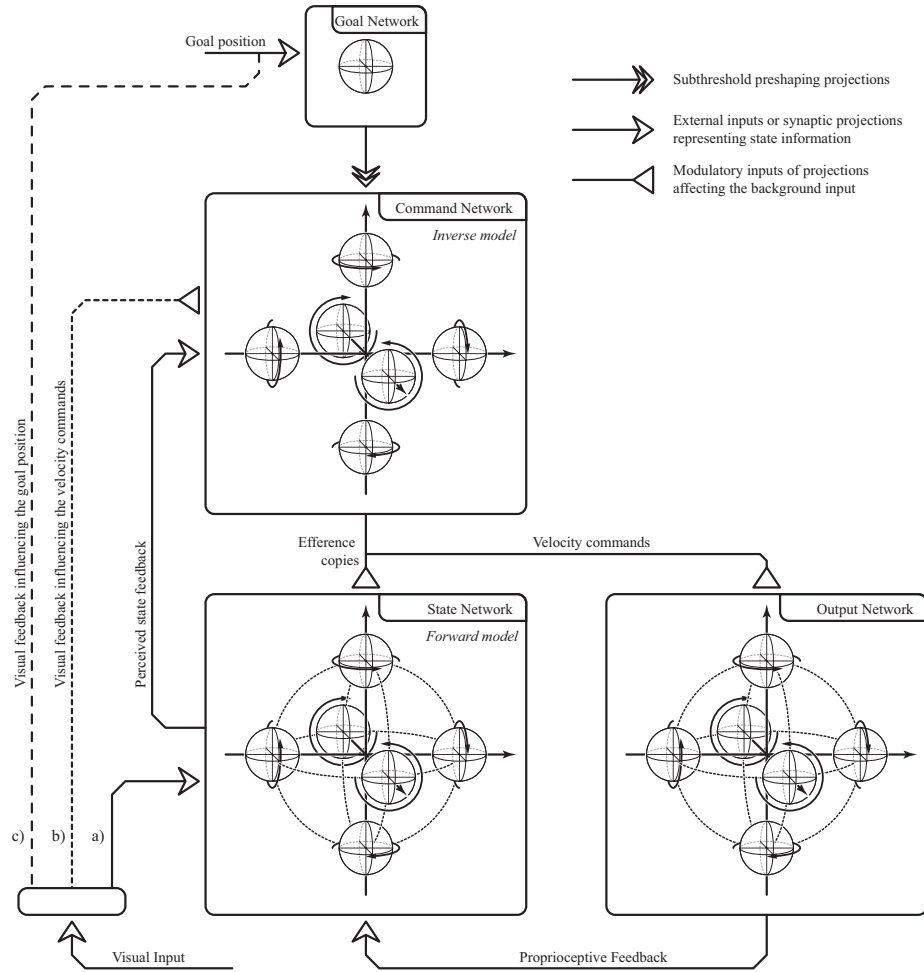


Figure 8.6: Illustration of the architecture of the model used in this study. The goal network which encodes the target position creates a subthreshold preshaping in the command network by means of synaptic projections. This network is then reciprocally connected to the state network. The latter neural field sends the current state of the system to the command network, which projects back the velocity commands needed to reach the goal. Simultaneously, the commands are also sent to another network, which serves here only to integrate these commands. The output state of this network is intended to correspond to the real end-effector position. This information is then used as a proprioceptive feedback to the state network. It is important to note that the state network represents the state that is perceived by the system, and not the real one. Indeed, in addition to that proprioceptive feedback, this network can also, through channel **a)** receive a visual feedback. A multisensory integration of information thus occurs here. Finally, as will be described in the second experiment (Section 8.3.2), visual feedback may also influence the network dynamics by affecting the goal position or the velocity commands through channels **b)** and **c)**, respectively.

tant to stress that this work does not take into account the time delay which is known to arise from the sensorimotor loop which sends motor commands to the muscles and which brings back the proprioceptive signals to the cortex. On the other hand, the considered visual input signal may consist of different types of information. It can be either the direct visual representation of the movement belonging to the self which is, in normal conditions, identical to the proprioceptive input, or a biased visual feedback of that same information, or finally, a visual signal of the same body part, but which belongs to another individual. The exact mathematical form as to how these types of visual inputs are fed to the network will be given when appropriate, in the description of the setup of each experiment.

8.3 EXPERIMENTS AND RESULTS

In this section, two experiments are proposed to validate the model, as well as to provide experimental predictions regarding some of the current hypotheses concerning how imitative interferences are induced in the brain when observing the actions of others. The description of each experiment will give details about their setup and simulation results. In addition, the values taken by the network parameters in all these experiments can be found in Appendix B.4.

8.3.1 MOTOR ADJUSTMENTS CAUSED BY MULTISENSORY CONFLICTS

The first experiment attempts to provide an explanation at the neural network level of the behavioral evidence that the multisensory integration of self-related information is broadly sensitive to the spatial congruency of the feedback signals. Indeed, this integration has been shown to be performed up to a certain level of discrepancy between these multisensory signals. The experiment that is described here is very similar to those proposed by Fournieret and Jean-nerod (1998) and Slachevsky et al. (2001). In their experimental setup, shown in Figure 8.1 and described in Section 8.1, a bias was artificially applied to the visual feedback, which is meant to indicate the actual location of the subjects' hand. Consequently, in order to reach for the target, subjects had to correct their movement by applying a motor bias in the opposite direction. However, since the amount of the bias could not be inferred instantaneously, the subjects had to apply their correction on the fly. Interestingly, for small bias, subjects mostly corrected unconsciously their movements, whereas for large biases, they had to consciously monitored their movements.

In comparison to this work, the the experimental setup described here differs slightly from of the task instructions. In order to avoid considering the change

of movement strategy resulting from large visual bias, and hence to avoid developing an additional neural module in charge of it, the task here is no more to reach for the target visually, but at the proprioceptive level only, leaving the visual feedback with the role of a distractor. An experimental trial is decomposed as follows. The goal network initially receives an input which indicates the starting position \vec{s}_0^G so that the system converges toward that initial location. The angular bias θ , centered around the starting position, is then introduced in the visual feedback. That feedback, coding for the state $\vec{s}^V(t)$ of the system in visual terms is fed into the state network using Equ. (7.11). Note that the amplitude of both the visual and proprioceptive signals were set to the same value, so that none of them was considered more important than the other. Further, the goal is set to the target position \vec{s}_1^G , which drives the system to reach for it.

Illustrations of the typical behavior of the system during this task is given in Figure 8.7a for two different conditions, one with a relatively small bias, and another with a larger one. This figure shows the population vector response of several neural populations of the model, especially those corresponding to the actual and perceived states of the system. Moreover, since the simulation were performed on a spherical space, in order to project the neural responses on a plane, the transformation shown in Figure 8.7b was applied to the data. In the small bias condition, it can be noticed that the real movement is initially correct, i.e., it moves straight to the target. However, since the visual feedback is slightly biased, the perceived multisensory representation is shifted between the proprioceptive and the visual feedback. Moreover, since both sensory representations do not gets far enough from each other, the internal representation keeps integrating both input sources. Consequently, although the perceived internal state accurately reach the target, it is not reached neither visually nor proprioceptively. At the beginning of the other trial, where a large bias was induced, the same shifting effect is observed. However, when the visual feedback becomes too different from the proprioceptive signal, it enters in conflict with the network intrinsic dynamics. Consequently, it is removed from the multisensory integration process, allowing the model to finally reach the target without paying attention to the bias induced in the visual feedback.

Other simulation results are given in Figure 8.8. They show that up to a certain amount of the visually induced bias, the proprioceptive and the visual inputs are merged together in the multisensory state perceived by the system. Since both inputs are set with an equal strength, the perceived state represents the average location conveyed by both inputs, and consequently, the output network does not perfectly reach the target but stays at half of the angular distance determined by the bias. However, over the discrimination threshold, the state network selects the input which corresponds best to the state predicted by its internal dynamics. The visual input gets thus inhibited and the proprioceptive signals becomes the only ones on which the internal representation relies.

Further, in order to determine the angle of deviation at which, for a given

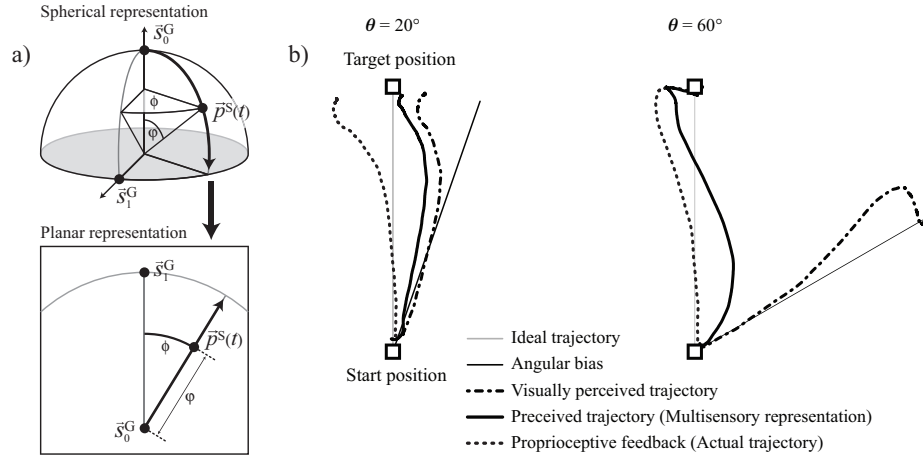


Figure 8.7: **a)** The transformation used to convert simulation data encoded in spherical coordinates to planar ones is given. **b)** Illustration of two trajectories followed by the network where two different visual bias θ were applied. In both cases, since the trajectory of the perceived state ends on the target, the model believes that the goal is reached and hence stops moving. On the left, the induced bias is small. Consequently, the perceived state lies between the actual state, given by proprioceptive signals, and that provided by the visual feedback. A different behavior can be observed on the right where the artificially induced visual bias is much larger. Indeed, although the perceived state is also biased at the beginning of the trial, the system finally selects the proprioceptive feedback because it corresponds best to the internal dynamic of the system, and also because the two sensory signals are too divergent. After a correction phase, the perceived state finally converges to the information provided by proprioception only.

breadth σ of the internal representation, the system becomes capable of clearly perceiving a difference between visual and proprioceptive signals, a model function f , linear by parts, was fitted to these simulation results. It is given by

$$f(\theta) = \begin{cases} a_0\theta & \theta \leq \theta_0 \\ a_0\theta_0 - a_1(\theta - \theta_0) & \theta_0 < \theta \leq \theta_1 \\ a_0\theta_0 - a_1(\theta_1 - \theta_0) & \theta_1 < \theta \end{cases} \quad (8.9)$$

where a_0 and a_1 correspond, respectively, to the raising and falling slope of the function, and θ_0 and θ_1 to the boundaries where the strength of the conflicts in the multisensory integration begins to decrease and when it reaches almost zero, respectively. The discrimination threshold $\hat{\theta}$ is then defined such that $\hat{\theta} = (\theta_0 + \theta_1)/2$. Results of this fitting procedure² as well as values of the discrimination threshold $\hat{\theta}$ with respect to σ are given in Figures 8.8b and c. As can be observed, the dependence of the breadth of the representation on the discrimination threshold is almost linear.

Although the measure of the error is technically different between this study and that of Slachevsky et al. (2001), there is an important qualitative similarity between the simulation results presented above and theirs. In both studies, the

²The fit was obtained by estimating the coefficients of the nonlinear function f using least squares.

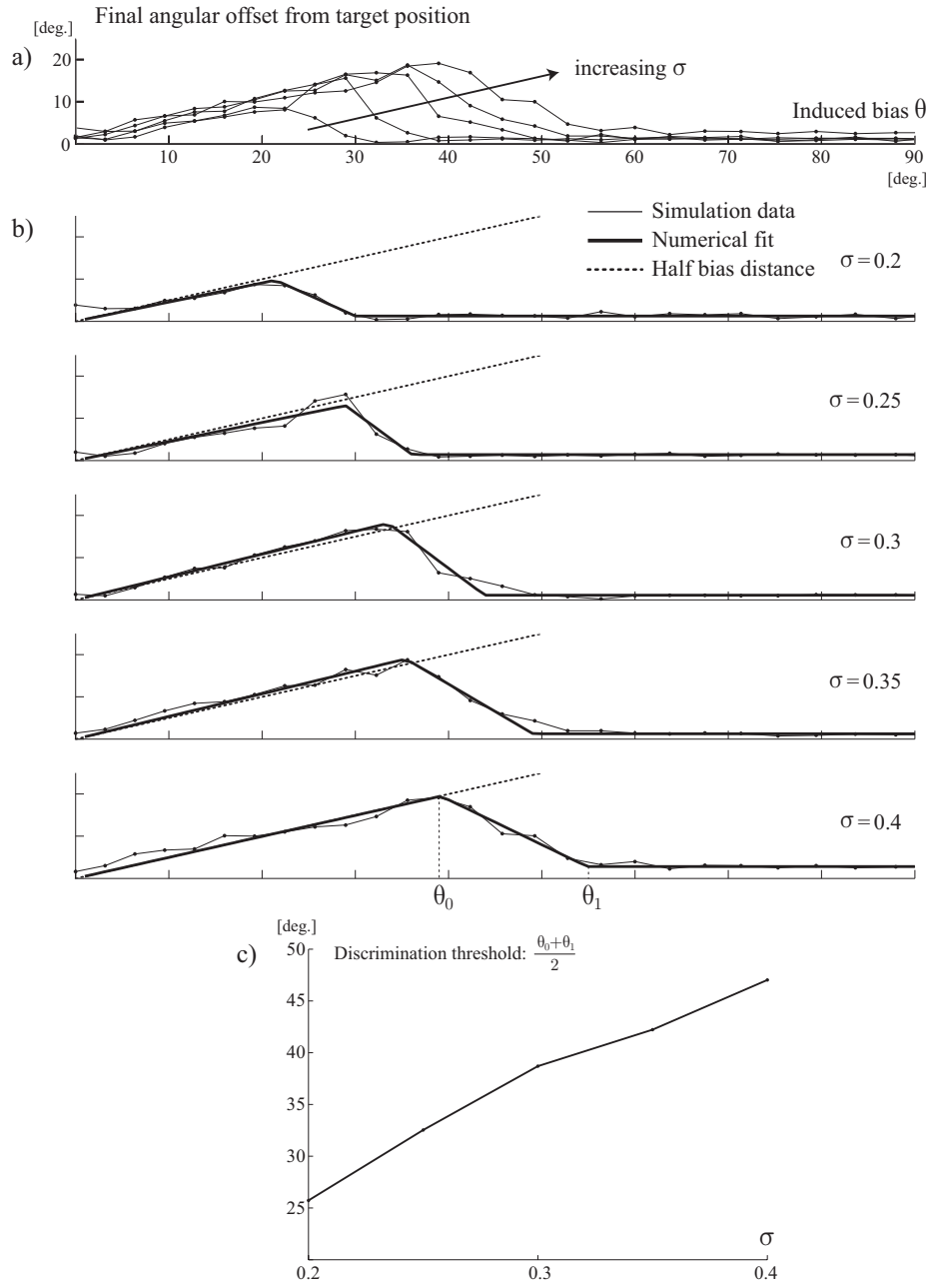


Figure 8.8: **a)** Simulation results showing the final angular offset from the desired position measured on the output network for different bias values and different breadth σ of the neural representation. **b)** The result of the numerical fit on simulation data using the function $f(\theta)$ given in Equ. (8.9) are shown. In addition, for small θ , since the amplitude of the proprioceptive and visual feedbacks were set with equal amplitude, the final angular bias of the output network lies approximatively in the middle distance between the two sensory signals. In contrast, for large θ , the network always selects the correct sensory input, and is hence no more influenced by the visual bias. **c)** The value of the discrimination threshold θ for different values of σ is reported. It constantly increases with σ .

error increases almost linearly until the bias reaches the discrimination threshold, after which, the error stays more or less constant. However, because of the difference in the task instructions as well as the way errors were measured, in this work, the error decreases until almost zero around the discrimination threshold. Nevertheless, this modeling study replicated the fact that sensory discrimination is not accurate until a certain bias level.

8.3.2 INTERFERENCES DURING SIMULTANEOUS EXECUTION AND OBSERVATION OF MOVEMENTS

In the second experiment, an analysis of the possible interference patterns on self movements during the simultaneous observation of the movements of others is considered. The setup of this task is very similar to that used by Kilner et al. (2003), where a subject and an experimenter are simultaneously performing oscillatory arm movements in the same or different directions (See Figures 2.1c and d). In this experimental configuration, the presence of interferences were revealed by an increase of variability of the subjects' movements, which was higher for incongruent pairs of observed and executed movements than for congruent ones (Kilner et al., 2003; Chaminade et al., 2005).

The idea behind the experiment proposed here, is to analyze the interference patterns into more details. Instead of considering only movement variability as a measure of the interference (Kilner et al., 2003; Chaminade et al., 2005), this work proposes to look for interference patterns at the level of the movement trajectory. As will be shown later, this analysis may be useful for determining which components of observed movements are inducing interferences, and subsequently, which cortical pathways could be responsible for such an effect. Therefore, in the present study, the task was slightly modified. In order to simplify the analysis of the interference patterns, movements are restricted to a single motion plane. The subject, standing in front of the experimenter, is requested to raise his/her arm in the horizontal plane and to keep it straight toward the experimenter, whereas the latter produces only vertical oscillatory arm movements in the direction of the subject.

Considering now the model connectivity, recall that it incorporates several key elements associated with the movement goal, velocity commands as well as an internal shared representation. Therefore, the response of the model to visual signals interfering with these three different system pathways can be simulated. Effects resulting from interferences induced in these three regions of the model are separately tested in three different conditions.

First, as illustrated on Figure 8.6, the actual position of the hand of the experimenter may influence the goal representation of the subject. In this con-

dition, the variable encoded by the input of the goal network becomes

$$\vec{s}^G(t) = \frac{\vec{s}_0 + \rho \vec{s}^{\text{Arm}}(t)}{\|\vec{s}_0 + \rho \vec{s}^{\text{Arm}}(t)\|} \quad (8.10)$$

where \vec{s}_0 and $\vec{s}^{\text{Arm}}(t)$ correspond, respectively, to the subject's target arm position and the position of the arm of the demonstrator as given in Equ. (6.5). ρ denotes the strength with which the representation of the goal is biased by the observed movement. Secondly, if the interferences are considered to come from a pathway feeding the observed velocity of the experimenter's hand into the inverse model of the network, the imitative mechanism is assumed to try to drive the actual state of the system toward that speed. This can be realized by acting on the background input $h^C(\vec{r}_B, t)$ of the command network, which gives

$$h^C(\vec{r}_B, t) = \rho ((\dot{\vec{s}}^{\text{Arm}}(t) - \dot{\vec{p}}^S(t)) \cdot \vec{r}_B) \quad (8.11)$$

where, ρ corresponds here to another factor modulating the strength of that external input. Finally, if the reported interferences are caused by an integration of sensory information related to both self and others into the same neural substrate, the external input of the state network, in addition to receive proprioceptive feedback, also receives an input $\vec{s}^V(t)$ equal to the orientation of the experimenter's arm $\vec{s}^{\text{Arm}}(t)$. Note that, similarly to the work in Chapter 4, the sensory information which flows along the system is assumed to be encoded in the same frame of reference.

Interference Patterns

Figure 8.9 reports, for each condition, a typical interference pattern which consists of the trajectory followed by the model's effector while observing the experimenter's oscillatory movement. In the condition where movement observation influences the goal of the executed movement, the interference patterns closely resembles to the interfering movement. It is however smaller in amplitude and lags behind the observed motion. In contrast, in condition where the movement velocity is in charge of the interferences, the motion pattern leads the trajectory of the observed movement. By contrasting these two conditions, on one hand, the lag is primarily produced by the time needed for the goal representation to be updated and then for propagating this information into actual motor commands. On the other hand, the lead is caused by the inverse model, which reacts against the perturbation by producing commands in the opposite direction to that of the observed movement. Indeed, at the beginning of the movement period, the currently perceived state approximatively indicates the goal position, and thus, no motor commands are produced. However, as the demonstrator moves, this induces a movement of the model which is then counteracted by the self-generated commands which tries to keep the straight position.

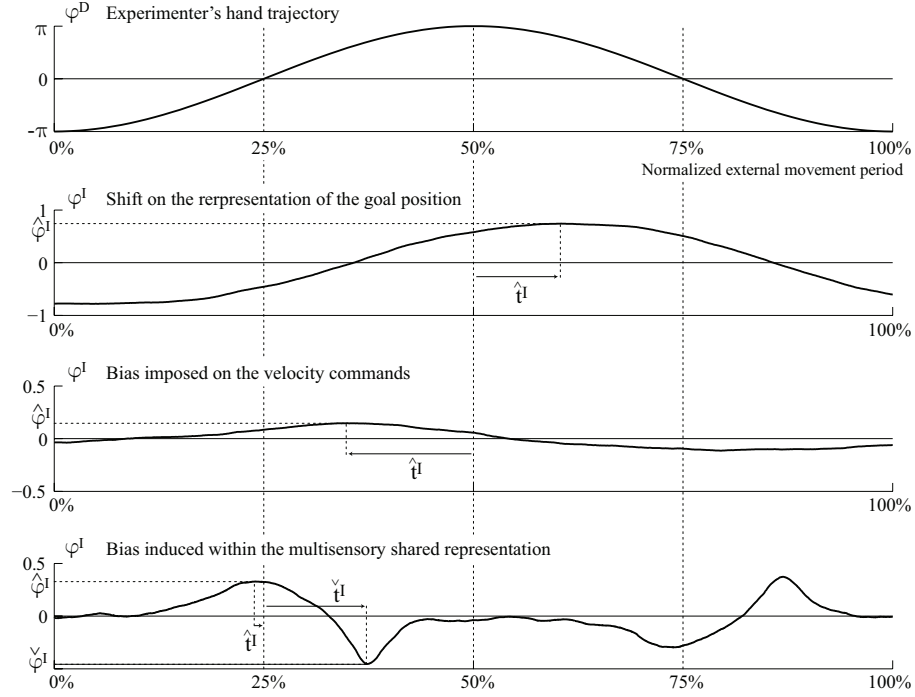
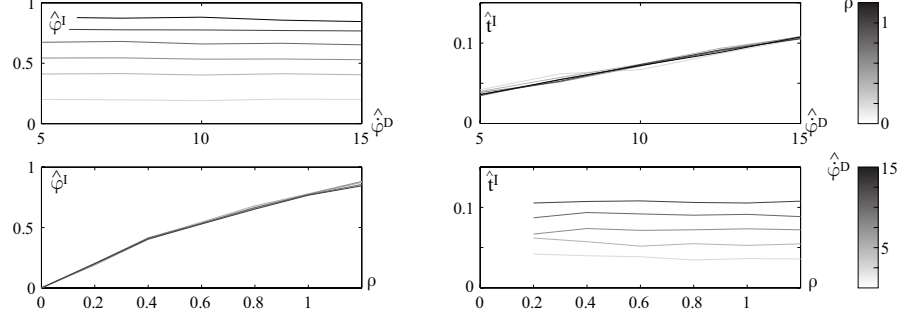


Figure 8.9: Typical interference patterns are shown aligned to a single period of the oscillatory movement performed by the experimenter. **a)** The movement performed by the demonstrator follows a cosine shaped trajectory. **b)** Interference pattern observed in the first condition where the representation of the static target position is influenced by the position of the experimenter's hand. **c)** In the second condition, the network receives conflicting velocity commands which try to make the velocity profile of the observer's movement to match that of the experimenter. **d)** Typical trajectory followed by the observer's hand in the third condition where the conflict occurs at the level of the internal representation of the system's state.

Further, in the last condition where information related to self and others are merged together, the typical interference pattern is quite different. When the position of the experimenter's hand becomes similar to that of the subject, the perceived state of the system becomes less precise by representing a sort of averaging of this input with the proprioceptive feedback. Consequently, the velocity commands in the opposite direction is produced. Then, as the experimenter's hand continues its trajectory, the internal integration process of multisensory signals selects only the input which corresponds to the subject actual effector. Since at this particular moment, it is located away from the target position, a velocity command in the opposite direction is initiated, which produces a deviation in the opposite direction in order to recover the desired position. This phenomenon is finally reproduced during the second part of the experimenter's movement, when his/her hand comes again in the vicinity of the subject's hand position.

a) Effect of biasing the goal representation



b) Effect of velocity induced interferences

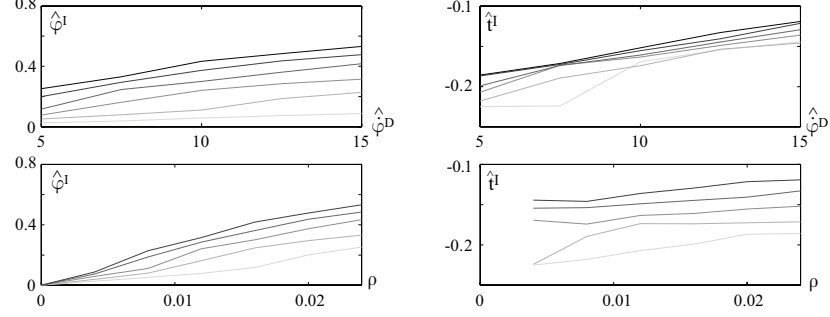


Figure 8.10: Results of the numerical fit of the trajectories followed by the population vector of the output network as recorded in the **a)** first and in the **b)** second condition. The response variables shown here correspond to the results of the fit which used the function described in Equ. (8.12). The signification of these variables are also shown in Figure 8.9.

Effects of Movement Velocity and Input Strength

Since the velocity of the observed movement, given by its maximal angular speed $\hat{\varphi}^D$ as well as the strength ρ of its influence on the system were systematically varied, their effects on the interference patterns were analysed. In order to describe them, the amplitude and the temporal phase of the maximal deviation produced by the interferences on the overt motion trajectory were considered. As can be noticed in Figure 8.9, the interference patterns in the two first conditions can easily be fitted by means of a parameterized cosine function $f(t)$ given by

$$f(t) = \hat{\phi}^I \cos \left(2\pi \left(t - \hat{t}^I - \frac{1}{2} \right) \right) \quad (8.12)$$

where $\hat{\phi}^I$ is the maximal amplitude of variation and \hat{t}^I its phase. This phase is measured relative to the half period time of the observed trajectory so that a zero phase indicates that both the observed movement and the interference patterns are aligned. In contrast, the interference pattern produced by the integration of self- and others-related sensory information is more likely to be described by the difference between two gaussian functions, which is repeated

Effect of multisensory integration within a shared representation

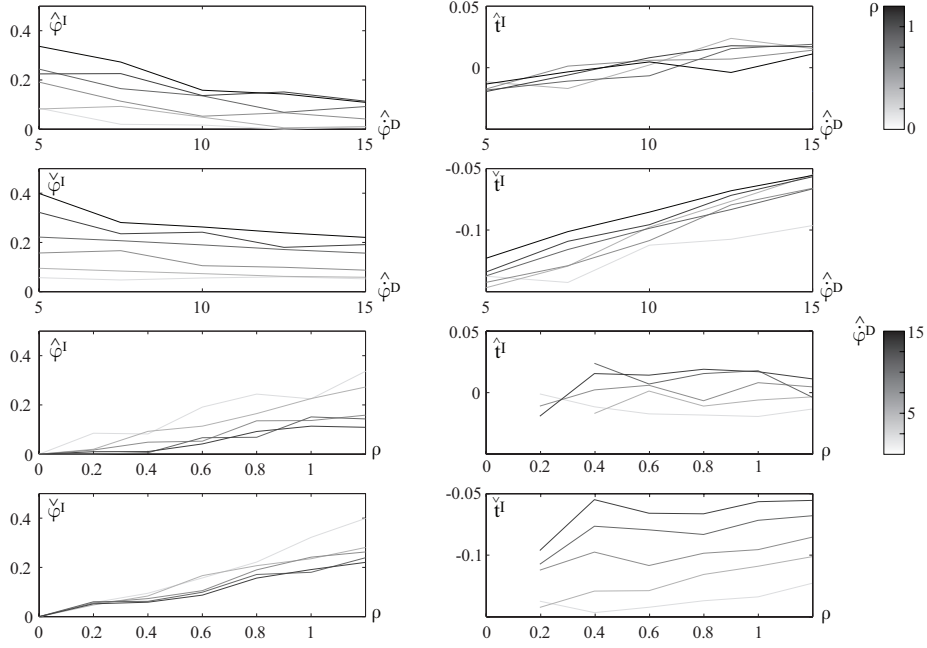


Figure 8.11: Results of the numerical fit of the trajectories followed by the population vector of the output network as recorded in the third condition. The same notation as in Figure 8.10 is used. However, the fitting function corresponds here to a difference of two gaussian functions as described in Equ. (8.13).

during the second half of the movement period with inverted amplitudes. In this case, the model function f to be fitted is given by

$$f(t) = f_0 \left(t - \frac{1}{4} \right) - f_0 \left(t - \frac{3}{4} \right) \quad \text{where}$$

$$f_0(t) = \hat{\varphi}^I \exp \left(-\frac{(t - \hat{t}^I)^2}{2\hat{\sigma}^2} \right) + \check{\varphi}^I \exp \left(-\frac{(t - \check{t}^I)^2}{2\check{\sigma}^2} \right) \quad (8.13)$$

where $\hat{\varphi}^I$ and $\check{\varphi}^I$ are the maximal amplitude of variation in the sensory merging phase and in the discrimination phase, respectively. Similarly, \hat{t}^I and \check{t}^I correspond to the associated time at which the maximum deviation is attained.

In Figures 8.10 and 8.11, the results of the fit performed on the simulated data during the three conditions are shown. A summary of the effects of each external parameters on the amplitude and phase of the interference patterns is provided in Table 8.1. In the first condition, where the representation of the goal is biased, the velocity of the experimenter's movement has no influence on the amplitude of the interference pattern but tends to shift the phase to a later time. The faster is the observed movement, the larger is the phase difference between observed movement and the interference pattern. In contrast, the influence of the strength ρ of the interfering input is opposite, in that it is proportional to the amplitude of deviation and has no influence on the phase. In the second

	Goal representation		Bias influencing the Velocity commands		Shared representation			
	$\hat{\varphi}^I$	\hat{t}^I	$\hat{\varphi}^I$	\hat{t}^I	$\hat{\varphi}^I$	\hat{t}^I	$\check{\varphi}^I$	\check{t}^I
$\hat{\varphi}^D \nearrow$	—	\nearrow	\nearrow	\nearrow	\searrow	\nearrow	\searrow	\nearrow
$\rho \nearrow$	\nearrow	—	\nearrow	\nearrow	\nearrow	—	\nearrow	—

Table 8.1: Summary of the effects on the interference patterns produced by increasing the velocity $\hat{\varphi}^D$ of the observed movement and the strength ρ of its internal representation. Results are shown for the three suggested hypotheses.

condition, the effect of these parameters is different. Both the velocity and the strength of the input increase the amplitude of deviation as well as the phase of the interference patterns. The faster is the observed movement, the stronger is the interference. In addition, a strong coupling value leads the system to match more closely the observed movement. Finally, in the last condition, where the interfering signal related to the experimenter is integrated into the representation of the state of the system, the effect of the observed movement velocity is completely different to the other conditions. Indeed, the amplitude of deviation is reduced as that velocity increases. Indeed, when a faster movement is observed, the time during which the internal representation of the self and that of others overlap is reduced, and consequently, the duration of the interference too. The involvement of the sensory discrimination is here clearly at play in that, as the two sensory inputs get distant, the wrong input is inhibited and has no more influence on the system's state. Next, the effect of the velocity on the phase of the maximal deviations is however similar to those in the other conditions, i.e., the phase increases with an increasing observed movement velocity. Next, as could have been expected, the strength of the external signal affects the amplitude of the interference pattern in a proportional manner. Indeed, as ρ grows, the bias induced in the shared representation moves the perceived state more toward the that of the observed movement. Finally, the effect of this parameter on the phase of the deviation is quite small. It seems that this is the consequence of the model ability to select the correct sensory input with an almost equal latency for any values of ρ , while the amplitude of the visual signal stays below that of the proprioceptive feedback.

8.4 DISCUSSION

In this chapter, a neural network model based on a paired inverse and forward model for the control of goal-directed movements was shown to be capable of controlling the kinematic of simple movements. The forward model consists of a network which integrates the commands provided by the inverse model on the currently perceived state. This state is the result of the integration of multisensory signals coming from proprioceptive and visual feedbacks. Impor-

tantly, the fundamental property of the internal dynamics of the system is to constrain the representation of the state to stay coherent with the consequence of self-generated motions. Indeed, when biased or contradictory feedback signals are perceived, they are removed from the integration process. In addition, the model was also designed so as to be capable to investigate how self- and others-generated movements are simultaneously integrated together within a shared representation of movements.

While the biological plausibility of the part of the model in charge of integrating velocity commands has already been discussed in Chapter 7, that of the inverse model is discussed here. First of all, the biological plausibility of its neural structure, by its similarity to that of the forward network, can easily stand on the same arguments (See Section 7.4). However, the plausibility of the principle as to how goal-dependent commands are produced may be more difficult to assess. Indeed, the subthreshold activation patterns, which were suggested to control velocity commands, can not, by their subthreshold nature, be directly compared to neural recording data. Nevertheless, they may be indirectly observed by monitoring how the firing amplitude of the activity of neural populations, which represent current state variables, may be influenced by different goal locations. Moreover, since the activity profile of these neurons can either be positively or negatively modulated by the distance to the goal, an effect which depends on the preferential motion direction of the neurons, it would thus be interesting to see whether the activity of groups of neurons within cortical regions responsible for reaching movements respects this correlation.

Behavioral Responses to Multisensory Conflicts

In this modeling study, two behavioral experiments were proposed and simulated. In the first experiment, the model was requested to perform a reaching task, while relying on proprioceptive and visual feedback signals in order to estimate its current state. In this apparently simple task, a bias in the visual feedback was introduced which produced conflicts in the multisensory integration process. Simulation results have shown that below a certain amount of bias, the state perceived by the model consists of a mixture of both sensory information, indicating that, despite of their divergence, both inputs were still considered similar. However, above a certain threshold, the incongruent visual feedback was progressively getting inhibited during the course of the movement. At the end of the trial, the internal state of the system was finally relying only on proprioceptive information. This allowed the model to perform the task correctly. This general behavioral effect is concordant with results obtained in experimental studies which reported that under a certain discrepancy level, multiple sources of information are unconsciously merged together, while above, a conscious monitoring of the sensed discrepancy allows humans to change their control strategy accordingly (Fournieret & Jeannerod, 1998; Slachevsky et al., 2001).

In this experiment, the breadth of the internal representation of the system was also systematically varied. The analysis of the effect produced by this parameter on the discrimination threshold showed that the latter grows almost linearly with the unreliability of the sensory representation³. Interestingly, this may bring one to postulate that the sensitivity of the discrimination process as observed in the related experiments may be less accurate in children than it is in adults. Indeed, the neural representation of spatial location has been suggested to become more precise along developmental stages (Schutte et al., 2003).

Finally, the first experiment did not address the effect of attentional processes. Indeed, since the instructed task was to reach for the target in proprioceptive terms, one may argue that humans may simply inhibit the visual feedback signal and always succeed in the task. However, several behavioral experiments have shown that the inhibition of visual responses is hard, and that interferences are still observed even when subjects are aware that visual information is purely distractive in nature (Simon & Berbaum, 1990; Kilner et al., 2003; Bertenthal et al., 2006). Nevertheless, it is rational to consider that the strength of the visual feedback can be slightly inhibited by a top-down modulation process. In such a case, by extension to the analysis of the integration model given in Section 7.3.4, an interfering signal with a smaller amplitude should affect the integration process to a lesser extent. Therefore, by consciously trying to avoid paying attention to the visual feedback, the final distance between the location of the actual movement and that of the target, as well as the discrimination threshold may be reduced.

Interferences during Simultaneous Movement Execution and Observation

In the second experiment, the question as to how interferences on one's own movements are produced while observing another individual which is also moving was addressed. With respect to the related experimental studies which have reported this effect (Kilner et al., 2003; Chaminade et al., 2005), this work tries to go one step further, first by considering the possible factors which are responsible for this behavioral effect, and then by offering model predictions to be compared to real experimental data. Indeed, these experimental studies have only considered the movement variance as a measure of the interferences, which does not provide much information as to which factors are effectively influencing this behavioral property.

In this work, in addition to have reproduced the increase of movement variability in conflicting situations, three hypotheses related to the nature of the visual signals interfering with the internal representation of movements were suggested. The first two hypotheses consider that automatic imitative behaviors tend to make one's own body to match either the position or the velocity of

³It has been suggested that the breadth of the sensitivity profile of neurons is inversely proportional to the reliability of the information they convey (Pouget et al., 2003)

the observed individual in a body-centered frame of reference⁴. The third hypothesis assumes that interferences are produced by the integration of self- and others- related signals within the same neural substrate. Simulation results then showed that these three forms of interfering signals produce different stereotypical interference patterns. While the reported patterns in the first two conditions highly resemble the observed movement, that in the third condition is clearly different. Since the representation of others' movements interferes with one's own self representation, instead of mimicking the experimenter's movements, the recorded trajectory reflects counteracting forces to bring the perceived state back to the goal position.

Further, the effect of the velocity of the observed movement was tested against these three conditions. Interestingly, the third hypothesis predicts, in contrast to the other ones, that the amplitude of the interference should decrease with increasing velocity. In addition, the strength of interfering input to the system was also systematically varied. In relation to neurobiology, this parameter could be seen as the attentional level that is focused on the observed motion. As could have been expected, it increases the amplitude of the interferences in all conditions. However, in addition to the hypothesis that this parameters may represent an attentional level, one may also suggest that this parameter may also reflect how much the observer identifies himself/herself with the experimenter. Indeed, it has been shown that arm movements performed by a robotic arm does not seem to produce interferences (Kilner et al., 2003). Nevertheless, in a subsequent experiment, Chaminade et al. (2005) showed that the level of similarity between the movements performed by a humanoid robot and those performed by real humans influences the amount of interference. Thus, this may suggests that the more biological observed movements are, the more one sensorimotor system gets excited, and hence, the more interferences are produced. This behavioral characteristic may thus be used to help determine the visual and neural factors that influence one's own movements. Moreover, this parameter may also be affected experimentally by influencing the observer's perception of the observed body as if it was actually belonging to him/her, and as if he/she was actually controlling it. This artefact on the sense of agency of the observer can be produced by means of the *rubber-hand illusion* which is an established method for manipulating the sense of body ownership (Schütz-Bosbach, Mancini, Aglioti, & Haggard, 2006).

Importantly, it should be mentioned that, these three hypotheses are not exclusive. Indeed, it is even more probable that all the considered factors participate in the interfering process. It would however be interesting to quantify experimentally whether human responses to conflicting movements could be decomposed into several independent components such as those described here,

⁴In order to avoid considering possible conflicts in the imitative strategies such as those described in Chapter 6, the experimental setup was chosen so that the response of these strategies, i.e., mirror or anatomical imitation, is always congruent with the other.

and also how the velocity of observed movements may affects the interferences. This would deepen our understanding of the neural mechanisms and brain pathways responsible for automatic imitative behaviors. Indeed, as will be described below, brain regions can be associated to the presented model.

Cortical Routes

Finally, several associations can be suggested between the parts of the model and cortical regions. First, according to Miall (2003), the computations responsible for transforming goal locations into motor commands, i.e., inverse computations, are produced in the brain along two separate pathways: one along a cortico-cortical route from the posterior parietal cortex to the ventral premotor cortex, and the other through the cerebellum. Conversely, the predictive computations are suggested to be processed along the reverse direction. Miall (2003) does not clearly mention specific regions where each computation occurs but rather brain pathways along which they are processed. This suggests that, instead of trying to separate the forward and inverse processes into different brain regions, single brain regions may simultaneously be responsible for both computations. Consequently, the system controlling motor actions may in fact be distributed across the whole brain, where each sub-regions could be in charge of computing action commands related to the sensory inputs it receives while simultaneously maintaining a coherent representation of that multisensory state. An example of functional cognitive model which considers this paired inverse and forward network to be localized within a single brain area was proposed by Oztot et al. (2005). They suggest that both computations occur in monkey area F5. But nevertheless, they also propose a separation of these two processes into two known subclasses of F5 neurons, i.e., the canonical and the mirror neurons. While canonical neurons may encode how an action should be performed in motor terms, the mirror neurons may translate this information into a multisensory representation useful for monitoring that action (Oztot et al., 2005). The computational modeling approach described here is in line with this hypothesis of a gradual computation of forward and inverse commands across brain pathways. Indeed, the proposed network is very compact and simultaneously computes commands and estimates their consequences. Moreover, it is important to note that the forward and inverse network proposed here may be reduced to a single network with the help of several technical modifications. However, the dynamics of the system was considered to be more easily describable in two parts.

As mentioned above, one important computational property brought to light by the present study, is that this neural field model can discriminate self-generated movements from those produced by others. This property naturally emerges from the competition between the network's intrinsic dynamics with the sensory inputs it receives. While compatible inputs are enhanced, incompatible ones get inhibited. As a consequence, this system can simultaneously integrate sensory information coming from multiple sensory source, while selecting those

that are coherent with self motions. A link between this model and the brain areas responsible for discrimination between self and others is then discussed. As described by several neurophysiological studies, one important brain area involved in this process is the right tempo-parietal junction (rTPJ). This area is reciprocally connected with the parietal cortex known to represent the body schema, i.e., the state of the body (Berlucchi & Aglioti, 1997; Farrer & Frith, 2002). In addition, since this area also receives the information related potential action goals, it can thus also represents the acting body in relation to its environment. Furthermore, rTPJ is also suggested to receive visual inputs related to observed bodies, including one's own body, from the extra-striate body area (EBA) (Downing et al., 2001; Astafiev et al., 2004). Therefore, rTPJ is an ideal candidate to be the area which performs multisensory integration for discriminating between self- and others-related visual feedback, and subsequently, to implement the neural network associated with the forward model proposed in this study. Indeed, as suggested by Decety and Sommerville (2003), a method for achieving this discrimination is to compare the sensory outcomes with those predicted by a forward model. Importantly, the neural network described here is capable of performing both operations simultaneously.

Summary

This chapter investigated the neural mechanisms responsible for multisensory integration and for discriminating visual feedback signals caused by self-generated movements as opposed to those generated by other individuals. The neural model is an extension of that presented in Chapter 7, which was itself capable of integrating velocity commands for predicting their outcome. The neural network in this chapter, when coupled with that described in Chapter 7, resulted in a simple coupled inverse and forward controller. This model is simultaneously capable of generating movement commands, merging multisensory information, and keeping its internal representation coherent with the predicted consequence of self-generated movements. This suggests that the brain may, by means of the same neural structure, combine a representation of others-generated motions with that of self-generated motions while still being capable of discriminating between the two sets of actions.

The model was also tested against two experiments derived from two real behavioral studies. The first simulation study showed that the model can qualitatively account for how humans control movements when proprioceptive and visual feedback conflict, i.e., when an artificial bias is applied to one modality. Below a certain discrepancy level, both sensory signals are treated similarly, while above that level the signal with a non-predictable behavior is suppressed. In the second experiment, the same model was tested against an interference paradigm where movement execution and movement observation occur simultaneously. There were several hypotheses concerning how the visual information associated with the movements of others interferes with normal behavior. Simulations demonstrated that, in all cases, the model reproduced the interference effect as reported experimentally: movement variability increased when observing others. By investigating which of the movement parameters are inducing interferences, this work proposed several new experimental predictions. If tested, these predictions would help better understand the neural mechanisms underlying the multisensory representation of the self, as well as the influence of the representation of others.

This chapter concludes the description of the work contained in this thesis. Several neural models describing either specialized circuits or cortical networks involved in the brain processes of imitation were developed. Although these models were separately capable of reproducing neurophysiological and behavioral data, each modeling study was restricted to modeling only a subpart of the cortical network responsible for imitation. A unified and global view of the brain - in which the specific neural networks developed here may be embedded - remains to be provided. This synthesis will be described in the next chapter, which will also discuss the contributions of this thesis.

SYNTHESIS: TOWARD A UNIFIED CORTICAL MODEL OF IMITATION

This chapter begins with a rapid summary of the work accomplished in this thesis along with its principal contributions. A synthesis based on a high-level model of brain organization will then be presented in order to provide a general overview of the modeling studies described across this dissertation, as well as the cortical mechanisms and pathways of imitation in general.

9.1 CONTRIBUTIONS OF THIS THESIS

This thesis aimed at shedding some light on our understanding of several important neural mechanisms and cortical pathways related to human imitative behaviors by means of neural modeling studies. The key computational issues which drove this thesis concerned the neural mechanisms responsible for the way the visual representation of other individuals is changed into a frame of reference useful for imitation, and for the integration of this representation of others into the representation of one's own movements. How these processes influence human behaviors were discussed along the development of the neural models which can account for these cognitive abilities. In particular, this work sought: the cortical mechanisms and pathways which are responsible for the behavioral expression of the ideomotor principle; the existence of competitive interactions between different imitative strategies; and, finally, the ability to discriminate between self-generated movements and those produced by others.

This work started with a computational investigation of the cortical routes involved in automatic imitative behaviors. It mainly provided experimental predictions of two models which suggest, respectively, that imitative behaviors are either primarily mediated by a high-level cognitive decisional process or by a direct sensorimotor pathway. Since this study assumed that observed and self-generated movements are encoded within the same frame of reference, several biologically plausible neural mechanisms of frame of reference transformations were developed in order for this mapping to be possible. Two different mechanisms were proposed to direct the neural processes of sensorimotor transformations involving proprioceptive and visual modalities, respectively. Furthermore, in order to relate these transformations to human behavioral data, a neural model was developed to propose how spatial and anatomical imitative

strategies are encoded and interact. An experimental paradigm along with the predictions of the model were presented so as to validate or refute the working hypotheses of this modeling study. Next, since the brain has to integrate the information related to the movements of others into the representation of one's own movements, while simultaneously keeping track of the ownership of sensory information, a neural model capable of discriminating between these two sets of sensory information has been developed. By predicting the consequences of one's own movements, the competitive dynamics of the model can then select which sensory input is actually self-controlled. In order to describe the expression of this neural mechanism on human behaviors, a neural model capable of controlling and monitoring the execution of movements was tested against two experiments. These addressed the processes of multisensory integration: one case where biases were introduced on visual feedback, and one where movement observation and movement execution were simultaneous. This study also analyzed the possible components of others' movements which could be a source of interferences on one's own movements, and thus, it provides predictions which, if compared to real behavioral data, could give new insights into the neural mechanisms and cortical pathways mediating imitative behaviors.

Although all these models were capable of reproducing behavioral data as well as providing new experimental predictions, one will have to relate these models in order to provide a global picture of the cortical routes responsible for imitation. Thus, a synthetic model of brain organization is presented in this chapter. Importantly, the computational models developed in this thesis will be shown to correspond to fundamental building blocks of this model. In addition, this model will also highlight the numerous streams of information processing which are involved in imitation. Finally, based on these cortical pathways, the contributions of the developed models of human imitative behaviors on the deciphering of the cortical routes of imitation will be given.

9.2 TOWARD A UNIFIED CORTICAL MODEL

Along the development of the computational models presented in this thesis, a global model of brain architecture has emerged. Although this idea was not explored as much as it could have been, it nevertheless appeared to be an interesting model via which the various parts of this work could be unified. In addition, since each modeling study of this thesis focused on a precise topic, it is important to go one step further in order to provide a global description as to how they contribute to a better understanding of the neural mechanisms and cortical pathways involved in imitation and related behaviors.

The cortical model that is presented next is based on the brain as a highly distributed system, not only at the level of small cortical ensembles, but also at the cortical level. The core of the model consists of a complete representation of

the body encoded along a gradient of reference frames. Three instances of this representation are respectively associated with large portions of the parietal, premotor and temporal cortices. These cortical areas are assumed to process a representation of the body and its interactions with objects in multisensory, motor and visual terms, respectively. Importantly, by considering these three cortices simultaneously, the model naturally encompasses the neural pathways of imitation.

9.2.1 CORTICAL ORGANIZATION ALONG MULTIPLE FRAMES OF REFERENCE

The neural representation of one's own body is known to be widely distributed along multiple frames of reference (Johnson et al., 1996; Burnod et al., 1999; Battaglia-Mayer et al., 2003; Avillac et al., 2005). This property primarily facilitates the combination of multisensory and motor information that are intrinsically encoded within different reference frames. Multisensory integration processes usually result in coherent representations. Moreover, by acting on multiple coordinate frames such as body part-centered or goal-centered, the control over the body is simplified (Rizzolatti et al., 1990; Burnod et al., 1999; Battaglia-Mayer et al., 2003). With respect to this neurophysiological evidence, a theoretical model has already been proposed by Burnod et al. (1999), who proposed a model of the parieto-frontal brain network responsible for controlling reaching movements in which the neural representation of the body is distributed along a gradient of reference frames. The model, which is presented here, extends the above mentioned model by including abstract multisensory representations such as goal-directed ones, and areas of the temporal cortex involved in the visual recognition of bodies and objects. As will be shown later, these additions make it possible to consider the cortical areas and pathways related to imitation, such as those described in this thesis. But first, the next paragraphs describe how bodies and objects could be represented in multiple frames of reference within a multisensory and hierarchical organization.

A Hierarchical Representation of Frames of Reference in the Cortex

A simple kinematic chain representing one of the upper sides of the body is considered in order to show how to gradually merge visual and proprioceptive information together. It starts from the eyes and then continues: head, torso, upper arm, forearm, hand and fingers. As shown in Figure 9.1, the visual representation of the body parts can be displayed horizontally in a sequence of representations in an eye-centered frame of reference. Then, by proprioception, the angle formed by the eyes in their orbit with respect to the head and the distance between the head and the eyes is known. Thus, the representation of each body part can be transformed into a head-centered representation, which,

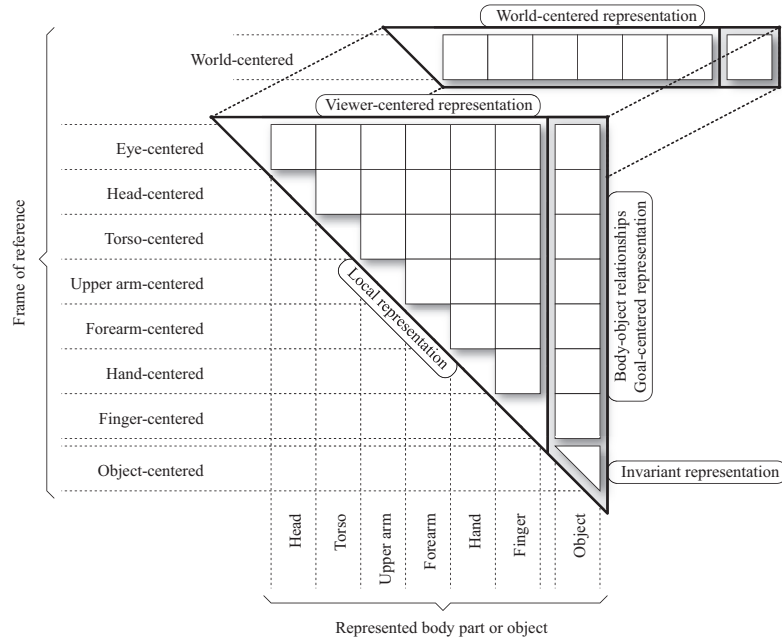


Figure 9.1: A hierarchical representation of frames of reference for representing the body and objects is shown. It encompasses both visual and proprioceptive representation, i.e., viewer-centered and local representations. Each block within the white triangle corresponds to the representation of a body part (horizontal axis) in a frame of reference centered on another body part (vertical axis). This white triangle can thus be associated the representation of the body. The gray extending area on the right of the body representation corresponds to the representation of an object in all possible frames of reference. Since the small area at the bottom of this gray column is an object representation in an object-centered reference frame, it corresponds to an invariant representation of objects. Finally, on top of the figure, in parallel to the eye-centered representation, a world-centered representation of the body in its environment can be found.

in turn, can be represented by the second row in Figure 9.1. Importantly, this top-down relationship is reversible. Furthermore, A complete hierarchical representation of the body within multiple frames of reference can be constructed following the same reasoning for each other body parts. As shown in Figure 9.1, this representation resembles a triangle where each side corresponds respectively to a purely viewer-centered representation of the body, a local representation of each body part with respect to its adjacent ones, and, finally, a representation of the end-effector in the frame of reference of all the other body-parts. From this structure, any information related to a body part can be transferred sequentially into any reference frame.

In addition, a complete description of an object in all the reference frames can be obtained by considering objects as extensions of the kinematic chain. Importantly, since the way objects are manipulated can be encoded, one may consider this extension of the body representation as a goal-centered representation. This model thus suggests that goal-centered representations naturally emerge from a gradual combination of proprioceptive and visual sensory modalities into a co-

herent and complete representation of the body, i.e., the body schema. Next, as indicated by the small triangle at the bottom of the object-related structure, an invariant representation of objects such as those found in the inferotemporal cortex (Booth & Rolls, 1998) can be produced. Indeed, in an object-centered frame of reference, the object representation is invariant. Moreover, since neural representations of the body in space also exist (Snyder et al., 1998; Redish, 1999), a world-centered representation can also be constructed. However, because the vestibular signals associated with this form of representation originate from sensors located in the head, this world-centered representation must be considered parallel to that in eye-centered coordinates.

Contributions of this Thesis

In order to process the information conveyed by each neural representation, transformations across frames of reference are necessary. Chapter 5 of this thesis presented plausible mechanisms for sensorimotor mappings between distinct frames of reference. This modeling study contributed to the understanding and to the modeling of large-scale cortical representations, by proposing models for transferring information between brain areas sharing similar information within different reference frames. In the next section, we will show how such a hierarchical model of frame of reference transformations can be embedded in an integrated and synthetic view of the brain organization to account for all the cortical pathways activated during imitative and related behaviors.

9.2.2 THE FRONTO-PARIETO-TEMPORAL MODEL

This section presents the fronto-parieto-temporal model, which represents the global organization of the cortex in terms of frames of reference. As shown in Figure 9.2, it is composed of three instances of the hierarchical structure described above. Each instance, or cortical schema, roughly incorporates the brain areas located within the parietal cortex, the premotor and motor cortices, and the temporal cortex. Respectively, they are assumed to be involved in the processing of the multisensory, motor and visual representation of the body and its possible relations with objects. In agreement with neurophysiological data, Figure 9.2 also shows how cortical areas are associated with these hierarchical structures. This segregation is not strict but smooth, i.e., a brain region can be mostly tuned to a given frame of reference, but a gradual sensitivity to the neighboring reference frames and body parts is also possible. Furthermore, the arrangement of the hierarchical structures reflects that of the cortex, where the axes of symmetry correspond to the central sulcus and the sylvian fissure. This visual representation allows a quasi-direct correspondence between the information pathways of this model and those of the cortex.

The principal connectivity patterns across brain areas are superimposed on

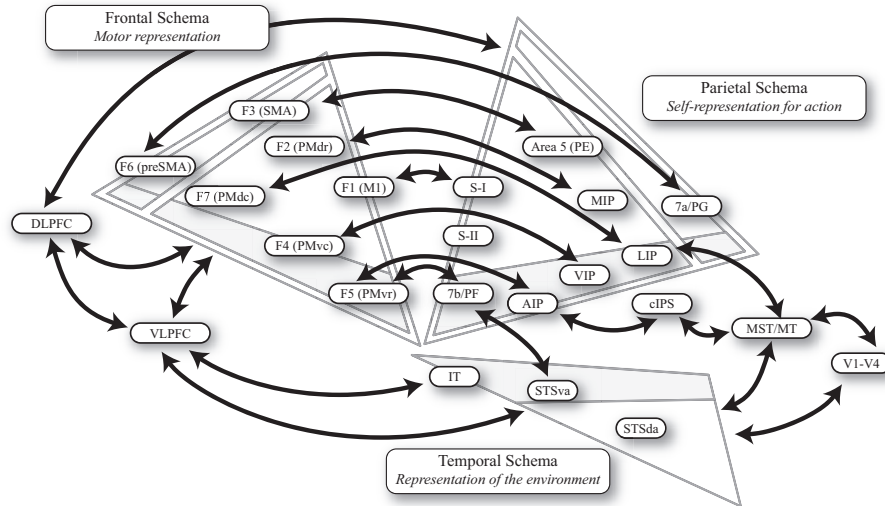


Figure 9.2: This figure presents the Parieto-Fronto-Temporal Model. It represents the relationships between the brain areas with respect to the frame of reference in which they encode information. The model is composed of three schemas. The **parietal schema** comprises the parietal areas responsible for a multisensory and distributed representation of the self. The **frontal schema** contains the premotor and motor regions in charge of the control of the body into multiple reference frames. Finally, **temporal schema** includes the complex visual recognition areas which represent the other individuals as well as the environment. The gray arrows across the schemas indicates known connectivity patterns between cortical regions.

this figure. In addition to the connectivity between neighboring areas - which for clarity were not drawn - the connectivity patterns across the parieto-frontal and parieto-temporal networks show links between areas which represent information of similar body parts in similar frames of reference. However, the connectivity across the parieto-temporal network is relatively weak. Indeed, little evidence, at least in monkeys, suggests the existence of important direct projections from the temporal cortex to the parietal cortex.

Several important brain areas were not assigned to a specific cortical schema of the model. These areas belong to two different categories. First, the primary visual areas which process only basic visual features cannot be associated with any of the cortical schemata. Secondly, since prefrontal areas are believed to process contextual information for decision-making and for the high-level control of sensorimotor streams, they embrace the whole structure rather than being part of it.

Contributions of this Thesis

In this model of cortical organization, each cortical region represents a specific body-part or object in a specific frame of reference. However, the cortex does not only represent static information. It also produces actions which can have consequences on this internal representation of the body and environment. Thus, neural processes in charge of producing movements, predicting

their consequences while maintaining a coherent internal representation have to be considered. Importantly, plausible neural models of these processes have been developed along the modeling studies presented in Chapters 7 and 8 of this thesis.

These modeled neural mechanisms can be associated with those that are required by each cortical schema of the fronto-parieto-temporal model. For instance, the neural structure described in Chapter 8, which is capable of generating goal-directed motion commands according to the state of the body, such as the current position of the arm, may correspond to a general mechanism by which the frontal schema produces motor commands. The neural structure described in Chapter 7, which ensures coherence within a multisensory representation of the body even in the presence of conflicting sensory signals, is a fundamental component of the neural processing occurring within the parietal schema. The frontal and parietal schemata are extensively interconnected. As described in the experiments of Chapter 8, this interconnectedness allows the frontal schema to have access to an accurate representation of the body, which is constantly kept up to date by the parietal schema by means of the motor efference copies received from the former. Finally, the neural mechanisms of sensory prediction described in Chapter 7 can also be associated with the function of the temporal schema. Indeed, although the temporal areas are usually thought of being involved only in the processing of visual information, it has also been shown that those can play a role in predicting the consequences of self body motion. To conclude, the neural mechanisms developed in Chapters 5, 7 and 8 of this thesis, form fundamental components of a global model of brain organization and operation.

9.2.3 THE CORTICAL ROUTES INVOLVED IN IMITATION

It must by now be clear that human imitative behavior results from the interactions across several important cortical pathways. In this section, we will draw those pathways onto our fronto-parieto-temporal model and highlight several of their fundamental characteristics and role in driving imitation in humans. Then, the contributions of the modeling studies of this thesis which directly address imitative behaviors will be also be described in light of this model.

Goal-directed Imitation

Let us first consider the cortical routes responsible for goal-directed imitation. The most important areas usually associated with this behavior belong to the monkey mirror system. These are: the parietal area 7b/PF included in the intra-parietal sulcus (IPS), the area F5 located in the ventral premotor cortex (PMv), and the ventral part of the superior temporal sulcus (STSva). Basically, these areas are thought to compute possible motor responses to objects or goals

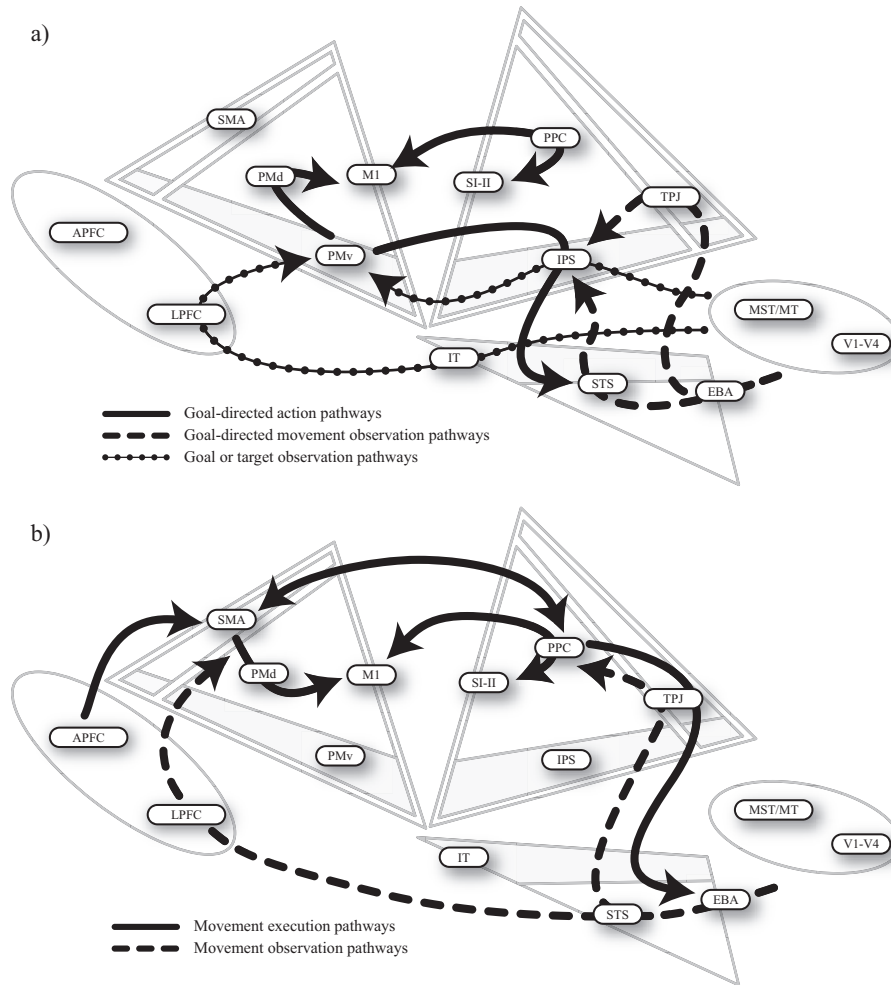


Figure 9.3: These figures illustrate the major cortical pathways which may be involved in **a)** goal-directed imitation, and **b)** the imitative of intransitive movements. Both figures were drawn according to the brain pathways described in Section 2.3.3 and shown in Figures 2.12 and 2.13. On this model, a clear cut between the areas involved in each imitative behavior can be observed.

and to associate the movements of others with these responses (Rizzolatti et al., 2001). As reported in Figure 9.2 and 9.3a, these three areas encode hand-object relationships in a goal-centered frame of reference. Moreover, because of their distinct roles, they are located on a different cortical schema. Interestingly, despite their distance on the cortical surface (Figure 2.5), this common property could explain the existence of strong reciprocal projections between them, projections which have led to the development of a goal-directed mirror system in both monkeys and humans (Fogassi & Gallese, 2002).

Goal-directed imitation pathways also involve two parallel streams. The first route flows along the parieto-prefrontal network, which represents the acting body and is purely sensorimotor (Johnson et al., 1996). The second route is a more cognitive pathway in the sense that reasoning centers are involved. This

stream flows along the temporo-frontal network and recruits both visual recognition areas such as the inferotemporal cortex (IT) and STS (Booth & Rolls, 1998; Jellema et al., 2004) and context-dependent decisional areas located within the lateral prefrontal cortex (LPFC) (Sakagami et al., 2006). While the former stream seems to be responsible for rapid and automatic imitative responses, the latter route is involved in controlling the overt production of these responses (Lhermite et al., 1986; Shimomura & Mori, 1998).

Imitation of Intransitive Movements

As illustrated in Figure 9.3b, the brain pathways mediating the imitation of intransitive actions recruit different cortical regions and routes than those in charge of goal-directed imitation. A clear distinction is evident between the brain areas involved in each imitative behavior. This separation shows that while goal-directed behaviors engage primarily goal-centered representations, intransitive actions recruit more body-centered ones. Moreover, just as the pathways of goal-directed imitation, the cortical routes mediating the imitation of intransitive movements are also separated into two major streams. While the dorsal route integrates the representation of others' movements into one's own sensorimotor representation of movements, the ventral route is instead concerned with the precise analysis of observed movements and with the control of overt movement execution.

CONTRIBUTIONS OF THE CORTICAL MODELS DEVELOPED IN THIS THESIS

By means of the fronto-parieto-temporal model, a clear separation between the routes mediating goal-directed and intransitive imitation has been put forward. However, because of the nature of cortical organization, which exhibits a gradient of regions sharing similar information encoded within similar frames of reference, these pathways should not be seen as distinct as they have often been considered in the literature. As will be shown next, the modeling studies of this thesis which addressed the cortical networks of imitation mostly focused on the possible links between these pathways.

Automatic Imitative Behaviors

In Chapter 4, the neural information pathways responsible for imitative behaviors in response to the observation of intransitive actions were investigated. In particular, an interference paradigm was used in order to produce conflicts between the cortical processes responsible, respectively, for controlling actions in response to non-biological visual cues, and for intransitive imitative behaviors. Figure 9.4 shows the cortical routes considered in this study. While the former process was suggested to flow along the goal-centered visuomotor pathway for action, the latter was thought to involve the imitative pathways for intransitive

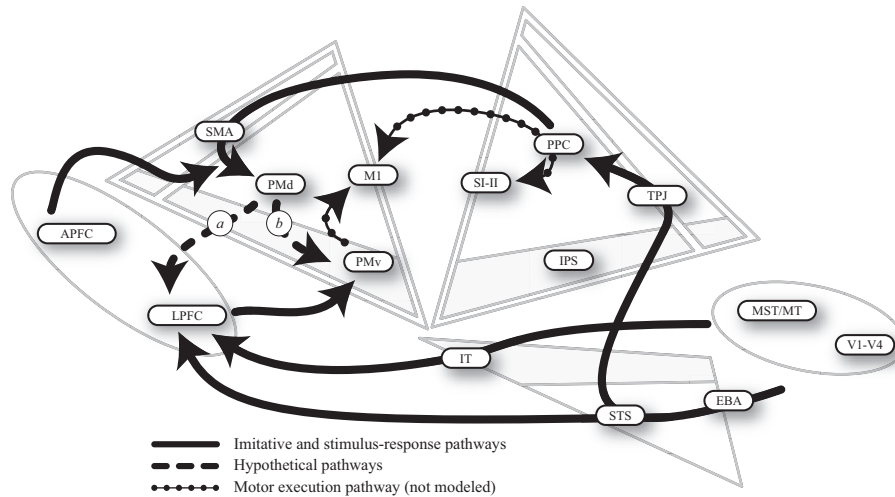


Figure 9.4: The cortical pathways modeled in the study described in Chapter 4 are shown on the proposed model of brain organization. This study asked for the existence of the cortical links shown by the labels (a) and (b) which might convey information related to automatic and intransitive imitative behaviors.

actions. This work primarily investigated the precise role of the dorsal sensorimotor route in facilitating motor execution. Does it influence the decisional process occurring in the lateral prefrontal cortex (LPFC), or does it influence directly the motor programs of the ventral premotor cortex (PMv)? As proposed by this study, one method to clarify this issue would be to perform the behavioral experiment developed in Section 4.3.2, which is suggested to increase the computational load in LPFC.

Imitative Strategies

The study presented in Chapter 6 investigated the interference effects that were reported between anatomical and spatial imitative strategies. In particular, the developed model suggested the existence of two distinct cortical circuits, one for each transformation, whose results are then merged within a selection network. These two circuits may be processed in parallel within the same regions: the posterior parietal cortex (PPC) and the supplementary motor area (SMA) (Koski et al., 2003).

However, despite this attractive and simple hypothesis, one may also envision an alternative, which suggests the existence of competitive interactions between the goal-directed and the intransitive imitative route. Indeed, goal-directed actions usually imply the respect of the spatial relationship between targets and the direction of movements. Thus, as illustrated in Figure 9.5, the corresponding pathway recruiting STS, IPS and PMv could be responsible for spatial imitation. In contrast, anatomical imitation may instead involve the network formed by EBA, STS and the dorsal part of the parietal cortex (PPC), i.e., the intransitive imitative route. To recall, PPC encompasses a faithful rep-

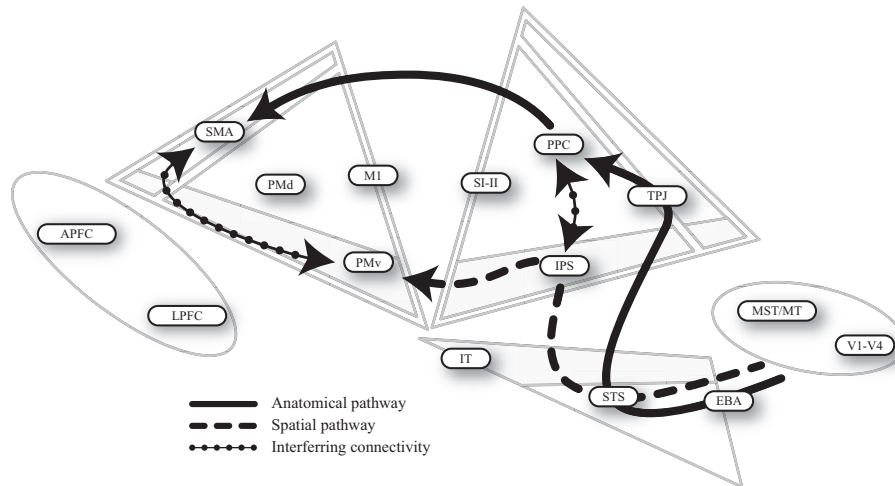


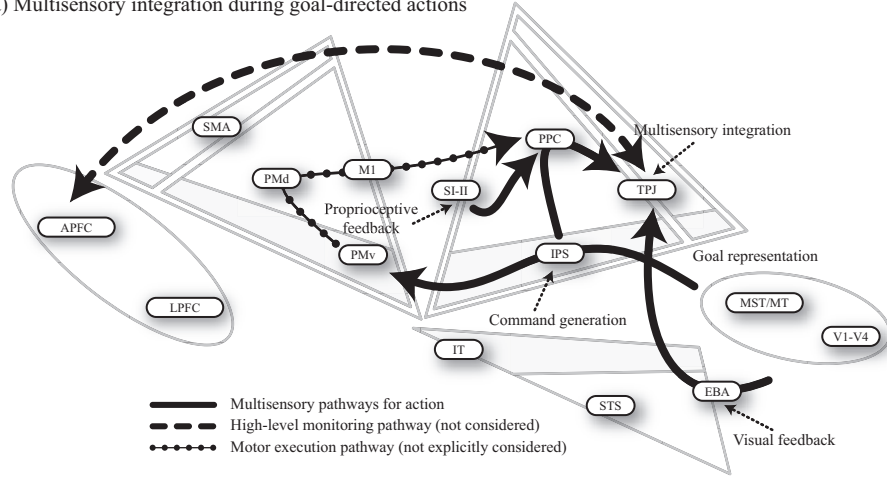
Figure 9.5: One hypothetical cortical pathway modeled in the study described in Chapter 6 is shown along the proposed model of brain organization. This hypothesis suggests that specular and anatomical imitation are grounded on the separate pathways for goal-directed and intransitive imitation. Cortical connectivity which could be at the origin of the behavioral interferences produced by these two strategies are also shown.

resentation of the body configuration. If this two-route hypothesis was true, interfering cortical pathways linking transitive and intransitive imitative pathways should exist in either the frontal cortex, the parietal cortex, or both. This may consequently explain the increased activation of PPC and SMA which was reported by an fMRI study under such conditions (Koski et al., 2003). Therefore, since this hypothesis and the one described in the previous paragraph are both plausible, further computational and experimental studies are needed in order to disentangle these two hypotheses.

Multisensory Integration and Discrimination

In Chapter 8, two experiments investigated the neural processes underlying the integration of multisensory signals within the representation of one's own body. The cortical pathways associated with the model of the first experiment, which focused on the human ability to detect discrepancies between visual and proprioceptive feedback, are provided in Figure 9.6a. Since the aim of the task was to reach for a target, the goal-directed route involving the intraparietal sulcus (IPS) and PMv is considered for the production of motor commands. Then, the efferent copies are sent back to the parietal area PPC and TPJ, which integrate them and predict their consequences within the body representation. In parallel, proprioceptive and visual feedback are processed and relayed by the somatosensory area (SI) and EBA, respectively. Finally, the integration of the latter signal is putatively gated by TPJ, which, in case of a large discrepancy, inhibits this input. In addition, TPJ may also inform the frontal cortex, which can, if necessary, modify the control strategy. The integration of multisensory

a) Multisensory integration during goal-directed actions



b) Multisensory discrimination during intransitive actions

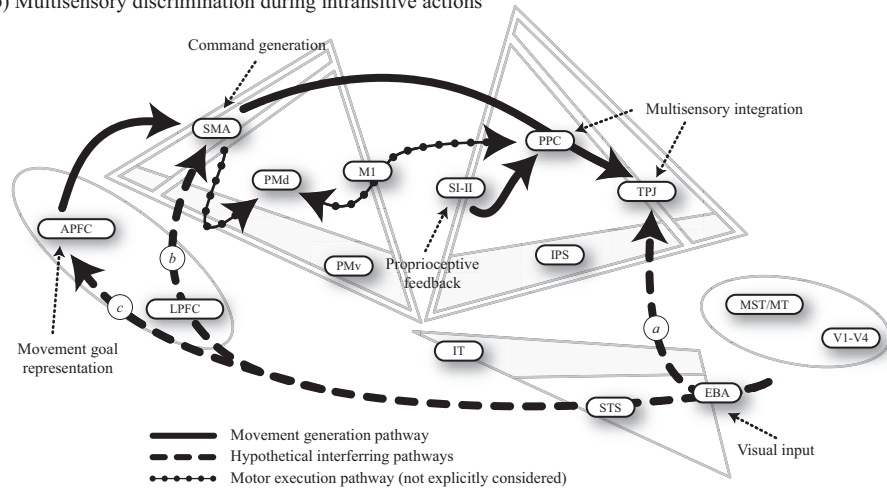


Figure 9.6: The modeling studies described in Chapter 8 used the same neural architecture in order to address the mechanisms of multisensory integration. They also allow the discrimination between self and others. **a)** It is suggested that the cortical pathways are responsible of multisensory integration in the experiment described in Section 8.3.1 are shown. In this figure, the functions assigned to the cortical areas correspond to those attributed to the network model given in Figure 8.6. **b)** The cortical pathways might be responsible for discriminating between visual inputs related to self and others are shown. In Section 8.3.2, several hypotheses concerning the components of observed movements which may be involved in imitative interferences were proposed. Their associated brain pathways are shown here by the labels (a)-(c) which correspond to those given in Figure 8.6.

signals is clearly supposed to be gated by TPJ, which is in line with current neurophysiological data (Blanke et al., 2002; Decety & Sommerville, 2003; Decety & Chaminade, 2003).

The second experiment focused on how the observation of movements performed by other individuals is integrated into the representation of one's own body. Complementarily, it also addressed the cognitive ability to discriminate between self-generated movements and those produced by others. As shown in

Figure 9.6b, since the task of this experiment concerned arm movements, the dorsal intransitive pathway is activated. This pathway includes TPJ and SMA, which are involved in the integration of multisensory information for monitoring the representation of the body, and in the generation of motor commands, respectively. Moreover, the prefrontal cortex is also recruited to decide which movement to execute.

Finally, this study also raised three hypotheses concerning how observed movements may influence one's own movements. The pathways associated with each of these hypotheses are depicted in Figure 9.6b. First, if the interferences are produced by the location of the demonstrator's hand, the activation of the antero-medial frontal cortex (AMFC), which represents movement goals, should be increased. Second, if the interferences are driven by the dynamics of observed movements, conflicts within SMA should appear. Third, if the visual representation of others' movements is integrated like that of self-generated movements, conflicts at the level of TPJ should be observed. Lastly, it is important to note that these three hypotheses are not mutually exclusive. Indeed, it seems more likely that all corresponding pathways influence behavior. As mentioned in Chapter 8, it would be interesting to measure the level of interference that each of these pathways has on imitative behaviors.

9.3 DISCUSSION

In this chapter, a global and integrated view of the cortical processes and pathways responsible for imitative behaviors has been presented. In particular, the described model suggests that a distributed representation of the body across different frames of reference may be a key characteristic of brain organization. This property may have significant consequences on our understanding of the imitation mechanisms.

The core of the model consists of a complete and coherent representation of the body, encoded along a gradient of reference frames. Interestingly, goal-centered representations emerge naturally from the gradual combination of the information conveyed by both proprioceptive and visual sensory modalities. Furthermore, three interconnected instances of such a complete representation were associated with large portions of the parietal, premotor and temporal cortices, where each instance is respectively encoding the multisensory body schema and its possible interactions with objects, its control of movement, and the visual processing of the environment. By integrating all the necessary cortical areas thought to be in charge of the sensorimotor control of the body, parallels were drawn between this model and the cortical processes and pathways involved in imitation.

The cortical routes commonly suggested to mediate goal-directed and intransitive imitation strategies were shown to recruit different networks of brain areas.

While goal-directed imitation primarily involves cortical regions representing goal-centered information, intransitive imitation involves more body-centered ones. Then, in showing which pathways were addressed by the modeling studies of this thesis, the existence of links between the cortical areas involved in goal-directed and intransitive imitation have been put forward. These links clearly suggest that these imitative behaviors are not as distinctively processed in the brain, as is commonly believed.

Several supplementary arguments to this proposal can also be given. Firstly, sensorimotor processing has clearly been shown to recruit large portions of complementary brain areas along the fronto-parietal cortex. Since this cognitive process and imitation are tightly linked, imitation may also involve this complete network. Secondly, the intrinsic nature of cortical organization, highlighted by the fronto-parieto-temporal model, is endowed with many interconnected regions representing information which is transferred along a gradient of frames of reference, such as from body-centered representations to goal-centered representations and vice versa. This intrinsic organization makes obvious the existence of links between goal-directed and intransitive imitative pathways.

Therefore, this involvement of the whole sensorimotor brain in the processing of imitative behaviors may explain why it has been so difficult to identify only a small set of brain regions which precisely mediate imitation. This hypothesis may, however, raise a question related to the behavioral differences reported between humans and monkeys. Indeed, similarly to humans, monkeys have a highly developed cortical circuitry for sensorimotor processing, and also have strong abilities to recognize the actions of their conspecifics. Why, then, can't they display as much imitative skill as humans do?

Monkeys Vs. Humans

Following from the description of the cortical pathways which mediate imitative behaviors, a hypothesis related to the cortical differences between humans and monkeys can be put forth. Monkeys may lack the connectivity from STS to the dorsal part of the parietal cortex, potentially due to the lack of a homologous region to the human extra-striate body area (EBA). Indeed, EBA has been suggested to represent body parts irrespective of the parts' owner, whether they belong to the self or to others (Ruby & Decety, 2001; Astafiev et al., 2004; Chan et al., 2004). Therefore, a direct visuomotor mapping between the representation of others' movements with one's own internal representation seems to be a key ingredient of human ability to imitate intransitive actions. In line with the results of ethological studies (Tomasello et al., 1993; Whiten et al., 1991), this hypothesis suggests that the cortical network responsible for human imitative abilities extends that found in its ancestor, in particular by means of additional connectivity patterns between temporal and parietal cortices. Therefore, the putative lack of an intentional system for imitation in monkeys (Rizzolatti & Luppino, 2001) should not be considered as the only reason by which they are

not capable of true imitation. Monkeys may also lack cortical areas such as EBA, as well as temporal-to-parietal connections between the regions processing information in a body-centered frame of reference.

Summary

In this chapter, a synthetic model of brain organization has been proposed, namely, the fronto-parieto-temporal model. This model has laid the groundwork for several important points related to the modeling studies accomplished throughout this thesis. First of all, the model shows that the neural structures described in Chapters 5, 7, and 8 provide keystones for the general understanding of both the organization of and the functioning of the cortex. This model also provides a global description of the neural pathways responsible for goal-directed and intransitive imitative behaviors in terms of the frames of reference recruited by both strategies. Then, although goal-directed and intransitive movement pathways are usually considered distinct, the modeling studies described in Chapters 4 and 6 proposed that cortical projections between these routes do exist. Furthermore, the results presented in Chapter 8 suggest that the processes of multisensory integration and discrimination for either goal-directed or intransitive imitation seem to be mediated by the same set of parietal areas which represent body information in visual terms. Finally, this chapter also provides a global and integrated view on the cortical mechanisms and pathways involved in imitative behaviors. The strong proposal raised is that imitation should not be considered as recruiting specific brain areas, but rather the whole sensorimotor cortex.

The next chapter provides a discussion related to the general limitations of the approach followed in this thesis. It also develops future research directions which could complement this work.

LIMITATIONS AND FUTURE WORK

WITH respect to the work accomplished throughout this thesis, several important limitations can be brought to light along with their possible improvements. They are described next.

GRANULARITY OF THE MODELING APPROACH

In this thesis, the dynamic neural field approach has been adopted in order to model the neural mechanisms underlying imitation. An important advantage of this methodology is that it allows an interesting trade-off between realistic models of neurons and purely connectionist models. Indeed, while connectionist models may not capture some of the important dynamical properties of neural ensembles, the use of models of spiking neurons would certainly have put too much complexity in the models, and thus, would have driven this thesis too far away from its central objective.

LEARNING ISSUES

Although this approach has successfully allowed the modeling of several cortical pathways responsible for imitative behaviors, the important issue concerning the processes of learning neural representations and inter-cortical connectivity have not been tackled in this thesis. The neural structures and connectivity patterns were always assumed fixed and known in advance. Nevertheless, this lack is believed to not have a strong incidence on the predictions provided by the presented modeling studies. Indeed, although imitation may be seen as mandatorily requiring learning, the aspects of imitative behaviors addressed here do not involve this cognitive skill. In particular, the focus was primarily on automatic imitative behaviors which recruit cortical networks that are already structured and functioning. If the aim of this work would have been to model either the developmental stages of imitation in children, or the processes of learning by imitation, learning mechanisms would certainly have been central to this thesis.

Still, this issue is fundamental, which explain the huge interest in learning mechanisms from the computational neuroscience community. Indeed, compu-

tational studies endowing neural fields with learning abilities have already been described in the literature. For instance, it has been shown that the recurrent connectivity patterns of neural fields could be built using learning rules based on the Hebb's rule (Zhang, 1996; Compte et al., 2000; Stringer et al., 2004). In addition, both symmetric and asymmetric profiles of the synaptic weights could also be learned, based on techniques derived from the principle of spike timing dependent plasticity (Roberts, 1999; Melamed et al., 2004).

Another important reason why learning should be considered in future work, is that it could allow the number of parameters to be extended, which can be simultaneously encoded within a neural field without increasing its dimension. For instance, Meñard and Frezza-Buet (2005), by means of a mixture of a Kohonen-based algorithm with a neural field, were able to show that associations between various high-dimensional sensory inputs and motor outputs can be learned. However, this model is not capable of displaying the computational properties that the studies developed in Chapters 7 and 8 have shown to be extremely useful for multisensory integration as well as for controlling movements. Therefore, one may suggest that both models be integrated together, which may be expected to increase the computational power of this class of models.

DEALING WITH COMPLEXITY

The modeling studies of this thesis that addressed the cortical networks involved in imitation, considered relatively few internal and behavioral variables. For instance, only mono-articular movements were used in order to keep the models simple and tractable. In addition, these variables were also first restricted to positional information, and then extended to velocity. This form of encoding, by its attractive simplicity, has often been reported in the neurophysiological and modeling literature, but a more complete system should nevertheless take into account supplementary variables such as force and acceleration, as well as supplementary body parts and articulations. Similarly, more complex representations of visual stimuli should be considered, and in particular, those that encode biological motions. Therefore, in order to consider these visuomotor aspects of cortical processing, future work should attempt to model how visual and motor representations can simultaneously be learned and self-organized within compact neural ensembles.

EMBODIMENT

Another limitation of this work is that the modeled systems are not embodied. This was not made possible as control of complex body dynamics was not considered together with the explicit recognition of biological motions. As a simplification, system variables were calculated in a computer-generated en-

vironment simulating only body kinematics. In addition, embodiment requires a body, which is often materialized by means of a robotic platform. And because of the computational resources required to process the dynamics of high-dimensional neural fields, real-time control of a complex robotic platform - such as a humanoid robot - is hardly realistic. Nevertheless, the development of hardware devices taking advantage of the distributed nature of these networks might help endowing robots with bio-inspired cognitive abilities (Hahnloser, Sarpeshkar, Mahowald, Douglas, & Seung, 2000).

TOWARD A LARGE-SCALE CORTICAL MODEL

Another important limitation concerns the well-known existence of several parallel processing loops in the brain. This thesis primarily considered feed-forward processing pathways when addressing the cortical networks mediating imitation. In order to resolve this issue, one may consider implementing the fronto-parieto-temporal model described in Section 9.2.2 which is distributed and recurrent by nature. In addition, this framework has been suggested to incorporate an important set of the cortical pathways responsible for a large palette of imitative behaviors. Furthermore, implementing this large scale model also implies considering more system variables such as those needed to represent complete kinematic chains. Therefore, additional imitative behaviors than those described in this thesis could be modeled. For instance, the goal-directed nature of imitation (Wohlschläger et al., 2003) could be investigated in more detail.

Furthermore, the fronto-parieto-temporal model also described three distinct complete representation of bodies and their relation with the environment. Importantly, it proposed a visual to motor gradient of neural representations flowing from the temporal lobe, through the parietal one, up to the frontal cortex. From this, one may imagine this gradient being an evolutionary solution to the curse of dimensionality, by which the brain reduces the amount of information that each cortical region should process simultaneously, while keeping a coherent representation of the body in multiple and complimentary modalities. This remains, however, to be investigated more extensively.

Finally, as reported in Section 2.3.4, an hemispheric specialization could also be considered, where the left hemisphere would be more prone to process-shared representations, whereas the right hemisphere would contain more information focusing on self-representation (Barresi & Moore, 1996; Decety & Chaminade, 2003). Therefore, the development of a bi-hemispheric cortical model could also provide more insights as to how one represents the self and the others, and by extension, on the mechanisms of imitation. Indeed, shared representations are only part of the complex neural processes mediating imitative behaviors, which also recruit the sensorimotor cortex in its whole. In order not to confound oneself with the others, it seems natural to have cortical regions restricted to one's own representation, whereas shared neural structures would be required

to relate oneself with the others, or to represent others' movements within one's own sensorimotor repertoire.

DEVELOPMENTAL AND BIO-INSPIRED ENGINEERING APPROACHES

Building such a large-scale system could seem huge and difficult to handle. Because of the large amount of neural representations and their connectivity patterns, the complexity of the model can grow very fast. Consequently, the system may become difficult to analyze. In addition, if learning abilities are taken into account, its analysis may become even harder. Several solutions to these problems may nevertheless be proposed. First, a developmental approach, in which the size of the model could be progressively increased, may allow to better track and understand its dynamics.

Furthermore, another approach might be to adopt a bio-inspired engineering-based methodology. Importantly, this approach may be a good compromise to understand the principles of cognitive processing while reducing the computational cost and the complexity that the neural field approach produces. For example, bio-inspired engineering techniques have already been applied in order to learn the kinematic chain of a robotic manipulator, i.e., its body schema, across distributed and local representations of the body joints (Hersch, Sauser, & Billard, 2007). In addition, in tasks where one has to imitate novel actions, it appeared that taking advantage of information encoded within multiple reference frames is useful and even mandatory to allow an algorithm to reproduce accurately actions involving the manipulation of several objects (Calinon et al., 2007). Therefore, rather than only focusing on a particular approach to understand how cognitive processes may lead to imitative abilities, one should instead adopt a multidisciplinary philosophy for research. In such research, each sub-field should participate, at its own level and provide coherent answers to the important questions related to this field.

A MULTIDISCIPLINARY FIELD OF RESEARCH

In relation to this proposal, the last point which is worth mentioning is the absence of real experimentations which could either confirm or refute the hypotheses raised throughout this thesis. Indeed, the approach which was adopted here consisted first in developing neural models capable of replicating current data related to the behavioral expression of imitative behaviors, and then in drawing new hypotheses in order to propose new research directions. However, as long as actual experiments are not performed on human subjects, it is hard to proceed with hypotheses previously raised which were not validated. Although this absence of results on the actual validity of the modeling hypotheses may seem somewhat odd to the reader, the stance that was taken here was to

propose model-based predictions and call interested people from the scientific community to make the experiments and verify their validity. Indeed, model predictions are often presented along with their experimental counterparts in the scientific literature. However, this can cast some doubt on the methodology, predictions or experimental results, which ever come first, are then difficult to assess. Experimentalists and modelers basically follow the same ultimate goal, but use different approaches. Thus by sharing the competencies of specialists from both fields of research some of the bias always present may be removed. In conclusion, one regret is that a collaborative experience with experimentalists, despite having certainly been a long-drawn-out job, might have been very constructive and fruitful.

CONCLUSION

IN this thesis, models of several neural mechanisms and cortical pathways involved in imitative behaviors were developed. The major issues tackled here included the brain mechanisms responsible for: *a)* ideomotor behaviors, which are the overt expression of the influence of others' actions on the quality of one's own performance; *b)* visuomotor transformations, which allow one to map the movements of others to a coordinate system centered on one's own body; and *c)* multisensory integration, prediction and discrimination, which are important for disambiguating the visual signals related to either the self or others, and consequently, for distinguishing oneself from others. Along these lines of research, the contributions of this thesis fall into two major categories.

First of all, this thesis contributes to a better understanding of the neural mechanisms underlying human imitative behaviors. This work questioned the hypothesis that sensorimotor pathways are directly responsible for automatic imitative behaviors, by providing an alternative which puts forward the involvement of decisional circuits located in the frontal cortex. In order to remove all doubt, a novel experimental paradigm has been proposed along with model predictions. Next, this work proposed that the neural processes of spatial and anatomical imitation compete for the control of imitative movements. The results of this study showed that the human preference for spatial imitation may be caused by the higher computational load required by the anatomical imitation, and by the competitive interactions present between the two strategies. Model predictions were provided for future verification of this hypothesis. This work also developed a model of the mechanisms responsible for the integration of the representation of others' movements into one's own representation of movements. Importantly, this study went beyond current behavioral studies as it provided experimental predictions which could help elucidate how and where in the brain the two representations interact and interfere with each other. Finally, in contrast to the neural models of imitation reported in the literature, which usually focus only on small cortical networks, a large-scale model of imitation - which encompasses the whole sensorimotor brain - has been proposed. This model shows that the cortical pathways which mediate goal-directed and intransitive imitation may not be as distinct as commonly believed. According to this model, the behavioral differences reported between humans and mon-

keys were suggested to be attributed to the existence, in humans, of stronger connectivity patterns between the cortical areas respectively responsible for the visual recognition of biological motion and for the representation of the self.

Secondly, this thesis contributes to the field of computational neuroscience by providing new developments concerning the class of artificial neural networks known as the neural fields. While investigating the possible mechanisms of frame of reference transformations, in contrast to other models, this work paid special attention to the amplitude of the signal conveyed by neural populations, and then showed how errors arising from transferring this information could be reduced and avoided. This work, inspired by related modeling studies, also proposed a neural structure capable of modifying its internal dynamics to update its internal representation by means of external commands. Importantly, since this model was shown to exhibit preferential tuning to external inputs whose dynamics are coherent with its own, this neural model has also been shown to be capable of discriminating sensory inputs resulting from self-generated actions, and thus capable of enforcing the coherence of multisensory representations. Finally, the generality of these neural mechanisms devoted to the transformation and integration of information suggests that these processes could be widespread across the whole cerebral cortex.

To conclude, one hopes that the computational models and the hypotheses raised throughout this thesis will inspire further research, and will consequently help the scientific community to understand better the cognitive processes underlying the human ability to imitate.

TECHNICAL AND IMPLEMENTATION DETAILS

IN this appendix, several additional useful information can be found. In the first section, implementation details as to how quasi uniform distributions of preferred directions could be obtained on a spherical surface are given. This method mainly concerns the simulation results presented in Chapters 5, 6 and 8, where neural fields defined on a spherical surface were used as modeling tools. In the second section, the detailed mathematical developments of the modeling study presented in Chapter 7 are provided.

A.1 SPHERICAL REPRESENTATION

In order to simulate the neural fields in which the neural sensitivity, i.e., the parameter space Γ , is defined on a spherical surface, a quasi uniform distribution of N preferred directions, where N corresponds to the number of neurons within a population, has to be generated. An iterative algorithm, inspired by the mechanism governing the **self organizing maps** (Kohonen, 1990) was developed. Indeed, in contrast to torus-like manifolds, where a regular distribution of preferred directions is straightforward to produce, this is not the case for an uniformly distributed set of unitary vector on a sphere, except for a finite set of numbers of points, that correspond to the corners of the regular polyhedra.

For a population constituted of N neurons, during an initialization phase, N vectors \vec{r}_i , $i \in \{1..N\}$ on the unit sphere are randomly generated by choosing two random values $\lambda_1 \in [0, 2\pi[$ and $\lambda_2 \in [-1, 1[$, such that

$$\vec{r}_i = \begin{pmatrix} \sqrt{1 - (\lambda_2)^2} \cos(\lambda_1) \\ \sqrt{1 - (\lambda_2)^2} \sin(\lambda_1) \\ \lambda_2 \end{pmatrix}. \quad (\text{A.1})$$

This vector generation method guarantees a statistical uniform distribution on the unit sphere that does not lead to a concentration of points at the poles (Marsaglia, 1972). An iterative processes is then followed. At each time step, using the same technique as in Equation (A.1), a training random input vector

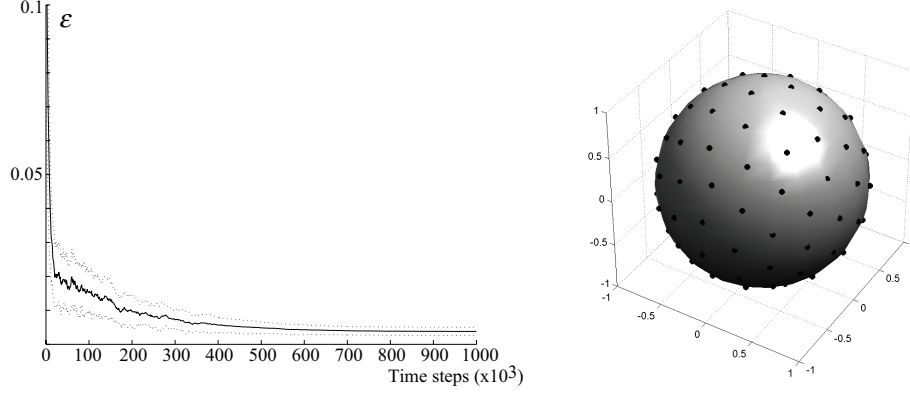


Figure A.1: Results produced by the iterative algorithm that generates a quasi uniform distribution of points on a sphere. (left) Evolution of the mean (filled line) and standard error (dotted line) of uniformity error ϵ along training time for the generation of 100 populations of 100 neurons. (right) 100 points distributed on a sphere using this algorithm.

\vec{p} is generated followed by an update of each \vec{r}_i such that

$$\vec{r}_i(t+1) = \begin{cases} \vec{r}_i(t) + \delta(t)(\vec{p} - \vec{r}_i(t)) & \text{if } i = i^* \\ \vec{r}_i(t) & \text{if } i \neq i^* \end{cases} \quad (\text{A.2})$$

where $\delta(t) \in]0, 1]$, a learning rate, decreases exponentially over training time, and i^* corresponds to the index of the closest preferred direction to \vec{p} such that

$$\forall i, \quad \vec{r}_{i^*}(t) \cdot \vec{p} \geq \vec{r}_i(t) \cdot \vec{p}.$$

This mechanism lets the set $\{\vec{r}_i\}$ of preferred directions converge toward an uniform distribution of its input space (Kohonen, 1990), which, by Equation (A.1), is uniform over the unit sphere. To quantify the uniformity of the resulting distribution, the error $\epsilon = \|\frac{1}{N} \sum_i \vec{r}_i\|$ is defined. Figure A.1 (left) shows the evolution of ϵ for several trials, while Figure A.1 (right) shows an example of a resulting set of preferred directions.

A.2 A COUPLED ATTRACTOR MODEL

In Chapter 7 a neural field model capable of integrating motion commands in order to update its internal representation has been presented. However, few mathematical details were provided to lighten the text. They are described in the following section.

As introduced in Chapter 7, the following continuous attractor network dynamics is considered

$$\begin{aligned} \tau \dot{u}(\vec{r}_A, \vec{r}_B, t) = & -u(\vec{r}_A, \vec{r}_B, t) + h(\vec{r}_B, t) + x(\vec{r}_A, \vec{r}_B, t) + \\ & \oint [W^R(\vec{r}'_A, \vec{r}_A) - \lambda \nabla W^R(\vec{r}'_A, \vec{r}_A) \cdot \vec{r}'_B] \\ & f(u(\vec{r}'_A, \vec{r}'_B, t)) d\vec{r}'_A d\vec{r}'_B \end{aligned} \quad (\text{A.3})$$

where $u(\vec{r}_A, \vec{r}_B, t)$ corresponds to the membrane potential of a neuron with time constant τ , which is preferentially tuned to \vec{r}_A , a variable in stimulus space Γ^A and to $\vec{r}_B \in \Gamma^B$. As definition spaces, ring and torus spaces were considered (see Table 7.1). $f(u)$ is the activation function chosen to be the linear threshold function: $\max(0, u)$. W^R is a center-surround, symmetric and Gaussian-like recurrent weights kernel, $\lambda > 0$ a scaling factor, and ∇ corresponds to the gradient operation along \vec{r}_A . As will be shown further in the text, the main effect of the second term of the recurrent connectivity is to make the neuron's sensitivity to the variable \vec{r}_B implicitly correspond to a preferred movement direction along the other variable \vec{r}_A . The external inputs are decomposed into two parts: the background input $h(\vec{r}_B, t)$ and the stimulus input $x(\vec{r}_A, \vec{r}_B, t)$.

In the following, the dynamics of the interactions between the network and its external inputs is considered. First of all, the case where the background input alone is sufficient to drive the network marginal attractor states to become marginal linear trajectories in the space Γ^A will be presented. Afterwards, a stimulus input will be defined so that it can drive the network toward its own motion dynamics. A non-linear form of it will first be provided, since its derivation from the background input form is the most straightforward. Then, a linear form of it will be derived. Importantly, the latter is more suitable for transferring information across neural populations, and therefore, this form will help derive adequate synaptic projections. Consequently, within a larger network of neural populations, instances of the model could transfer their information from one to another.

Before starting, it is worth mentioning that, since such network dynamics can only be described fully analytically for certain rare special cases (Amari, 1977; Xie et al., 2002; Zhang, 1996), linear approximations around equilibrium points will be considered. Importantly, general stable solutions are assumed to exist. Indeed, the nature of the marginally stable solutions of such systems, known as activity bumps, has already been long described in the literature (Wilson

& Cowan, 1973; Amari, 1977; Zhang, 1996; Salinas & Thier, 2000; Xie et al., 2002).

STATIC SOLUTIONS

First of all, static solutions are considered, which can be obtained when the background input is balanced, and when no spatial input is applied to the network, i.e., respectively, $h(\vec{r}_B, t) = h_0$ and $x(\vec{r}_A, \vec{r}_B, t) = 0$. Omitting the system variables, a marginally stable and static solution given by u^* and $f^* = f(u^*)$ are considered without loss of generality. In this case, since the system is constant along \vec{r}_B , the second term of the recurrent connectivity vanishes through the closed integral, which, by rewriting Equ. (A.3), gives

$$u^* - \oint W^R * f^* d\vec{r}'_B = h_0 \quad (\text{A.4})$$

where $*$ denotes the convolution operation along \vec{r}_A . Then, in order to extract the homogeneous term h_0 from the solutions, a normalization of these solutions is defined by the following substitution

$$u^* = h_0 U_0^* \quad \text{and} \quad f^* = h_0 F_0^* \quad (\text{A.5})$$

where the $U_0^*(\vec{r})$ and $F_0^*(\vec{r})$ are solutions only defined on \vec{r}_A . Further, rewriting Equ. (A.4) gives

$$U_0^* - 1 = \oint W^R * F_0^* d\vec{r}'_B \quad (\text{A.6})$$

From this, the multiplicative effect of the background homogeneous input h_0 on the solutions of the system can be noticed. This property is at the basis of the non-linear behavior of the neural fields in general (Salinas and Thier (2000) and Chapter 5 of this thesis). This implies that, since h_0 only acts as a scaling factor, the following analysis can be restricted, without loss of generality, by considering only normalized solutions of the system.

A.2.1 VELOCITY INTEGRATION

This section now describes how the background input can drive a stable field activity packet toward a similar but traveling bump with velocity \vec{v}^* along the space defined on \vec{r}_A . A usual transformation performed in such a situation is to focus on a frame of reference moving with the bump so that, in this new frame, it appears static (Xie et al., 2002; Zhang, 1996). According to the following variable substitution

$$\tilde{\vec{r}}_A = \vec{r}_A - \int_0^t \vec{v}^*(t') dt' \quad (\text{A.7})$$

gives, $\dot{u}(\vec{r}_A, \vec{r}_B, t) = -\vec{v}^*(t) \cdot \nabla u(\vec{r}_A, \vec{r}_B, t) + \dot{u}(\vec{r}_A, \vec{r}_B, t)$, where ∇ is the gradient operator along \vec{r}_A . This leads Equ. (A.3) to become

$$\begin{aligned} -\tau \vec{v}^*(t) \cdot \nabla u(\vec{r}_A, \vec{r}_B, t) \\ + \tau \dot{u}(\vec{r}_A, \vec{r}_B, t) = & -u(\vec{r}_A, \vec{r}_B, t) + h(\vec{r}_B, t) + x(\vec{r}_A, \vec{r}_B, t) + \\ & \oint [W^R(\vec{r}'_A, \vec{r}_A) - \lambda \nabla W^R(\vec{r}'_A, \vec{r}_A) \cdot \vec{r}'_B] \\ & f(u(\vec{r}'_A, \vec{r}'_B, t)) d\vec{r}'_A d\vec{r}'_B \end{aligned} \quad (\text{A.8})$$

As mentioned earlier in the static case, a constant input $h(\vec{r}_B, t) = h_0$ leads the system to be constant along \vec{r}_B , which suppresses the second term of the recurrent connectivity. Moreover, as can be guessed, this term will be responsible for compensating the new term on the left hand-side of Equ. (A.8). Therefore, this symmetry needs to be broken, by adding an asymmetric term $\vec{h}_0^d(t) \cdot \vec{r}_B$ to the driving background input such that

$$h(\vec{r}_B, t) = h_0 [1 + \vec{h}_0^d(t) \cdot \vec{r}_B] \quad (\text{A.9})$$

where \vec{h}_0^d corresponds to the strength of the asymmetry breaking relative to the constant term h_0 . If, for a while, the terms containing a gradient expression in Equ. (A.8) are omitted, an approximate solution \tilde{u}^* defined along both \vec{r}_A and \vec{r}_B can be found. It is equivalent to that given in Equ. (A.5) up to a constant given by the asymmetric term in Equ. (A.9). In a compact form, the solution of the system can be written as

$$\tilde{u}^*(\vec{r}_A, \vec{r}_B, t) = h_0 [\vec{h}_0^d(t) \cdot \vec{r}_B + U_0^*(\vec{r}_A)] \quad (\text{A.10a})$$

where $U_0^*(\vec{r}_A)$ is given by Equ. (A.6). Then, from Equ. (A.5), where the output function f^* was proportional to the background homogenous input, an approximate \tilde{f}^* is also assumed to be obtainable by such a non-linear transformation. By means of a linear approximation around the equilibrium point, i.e., the static case,

$$\tilde{f}^*(\vec{r}_A, \vec{r}_B, t) = h_0 [1 + \gamma_0 \vec{h}_0^d(t) \cdot \vec{r}_B] F_0^*(\vec{r}_A) \quad (\text{A.10b})$$

where γ_0 corresponds to the linear approximation factor. It corresponds to the slope of the relation between the asymmetry of the background input and the velocity of the network response. It strictly depends on the recurrent weights profile, whose convolution is hard to solve analytically (Zhang, 1996).

Now, coming back to the full form of Equ. (A.8), the aim is here to show how the background input should be set in order for the network representation to exhibit the desired traveling activity profile. By first considering the symmetric

part of the recurrent synaptic drive, it gives

$$\begin{aligned}
\oint W^R * \tilde{f}^* d\vec{r}'_{\mathcal{B}} &\stackrel{(A.10b)}{=} h_0 \oint W^R * [1 + \gamma_0 \vec{h}_0^d \cdot \vec{r}'_{\mathcal{B}}] F_0^* d\vec{r}'_{\mathcal{B}} \\
&= h_0 \left[\oint W^R * F_0^* d\vec{r}'_{\mathcal{B}} + \gamma_0 \oint W^R * F_0^* [\vec{h}_0^d \cdot \vec{r}'_{\mathcal{B}}] d\vec{r}'_{\mathcal{B}} \right] \\
&\stackrel{(A.6)}{=} h_0 [(U_0^* - 1) + 0] = h_0 (U_0^* - 1)
\end{aligned}$$

Similarly, by considering its asymmetric component, it gives

$$\begin{aligned}
-\oint \lambda \nabla W^R \cdot \vec{r}'_{\mathcal{B}} * \tilde{f}^* d\vec{r}'_{\mathcal{B}} &\stackrel{(A.10b)}{=} -h_0 \lambda \oint \nabla W^R \cdot \vec{r}'_{\mathcal{B}} * [1 + \gamma_0 \vec{h}_0^d \cdot \vec{r}'_{\mathcal{B}}] F_0^* d\vec{r}'_{\mathcal{B}} \\
&= -\lambda h_0 \left[\oint \nabla W^R \cdot \vec{r}'_{\mathcal{B}} * F_0^* d\vec{r}'_{\mathcal{B}} + \right. \\
&\quad \left. \gamma_0 \oint \nabla W^R \cdot \vec{r}'_{\mathcal{B}} * F_0^* [\vec{h}_0^d \cdot \vec{r}'_{\mathcal{B}}] d\vec{r}'_{\mathcal{B}} \right] \\
&= -\lambda h_0 \left[0 + \gamma_0 \vec{h}_0^d \cdot \nabla [W^R * F_0^*] \right] \\
&\stackrel{(A.6)}{=} -\lambda \gamma_0 h_0 \vec{h}_0^d \cdot \nabla (U_0^* - 1) \\
&= -\lambda \gamma_0 h_0 \vec{h}_0^d \cdot \nabla U_0^*
\end{aligned}$$

And finally, the complete recurrent synaptic drive is given by

$$\oint [W^R - \lambda \nabla W^R \cdot \vec{r}'_{\mathcal{B}}] * \tilde{f}^* d\vec{r}'_{\mathcal{B}} = h_0 [U_0^* - 1 - \lambda \gamma_0 \vec{h}_0^d \cdot \nabla U_0^*] \quad (A.11)$$

Then, by substituting Eqs. (A.9), (A.10a) and (A.11) into the system Equ. (A.8), it can be found that the velocity of the activity blob is determined by the asymmetric component of the background input, following

$$\vec{v}^*(t) \approx \frac{\lambda \gamma_0}{\tau} \vec{h}_0^d(t) \quad \Leftrightarrow \quad \vec{h}_0^d(t) \approx \frac{\tau}{\lambda \gamma_0} \vec{v}^*(t) \quad (A.12)$$

Further, because the approximate linear relationship between the external commands and the network response is unbounded, the integration limit of the network has to be determined. Since the network integration property relies on a competition across the sub-layers sharing the same preferential tuning to the variable $\vec{r}_{\mathcal{B}}$, the maximum intrinsic speed is reached when a single sub-layer has a positive activation. In the ring attractor case, this special network state occurs when the asymmetry in the background input $h(\vec{r}_{\mathcal{B}}, t)$ is sufficiently strong to completely inhibit one of the two sub-layers. However, in the torus case, because of the continuous nature of the space $\Gamma^{\mathcal{B}}$ spanned by the sub-layers, one sub-layer may eventually drive the network alone only for $\|\vec{h}_0^d\| \rightarrow \infty$. In these cases, the closed integral along $\vec{r}_{\mathcal{B}}$ of the recurrent connectivity may be removed. The full recurrent synaptic drive thus becomes

$$[W^R(\vec{r}'_{\mathcal{A}}, \vec{r}'_{\mathcal{A}}) - \lambda \nabla W^R(\vec{r}'_{\mathcal{A}}, \vec{r}'_{\mathcal{A}}) \cdot \vec{r}'_{\mathcal{B}}] * \tilde{f}^*(\vec{r}'_{\mathcal{A}}, \vec{r}'_{\mathcal{B}}, t) \quad (A.13)$$

where $\vec{r}'_{\mathcal{B}}$ denotes the only active sub-layer given by $\vec{r}'_{\mathcal{B}} = \vec{h}_0^d / \|\vec{h}_0^d\|$. This recurrent drive has already been shown to lead a traveling activity peak to move with a constant velocity equal to λ/τ and with direction $\vec{r}'_{\mathcal{B}}$ (Zhang, 1996). This is thus the integration limit that the model can reach.

A.2.2 STIMULUS INPUT

In this section, the interest is on how a stimulus input which conveys information related to the spatial localization of a stimulus may drive the network representation to reflect the motion dynamics of that input. A stimulus located at $\vec{s}(t)$ in stimulus space and moving in phase with the network intrinsic velocity $\vec{v}^*(t)$, i.e., $\dot{\vec{s}}(t) = \vec{v}^*(t)$, is considered. The shape G of the stimulus input is assumed to be a Gaussian-like shape centered on its current location $\vec{s}(t)$ with breadth σ_s . Moreover, in order to compensate for the neural dynamics and for the effect of the network recurrent connections, respectively, an additional differential term and a scaling factor are needed to define a "well-behaving" input, i.e., one which does not turn into an asymmetric activity peak in the network representation when it moves within the neural field. It is defined by

$$x(\vec{r}_{\mathcal{A}}, \vec{r}_{\mathcal{B}}, t) = h_1 \left[G(\vec{r}_{\mathcal{A}}, \vec{s}(t), \sigma_s) - \tau \dot{\vec{s}}(t) \cdot \nabla G(\vec{r}_{\mathcal{A}}, \vec{s}(t), \sigma_s) \right] \left[1 + \epsilon \dot{\vec{s}}(t) \cdot \vec{r}_{\mathcal{B}} \right] \quad (\text{A.14})$$

where ϵ is a temporary constant. As can be noticed while looking at the network dynamics, the second differential term here compensates for the similar term in Equ. (A.8). Then, considering the asymmetry in the strength of the input relative to the neuron preferred movement direction, it modifies the network intrinsic dynamics in a similar way to that of the background input (Equ. (A.9)). Moreover, the velocity term $\dot{\vec{s}}(t)$ of the stimulus input is also assumed to be in a form comparable to that of the background homogenous input given by Eqs. (A.9) and (A.12) such that

$$\dot{\vec{s}}(t) = \frac{\lambda \gamma_1}{\tau} \vec{h}_1^d(t) \quad (\text{A.15})$$

where $\vec{h}_1^d(t)$ is a directional vector similar to $\vec{h}_0^d(t)$ defined previously. It determines the strength of the stimulus asymmetry and, thus, the stimulus velocity. γ_1 is like γ_0 and corresponds to a linear approximation factor. Then, rewriting Equ. (A.14) using (A.15) and writing $\epsilon = \tau/\lambda\gamma_1$ gives

$$x(\vec{r}_{\mathcal{A}}, \vec{r}_{\mathcal{B}}, t) = h_1 \left[G(\vec{r}_{\mathcal{A}}, \vec{s}(t), \sigma_s) - \lambda \gamma_1 \vec{h}_1^d(t) \cdot \nabla G(\vec{r}_{\mathcal{A}}, \vec{s}(t), \sigma_s) \right] \left[1 + \vec{h}_1^d(t) \cdot \vec{r}_{\mathcal{B}} \right] \quad (\text{A.16})$$

In the moving frame of reference, using the substitution given in Equ. (A.7),

the stimulus input equation becomes

$$x(\vec{r}_A, \vec{r}_B, t) = h_1 [G(\vec{r}_A, \vec{s}, \sigma_s) - \lambda \gamma_1 \vec{h}_1^d(t) \cdot \nabla G(\vec{r}_A, \vec{s}, \sigma_s)] [1 + \vec{h}_1^d(t) \cdot \vec{r}_B] \quad (\text{A.17})$$

where \vec{s} is constant since the stimulus is assumed to move according to $\dot{\vec{s}}(t) = \vec{v}^*(t)$. Again, similarly to the background input case described before (Eqs. (A.10a) and (A.10b)), a linear approximation of the network activity around the equilibrium point is considered. In addition, a superposition of solutions may result from both forms of inputs, i.e., $h(\vec{r}_B, t)$ and $x(\vec{r}_A, \vec{r}_B, t)$ is also assumed. This leads to a family of solutions given by

$$\tilde{u}^* = h_0 [U_0^* + \vec{h}_0^d \cdot \vec{r}_B] + h_1 [U_1^* + \vec{h}_1^d \cdot \vec{r}_B G] \quad (\text{A.18a})$$

$$\tilde{f}^* = h_0 [1 + \gamma_0 \vec{h}_0^d \cdot \vec{r}_B] F_0^* + h_1 [1 + \gamma_1 \vec{h}_1^d \cdot \vec{r}_B] F_1^* \quad (\text{A.18b})$$

where

$$-1 + U_0^* = \oint W^R * F_0^* d\vec{r}_B' \quad (\text{A.19a})$$

$$-G + U_1^* = \oint W^R * F_1^* d\vec{r}_B' \quad (\text{A.19b})$$

From these equations, the value that the parameters of the stimulus input should take in order to respect the network dynamics can be determined. Again developing the system Equ. (A.8) results here in

$$\begin{aligned} -\tau \vec{v} \cdot \nabla \tilde{u}^* + \tilde{u}^* &= h + x + \oint [W^R - \lambda \nabla W^R \cdot \vec{r}_B'] * \tilde{f}^* d\vec{r}_B' \\ &\stackrel{(\text{A.9}), (\text{A.12}), (\text{A.16}), (\text{A.18}), (\text{A.19})}{\Leftrightarrow} \\ -\lambda \gamma_0 \vec{h}_0^d [h_0 \nabla U_0^* + h_1 \nabla [\vec{h}_1^d \cdot \vec{r}_B G + U_1^*]] + \\ h_0 [\vec{h}_0^d \cdot \vec{r}_B + U_0^*] + h_1 [\vec{h}_1^d \cdot \vec{r}_B G + U_1^*] &= \\ h_0 [1 + \vec{h}_0^d \cdot \vec{r}_B] + h_1 [G - \lambda \gamma_1 \vec{h}_1^d \cdot \nabla G] [1 + \vec{h}_1^d \cdot \vec{r}_B] + \\ h_0 [U_0^* - 1 - \lambda \gamma_0 \vec{h}_0^d \cdot \nabla U_0^*] + h_1 [U_1^* - G - \lambda \gamma_1 \vec{h}_1^d \cdot \nabla [U_1^* - G]] \\ &\Leftrightarrow \\ \lambda \gamma_0 \vec{h}_0^d [h_1 \nabla [\vec{h}_1^d \cdot \vec{r}_B G + U_1^*]] &= \lambda \gamma_1 \vec{h}_1^d [h_1 \nabla [\vec{h}_1^d \cdot \vec{r}_B G + U_1^*]] \end{aligned}$$

And thus, the following constraints can be found, i.e.,

$$\gamma_1 \stackrel{!}{=} \gamma_0 = \gamma \quad \text{and} \quad \vec{h}_1^d \stackrel{!}{=} \vec{h}_0^d \quad (\text{A.20})$$

Therefore, in order for the network intrinsic dynamics to follow its input, the asymmetric breaking of the network along \vec{r}_B must match that the stimulus inputs.

From now on, the form of the stimulus input is defined using a velocity-dependent scaling factor, which needs a non-linear operation in order to be constructed. In the following, a different form of input containing only linear terms is defined. Indeed, as will be shown further, this form of input will help derive an expression for the synaptic projections across instances of the model, which will allow them to completely transfer information they are encoding. By rewriting the stimulus input in the moving frame of reference, as defined in Equ. (A.17), and by replacing the velocity-dependent non-linear factor with a velocity-dependent linear term, it gives

$$x(\tilde{\mathbf{r}}_{\mathcal{A}}, \tilde{\mathbf{r}}_{\mathcal{B}}, t) = h_1 [G(\tilde{\mathbf{r}}_{\mathcal{A}}, \tilde{\mathbf{s}}(t), \sigma_s) - \lambda \gamma \tilde{\mathbf{h}}_0^d(t) \cdot \nabla G(\tilde{\mathbf{r}}_{\mathcal{A}}, \tilde{\mathbf{s}}(t), \sigma_s) + \eta \tilde{\mathbf{h}}_0^d(t) \cdot \tilde{\mathbf{r}}_{\mathcal{B}}] \quad (\text{A.21})$$

where $\eta > 0$ is a scaling factor. Then, similarly to the previous sections, a family of solutions is also considered. In this case, they are given by

$$\tilde{u}^* = h_0 [U_0^* + \tilde{\mathbf{h}}_0^d \cdot \tilde{\mathbf{r}}_{\mathcal{B}}] + h_1 [\eta \tilde{\mathbf{h}}_0^d \cdot \tilde{\mathbf{r}}_{\mathcal{B}}] + h_1 U_1^* \quad (\text{A.22a})$$

$$\tilde{f}^* = h_0 [1 + \gamma \tilde{\mathbf{h}}_0^d \cdot \tilde{\mathbf{r}}_{\mathcal{B}}] F_0^* + h_1 [\gamma \eta \tilde{\mathbf{h}}_0^d \cdot \tilde{\mathbf{r}}_{\mathcal{B}}] F_0^* + h_1 F_1^* \quad (\text{A.22b})$$

where the system of Eqs. (A.19) still apply. Again, replacing these equations into the system dynamics (A.8) gives

$$\begin{aligned} -\tau \tilde{\mathbf{v}}^* \cdot \nabla \tilde{u}^* + \tilde{u}^* &= h + x + \oint_{(\text{A.9}), (\text{A.12}), (\text{A.19}), (\text{A.21}), (\text{A.22})} [W^R - \lambda \nabla W^R \cdot \tilde{\mathbf{r}}_{\mathcal{B}}'] * \tilde{f}^* d\tilde{\mathbf{r}}_{\mathcal{B}}' \\ &= -\lambda \gamma \tilde{\mathbf{h}}_0^d [h_0 \nabla U_0^* + h_1 \nabla U_1^*] + \\ &h_0 [U_0^* + \tilde{\mathbf{h}}_0^d \cdot \tilde{\mathbf{r}}_{\mathcal{B}}] + h_1 [\eta \tilde{\mathbf{h}}_0^d \cdot \tilde{\mathbf{r}}_{\mathcal{B}}] + h_1 U_1^* = \\ &h_0 [1 + \tilde{\mathbf{h}}_0^d \cdot \tilde{\mathbf{r}}_{\mathcal{B}}] + h_1 [G - \lambda \gamma \tilde{\mathbf{h}}_0^d \cdot \nabla G + \eta \tilde{\mathbf{h}}_0^d \cdot \tilde{\mathbf{r}}_{\mathcal{B}}] \\ &+ h_0 [U_0^* - 1 - \lambda \gamma \tilde{\mathbf{h}}_0^d \cdot \nabla U_0^*] - \lambda \gamma \eta h_1 \tilde{\mathbf{h}}_0^d \cdot \nabla U_0^* + h_1 [U_1^* - G] \\ &\Leftrightarrow \\ &\nabla [U_1^* - G] \stackrel{!}{=} \eta \nabla U_0^* \end{aligned} \quad (\text{A.23})$$

Therefore, in order for the network intrinsic dynamics to follow that of the input, the gradient of the stimulus-related solution $[U_1^* - G]$ has to be proportional to that of the homogeneous background input U_0^* . However, depending on the shape of G , this may not necessarily be the case. Nevertheless, since the solutions U_i^* are the result of an equilibrium between the input and the drive of the recurrent weights, one can assume that the latter dominate. As a result, both sides of Equ. (A.23) can be considered to be very similar and hence

proportional. An approximate value for η can thus be found such that

$$\nabla[U_1^* - G] \approx \eta \nabla U_0^* \quad (\text{A.24})$$

which, subsequently, using the system of Eqs. (A.19) is equivalent to

$$\nabla \oint W^R * F_1^* d\vec{r}'_B \approx \eta \nabla \oint W^R * F_0^* d\vec{r}'_B \quad (\text{A.25})$$

A.2.3 SYNAPTIC PROJECTIONS

In this section, synaptic projections across populations are considered. Since the proposed model might be integrated in a larger network of neural populations, the way how these populations may transfer the information they are encoding must be determined. Two neural populations, A and B, are considered. In the following, these names will be used as indices in order to clearly identify the corresponding variables and parameters of each population. The synaptic projections W^{AB} refer to a full directional connectivity between populations A and B. It is given by

$$x^B(\vec{r}_A, \vec{r}_B, t) = \oint\!\!\!\oint W^{AB}(\vec{r}'_A, \vec{r}'_B, \vec{r}_A, \vec{r}_B) f(u^A(\vec{r}'_A, \vec{r}'_B, t)) d\vec{r}'_A d\vec{r}'_B \quad (\text{A.26})$$

such that the input x^B is fed to the population B. Moreover, W^{AB} will be defined so that the information related to both position and velocity will be transferred from population A to B, such that $\vec{p}^A = \vec{p}^B$ and $\dot{\vec{p}}^A = \dot{\vec{p}}^B$, where \vec{p}^i is the population vector of population i defined in Chapter 7 by Equ. (7.2). Throughout this section, the following convolution weights is considered

$$W^{AB}(\vec{r}'_A, \vec{r}'_B, \vec{r}_A, \vec{r}_B) = \left[W^T(\vec{r}'_A, \vec{r}_A) - \lambda^A \nabla W^T(\vec{r}'_A, \vec{r}_A) \cdot \vec{r}'_B + \mu^{AB} \vec{r}_B \cdot \vec{r}'_B \right] \quad (\text{A.27})$$

where W^T is a center-surround, Gaussian-like convolution kernel (See Appendix B.3 for a precise definition). Similarly to the fundamental input shape G of the linear input form (Equ. (A.21)), the convolution through the projection weights will produce a fundamental input shape, denoted by G^{AB} . Moreover, since the recurrent weights preferentially link neurons sharing similar tuning properties, this guarantees that the position-related information \vec{p}^A of the population A is correctly transferred to population B. Then, since the activity profile \tilde{f}^{A*} of the population A (Equ. (A.22)b) is composed of a linear combination of two terms F^{A0*} and F^{A1*} , the input shape G^{AB} is assumed to be written as

$$G^{AB} = h_0 \underbrace{\oint W^T * F^{A0*} d\vec{r}'_B}_{G^{AB0}} + h_1 \underbrace{\oint W^T * F^{A1*} d\vec{r}'_B}_{G^{AB1}} \quad (\text{A.28})$$

where G^{AB_0} and G^{AB_1} correspond, respectively, to the contribution of the background and the stimulus input of population A to the input x^B feeding population B. Further, from Equ. (A.25), there exists a η^A such that

$$\nabla G^{AB_1} \approx \eta^A \nabla G^{AB_0} \quad (\text{A.29})$$

Then, by constraining the weights convolution kernel such that

$$\oint W^T(\vec{r}) d\vec{r} = 0 \quad \Rightarrow \quad \oint G^{AB_0} d\vec{r} = \oint G^{AB_1} d\vec{r} = 0 \quad (\text{A.30})$$

both sides of Equ. (A.29) can be integrated, which gives

$$G^{AB_1} \approx \eta^A G^{AB_0} \quad (\text{A.31})$$

which, by Equ. (A.28) then implies

$$G^{AB} \approx (h_0 + \eta^A h_1) G^{AB_0} \quad (\text{A.32})$$

Further, an exact input form equivalent to Equ. (A.21) can be defined in order to match the projection through the weights W^{AB} . It is given by

$$x^B(\vec{r}_A, \vec{r}_B, t) = (h_0 + \eta^A h_1) \left[G^{AB_0}(\vec{r}_A, \vec{p}^A(t), \sigma_s) - \lambda^B \gamma^B \vec{h}_0^{B^d}(t) \cdot \nabla G^{AB_0}(\vec{r}_A, \vec{p}^A(t), \sigma_s) + \eta^B \vec{h}_0^{B^d}(t) \cdot \vec{r}_B \right] \quad (\text{A.33})$$

where η^B may be different from η^A since these constants depend on the recurrent weights of their respective populations and on the shape of their respective inputs, i.e., G and G^{AB} . Similarly, the population's intrinsic parameters, such as γ and λ , may also be different¹. Thus, the relationship between them using the constraint on the velocity information transfer is then defined by

$$\dot{\vec{p}}^A = \dot{\vec{p}}^B \quad (\text{A.12}) \quad \Leftrightarrow \quad \frac{\lambda^A \gamma^A}{\tau} \vec{h}_0^{A^d} = \frac{\lambda^B \gamma^B}{\tau} \vec{h}_0^{B^d} \quad (\text{A.34})$$

Further, a value for μ_{AB} which satisfies the following equation

$$\oint W^{AB} * \tilde{f}^{A*} d\vec{r}'_B = (\text{A.33}) \quad (\text{A.35})$$

can be found, where \tilde{f}^{A*} is given by rewriting Equ. (A.22)b with the corresponding population indices, i.e.,

$$\begin{aligned} \tilde{f}^{A*} &= h_0 [1 + \gamma^A \vec{h}_0^{A^d} \cdot \vec{r}_B] F^{A_0*} + \\ &h_1 [\gamma^A \eta^A \vec{h}_0^{A^d} \cdot \vec{r}_B] F^{A_0*} + h_1 F^{A_1*} \end{aligned} \quad (\text{A.36})$$

¹For simplicity, τ was assumed to be constant across the populations.

Then, developing the left-hand side of Equ. (A.35) gives

$$\begin{aligned}
& \oint [W^T - \lambda^A \nabla W^T \cdot \vec{r}'_B + \mu_{AB} \vec{r}_B \cdot \vec{r}'_B] * \tilde{f}_A^* d\vec{r}'_B \stackrel{(A.36)}{=} \\
& h_0 [G^{AB_0} - \lambda^A \gamma^A \vec{h}_0^{A^d} \nabla G^{AB_0} + \mu^{AB} \gamma^A \vec{h}_0^{A^d} \cdot \vec{r}_B \oint F^{A_0*}] \\
& + h_1 [G^{AB_1} - \lambda^A \gamma^A \vec{h}_0^{A^d} \nabla G^{AB_1} + \eta^A \mu^{AB} \gamma^A \vec{h}_0^{A^d} \cdot \vec{r}_B \oint F^{A_0*}] \\
& \stackrel{(A.32)}{\approx} (h_0 + h_1 \eta^A) [G^{AB_0} - \lambda^A \gamma^A \vec{h}_0^{A^d} \nabla G^{AB_0} + \mu^{AB} \gamma^A \vec{h}_0^{A^d} \cdot \vec{r}_B \oint F^{A_0*}] \\
& \stackrel{(A.34)}{\approx} (h_0 + h_1 \eta^A) [G^{AB_0} - \lambda^B \gamma^B \vec{h}_0^{B^d} \nabla G^{AB_0} + \mu^{AB} \frac{\lambda^B \gamma^B}{\lambda^A} \vec{h}_0^{B^d} \cdot \vec{r}_B \oint F^{A_0*}]
\end{aligned} \tag{A.37}$$

And finally, by comparing Eqs. (A.33) and (A.37), it remains

$$\mu^{AB} \approx \frac{\lambda^A}{\lambda^B} \frac{\eta^B}{\gamma^B \oint F^{A_0*}} \tag{A.38}$$

Therefore, using the synaptic projections W^{AB} described by Equ. (A.27), a population A can transfer its encoded information related to both position and velocity to another population B.

SIMULATION PARAMETERS

B.1 BRAIN PATHWAYS OF IMITATION AND THE IDEOMOTOR PRINCIPLE

The parameters used in the simulations of the two models described in Chapter 4 and illustrated in Figure 4.2, can be found in Table B.1. Concerning the choice of these simulations parameters, they were initially tuned according to the work hypotheses as presented in Section 4.2.1. Then, they were fine-tuned using a gradient descent method in order to minimize the error

$$E = \|Y - \hat{X}\|^2 \quad (\text{B.1})$$

between the behavioral data Y and the simulation results \hat{X} . Simultaneously, simulated RTs X were fitted to the original data Y using a first order least squares error regression method. The estimated RTs \hat{X} are given by

$$\hat{X} = c_1 X + c_2 \quad (\text{B.2})$$

where the constants c_1 and c_2 were determined so that the error given in Equ. (B.1) is also minimized.

	Parameter	Single-Route	Direct-Matching
Perceptual Layer			
IT Region			
Spatial Cue	α	1.5	1.53
	β	0.88	0.9
	φ	$\pm\pi/2$	$\pm\pi/2$
STS Region			
Movement Cue	α	1.5	1.53
	β	0.97	1.0
	φ	$\pm\pi/2$	$\pm\pi/2$
Top-down Modulation			
Reciprocal Inhibition	h^T	0.26	0.2
	$W^{\text{SpaCue,MvtCue}}$	24.11	16.65
	$W^{\text{MvtCue,SpaCue}}$	24.11	16.65
STS Region			
Movement Observation	α	0.0	0.0
	β	0.79	0.5
	φ	$\pm\pi/2$	$\pm\pi/2$
Decision Layer			
LPFC Region			
Cue Integration	α	1.02	2.0
	$\alpha^{\text{SpaCue,CueInt}}$	1.16	1.65
	$\alpha^{\text{MvtCue,CueInt}}$	1.16	1.65
	$\gamma^{\text{IdeInt,CueInt}}$	8.95	
Motor Preparation Layer			
AMFC, SMA Regions			
Motor Plan (L/R)	α	0.0	0.0
	β	1.0	1.0
	φ	$\pm\pi/2$	$\pm\pi/2$
SMA, PMd Regions			
Ideomotor Integration (L/R)	α	1.98	1.81
	$\alpha^{\text{MotPlan,IdeInt}}$	1.08	1.19
	$\alpha^{\text{MotObs,IdeInt}}$	1.08	1.19
PMv Region			
Motor Selection (L/R)	α	0.15	0.09
	$W^{\text{CueInt,MotSel}}$	3.91	4.02
	$\alpha^{\text{MotPlan,MotSel}}$	1.58	
	$\alpha^{\text{IdeInt,MotSel}}$		1.58
Reciprocal Inhibition			
	$W^{\text{MotSel(L),MotSel(R)}}$	7.71	7.34
	$W^{\text{MotSel(R),MotSel(L)}}$	7.71	7.34
Execution Threshold			
	E_0	0.08	0.08
Other Constants			
Variance Profile	σ	0.3	0.3
	$\sigma^{\text{CueInt,MotSel}}$	0.5	0.5
Time Constant	τ	0.1	0.1
Additional time delay [ms]			
Compatible mapping condition	Δ	0.0	0.0
		73.0	73.0
Regression constants			
	c_1	3292.19	-263.32
	c_2	1770.54	-69.21

Table B.1: Simulation parameters used in the study presented in Chapter 4.

B.2 INTERFERENCES IN THE TRANSFORMATION OF FRAMES OF REFERENCE

The parameters used in the simulations of the model described in Chapter 6 and illustrated in Figure 6.4, can be found in the following. The respective range of arm and body orientations are $\theta^D \in \{k \cdot 22.5^\circ | k \in \{0..8\}\}$ and $\phi^D \in \{k \cdot 22.5^\circ | k \in \{0..15\}\}$. The amplitude of postural change in Experiment 2 is in the range $\Delta\theta^D \in \{k \cdot 22.5^\circ | k \in \{1..8\}\}$. The additional inhibition Δh was varied such that $\Delta h \in \{0.5 + k \cdot 0.125 | k \in \{0..6\}\}$. According to the simulation parameters, the complete inhibition of the irrelevant transformation was reached when $\Delta h = \Delta h_0 = 1.25$. This case was considered to be the baseline condition. The model parameters are: the amplitudes of the weights, $\alpha = 12$, $\alpha^{\text{SArm, SOut}} = \alpha^{\text{AArm, GF}\phi} = \alpha^{\text{GF}\phi, \text{AOut}} = 5.4$, $\alpha^{\text{ABody, GF}\phi} = 8.0$, and $\alpha^{\text{AOut, Sel}} = \alpha^{\text{SOut, Sel}} = 5.0$, the breadth of the weights profiles and their offset, unless specified, $\sigma = 0.5$ and $\eta = \oint g(\vec{r}, \vec{r}') d\vec{r}$, then $\sigma^{\text{ABody, GF}\phi} = \infty$ and $\eta^{\text{AArm, GF}\phi} = 1.0$; the amplitude of the inputs, $\beta^{\text{SArm}} = \beta^{\text{AArm}} = \beta^{\text{ABody}} = 0.5$. These parameters were chosen so that the response energy of both transformations are equal for an equivalent task modulation. Finally, the task modulatory inputs and go signal are, $h^{\text{Task}} = 0.5$, $\Delta h_0 = 0.75$ and $h^{\text{Go}} = 1.5$.

B.3 MOTION INTEGRATION, SENSITIVITY AND SENSORY DISCRIMINATION

In order to implement the model described in Chapter 7, the recurrent weights followed Equ. (3.8), and the shape G of the stimulus input was defined according to (3.10). Moreover, the simulations were performed using the fourth-order Runge-Kutta numerical algorithm for the resolution of dynamic systems. Except where explicitly defined in the description of each experiment, the simulation parameters that were used in Chapter 7 are given in the following table.

	Value of the parameters
Ring attractor	$\alpha = 2$, $\alpha^{\text{T}} = 2$, $\sigma = 0.3$, $\delta = 1$, $\lambda = 0.1$, $\beta = 1$ $h_0 = 1$, $h_1 = 1$, $\tau = 0.1$ Discretization along Γ_1^{A} : 128 neurons Γ_1^{B} : 2 sub-layers
Torus attractor	$\alpha = 1$, $\alpha^{\text{T}} = 2$, $\sigma = 0.3$, $\delta = 1$, $\lambda = 0.1$, $\beta = 1$, $h_0 = 1$, $h_1 = 1$, $\tau = 0.1$ Discretization along Γ_2^{A} : 32×32 neurons Γ_2^{B} : 8 sub-layers

B.4 MOVEMENT GENERATION, SENSORY DISCRIMINATION AND INTERFERENCES

The parameters used in the simulations of the model described in Chapter 8 and illustrated in Figure 8.6, can be found in the following table.

	Value of the parameters
General parameters	$\tau = 0.1$ Discretization along Γ^A : 256 neurons Γ^B : 6 sub-layers
Recurrent Connectivity	The goal network: $\alpha = 4, \sigma = 0.3, \delta = 1$ The other networks: $\alpha = 2, \sigma = 0.3, \delta = 1, \lambda = 1, \gamma = 0.91$
Synaptic Projections	$\alpha^{G,C} = 2, \sigma^{G,C} = 0.3, \delta^{G,C} = 4$ $\alpha^{S,C} = 2, \sigma^{S,C} = 0.3, \delta^{S,C} = 1$ $W^{C,S} = W^{C,O} = 0.55$
External inputs	$h_0^S = h_0^O = 1, h_0^G = h_0^C = 0$ $\beta^G = \beta^S = \beta^V = 1, \sigma_s = 0.3$

PUBLICATIONS OF THE AUTHOR

THIS appendix lists the publications of the author related to the work described in this thesis. Along this enumeration, a brief comment concerning the relationship of each paper with this thesis is given. The articles are listed according to the date of their publication.

Sauser, E. L. and Billard, A. G.. Three dimensional frames of reference transformations using gain modulated populations of neurons. *In Proceedings of 12th European Symposium on Artificial Neural Networks (ESANN), Bruges, Belgium.* 543-548, 2004.

This paper addresses the problem of frames of reference transformations as is described in Chapter 5 of this thesis. Importantly, a formal demonstration of how population vector coding can proceed arbitrary three dimensional rotations and translations is provided. The described method for three dimensional rotations consists of a decomposition of the transformation into a projection, a planar rotation and a translation.

Sauser, E. L. and Billard, A. G.. Three dimensional frames of references transformations using recurrent populations of neurons. *Neurocomputing.* 64:5-24, 2005.

The model presented in this paper extends that mentioned above, mostly by proposing a method by which a network can preserve the amplitude of the information to be transformed across the transformation process. It consists of a two-layer architecture which is capable to remove the unnecessary amplification term produced by the recurrent interactions of the considered neural field model. This work also contributed significantly to Chapter 5.

Sauser, E. L. and Billard, A. G.. View sensitive cells as a neural basis for the representation of others in a self-centered frame of reference. *In Proceedings of the Third International Symposium on Imitation in Animals and Artifacts (AISB), Hatfield, UK.* 119-127, 2005.

This paper presents a neural mechanism for frames of reference transformations which performs projections of the input vector on the principal axes of the referential. It primarily aims at explaining the transformation processes occurring when considering visual information. An application proposed by this paper is a simple mimicry task where a robot can reproduces hand movements performed by a human subject standing in various orientations relative to the robot. This work was also reported in Chapter 5 of this thesis.

Sauser, E. L. and Billard, A. G.. Parallel and distributed neural models of the ideomotor principle: An investigation of imitative cortical pathways. *Neural Networks*. 19(3):285-298,2006.

The work presented in this paper corresponds to the modeling investigation described in Chapter 4 of this thesis. Its aim is to understand and decipher the neural processes and the cortical pathways responsible for automatic imitative behaviors as reported in experimental psychology. Importantly, this work suggests two plausible cortical pathways which could mediate the related behaviors. In addition, a novel experimental paradigm which could validate or refute either of the two hypotheses is also proposed along with its predictions.

Sauser, E. L. and Billard, A. G.. Biologically inspired multimodal integration: Interferences in a human-robot interaction game. *In proceedings of the International Conference on Intelligent Robots and Systems (IROS), Beijing, China*. 2006.

This paper illustrates how the principle of ideomotor compatibility could be applied to a robotic platform in order to provide robots with human-like behaviors.

Sauser, E. L. and Billard, A. G.. Dynamic updating of distributed neural representations using forward models. *Biological Cybernetics*. 95(6):567-588, 2006.

This paper presents an investigation as to how a neural field can integrate external commands in order to update dynamically its internal representation. Importantly, the analysis of the developed network when confronted to external inputs showed that the model can exhibit several interesting properties, almost all of which were reported in the experimental literature. They includes neural tuning to the velocity of visual stimuli and abilities for sensory discrimination. Moreover, such mechanism has also been suggested to have important implications regarding sensorimotor transformations, motor control, motor imagery, and imitation. This work was reported in Chapter 7 of this thesis.

Sauser, E. L. and Billard, A. G.. Interferences in the transformation of reference frames during a posture imitation task. *In Proceedings of the International Conference on Artificial Neural Networks (ICANN), Porto, Portugal*. 2007.

The modeling study presented in this paper was reported in Chapter 6 of this thesis. It describes a model which illustrates how the mechanisms of transformations across reference frames may be applied in order to model the problem of transformations across frames of reference in a posture imitation task. This work proposes that, in such a task, imitation is mediated by two concurrent transformations selectively sensitive to spatial and anatomical cues. In addition, an experimental paradigm was devised, which allowed the modeling of the interference patterns caused by the interaction between the anatomical, and the spatial imitative strategy. The predictions provided by this simulation study are finally suggested to be confronted to real behavioral data in order to validate or refute the model.

REFERENCES

- Abbott, L. F., & Nelson, S. B. (2000). Synaptic plasticity: Taming the beast. *Nature Neuroscience Supplement*, 3, 1178-1183.
- Aflalo, T. N., & Graziano, M. S. A. (2006a). Partial tuning of motor cortex neurons to final posture in a free-moving paradigm. *Proceedings of the National Academy of Sciences USA*, 103(8), 2909-2914.
- Aflalo, T. N., & Graziano, M. S. A. (2006b). Possible origins of the complex topographic organization of motor cortex: Reduction of a multidimensional space onto a two-dimensional array. *Journal of Neuroscience*, 26, 6288-6297.
- Aflalo, T. N., & Graziano, M. S. A. (2007). Relationship between unconstrained arm movements and single-neuron firing in the macaque motor cortex. *Journal of Neuroscience*, 27, 2760-2780.
- Afraz, S.-R., Kiani, R., & Esteky, H. (2006). Microstimulation of inferotemporal cortex influences face categorization. *Nature*, 442, 692-695.
- Alexander, G. E., & Crutcher, M. D. (1990). Preparation for movement: Neural representations of intended direction in three motor areas of the monkey. *Journal of Neurophysiology*, 64(1), 133-150.
- Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27, 77-87.
- Amirikian, B., & Georgopoulos, A. P. (2000). Directional tuning profiles of motor cortical cells. *Neuroscience Research*, 36(1), 73-79.
- Andersen, R. A., Asanuma, C., Essick, G., & Seigel, R. M. (1990). Corticocortical connections of anatomically and physiologically defined subdivisions within the inferior parietal lobule. *Journal of Comparative Neurology*, 296, 65-113.
- Andersen, R. A., Snyder, L. H., Bradley, D. C., & Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual Reviews Neuroscience*, 20, 303-330.
- Andry, P., Gaussier, P., Nadel, J., & Hirsbrunner, B. (2004). Learning invariant sensorimotor behaviors: A developmental approach to imitation mechanisms. *Adaptive behavior*, 12(2), 117-138.
- Arbib, M., Billard, A., Iacoboni, M., & Oztop, E. (2000). Mirror neurons, imitation and (synthetic) brain imaging. *Neural Networks*, 13, 953-973.
- Arbib, M. A. (2002). The mirror system, imitation and the evolution of language. In *Imitation in Animals and Artefacts, A Bradford Book, MIT Press*, 229-280.
- Arbib, M. A., Bischoff, A., Fagg, A., & Grafton, S. (1995). Synthetic pet: Analyzing large-scale properties of neural networks. *Human Brain Mapping*, 2, 225-233.
- Arbib, M. A., & Bota, M. (2003). Language evolution: Neural homologies and neuroinformatics. *Neural Networks*, 16, 1237-1260.
- Arzy, S., Thut, G., Mohr, C., Michel, C. M., & Blanke, O. (2006). Neural

- basis of embodiment: Distinct contributions of tempoparietal junction and extrastriate body area. *Journal of Neuroscience*, 26(31), 8074-8081.
- Asada, M., McDorman, K. F., Ishiguro, H., & Kuniyoshi, Y. (2001). Cognitive developmental robotics as a new paradigm for the design of humanoids robots. *Robotics and Autonomous Systems*, 37, 185-193.
- Ashbridge, E., Perrett, D. I., Oram, M. W., & Jellema, T. (2000). Effect of image orientation and size on object recognition: Responses of single units in the macaque monkey temporal cortex. *Cognitive Neuropsychology*, 17, 13-34.
- Astafiev, S. V., Stanley, C. M., Shulman, G. L., & Corbetta, M. (2004). Extrastriate body area in human occipital cortex responds to the performance of motor actions. *Nature Neuroscience*, 7, 542-547.
- Avillac, M., Denève, S., Olivier, E., Pouget, A., & Duhamel, J. R. (2005). Reference frames for representing the location of visual and tactile stimuli in the parietal cortex. *Nature Neuroscience*, 8(7), 941-949.
- Baker, C., Keysers, C., Jellema, T., Wicker, B., & Perrett, D. (2001). Neuronal representation of disappearing and hidden objects in temporal cortex of the macaque. *Experimental Brain Research*, 140, 375-381.
- Baldwin, J. (1902). Development and evolution. *Macmillan, New-York, USA*.
- Baraduc, P., & Guigon, E. (2002). Population computation of vectorial transformations. *Neural Computation*, 14(4), 845-871.
- Barresi, J., & Moore, C. (1996). Intentional relations and social understanding. *Behavioral and Brain Sciences*, 19, 107-154.
- Batista, A. P., Bruneo, C. A., Snyder, L. H., & Andersen, R. A. (1999). Reach plans in eye-centered coordinates. *Science*, 285, 257-260.
- Battaglia-Mayer, A., Caminiti, R., Lacquaniti, F., & Zago, M. (2003). Multiple levels of representation of reaching in the parieto-frontal network. *Cerebral Cortex*, 13, 1009-1022.
- Bekkering, H., Wohlschälger, A., & Gattis, M. (2000). Imitation of gestures in children is goal-directed. *Quarterly Journal of Experimental Psychology*, 53, 153-164.
- Ben Hamed, S., Duffy, C., & Pouget, A. (2003). MSTd neuronal basis functions for the population encoding of heading direction. *Journal of Neurophysiology*, 90, 549-558.
- Ben-Yishai, R., Bar-Or, R. L., & Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences USA*, 92, 3844-3848.
- Ben-Yishai, R., Hansel, D., & Sompolinsky, H. (1997). Traveling waves and the processing of weakly tuned inputs in a cortical network module. *Journal of Computational Neuroscience*, 4(1), 57-77.
- Berlucchi, G., & Aglioti, S. (1997). The body in brain: Neural bases of corporeal awareness. *Trends in Neuroscience*, 20, 560-564.
- Bertenthal, B. I., Longo, M. R., & Kosobud, A. (2006). Imitative response tendencies following observation of intransitive actions. *Journal of Experimental Psychology: Human Perception and Performance*, 32(2), 210-225.
- Bienenstock, E. L., Cooper, L. N., & Munro, P. W. (1982). Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, 2(1), 32-48.
- Billard, A. (2002). Imitation. In M. A. Arbib (ed.), *The Handbook of Brain Theory and Neural Networks*, MIT Press, 566-569.
- Billard, A., Epars, Y., Calinon, S., Cheng, G., & Schaal, S. (2004). Discovering optimal imitation strategies. *Robotics and Autonomous Systems*, 47(2-3), 69-77.

- Billard, A., & Hayes, G. (1999). DRAMA, a connectionist architecture for control and learning in autonomous robots. *Adaptive Behavior Journal*, 7(1), 35-64.
- Billard, A., & Mataric, M. (2001). Learning human arm movements by imitation: Evaluation of a biologically-inspired connectionist architecture. *Robotics and Autonomous Systems*, 941, 1-16.
- Blakemore, S.-J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, 2, 561-567.
- Blakemore, S.-J., & Frith, C. (2005). The role of motor contagion in the prediction of action. *Neuropsychologia*, 43, 20-267.
- Blanke, O., & Mohr, C. (2005). Out-of-body experience, heautoscopy, and auto-scopic hallucination of neurological origin: Implications for neurocognitive mechanisms of corporeal awareness and self-consciousness. *Brain Research Reviews*, 50(1), 184-199.
- Blanke, O., Ortigue, S., Landis, T., & Seeck, M. (2002). Stimulating illusory own body perceptions. *Nature*, 419, 269-270.
- Bonaiuto, J., Rosta, E., & Arbib, M. (2007). Extending the mirror neuron system model, I. audible actions and invisible grasps. *Biological Cybernetics*, 96(1), 9-38.
- Booth, M. C. A., & Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal cortex. *Cerebral Cortex*, 8, 510-523.
- Botvinick, M., & Cohen, J. (1998). Rubber hand feel touch that eyes see. *Nature*, 391, 756.
- Brass, M., Bekkering, H., & Prinz, W. (2001). Movement observation affects movement execution in a simple response task. *Acta Psychologica*, 106, 3-22.
- Brass, M., Bekkering, H., Wohlschläger, A., & Prinz, W. (2000). Compatibility between observed and executed finger movements: Comparing symbolic, spatial and imitative cues. *Brain and Cognition*, 44, 124-143.
- Brass, M., Derrfuss, J., & von Cramon, D. Y. (2005). The inhibition of imitative and overlearned responses: A functional double dissociation. *Neuropsychologia*, 43, 89-98.
- Brass, M., & Heyes, C. (2005). Imitation: Is cognitive neuroscience solving the correspondence problem? *Trends in Cognitive Sciences*, 9(10), 489-495.
- Bremmer, F., Graf, W., Ben Hamed, S., & Duhamel, J. R. (1999). Eye position encoding in the macaque ventral intraparietal area (VIP). *NeuroReport*, 10, 873-878.
- Bruneo, C. A., Jarvis, M. R., Batista, A. P., & Andersen, R. A. (2002). Direct visuomotor transformations for reaching. *Nature*, 416, 632-636.
- Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., et al. (2001). Action observation activated premotor and parietal area in a somatotopic manner: An fMRI study. *European Journal of Neuroscience*, 13, 400-404.
- Burnod, Y., Baraduc, P., Battaglia-Mayer, A., Guigon, E., Koechlin, E., Ferraina, S., et al. (1999). Parieto-frontal coding of reaching: An integrated framework. *Experimental Brain Research*, 129, 325-346.
- Byrne, R., & Russon, A. (1998). Learning by imitation: A hierarchical approach. *Behavioral and Brain Science*, 21, 667-721.
- Byrne, R. W. (1999). Imitation without intentionality. using the string parsing to copy the organization of behaviour. *Animal Cognition*, 63-72.
- Calinon, S., & Billard, A. (2007). What is the teacher's role in robot programming by demonstration? - Toward benchmarks for improved learn-

- ing. *Interaction Studies: Special Issue on Psychological Benchmarks in Human-Robot Interaction*.
- Calinon, S., Guenter, F., & Billard, A. (2007). On learning, representing and generalizing a task in a humanoid robot. *IEEE Transactions on Systems, Man and Cybernetics, Part B. Special issue on robot learning by observation, demonstration and imitation*, 37(2), 286-298.
- Calvo-Merino, B., Glaser, D., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex*, 15, 1243-1249.
- Casabona, A., Valle, M. S., Bosco, G., & Perciavalle, V. (2004). Cerebellar encoding of limb position. *The Cerebellum*, 3, 172-177.
- Chagnac-Amitai, Y., & Connors, B. W. (1989). Horizontal spread of synchronized activity in neocortex and its control by GABA-mediated inhibition. *Journal of Neurophysiology*, 61, 747-758.
- Chaminade, T., & Decety, J. (2002). Leader of follower. Involvement of the inferior parietal lobule in agency. *NeuroReport*, 13, 1975-1978.
- Chaminade, T., Franklin, D., Oztop, E., & Cheng, G. (2005). Motor interference between humans and humanoid robots: Effect of biological and artificial motion. *In proceedings of the 4th IEEE International Conference on Development and Learning, Osaka, Japan*, 96-101.
- Chan, A. W., Peelen, M. V., & Downing, P. E. (2004). The effect of viewpoint on body representation in the extrastriate body area. *NeuroReport*, 15, 2407-2410.
- Chapman, C. E., Spidalieri, G., & Lamarre, Y. (1984). Discharge properties of area 5 neurones during arm movements triggered by sensory stimuli in the monkey. *Brain Research*, 309, 63-77.
- Cheng, K., Hasegawa, T., Kadarbata, S. S., & Tanaka, K. (1994). Comparison of neuronal selectivity for stimulus speed, length, and contrast in the prestriate visual cortical areas V4 and MT of the macaque monkey. *Journal of Neurophysiology*, 71(6), 2269-2280.
- Chey, J., Grossberg, S., & Mingolla, E. (1998). Neural dynamics of motion processing and speed discrimination. *Vision Research*, 38, 2769-2786.
- Chiavarino, C., Apperly, I. A., & Humphreys, G. W. (2007). Exploring the functional and anatomical bases of mirror-image and anatomical imitation: The role of the frontal lobes. *Neuropsychologia*, 45, 784-795.
- Cisek, P. (2006). Inneural processes for defining potential actions and deciding between them: A computational model. *Journal of Neuroscience*, 26(38), 9761-9770.
- Cisek, P., & Kalaska, J. F. (2005). Neural correlates of reaching decision in dorsal premotor cortex: Specification of multiple direction choices and final selection of action. *Neuron*, 45, 801-814.
- Clark, S., Tremblay, F., & St-Marie, D. (2003). Differential modulation of the corticospinal excitability during observation, mental imagery and imitation of hand actions. *Neurophysiologia*, 42, 115-112.
- Clower, D. M., West, R. A., Lynch, J. C., & Strick, P. L. (2001). The inferior parietal lobule is the target of output from the superior colliculus, hippocampus, and cerebellum. *Journal of Neuroscience*, 21(16), 6283-6291.
- Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, 22, 319-349.
- Compte, A., Brunel, N., Goldman-Rakic, P. S., & Wang, X.-J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex*, 10, 910-923.
- Connor, C. E., Preddie, D. C., Gallant, J. L., & VanEssen, D. C. (1997).

- Spatial attention effects in macaque area V4. *Journal of Neuroscience*, 17, 3201-3214.
- Constantindis, C., & Steinmetz, M. A. (1996). Neuronal activity in posterior area 7a during the delay periods of a spatial memory task. *Journal of Neurophysiology*, 76, 1352-1355.
- Constantindis, C., & Steinmetz, M. A. (2001). Neuronal responses in area 7a to multiple stimulus displays: I. neurons encode the location of the salient stimulus. *Cerebral Cortex*, 11, 581-591.
- Cordo, P. J., Flores-Vieira, C., Verschueren, S. M. P., Inglis, J. T., & Gurfinkel, V. (2002). Position sensitivity of human muscle spindles: Single afferent and population representations. *Journal of Neurophysiology*, 87(3), 1186-1195.
- Crammond, D. J., & Kalaska, J. F. (1994). Modulation of preparatory neuronal activity in dorsal premotor cortex due to stimulus-response compatibility. *Journal of Neurophysiology*, 71(3), 1281-1284.
- Cuijpers, R. H., Van Schie, H. T., Koppen, M., Erhlagen, W., & Bekkering, H. (2006). Goals and means in action observation: A computational approach. *Neural Networks*, 19, 311-322.
- De Bruyn, B., & Orban, G. A. (1988). Human velocity and direction discrimination measured with random dot patterns. *Vision Research*, 28(12), 1323-1335.
- Decety, J., & Chaminade, T. (2003). When the self represents the other: A new cognitive neuroscience view on psychological identification. *Consciousness and Cognition*, 12, 577-596.
- Decety, J., Chaminade, T., Grezes, J., & Meltzoff, A. N. (2002). A PET exploration of the neural mechanisms involved in reciprocal imitation. *Neuroimage*, 15, 265-272.
- Decety, J., Grezes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., et al. (1997). Brain activity during observation of action. influence of action content and subject's strategy. *Brain*, 120, 1763-1777.
- Decety, J., & Sommerville, J. A. (2003). Shared representations between self and others: A social cognitive neuroscience view. *Trends in Cognitive Science*, 7, 527-533.
- Demiris, Y., & Hayes, G. (2002). Imitation as a dual process featuring predictive and learning components: A biologically plausible computational model. *In Imitation in Animals and Artefacts, A Bradford Book, MIT Press*, 327-362.
- Demiris, Y., & Johnson, M. (2003). Distributed, predictive perception of actions: A biologically inspired robotics architecture for imitation and learning. *Connection Science*, 15(4), 231-243.
- Demiris, Y., & Simmons, G. (2006). Perceiving the unusual: Temporal properties of hierarchical motor representations for action perception. *Neural Networks*, 19(3), 272-284.
- Deneve, S., Latham, P. E., & Pouget, A. (1999). Reading population codes: A neural implementation of ideal observers. *Nature Neuroscience*, 2(8), 740-745.
- Deneve, S., Latham, P. E., & Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nature Neuroscience*, 4(8), 826-831.
- Deneve, S., & Pouget, A. (2003). Basis function for object-centered representations. *Neuron*, 37, 347-359.
- De Yoe, E. A., & Van Essen, D. C. (1988). Concurrent processing streams in monkey visual cortex. *TINS*, 11(5), 219-226.

- Dong, W., Chudler, E. H., Sugiyama, K., Roberts, V. J., & Hayashi, T. (1994). Somatosensory, multisensory and task-related neurons in cortical area 7b(PF) of unanesthetized monkeys. *Journal of Neurophysiology*, 72, 542-564.
- Douglas, R. J., Koch, C., Mahowald, M., Martin, K. A. C., & Suarez, H. H. (1995). Recurrent excitation in neocortical circuits. *Science*, 269, 981-985.
- Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, 293, 2470-2473.
- Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Networks*, 12, 961-974.
- Dumitru, D., King, J. C., Nandedkar, S. D., & Oja, E. (1997). The nonlinear PCA learning rule in independent component analysis. *Neurocomputing*, 17(1), 25-45.
- Eimer, M., Hommel, B., & Prinz, W. (1995). S-R compatibility and response selection. *Acta Psychologica*, 90, 301-313.
- Engbert, K., & Wohlschläger, A. (2007). Intentions and expectations in temporal binding. *Consciousness and Cognition*, 16(2), 255-264.
- Erlhagen, W. (2003). Internal models for visual perception. *Biological Cybernetics*, 88, 409-417.
- Erlhagen, W., & Bicho, E. (2006). The dynamic neural field approach to cognitive robotics. *Journal of Neural Engineering*, 3, 36-54.
- Erlhagen, W., Mukovskiy, A., Bicho, E., Panin, G., Kiss, C., Knoll, A., et al. (2006). Goal-directed imitation in robots: A bio-inspired approach to action understanding and skill learning. *Robotics and Autonomous Systems*, 54(5), 353-360.
- Erlhagen, W., Mukovsky, A., & Bicho, E. (2006). A dynamic model for action understanding and goal-directed imitation. *Brain Research*, 1083, 174-188.
- Erlhagen, W., & Schöner, G. (2002). Dynamics field theory of movement preparation. *Psychological Review*, 109(3), 545-572.
- Eskandar, E. N., & Assad, J. A. (1999). Dissociation of visual, motor and predictive signals in parietal cortex during visual guidance. *Nature Neuroscience*, 2, 88-93.
- Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, 73, 2608-2611.
- Fagg, A. H., & Arbib, M. A. (1998). Modelling parietal-premotor interactions in primate control of grasping. *Neural Networks*, 11, 1277-1303.
- Farné, A., Pavani, F., Meneghello, F., & Ládavas, E. (2000). Left tactile extinction following visual stimulation of a rubber hand. *Brain*, 123, 2350-2360.
- Farrer, C., & Frith, C. D. (2002). Experiencing oneself vs. another person as being the cause of an action: The neural correlates of the experience of agency. *Neuroimage*, 15, 596-603.
- Ferstl, E. C., & Von Crammond, D. Y. (2002). What does the frontomedian cortex contribute to language processing: Coherence of theory of mind. *Neuroimage*, 17, 1599-1612.
- Field, T., Field, T., Sanders, C., & Nadel, J. (2001). Children with autism display more social behaviors after repeated imitation sessions. *Autism*, 5(3), 317-323.
- Fitzgerald, P. J., Lane, J. W., Thakur, P. H., & Hsiao, S. S. (2006). Receptive

- field properties of the macaque second somatosensory cortex: Representation of orientation on different finger pads. *Journal of Neuroscience*, 26(24), 6473-6484.
- Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, 424, 769-771.
- Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: From action organization to intention understanding. *Science*, 308(5722), 662-667.
- Fogassi, L., & Gallese, V. (2002). The neural correlates of action understanding in non-human primates. In *Mirror Neurons and the Evolution of Brain and Language, Advances in Consciousness Research*, John Benjamins Publishing, 13-35.
- Fogassi, L., Gallese, V., Fadiga, L., Luppino, G., Matelli, M., & Rizzolatti, G. (1996). Coding of peripersonal space in inferior premotor cortex (area F4). *Journal of Neurophysiology*, 76, 141-157.
- Fourneret, P., & Jeannerod, M. (1998). Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia*, 36(11), 1133-1140.
- Franconeri, S. L., & Simons, D. L. (2005). The dynamic events that capture visual attention: A reply to Abrams and Christ. *Perception and Psychophysics*, 67, 962-966.
- Frith, C. D., Blakemore, S.-J., & Wolpert, D. M. (2000). Abnormality in the awareness and control of action. *Philosophical Transactions of the Royal Society London B*, 355, 1771-1788.
- Fu, Q.-G., Flament, D., Coltz, J. D., & Ebner, T. J. (1997). Relationship of cerebellar purkinje cell simple spike discharge to movement kinematics in the monkey. *Journal of Neurophysiology*, 78(1), 478-491.
- Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, 4, 14-21.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119, 593-609.
- Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2, 493-501.
- Gattass, R., Nascimento-Silva, S., Soares, J. G. M., Lima, B., Jansen, A. K., Diogo, A. C. M., et al. (2005). Cortical visual areas in monkeys: Location, topography, connections, columns, plasticity and cortical dynamics. *Philosophical Transactions of the Royal Society of London B*, 709-731.
- Gentilucci, M., Fogassi, L., Luppino, G., Matelli, M., Camarda, R., & Rizzolatti, G. (1988). Functional organization of inferior area 6 in the macaque monkey. I. somatotopy and the control of proximal movements. *Experimental Brain Research*, 71, 475-490.
- Georgopoulos, A. P. (1996). On the translation of directional motor cortical commands to activation of muscles via spinal interneuronal systems. *Cognitive Brain Research*, 3, 151-155.
- Georgopoulos, A. P., Kalaska, J. F., Caminiti, R., & Massey, J. T. (1982). On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *Journal of Neuroscience*, 2, 1527-1537.
- Georgopoulos, A. P., Kettner, R. E., & Schwartz, A. B. (1988). Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. coding of the direction of movement by a neuronal population. *Journal of Neuroscience*, 8, 2928 - 2937.
- Gergely, G., & Csibra, G. (2003). Teleological reasoning infancy: The naïve theory of rational action. *Trends in Cognitive Science*, 7(7), 287-292.

- Gerstner, W., & Kistler, W. M. (2002). Spiking neuron models: Single neurons, populations, plasticity. *Cambridge University Press*.
- Geyer, S., Matelli, M., Lupino, G., & Zilles, K. (2000). Functional neuroanatomy of the primate isocortical motor system. *Anatomy and Embryology*, 202, 443-474.
- Giese, M. A. (2000). Neural model for the recognition of biological motion. In: *Dynamische Perzeption 2*, G. Barattoff and H. Neumann (eds.), *Infix Verlag, Berlin*, 105-110.
- Goldenberg, G., Laimgruber, K., & Hermsdörfer, J. (2001). Imitation of gestures by disconnected hemispheres. *Neuropsychologia*, 39, 1432-1443.
- Goldman-Rakic, P. S. (1996). The prefrontal landscape: Implications of functional architecture for understanding human mentation and the central executive. *Philosophical Transactions of the Royal Society of London B*, 351, 1445-1453.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *TINS*, 15(1), 20-25.
- Goodridge, J. P., & Touretzky, D. S. (2000). Modeling attractor deformation in the rodent head-direction system. *Journal of Neurophysiology*, 83, 3402-3410.
- Goodwin, A. W., & Henry, G. H. (1975). Direction selectivity of simple striate cells: Properties and mechanisms. *Journal of Neurophysiology*, 38, 1524-1540.
- Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, 391, 481-484.
- Graziano, M. S. A., Cooke, D. F., & Taylor, C. S. R. (2000). Coding the localization of the arm by sight. *Science*, 290, 1782-1786.
- Graziano, M. S. A., Hu, X. T., & Gross, C. G. (1997). Visuospatial properties of ventral premotor cortex. *Journal of Neurophysiology*, 77, 2268-2292.
- Graziano, M. S. A., Taylor, C. S. R., & Moore, T. (2002). Complex movements evoked by microstimulation of the precentral cortex. *Neuron*, 34, 841-851.
- Greenwald, A. G. (1970). Sensory feedback mechanisms in performance control: With special reference to the ideo-motor mechanism. *Psychological Review*, 77(2), 73-99.
- Grezes, J., Armony, J. L., Rowe, J., & Passingham, R. E. (2003). Activation related to mirror and canonical neurones in the human brain: An fMRI study. *Neuroimage*, 18, 928-937.
- Gross, C. G., Bender, D. B., & Rocha-Miranda, C. E. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science*, 5, 1303 - 1306.
- Guenther, F., Hersch, M., Calinon, S., & Billard, A. (2007). Reinforcement learning for imitating constrained reaching movements. *RSJ Advanced Robotics, Special Issue on Imitative Robots, In press*.
- Haaland, K. Y., & Harrington, D. L. (1996). Hemispheric asymmetry of movement. *Current Opinions in Neurobiology*, 6, 796-800.
- Haaland, K. Y., Prestopnik, J. L., Knight, R. T., & Lee, R. R. (2004). Hemispheric asymmetries for kinematic and positional aspects of reaching. *Brain*, 127, 1145-1158.
- Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, 5, 382-385.
- Haggard, P., & Clarke, S. (2003). Intentional action: Conscious experience and neural prediction. *Consciousness and Cognition*, 12(4), 695-707.
- Hahnloser, R., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J., & Seung, H. S. (2000). Digital selection and analogue amplification coexist in a

- cortex-inspired silicon circuit. *Nature*, 405, 947-951.
- Hamilton, A., Wolpert, D., & Frith, U. (2004). Your own action influences how you perceive another person's action. *Current Biology*, 14, 493-498.
- Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Selenius, S., & Rizzolatti, G. (1998). Activation of human primary motor cortex during action observation: A neuromagnetic study. *Proceedings of the National Academy of Science USA*, 95, 15061-15065.
- Hasbroucq, T., & Guiard, T. (1991). Stimulus-response compatibility and the Simon effect: Toward a conceptual clarification. *Journal of Experimental Psychology: Human Perception and Performance*, 17(1), 267-266.
- Hebb, D. O. (1949). The organization of behavior: A neuropsychological theory. *Wiley, New York*.
- Hedge, A., & Marsh, N. W. A. (1975). The effect of irrelevant spatial correspondences on two-choice response-time. *Acta Psychologica*, 39, 427-439.
- Hermisdörfer, J., Blankenfeld, H., & Goldenberg, G. (2003). The dependence of ipsilesional aiming deficits on task demands, lesioned hemisphere, and apraxia. *Neuropsychologia*, 41, 1628-1643.
- Hersch, M., Sauser, E., & Billard, A. (2007). Visual and tactile online learning of the body schema. *International Journal of Humanoid Robotics*, Submitted.
- Heyes, C. (2001). Causes and consequence of imitation. *Trends in Cognitive Sciences*, 253-261.
- Heyes, C., Bird, H., G. Johnson, & Haggard, P. (2005). Experience modulates automatic imitation. *Cognitive Brain Research*, 22, 233-240.
- Heyes, C., & Dawson, G. R. (1990). A demonstration of observational learning in rats using a bidirectional control. *Quarterly Journal of Experimental Psychology Section B: Comparative and Physiological Psychology*, 42(1), 59-71.
- Heyes, C., & Ray, E. (2004). Spatial S-R compatibility effects in an intentional imitation task. *Psychonomic Bulletin & Review*, 11(4), 703-708.
- Hick, W. E. (1952). On rate of gain of information. *Quarterly Journal of Experimental Psychology*, 4, 11-26.
- Hietanen, J. K., & Perrett, D. I. (1993). Motion sensitive cells in the macaque superior temporal polysensory area: Lack of response to the sight of the animals own limb movement. *Experimental Brain Research*, 93, 117-128.
- Hietanen, J. K., & Perrett, D. I. (1996). Motion sensitive cells in the macaque superior temporal polysensory area: Response discrimination between self-generated and externally generated pattern motion. *Behavioral Brain Research*, 76, 155-167.
- Hirosaka, O., Nakamura, K., Sakai, K., & Nakahara, H. (2002). Central mechanisms of motor skill learning. *Current Opinion in Neurobiology*, 12, 217-222.
- Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerves. *Journal of Physiology*, 117, 500-544.
- Holmes, N. P., & Spence, C. (2004). The body schema and multisensory representation(s) of peripersonal space. *Cognitive Process*, 5, 94-105.
- Hommel, B., Musseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding: A framework for perception and action planning. *Behavioral and Brain Sciences*, 24, 849-937.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the USA*, 79(8), 2554-2558.

- Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees and children. *Aminal Cogntion*, 8, 164-181.
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2, 359-366.
- Hubel, D., & Wiesel, T. (1977). Ferrier lecture: Functional architecture of the macaque monkey visual cortex. *Proceedings of the Royal Society London B*, 198, 1-59.
- Hyvarinen, J. (1982). Posterior parietal lobe of the primate brain. *Physiological Reviews*, 62, 1060 - 1129.
- Iacoboni, M. (2006). Cause and consequence of the mirror system and its dysfunction. *Trends in Cognitive Science*.
- Iacoboni, M., Koski, L. M., Brass, M., Bekkering, H., Woods, R. P., Dubeau M.-C. Mazziotta, J. C., et al. (2001). Reafferent copies of imitated actions in the right superior temporal cortex. *Proceedings of the National Academy of Science USA*, 98, 13995-13999.
- Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, 286, 2526-2528.
- Ijspeert, A. J., Nakanishi, J., & Schaal, S. (2002). Movement imitation with nonlinear dynamical systems in humanoid robots. In *Proceedings of the IEEE International Conference on Robotics and Automation, ICRA 2002*, 1398-1403.
- Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuoscience*, 2, 193-203.
- Jackson, P. L., Meltzoff, A. N., & Decety, J. (2006). Neural circuits involved in imitation and perspective-taking. *Neuroimage*, 31, 429-439.
- Jeannerod, M. (2003). The mechanism of self-recognition in human. *Behavioral Brain Research*, 142, 1-15.
- Jeannerod, M., & Decety, J. (1995). Mental motor imagery: A window into the representational stages of action. *Current Opinion in Neurobiology*, 5(6), 727-732.
- Jellema, T., Maassen, G., & Perrett, D. I. (2004). Single cell integration and aminate form, motion and location in the superior temporal cortex of the macaque monkey. *Cerebral Cortex*, 14, 781-790.
- Johnson, M. H. (2005). Sub-cortical face processing. *Nature Reviews Neuroscience*, 6, 766-774.
- Johnson, P. B., Ferraina, S., Bianchi, L., & Caminiti, R. (1996). Cortical networks for visual reaching: Physiological and anatomical organization of frontal and parietal lobe arm regions. *Cerebral Cortex*, 6, 102-119.
- Takei, S., Hoffman, D. S., & Strick, P. L. (1999). Muscle and movement representations in the primary motor cortex. *Science*, 285, 2136-2139.
- Takei, S., & Hoffman, P. L., D. S. and Strick. (2001). Direction of action is represented in the ventral premotor cortex. *Nature Neuroscience*, 4(10), 1020-1025.
- Kalaska, J. F., Cohen, D. A. D., Prud'homme, M., & Hyde, M. L. (1990). Parietal area 5 neuronal activity encodes movement kinetics, not movement dynamics. *Neuroscience Research*, 40, 163-175.
- Kandel, E. R., & Jessell, T. M. (1991). Touch. In: *Principle of Neural Science, Third Edition, Appelton & Lange Publishing, Kandel, Schwartz and Jessel Eds., USA*, 367-384.
- Kettner, R. E., Schwartz, A. B., & Georgopoulos, A. P. (1988). Primate motor cortex and free arm movements to visual targets in three- dimensional space. III. positional gradients and population coding of movement direc-

- tion from various movement origins. *Journal of Neuroscience*, 8, 2938-2947.
- Keysers, C., & Perrett, D. I. (2004). Demystifying social cognition: A Hebbian perspective. *Trends in Cognitive Sciences*, 8(11), 501-507.
- Keysers, C., Wicker, B., Gazzola, V., Anton, J.-L., Fogassi, L., & Gallese, V. (2004). A touching sight: SII/PV activation during the observation and experience of touch. *Neuron*, 42(2), 335-346.
- Kilner, J. M., Paulignan, Y., & Blakemore, S. J. (2003). An interference effect of observed biological movement on action. *Current Biology*, 13, 522-525.
- Knoblich, G., & Flach, R. (2001). Predicting the effects of actions: Interactions of perception and action. *Psychological Science*, 12, 467-472.
- Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science*, 297, 846-848.
- Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78(9), 1464-1480.
- Kopocz, K., & Schöner, G. (1995). Saccadic motor planning by integrating visual information and pre-information on neural dynamic fields. *Biological Cybernetics*, 73, 49-60.
- Kornblum, S. (1994). The way irrelevant dimensions are processed depends on what they overlap with: The case of Stroop- and Simon-like stimuli. *Psychological Research*, 56, 130-135.
- Koski, L., Iacoboni, M., Dubeau, M.-C., Woods, R. P., & Mazziotta, J. C. (2003). Modulation of cortical activity during different imitative behaviors. *Journal of Neuropsychology*, 89, 460-471.
- Koski, L., Wohlschläger, A., Bekkering, H., Woods, R. P., Dubeau, M.-C., Mazziotta, J. C., et al. (2002). Modulation of motor and premotor activity during imitation of target-directed actions. *Cerebral Cortex*, 12(8), 847-855.
- Kuniyoshi, Y., Inaba, M., & Inoue, H. (1994). Learning by watching: Extracting reusable task knowledge from visual observation of human performance. *IEEE Transactions on Robotics and Automation*, 10(6), 799-822.
- Laing, C. R., & Troy, W. C. (2003). Two-bump solutions of Amari-type models of neuronal pattern formation. *Physica D*, 3-4, 190-218.
- Lange, F. P., Helmich, R. C., & Toni, I. (2006). Posture influences motor imagery: An fMRI study. *Neuroimage*, 33, 609-617.
- Lauwereyns, J., Koizumi, M., Sakagami, M., Hikosaka, O., Kobayashi, S., & Tsutsui, K. (2000). Interference from irrelevant features on visual discrimination by macaques (*Macaca fuscata*): A behavioral analogue of the human Stroop effect. *Journal of Experimental Psychology: Animal Behavior Processes*, 3, 352-357.
- Lazebnik, Y. (2002). Can a biologist fix a radio? - or, what I learned while studying apoptosis. *Cancer Cell*, 2, 179-182.
- Lhermite, F., Pillon, B., & Serdaru, M. (1986). Human autonomy and the frontal lobes, part I: Imitation and utilization behavior: A neuropsychological study of 75 patients. *Annual Neurology*, 19, 326-334.
- Lukashin, A. V., Amirikian, B. R., Mazhaev, V. L., Wilcox, G. L., & Georgopoulos, A. P. (1996). Modeling motor cortical operation by an attractor network of stochastic neurons. *Biological Cybernetics*, 74, 255-261.
- Luppino, G., & Rizzolatti, G. (2000). The organization of the frontal motor cortex. *News in Physiological Science*, 15, 219-224.
- Lyons, D. E., Santos, L. R., & Keil, F. C. (2006). Reflections of other minds: How primate social cognition can inform the function of mirror neurons. *Current Opinion in Neurobiology*, 16(2), 230-234.

- Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, 9, 1432-1438.
- Mac Kay, W. A., & Crammond, D. J. (1987). Neuronal correlates in posterior parietal lobe of the expectation of events. *Behavioral Brain Research*, 24, 167-179.
- Marsaglia, G. (1972). Choosing a point from the surface of a sphere. *Annual Math Statistics*, 43, 645-646.
- Mataric, M. J., & Pomplun, M. (1998). Fixation behavior in observation and imitation of human movement. *Cognitive Brain Research*, 7, 191-202.
- McCulloch, W., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*(5), 115-133.
- Melamed, O., Gerstner, W., Maass, W., Tsodyks, M., & Markram, H. (2004). Coding and learning of behavioral sequences. *Trends in Neurosciences*, 27(1), 11-14.
- Meltzoff, A. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31, 838-850.
- Meltzoff, A., & Decety, J. (2003). What imitation tells us about social cognition: A rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society of London B*, 358, 491-500.
- Meltzoff, A., & Moore, M. (1997). Explaining facial imitation: A theoretical model. *Early Development and Parenting*, 6, 179-192.
- Meltzoff, A., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, 198, 75-78.
- Meñard, O., & Frezza-Buet, H. (2005). Model of multi-modal cortical processing: Coherent learning in self-organizing modules. *Neural Networks*, 18(5-6), 646-655.
- Miall, R. C. (2003). Connecting mirror neurons and forward models. *Neuroreport*, 14(16).
- Miall, R. C., & Reckess, G. Z. (2002). The cerebellum and the timing of coordinated eye and hand tracking. *Brain and Cognition*, 48, 212-226.
- Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Networks*, 9, 1265-1279.
- Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nature Reviews Neuroscience*, 1, 59-65.
- Mineiro, P., & Zipser, D. (1998). Analysis of direction selectivity arising from recurrent local interactions. *Neural Computation*, 10, 353-371.
- Mountcastle, V. B., Lynch, J. C., Georgopoulos, A., Sakata, H., & Acuna, C. (1975). Posterior parietal association cortex of the monkey: Command functions for operations within extrapersonal space. *Journal of Neurophysiology*, 38(4), 871-908.
- Mühlau, M., Hermsdörfer, J., Goldenberg, G., Wohlschläger, A. M., Castrop, F., Stahl, R., et al. (2005). Left inferior parietal dominance in gesture imitation: An fmri study. *Neuropsychologia*, 43(1), 1086-1098.
- Murata, A., Fadiga, L., Gogassi, L., Gallese, V., Raos, V., & Rizzolatti, G. (1997). Object representation in the ventral premotor cortex (area F5) of the monkey. *Journal of Neurophysiology*, 78, 2226-2230.
- Myowa-Yamakoshi, M., & Matsuzawa, T. (1999). Factors influencing imitation of manipulatory actions in chimpanzees. *Journal of Comparative Psychology*, 113(2), 128-136.
- Nadel, J., Guerini, C., Peze, A., & Rivet, C. (1999). The evolving nature of imitation as a format for communication. In Nadel and Butterworth Eds.:

- Imitation in Infancy*, Cambridge University Press, 209-234.
- Nakamura, K., & Colby, C. L. (2002). Updating of the visual representation in monkey striate and extrastriate cortex during saccades. *Proceedings of the National Academy of Science USA*, 99(6), 4026-4031.
- Nehaniv, C., & Dautenhahn, K. (2002). The correspondence problem. In *Imitation in animals and artifacts*, MIT Press, 41-61.
- Oja, E. (1992). Principal components, minor components, and linear neural networks. *Neural Networks*, 5, 927-935.
- Oram, M. W., & Perrett, D. I. (1996). Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *Journal of Neurophysiology*, 76, 109-129.
- Orban, G. A., Kennedy, H., & Bullier, J. (1986). Velocity sensitivity and direction selectivity of neurons in areas V1 and V2 of the monkey. *Journal of Neurophysiology*, 56(2), 462-480.
- Oztop, E., & Arbib, M. A. (2002). Schema design and implementation of the grasp-related mirror neuron system. *Biological Cybernetics*, 87, 116-140.
- Oztop, E., Kawato, M., & Arbib, M. (2006). Mirror neurons and imitation: A computationally guided review. *Neural Networks*, 19(3), 254-271.
- Oztop, E., Wolpert, D., & Kawato, M. (2005). Mental state inference using visual control parameters. *Cognitive Brain Research*, 22, 129-151.
- Paccalin, C., & Jeannerod, M. (2000). Changes in breathing during observation of effortful actions. *Brain Research*, 862, 194-200.
- Panzeri, S., Rolls, E. T., Battaglia, F., & Lavis, R. (2001). Speed of feedforward and recurrent processing in multilayer networks of integrate-and-fire neurons. *Network: Computation in Neural Systems*, 12, 423-440.
- Perrett, D. I., Harries, M., Mistlin, A. J., & Chitty, A. J. (1989). Three stages in the classification of body movements by visual neurons. In H.B et al. Barlow (ed.), *Images and Understanding*, Cambridge University Press, 94-107.
- Perrett, D. I., Harries, M. H., Bevan, R., Thomas, S., Benson, P. J., Mistlin, A. J., et al. (1989). Frameworks of analysis for the neural representation of animate objects and actions. *Journal of Experimental Biology*, 146, 87-113.
- Perrett, D. I., Harries, M. H., Mistlin, A. J., Hietanen, J. K., Benson, P. J., Bevan, R., et al. (1990). Social signals analyzed at the single cell level: Someone is looking at me, something touched me, something moved! *International Journal of Comparative Psychology*, 4(1), 25-55.
- Petreska, B., Adriani, M., Blanke, O., & Billard, A. (2007). Apraxia. a review. *Progress in Brain Research: From Action to Cognition*.
- Petreska, B., & Billard, A. G. (2006). A neurocomputational model of an imitation deficit following brain lesion. In *Proceedings of 16th International Conference on Artificial Neural Networks (ICANN 2006)*. *Lecture Notes in Computer Science*, 4131, 770-779.
- Petrides, M., & Pandya, D. N. (2002). Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *European Journal of Neuroscience*, 16, 291-310.
- Piaget, J. (1978). La formation du symbole chez l'enfant: Imitation, jeu et rêve, image et représentation. *Delachaux et Niestlé S.A.*.
- Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature*, 400, 233-238.
- Porro, C. A., Francescato, M. P., Cettolo, V., Diamond, M. E., Baraldi, P., Zuiani, C., et al. (1996). Primary motor and sensory cortex activation

- during motor performance and motor imagery: A functional magnetic resonance imaging study. *Journal of Neuroscience*, 16(23), 7688-7698.
- Pouget, A., Dayan, P., & Zemel, R. S. (2003). Inference and computation with population codes. *Annual Reviews Neuroscience*, 26, 381-410.
- Pouget, A., Deneve, S., & Duhamel, J. R. (2002). A computational perspective on the neural basis of multisensory spatial representations. *Nature Review Neuroscience*, 3, 741-747.
- Pouget, A., & Snyder, L. H. (2000). Computational approaches to sensorimotor transformations. *Nature Neuroscience*, 3, 1192-1198.
- Proctor, R. W., & Pick, D. F. (2003). Display-control arrangement correspondence and logical recoding in the Hedge and Marsh reversal of the Simon effect. *Acta Psychologica*, 112, 259-278.
- Proske, U., Wise, A. K., & Gregory, J. E. (2000). The role of muscle receptors in the detection of movements. *Progress in Neurobiology*, 60, 85-96.
- Prud'homme, M. J., & Kalaska, J. F. (1994). Proprioceptive activity in primate primary somatosensory cortex during active arm reaching movements. *Journal of Neurophysiology*, 72(5), 2280-2301.
- Pulvermüller, F. (2002). A brain perspective on language mechanisms: From discrete neuronal ensemble to serial order. *Progress in Neurobiology*, 67, 85-111.
- Raiguel, S. E., Xiao, D.-K., Marcar, V. L., & Orban, G. A. (1999). Response latency of macaque area MT/V5 neurons and its relationship to stimulus parameters. *Journal of Neurophysiology*, 82, 1944-1956.
- Rao, R. P. N., Shon, A. P., & Meltzoff, A. N. (2004). A bayesian model of imitation in infants and robots. *Imitation and Social Learning in Robots, Humans, and Animals: Behavioural, Social and Communicative Dimensions*, K. Dautenhahn and C. Nehaniv (eds.), Cambridge University Press.
- Redish, A. (1999). Beyond the cognitive map: From place cells to episodic memory. Cambridge MA, MIT Press.
- Redish, D., Elga, A. N., & Touretzky, D. S. (1996). A coupled attractor model of the rodent head direction system. *Network: Computation in Neural Systems*, 7(4), 671-685.
- Ribot-Ciscar, E., Bergenheim, M., Albert, F., & Roll, J. P. (2003). Proprioceptive population coding of limb position in humans. *Experimental Brain Research*, 149, 512-519.
- Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neuroscience*, 21(5), 188-194.
- Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., & Matteli, M. (1996). Functional organization of inferior area 6 in the macaque monkey: II. area F5 and the control of distal movements. *Experimental Brain Research*, 71, 491-507.
- Rizzolatti, G., Camarda, R., Fogassi, M., L. Gentilucci, Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. II. area F5 and the control of distal movements. *Experimental Brain Research*, 71, 491-507.
- Rizzolatti, G., Craighero, L., & Fadiga, L. (2002). The mirror system in humans. *In Mirror Neurons and the Evolution of Brain and Language, Advances in Consciousness Research*, John Benjamins Publishing, 37-59.
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3(2), 131-141.
- Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding of actions. *Nature Reviews Neuroscience*,

2, 661-670.

- Rizzolatti, G., Gentilucci, M., Camarda, R., Gallese, V., Luppino, G., & Matelli, M. (1990). Neurons related to reaching-grasping arm movements in the rostral part of area 6 (area 6 $\alpha\beta$). *Experimental Brain Research*, 82, 337-350.
- Rizzolatti, G., & Luppino, G. (2001). The cortical motor system. *Neuron*, 31(7), 889-901.
- Rizzolatti, G., Luppino, G., & Matelli, M. (1998). The organization of the cortical motor system: New concepts. *Electroencephalography and Clinical Neurophysiology*, 106, 283-296.
- Roberts, P. D. (1999). Computational consequences of temporally asymmetric learning rules: I. differential hebbian learning. *Journal of Computational Neuroscience*, 7, 235-246.
- Rochat, P., & Hespos, S. J. (1997). Differential rooting response by neonates: Evidence for an early sense of self. *Early Development and Parenting*, 6, 105-112.
- Roitman, A. V., Pasalar, S., Johnson, M. T. V., & Ebner, T. J. (2005). Position, direction of movement, and speed tuning of cerebellar Purkinje cells during circular manual tracking in monkey. *Journal of Neuroscience*, 25(40), 9244-9257.
- Roll, J.-P., Bergenheim, M., & Ribot-Ciscar, E. (2000). Proprioceptive population coding of two-dimensional limb movements in humans: II. muscle-spindle feedback during "drawing-like" movements. *Experimental Brain Research*, 134, 311-321.
- Rolls, E. T., Stringer, S. M., & Trappenberg, T. P. (2002). A unified model of spatial and episodic memory. *Philosophical Transactions of the Royal Society of London B*, 269, 1087-1093.
- Rougier, N. (2006). Dynamic neural field with local inhibition. *Biological Cybernetics*, 94, 169-179.
- Ruby, P., & Decety, J. (2001). Effect perspective taking during simulation of action: A PET investigation of agency. *Nature Neuroscience*, 4, 546-550.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533-536.
- Rumiati, R. I., Weiss, P. H., Tessari, A., Assmus, A., Zilles, K., Herzog, H., et al. (2005). Common and differential neural mechanisms supporting imitation of meaningful and meaningless actions. *Journal of Cognitive Neuroscience*, 17(9), 1420-1431.
- Rushworth, M. F., Nixon, P. D., & Passingham, R. E. (1997). Parietal cortex and movement. I. movement selection and reaching. *Experimental Brain Research*, 117, 75-82.
- Sakagami, M., Pana, X., & Uttla, B. (2006). Behavioral inhibition and prefrontal cortex in decision-making. *Neural Networks*, 19(8), 1255-1265.
- Sakagami, M., Tsutsui, K., Lauwereyns, J., Koizumi, M., Kobayashi, S., & Hikosaka, O. (2001). A code of behavioral inhibition on the basis of color, but not motion, in ventrolateral prefrontal cortex of macaque monkey. *Journal of Neuroscience*, 21, 4801-4808.
- Sakata, H., Taira, M., Hsounoki, M., Murata, A., & Tanaka, Y. (1997). The parietal association cortex in depth perception and visual control of hand action. *Trends in Neuroscience*, 20, 350-357.
- Sakata, H., Taira, M., Kusunoki, M., Murata, A., Tanaka, Y., & Tsutsui, K. (1998). Neural coding of 3D features of objects for hand action in the parietal cortex of the monkey. *Philosophical Transactions of the Royal Society of London B*, 353, 1363-1373.

- Sakata, H., Taira, M., Kusunoki, M., Murata, A., Tsutsui, K., Tanaka, Y., et al. (1999). Neural representation of three-dimensional features of manipulation objects with stereopsis. *Experimental Brain Research*, 128, 160-169.
- Salinas, E. (2003a). Background synaptic activity as a switch between dynamical states in a network. *Neural Computation*, 15(7), 1439-1475.
- Salinas, E. (2003b). Self-sustained activity in networks of gain-modulated neurons. *Neurocomputing*, 52-54, 913-918.
- Salinas, E., & Abbott, L. F. (1994). Vector reconstruction from firing rates. *Journal of Computational Neuroscience*, 1, 89-107.
- Salinas, E., & Abbott, L. F. (1995). Transfer of coded information from sensory to motor networks. *Journal of Neuroscience*, 15, 6461-6474.
- Salinas, E., & Abbott, L. F. (1996). A model of multiplicative neural response in parietal cortex. *Proceedings of the National Academy of Science USA*, 93, 11956-11961.
- Salinas, E., & Abbott, L. F. (1997). Invariant visual responses from attentional gain fields. *Journal of Neurophysiology*, 77, 3267-3272.
- Salinas, E., & Thier, P. (2000). Gain modulation: A major computational principle of the central nervous system. *Neuron*, 27, 15-21.
- Samson, D., Apperly, I. A., Chiavarino, C., & Humphreys, G. W. (2004). Left tempoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience*, 7(5), 499-500.
- Sausser, E. L., & Billard, A. G. (2004). Three dimensional frames of reference transformations using gain modulated populations of neurons. In *Proceedings of 12th European Symposium on Artificial Neural Networks, Bruges, Belgium*, 543-548.
- Sausser, E. L., & Billard, A. G. (2005a). Three dimensional frames of references transformations using recurrent populations of neurons. *Neurocomputing*, 64, 5-24.
- Sausser, E. L., & Billard, A. G. (2005b). View sensitive cells as a neural basis for the representation of others in a self-centered frame of reference. In *Proceedings of the Third International Symposium on Imitation in Animals and Artifacts, Hatfield, UK*, 119-127.
- Sausser, E. L., & Billard, A. G. (2006a). Biologically inspired multimodal integration: Interferences in a human-robot interaction game. In *proceedings of IROS'2006*.
- Sausser, E. L., & Billard, A. G. (2006b). Dynamic updating of distributed neural representations using forward models. *Biological Cybernetics*, 95(6), 567-588.
- Sausser, E. L., & Billard, A. G. (2006c). Parallel and distributed neural models of the ideomotor principle: An investigation of imitative cortical pathways. *Neural Networks*, 19(3), 285-298.
- Sausser, E. L., & Billard, A. G. (2007). Interferences in the transformation of reference frames during a posture imitation task. In *Proceedings of the International Conference on Artificial Neural Networks, ICANN 07, Porto, Portugal*.
- Schaal, S. (1999). Is imitation the route to humanoid robots. *Trends in Cognitive Sciences*, 3, 233-242.
- Schadlen, M. N., & Newsome, W. T. (2001). Neural basis of perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, 86, 1916-1936.
- Scherberger, H., & Andresen, R. A. (2003). Sensorimotor transformations. In: *Chalupa LM, Werner JS (Eds): The Visual Neurosciences*. MIT Press, 1324-1336.

- Schieber, M. H. (2001). Constraints on somatotopic organization in the primary motor cortex. *Journal of Neurophysiology*, 86(5), 2125-2143.
- Schöner, G. (2002). Dynamical systems approaches to neural systems and behavior. *International Encyclopedia of Social & Behavioral Sciences*, 10571-10575.
- Schutte, A. R., Spencer, J. P., & Schöner, G. (2003). Testing the dynamic field theory: Working memory for locations becomes more spatially precise over development. *Child Development*, 74, 1393-1417.
- Schütz-Bosbach, S., Mancini, B., Aglioti, S. M., & Haggard, P. (2006). Self and other in the human motor system. *Current Biology*, 16, 1830-1834.
- Schwartz, A. B., Kettner, R. E., & Georgopoulos, A. P. (1988). Primate motor cortex and free arm movements to visual targets in three-dimensional space. I. relations between single cell discharge and direction of movement. *Journal of Neuroscience*, 8, 2913-2927.
- Schwartz, A. B., & Moran, D. W. (1999). Motor cortical activity during drawing movements: Population representation during lemniscate tracing. *Journal of Neurophysiology*, 82, 2705-2718.
- Schweigenhofer, N., Arbib, M. A., & Kawato, M. (1998). Role of the cerebellum in reaching movements in humans. I. distributed inverse dynamics control. *European Journal of Neuroscience*, 10, 86-94.
- Scott, S. H., Gribble, P. L., Graham, K. M., & Cabel, D. W. (2001). Dissociation between hand motion and population vectors from neural activity in motor cortex. *Nature*, 413, 161-165.
- Scott, S. H., & Kalaska, J. F. (1997). Reaching movements with similar hand paths but different arm orientations. I. activity of individual cells in motor cortex. *Journal of Neurophysiology*, 77(2), 826-852.
- Shalden, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, 86, 1916-1936.
- Sharp, P. E., Blair, H. T., & Cho, J. (2001). The anatomical and computational basis of the rat head-direction cell signal. *Trends in Neuroscience*, 24(5), 289-294.
- Shimomura, T., & Mori, E. (1998). Obstinate imitation behaviour in differentiation of frontotemporal dementia from Alzheimer's disease. *The Lancet*, 352, 623-624.
- Simon, J. R., & Berbaum, K. (1990). Effect of conflicting cues on information processing: The Stroop effect vs. the Simon effect. *Acta Psychologica*, 73, 159-170.
- Simon, J. R., Sly, P. E., & Vilapakkam, S. (1981). Effect of compatibility of S-R mapping on reactions toward the stimulus source. *Acta Psychologica*, 47, 63-81.
- Sirigu, A., Duhamel, J. R., Cohen, L., Pillon, B., Dubois, B., & Agis, Y. (1996). The mental representation of hand movements after parietal damage. *Science*, 273, 1564-1568.
- Slachevsky, A., Pillon, B., Fournier, P., Pradat-Diehl, P., Jeannerod, M., & Dubois, B. (2001). Preserved adjustment but impaired awareness in a sensory-motor conflict following prefrontal lesions. *Journal of Cognitive Neuroscience*, 13(3), 332-340.
- Snyder, L. H., Batista, A. P., & Andersen, R. A. (1997). Coding of intention in the posterior parietal cortex. *Nature*, 386, 167-170.
- Snyder, L. H., Grieve, K. L., Brotchie, P., & Andersen, R. A. (1998). Separate body- and world-referenced representations of visual space in parietal cortex. *Nature*, 394, 887-891.

- Stein, J. F. (1989). Representation of egocentric space in the posterior parietal cortex. *Quarterly Journal of Experimental Physiology*, 74, 583-606.
- Stringer, S. M., Rolls, E. T., & Trappenberg, T. P. (2004). Self-organizing continuous attractor networks with multiple activity packets and the representation of space. *Neural Networks*, 17, 5-27.
- Stringer, S. M., Trappenberg, T. P., Rolls, E. T., & Araujo, I. E. T. (2002). Self-organising continuous attractor networks and path integration: One-dimensional models of head direction cells. *Network: Computation in Neural Systems*, 13, 217-242.
- Stürmer, B., Aschersleben, G., & Prinz, W. (2000). Correspondence effects with manual gestures and postures: A study of imitation. *Journal of Experimental Psychology: Human Perception and Performance*, 26(6), 1746-1759.
- Swindale, N. V., & Bauer, H.-U. (1998). Application of kohonen's self-organizing feature map algorithm to cortical maps of orientation and direction preference. *Philosophical Transactions of the Royal Society of London B*, 256, 827-838.
- Taube, J. S., & Bassett, J. P. (2003). Persistent neural activity in head direction cells. *Cerebral Cortex*, 13, 1162-1172.
- Taylor, J. G. (1999). Neural bubble dynamics in two dimensions: Foundations. *Biological Cybernetics*, 80, 393-409.
- Thorndike, E. (1898). Animal intelligence: An experimental study of the associative process in animals. *Psychology Review Monograph*, 2(8), 551-553.
- Tillery, S. I., Soechting, J. F., & Ebner, T. J. (1996). Somatosensory cortical activity in relation to arm posture: Nonuniform spatial tuning. *Journal of Neurophysiology*, 76, 2423-2438.
- Todorov, E. (2000). Direct cortical control of muscle activation in voluntary arm movements: A model. *Nature Neuroscience*, 3(4), 391-398.
- Todorov, E. (2002). Cosine tuning minimizes motor errors. *Neural Computation*, 14, 1233-1260.
- Tomasello, M. (1990). Cultural transmission in the tool use and communicatory signaling of chimpanzees. In Parker and Gibson Eds.: *Language and intelligence in monkeys and apes*, Cambridge University Press, 274-311.
- Tomasello, M., Savage-Rumbaugh, S., & Kruger, A. C. (1993). Imitative learning of actions on objects by children, chimpanzees, and enculturated chimpanzees. *Child Development*, 64(6), 1688-1705.
- Trappenberg, T. (2005). Continuous attractor neural networks. In *Recent Developments in Biologically Inspired Computing*, de Castro & Von Zuben Eds., IDEA Group Publishing.
- Tsunoda, K., Yamane, Y., Nishizaki, M., & Tanifuji, M. (2001). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nature Neuroscience*, 4(8), 832-838.
- Turrigiano, G. G., & Nelson, S. B. (2004). Homeostatic plasticity in the developing nervous system. *Nature Reviews Neuroscience*, 5, 97-107.
- Uddin, L. Q., Kaplan, J. T., Molnar-Szakacs, I., Zaidel, E., & Iacoboni, M. (2005). Self-face recognition activates a frontoparietal "mirror" network in the right hemisphere: An event-related fMRI study. *Neuroimage*, 25(3), 926-935.
- Umiltà, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., et al. (2001). "I know what you are doing": A neuropsychological study. *Neuron*, 31, 155-165.
- Unema, M. M., & Goldberg, M. E. (1997). Spatial processing in the monkey frontal eye field. I. predictive visual responses. *Journal of Neurophysiology*,

78, 1373-1383.

- Van Beers, R. J., Baraduc, P., & Wolpert, D. M. (2002). Role of uncertainty in sensorimotor control. *Philosophical Transactions of the Royal Society of London B*, 357, 1137-1145.
- Van Den Bos, E., & Jeannerod, M. (2002). Sense of body and sense of action both contribute to self recognition. *Cognition*, 85, 177-187.
- Van Rossum, M. C. W., & Renart, A. (2004). Computation with populations codes in layered networks of integrate and fire neurons. *Neurocomputing*, 58-60, 265-270.
- Vercher, J. L., & Gauthier, G. M. (1988). Cerebellar involvement in the coordination control of the oculo-manual tracking system: Effects of cerebellar dentate nucleus lesion. *Experimental Brain Research*, 73, 155-166.
- Vercher, J. L., Gauthier, G. M., Guedon, O., Blouin, J., Cole, J., & Lamarre, Y. (1996). Self-moved target eye tracking in control and deafferented subjects: Roles of arm motor command and proprioception in arm-eye coordination. *Journal of Neurophysiology*, 76(2), 1133-1144.
- Vitay, J., & Rougier, N. P. (2006). Emergence of attention within a neural population. *Neural Networks*, 19(5), 573-581.
- Wadman, W. J., & Gutnick, M. J. (1993). Non-uniform propagation of epileptiform discharge in brain slices of rat neocortex. *Neuroscience*, 52, 255-262.
- Watanabe, M. (1986). Prefrontal unit activity during delayed conditional go/no-go discrimination in the monkey. II. relation to go and no-go responses. *Brain Research*, 382, 15-27.
- Weber, C., Wermter, S., & Elshaw, M. (2006). A hybrid generative and predictive model of the motor cortex. *Neural Networks*, 19, 339-353.
- Wersing, H., Steil, J. J., & Ritter, H. (2001). A competitive-layer model for feature binding and sensory segmentation. *Neural Computation*, 13, 357-387.
- Whiten, A., Custance, D. M., Gomez, J.-C., Teixidor, P., & Bard, K. A. (1991). Imitative learning of artificial fruit processing in children and chimpanzees. *Journal of Comparative Psychology*, 110, 3-14.
- Wilimzig, C., & Schöner, G. (2005). The emergence of stimulus-response associations from neural activation fields: Dynamic field theory. In *Proceedings of the XXVIII Annual Conference of the Cognitive Science Society (CogSci 2005)*, Stresa, Italy, 2359-2364.
- Wilson, H., & Cowan, J. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13, 55-80.
- Wise, S. P., Broussaoud, D., Johnson, P. B., & Caminiti, R. (1997). Premotor and parietal cortex: Corticocortical connectivity and combinatorial computations. *Annual Reviews in Neuroscience*, 20, 25-42.
- Wohlschläger, A., & Bekkering, H. (2002). The role of objects in imitation. In *Mirror Neurons and the Evolution of Brain and Language, Advances in Consciousness Research*, John Benjamins Publishing, 101-113.
- Wohlschläger, A., Gattis, M., & Bekkering, H. (2003). Action generation and action perception in imitation: An instance of the ideomotor principle. *Philosophical Transactions of the Royal Society of London*, 358, 501-515.
- Wolpert, D. (1998). Maintaining internal representations: The role of the superior parietal lobule. *Nature Neuroscience*, 1(6), 529-533.
- Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London*, 358, 593-602.
- Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, 11(7-8), 1317-1329.

- Wu, S., Amari, S., & Nakahara, H. (2002). Population coding and decoding in a neural field: A computational study. *Neural Computation*, 14, 999-1026.
- Xie, X., Hahnloser, R., & Seung, H. (2002). Double-ring network model of the head-direction system. *Physical Review E*, 66, 041902.
- Xing, J., & Andersen, R. A. (2000). Models of the posterior parietal cortex which perform multimodal integration and represent space in several coordinate frames. *Journal of Cognitive Neuroscience*, 12(4), 601-614.
- Yokochi, H., Tanaka, M., Kumashiro, M., & Iriki, A. (2003). Inferior parietal somatosensory neurons coding face-hand coordination in Japanese macaques. *Somatosensory & Motor Research*, 20(2), 115-125.
- Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: Analysis by synthesis? *Trends in Cognitive Sciences*, 10(7), 301-308.
- Zentall, T. R. (2003). Imitation by animals: How do they do it? *Current Directions in Psychological Science*, 12(3), 91-95.
- Zhang, H., Zhang, J., & Kornblum, S. (1999). A parallel distributed processing model of stimulus-stimulus and stimulus-response compatibility. *Cognitive Psychology*, 38, 386-432.
- Zhang, K. (1996). Representation of the spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. *Journal of Neuroscience*, 16(6), 2112-2126.
- Zhang, K., & Sejnowski, T. J. (1999). A theory of geometric constraints on neural activity for natural three-dimensional movement. *Journal of Neuroscience*, 19, 3122-2145.
- Zukow-Goldring, P. (2004). Caregivers and the education of the mirror system. *In Proceedings of the international conference on development and learning (ICDL), La Jolla, USA*, 96-103.

Eric Sauser

3 May 1979
Swiss Citizen
Single

Rue Baulacre, 9
1202 Geneva
Switzerland
☎ +41 76 3942769
✉ eric.sauser@a3.epfl.ch

Education

- 2003–2007 **PhD Student in Computational Neuroscience**, *Learning Algorithms and Systems Laboratory (LASA) at the Swiss Federal Institute of Technology Lausanne (EPFL)*, Switzerland.
► Neural Modeling of the Neural Mechanisms underlying Imitative Behaviors:
A Dynamic Neural Field Approach.
Thesis advisor: Prof. Aude Billard
- 1998–2003 **MSc-BSc in Computer Science**, *EPFL*, Switzerland.

Professional Experience

- 2003–2007 **Research Assistant**, *LASA, EPFL*, Lausanne.
► Supervision of undergraduate and master students
► Software development, system and web administration
- 2000–2003 **Student Assistant**, *EPFL*, Lausanne.
► Teaching assistantship in courses on *Software engineering*, *Advanced computer architecture*, and *Programming*

Research Interests

- Artificial Intelligence
- Computational Neuroscience
- Autonomous Robotics
- Dynamical and Adaptive Systems
- Imitation
- Machine Learning

Publications

- Journals Hersch, M. and Sauser, E. and Billard, A. G. Visual and tactile online learning of the body schema. *International Journal of Humanoid Robotics*, Submitted.
- Sauser, E. L. and Billard, A. G. Dynamic updating of distributed neural representations using forward models. *Biological Cybernetics*, 95(6):567-588, 2006.
- Sauser, E. L. and Billard, A. G. Parallel and distributed neural models of the ideomotor principle: An investigation of imitative cortical pathways. *Neural Networks*, 19(3):285-298, 2006.
- Sauser, E. L. and Billard, A. G. Three dimensional frames of references transformations using recurrent populations of neurons. *Neurocomputing*, 64:5-24, 2005.
- Floreano, D., Kato, T., Marocco, D. and Sauser, E. Coevolution of active vision and feature selection. *Biological Cybernetics*, 90(3):218-228, 2004.
- Peer-reviewed proceedings Sauser, E. L. and Billard, A. G. Interferences in the transformation of reference frames during a posture imitation task. *In Proceedings of the International Conference on Artificial Neural Networks, ICANN 07, Porto, Portugal*, 2007.

Sauser, E. L. and Billard, A. G. Biologically inspired multimodal integration: Interferences in a human-robot interaction game. *In Proceedings of IROS'2006, Beijing, China*, 2006.

Sauser, E. L. and Billard, A. G. View sensitive cells as a neural basis for the representation of others in a self-centered frame of reference. *In Proceedings of the Third International Symposium on Imitation in Animals and Artifacts, Hatfield, UK*, 119-127, 2005.

Sauser, E. L. and Billard, A. G. Three dimensional frames of reference transformations using gain modulated populations of neurons. *In Proceedings of 12th European Symposium on Artificial Neural Networks, Bruges, Belgium*, 543-548, 2004.

Chavarriaga, R., Sauser, E. and Gerstner, W. Modeling directional firing properties of place cells. *Twelfth Annual Computational Neuroscience Meeting CNS*2003, Alicante, Spain*, 2003.

Conference Abstracts Sauser, E. L. and Billard, A. G. Dynamic selectivity in a continuous attractor model of movement generation. *Computational Neuroscience Meeting, CNS*06, Workshop on Continuous Attractor Neural Networks, Edinburgh, UK.*, 2006.

Sauser, E. L. and Billard, A. G. Ideomotor compatibility: investigating Imitative Cortical Pathways. *EPFL-LATSIS Symposium*, 2006.

Sauser, E. L. and Billard, A. G. Neural models of the ideomotor principle: An investigation of imitative cortical pathways. *ESF Research Conference on Brain Development and Cognition in Human Infants, Maratea, Italy*, 2005.

Sauser, E. L. and Billard, A. G. Neural model of the transformation from allo-centric to ego-centric representation of motions. *Fourteenth Annual Computational Neuroscience Meeting CNS*2005, Madison, USA*, 2005.

Sauser, E. L. and Billard, A. G. Three dimensional frames of reference transformations using recurrent populations of neurons. *Eighth International Conference on Cognitive and Neural Systems, ICCNS, Boston, USA*, 2004.

Sauser, E. L. and Billard, A. G. View sensitive cells as a neural basis for frames of reference transformations: A computational model. *nEuro-IT.net workshop at NIPS: The Brain Mechanisms of Imitation Learning, Whistler, Canada*, 2004.

Languages

French	Mother Tongue
English	Fluent
German	Scholar knowledge
Chinese	Good spoken skills

Computer skills

Office	Word, Excel, Latex	Graphics	Illustrator, Photoshop
Languages	C++, Java, Matlab, OpenGL, HTML, PHP, SQL, VHDL, Lisp, Prolog	O.S.	Windows, Linux

Hobbies

Music	Guitar, Drum, Computer-assisted music
Sports	Swimming, Bicycle
Cooking	Chinese food