



HAL
open science

Apprentissage visuo-moteur, implication pour le développement sensorimoteur et l'émergence d'interactions sociales

Nils Beaussé

► **To cite this version:**

Nils Beaussé. Apprentissage visuo-moteur, implication pour le développement sensorimoteur et l'émergence d'interactions sociales. Intelligence artificielle [cs.AI]. Université de Cergy Pontoise, 2019. Français. NNT : 2019CERG1051 . tel-02945304

HAL Id: tel-02945304

<https://theses.hal.science/tel-02945304>

Submitted on 22 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université de Cergy-Pontoise - École doctorale Économie, Management, Mathématiques,
Physique et Sciences informatiques

THÈSE

APPRENTISSAGE VISUO-MOTEUR, IMPLICATION POUR LE DÉVELOPPEMENT SENSORI-MOTEUR ET L'ÉMERGENCE D'INTERACTIONS SOCIALES

par

Nils Beaussé

ETIS - ENSEA / Université de Cergy-Pontoise / CNRS UMR 8051
6 avenue du Ponceau, 95014 Cergy-Pontoise Cedex, France

Soutenue le 19/12/2019 devant le jury composé de :

SAMER ALFAYAD	Laboratoire LISV	Rapporteur
ARNAUD REVEL	Laboratoire L3i	Rapporteur
YVONNE DELEVOYE-TURRELL	Laboratoire SCALab	Examinatrice
PATRICK HÉNAFF	Neurosys - Laboratoire LORIA	Examineur
MATHIAS QUOY	Université de Cergy-Pontoise	Examineur
EMMANUEL GARDINETTI	Direction Générale de l'Armement	Membre Invité
PHILIPPE GAUSSIER	Université de Cergy-Pontoise	Directeur de thèse
GHILÈS MOSTAFAOUI	Université de Cergy-Pontoise	Co-encadrant de thèse

Remerciements

Cette thèse s'est déroulée sur un domaine très particulier, dans une équipe mêlant de nombreux aspects pluridisciplinaires à des techniques de pointe en robotique et en apprentissage. Je pense honnêtement que peu d'équipes en France mêlent autant d'aspects différents, et aussi intéressants, dans des travaux en robotique aussi diversifiés. De fait, je considère comme une grande chance le fait d'avoir pu mener cette thèse. Je remercie donc non seulement mon directeur de thèse Philippe Gaussier, et mon encadrant Ghilès Mostafaoui, qui ont été à mes côtés et m'ont fait l'honneur de me sélectionner comme candidat pour cette thèse, mais je remercie également Mathias Quoy le directeur actuel du laboratoire, et Inbar Fijalkow l'ancienne directrice du laboratoire.

Je tiens à remercier dans cet esprit les personnes qui ont bien voulu soutenir mon dossier de thèse à la sortie de l'école d'ingénieur ENSEA, notamment Laurence Hafemeister, l'ancienne directrice, que j'apprécie beaucoup et Sophie Olijnyk, ma professeuse principale, que je remercie au passage pour ses très bons cours de danse durant des années. Je remercie également mes deux anciens maîtres de stages au sein du laboratoire ETIS, Pierre Andry et Arnaud Blanchard, qui m'ont permis de m'intégrer à l'équipe et m'ont offert une base de connaissance en lien avec les travaux du laboratoire qui m'a beaucoup aidé dans les débuts de ma thèse. Je pense honnêtement que beaucoup de directions prises dans mes travaux sont le fruit de leurs visions initiales sur notre domaine. De même, je remercie la DGA d'avoir accepté de financer ce dossier de thèse, notamment Mr Didier Bazalgette qui était le responsable de l'époque et Mr Emmanuel Gardinetti qui a pris le relais et suivi ma thèse récemment.

De façon plus personnelle, je tiens à remercier mes collègues qui m'ont soutenu durant ces années, cette thèse aurait été bien plus morne sans leur humour et leur camaraderie. Je pense notamment à Arnaud Blanchard, Raphaël Braud, Julien Abrosimov, Aliaa Moualla et Éva Ansermin qui ont été au cœur de mes relations de thèse au laboratoire et qui sont devenus très chers à mes yeux. Mais je tiens à remercier également Lise Aubin, Marwen Belkaid, Souheil Hanoune, Ali Karaouzene, Caroline Lesueur-Grand, Artem Melnyk et Anna Pugach, des collègues que j'ai beaucoup appréciés durant leur période de passage au laboratoire ETIS.

Au sein de l'équipe Neurocyber' et des autres équipes du laboratoire ETIS j'ai également été entouré de personnes chaleureuses et intéressantes que j'ai beaucoup appréciées et avec qui j'ai eu de nombreuses discussions très riches professionnellement et personnellement. Je pense notamment à Jérôme Fellus, Pierre Andry, Boris Borzic, Alexandre Pitti et également à Sofiane Boucenna.

Je tiens à remercier chaleureusement certains stagiaires qui sont restés dans mon cœur malgré, pour certain, plusieurs années passées sans les voir. Je pense notamment à Oriane Dermey (aujourd'hui au laboratoire Loria), Julien Dupeyroux, Zinedine Galouz qui m'a beaucoup aidé sur la fin de ma thèse et qui est aujourd'hui ingénieur au laboratoire, et enfin, Sébastien Barré avec qui j'ai travaillé sur un prototype de main robotique (ces travaux ne sont malheureusement pas décrits dans cette thèse).

Je tiens à remercier ma famille, notamment mon père, Bruno Beaussé et ma mère, Bénédicte Pourquier, ainsi que mes sœurs Aéla et Thaïs pour leur affection et leur soutien. Je remercie également mes « beaux-parents » : Françoise et Jean-Marie Manceaux que j'apprécie beaucoup.

Pour finir, je tiens à remercier celle qui est sûrement pour moi la personne la plus importante durant cette thèse, ma compagne Justine Manceaux, qui a supporté durant toutes ces années les contraintes de la thèse et mon humeur. Je lui exprime ici tout mon amour et j'espère passer une longue et heureuse vie à ses côtés.

Résumé

Cette thèse tente d'apporter des éléments de réponse à la question de l'apprentissage et du développement sensorimoteur lors d'une interaction entre humain et robot dans un environnement réel non contraint. Pour ce faire nous défendons dans cette thèse le fait que les êtres humains interagissent non seulement de façon intentionnelle ou consciente, mais également que les propriétés de leur système moteur à bas niveau, de leur corps, et des boucles d'apprentissage sensorimoteur, leur permettent de faciliter implicitement et naturellement cette interaction. Ainsi, nous abordons cette question par l'étude des propriétés du système sensorimoteur chez l'humain et notamment des mécanismes de développement de certaines de ces propriétés chez l'enfant.

Dans un premier temps, nous étudions les propriétés du robot Tino qui est un prototype de robot humanoïde hydraulique unique en France et qui représente la plateforme expérimentale principale utilisée durant cette thèse. Nous avons ainsi analysé finement certaines des propriétés intéressantes du prototype qui peuvent avoir une analogie avec les propriétés du système moteur humain présentant un intérêt implicite dans un cadre d'interaction avec l'environnement ou avec un partenaire. En accord avec le paradigme de « l'intelligence du corps », nous montrons comment certaines de ces propriétés peuvent être exploitées pour simplifier la tâche au système de contrôle. Nous étudions également les limites de cette analogie et de son exploitation.

Nous étudions ensuite dans cette thèse la modélisation des boucles motrices bas niveaux et des propriétés du système musculaire humain afin d'en isoler les propriétés utiles à l'interaction. Nous en proposons une implémentation sur robot et analysons les propriétés du système de contrôle proposé en simulation et sur la plateforme robotique Tino en lien avec les caractéristiques spécifiques de cette dernière. Par la suite nous développons une architecture neuronale bio-inspirée et développementale capable d'apprendre des associations visuo-motrices lors de phases de babillage moteur (« babbling »). Nous montrons à l'aide de ce modèle implémenté sur le robot Tino l'émergence implicite d'interactions sociales grâce aux propriétés des boucles sensorimotrices.

Plus précisément, l'ambiguïté perceptive appliquée à cet algorithme permet l'émergence d'une imitation immédiate et d'un geste de pointage. Ce modèle utilise un système d'apprentissage associatif simple qui ne fournit qu'un « répertoire d'action ». Il est incapable de réagir finement aux contraintes de l'environnement ou d'un partenaire et il est notamment incapable de prendre en compte l'environnement pour planifier des trajectoires cohérentes dans celui-ci. Nous avons alors développé, à travers deux simulations de complexité croissante, un modèle d'apprentissage par renforcement bio-inspiré pour permettre à notre architecture d'apprendre à planifier la trajectoire permettant par exemple d'atteindre et de déplacer un objet présent dans l'environnement et posé sur une table.

Nous montrons ensuite les analogies entre ce modèle et les expériences sur la prise en compte de l'intention dans les actions motrices et les trajectoires produites chez l'humain.

Enfin, nous nous sommes intéressés aux dynamiques d'interactions entre humains et à l'apport que pourrait amener l'intégration de contrôleurs neuronaux oscillatoires dans les architectures de contrôle sensorimoteur, pour ce faire nous avons proposé plusieurs types de modèles oscillants capables d'apprendre et de s'adapter en fonction de la dynamique des informations entrantes, afin de pouvoir s'adapter à des architectures bio-inspirées se développant en interaction avec un environnement non contraint.

Abstract

This thesis tries to bring answers to the question of the sensorimotor learning and development in the context of human-robot interactions in a real non-constrained environment. To achieve this goal we defend in this thesis the fact that human being interacts through intentional and conscious strategy but also depends on the properties of their low-level motor system, their body, and of their sensorimotor learning loops, allowing these to facilitate implicitly this interaction.

We try to answer these questions through the study of the sensorimotor loops in humans, and through the study of the development of these properties in infants.

First, we study here the properties of our robot "Tino", which is a prototype of a humanoid hydraulic robot, unique in France and which is the main experimental platform used in this thesis. We analyse in this thesis the property of this robot and made analogies with the human motor system properties that are implied in the interaction between human, the environment and other humans. We show how certain of these properties could be used to simplify tasks for the control system. We study finally the limit of this analogy and of the exploitation of these properties.

After this part we study in this thesis the modelling of low-level motor loop and of the properties of the human muscular system in order to capture the main interesting properties for interactions. We propose an implementation on robots and analyses the properties of this control system in simulations and on the robotic platform Tino.

Then, we propose a bio-inspired and developmental neural architecture that is able to learn visuomotor association with babbling exploration of the environment. We show with this model implemented on the robot Tino that we can observe the emergence of implicit social interaction through the sensorimotor loops, such are imitation and pointing gestures.

But this model uses a simple associative learning which is able to construct an "action repertoire" but is unable to react to the environment and humans finely. To solve this problem we have developed, through two simulations, a learning model based on reinforcement learning to allow our system to produce coherent trajectory in order to act in an environment. We applied this in a simulated task of grasping and moving an object on a table. We show then the analogies between this model and historic experiments about the impact of intention on the motor actions and trajectories in humans.

Finally, we study in this thesis the dynamic of interactions and the interest of bringing oscillatory neural network in these sensorimotor architectures. To this end we propose in this thesis several oscillatory models able to learn and to adapt in the context of bio-inspired architecture that learn in interaction with a real environment.

Table des matières

Remerciements	iii
Résumé	v
Abstract	vii
Table des matières	ix
Table des figures	xiii
Introduction	1
I. Tino : Le robot humanoïde	7
1. Le choix d'un robot humanoïde hydraulique comme plateforme expérimentale	8
1.1. Introduction	9
1.2. Environnement simulé et plateforme robotique réelle : avantages et inconvénients	11
1.2.1. Les avantages d'un environnement simulé	11
1.2.2. L'apprentissage long sur robot	11
1.2.3. L'intelligence du corps	12
1.2.4. Le problème de la régularité en simulation	13
1.2.5. La physicalité de l'interaction	13
1.2.6. Conclusion	13
1.3. Pourquoi un robot humanoïde ?	15
1.3.1. La vallée de l'étrange	15
1.3.2. Des indices biomimétiques pointant la non-optimalité du choix d'une forme humanoïde	16
1.3.3. Les avantages d'un robot humanoïde	16
1.4. Pourquoi un contrôle hydraulique ?	18
1.4.1. Avantages et inconvénients des différents types de moteurs	18
1.4.2. La problématique de l'effort subit avec les moteurs électriques	19
1.4.3. La solution « compliant »	19
1.4.4. Hydraulique ou pneumatique : la question de la sécurité	20
1.4.5. Conclusion	20
1.5. Conclusion	21
2. Caractérisation du robot humanoïde hydraulique Tino	22
2.1. Introduction	23
2.2. Description des parties du robot Tino	24
2.2.1. Le torse	24
2.2.2. La tête expressive	26
2.2.3. Matériel de commande et de capture	28
2.3. Caractérisation expérimentale du robot Tino	30
2.3.1. Caractérisation de la tête du robot Tino	30
2.3.2. Précision des capteurs du robot Tino	33

2.3.3. Relation commande/mouvement	34
2.3.4. Capacité de mouvement du robot Tino	35
2.4. Phénomènes émergents de la mécanique et du contrôle	39
2.4.1. Validation du phénomène à vide	39
2.4.2. Variation du phénomène selon le poids	40
2.4.3. Validation du phénomène dans un cas plus général	40
2.4.4. Conclusion sur les phénomènes émergents du contrôle	41
2.5. Acceptabilité du robot Tino	43
2.5.1. Présentation résumée du protocole	43
2.5.2. Résultats et interprétations	43
2.6. Conclusion	45

II. Architecture bio-inspirée pour le contrôle d'un robot 46

3. État de l'art sur le contrôle sensorimoteur et positionnement 47

3.1. Introduction	48
3.2. Types de modèles existants	49
3.2.1. Contrôle optimal	49
3.2.2. Apprentissage et adaptabilité : existence de nombreux modèles	52
3.3. Propriétés du système nerveux central et contrôle moteur chez l'humain	57
3.3.1. Cortex visuel	57
3.3.2. Thalamus	58
3.3.3. Cortex associatif pariétal	58
3.3.4. Cortex moteur et prémoteur	58
3.3.5. Ganglions de la base	58
3.3.6. Hypothalamus	59
3.3.7. Hippocampe	59
3.3.8. Cervelet	60
3.3.9. Développement du contrôle moteur chez l'enfant	63
3.3.10. Positionnement et choix des boucles sensorimotrices à modéliser	64
3.4. Conclusion	66

4. Modèle développemental d'un contrôleur sensorimoteur 67

4.1. Première architecture naïve	68
4.1.1. Concept général	68
4.1.2. Protocole expérimental	68
4.1.3. Architecture	68
4.1.4. Résultats expérimentaux	71
4.2. Contrôle musculaire et modèle de contrôle bas niveau pour robot	75
4.2.1. Le muscle seul et ses systèmes de régulations	75
4.2.2. Le jeu des muscles agonistes-antagonistes : co-contraction (ou co-activation musculaire)	79
4.2.3. Point de vue du roboticien sur les modèles de muscles	80
4.2.4. Modèle de co-contraction musculaire pour robot	80
4.2.5. Conclusion	87
4.3. Modèle développemental pour l'apprentissage de liens perceptions-actions	88
4.3.1. Description du modèle	88
4.3.2. Résultats et convergence de l'algorithme	96
4.3.3. Imitation immédiate : effet émergent du modèle en situation d'ambiguïté perceptive	99
4.3.4. Analyse et critiques, de l'algorithme et perspective pour une évolution du modèle	105
4.4. Conclusion	119

III. Architecture de contrôle par renforcement **120**

5. Ganglions de la base : propriétés et modélisations	121
5.1. Introduction	122
5.2. Les noyaux des ganglions de la base et leurs influences mutuelles	123
5.2.1. Voie directe, indirecte et liens topologiques : du striatum au thalamus	124
5.3. Les boucles parallèles des ganglions de la base	127
5.4. L'apprentissage dans les ganglions de la base	128
5.4.1. Substance noire compacte et aire tegmentale ventrale	128
5.4.2. Dopamine, récompense négative et récepteur de type D1/D2	129
5.5. L'apprentissage dans les ganglions de la base, modèle et similitude avec l'apprentissage par renforcement	131
5.5.1. Lien entre les ganglions de la base et les algorithmes de type TD-Lambda	131
5.5.2. Les modèles de type acteur-critique : un modèle général de l'apprentissage dans les ganglions de la base	132
5.6. Conclusion	135
6. Apprentissage par renforcement via un modèle acteur-critique	136
6.1. Introduction	137
6.2. Modèle acteur-critique	138
6.3. Premier environnement simulé, contexte et résultats	142
6.3.1. Description de l'environnement	142
6.3.2. Validation du fonctionnement	143
6.3.3. Contexte et choix d'expérience	144
6.3.4. Premiers résultats et variante sans punition explicite	145
6.3.5. Contrôle en force : la problématique de l'apprentissage incrémental et du format des sorties	147
6.3.6. « Contexte social » : résultats préliminaires	155
6.3.7. Conclusion	158
6.4. Second environnement : simulation 2D d'un bras robotique	159
6.4.1. Description technique de la simulation du bras	159
6.4.2. Simulation du bras et résultats	165
6.5. Analyse et critiques	168
6.5.1. L'influence des méta-paramètres et leurs crédibilités neurobiologiques	168
6.5.2. Rapidité de convergence, un frein au développement sur robot ?	169
6.5.3. L'hypothèse sociale, une hypothèse simpliste qui mériterait d'être affinée	169
6.6. Conclusion	171

IV. Synchronisation non-intentionnelle et réseaux oscillants **172**

7. Synchronisation non-intentionnelle, impact chez l'humain et usage en robotique	173
7.1. Introduction	174
7.2. La synchronisation non-intentionnelle chez l'humain	175
7.2.1. Modèle de synchronisation pour la coordination	176
7.2.2. Diverses démonstrations de la validité du modèle HKB	177
7.2.3. L'importance du type de couplage	178
7.3. Synchronisation et robotique	180
7.3.1. Introduction : des applications et des questions très diversifiées	180
7.3.2. Travaux de caractérisation de la synchronisation non-intentionnelle chez l'humain	181
7.3.3. Les réseaux de neurones oscillants, un modèle pour la synchronisation en robotique	185
7.3.4. Discussion sur la synchronisation en robotique	188
7.4. Conclusion. L'étude des réseaux oscillants, un intérêt pour la robotique interactive ?	194

8. Apprentissage dans les réseaux de neurones oscillants appliqués à la robotique	195
8.1. Introduction	196
8.2. L'application d'oscillateurs neuronaux dans un cadre robotique : emplacement et contraintes	197
8.2.1. Problématique de l'emplacement	197
8.2.2. Problématique du lien entre unités inhibitrices et activatrices	197
8.2.3. Le problème pragmatique de la création de multiples couples oscillants avec les modèles classiques	198
8.3. Apprentissage et adaptation dans les réseaux oscillants	199
8.4. Différentes propositions de modèles	200
8.4.1. Perspectives pour le modèle, quelles applications?	212
8.5. Conclusion	216
Conclusion	217
Annexes	220
A. Protocole d'expérimentation au laboratoire partenaire Ureca de Lille 3	221
Bibliographie	228

Table des figures

1.1. Le robot Tino	9
<hr/>	
2.1. Buste du robot Tino avec sa tête	23
2.2. Buste du robot Tino sans sa tête	24
2.3. Détails du bras du robot Tino	25
2.4. Schéma du fonctionnement des servovalves	26
2.5. Tête expressive complète du robot Tino	27
2.7. Tête de Tino sans expression faciale	27
2.6. Schéma de la tête de Tino	28
2.8. Architecture de contrôle des yeux et du cou simplifiée	30
2.9. Courbe de réponse des yeux face à un créneau	31
2.10. Protocole de caractérisation de la réactivité du cou et des yeux	31
2.11. Résultats extremum des expériences de caractérisation de la tête du robot Tino	32
2.12. Résultats des expériences de caractérisation de la tête de Tino : Participation œil vs cou	32
2.13. Courbes comparant les données reconstruites depuis les caméras 3D et les capteurs pour l'épaule droite	33
2.14. Courbes comparant les données reconstruites depuis les caméras 3D et les capteurs pour le coude droit	34
2.15. Courbes illustrant le lien entre commande et résultat observé sur le bras de Tino : courbe 1	35
2.16. Courbes illustrant le lien entre commande et résultat observé sur le bras de Tino : courbe 2	36
2.17. Courbes illustrant le lien entre commande et résultat observé sur le bras de Tino : courbe 3	37
2.18. Courbes illustrant le lien entre commande et résultat observé sur le bras de Tino : courbe 4	37
2.19. Courbes de réponse angulaire à différents créneaux	38
2.20. Courbes de comparaison entres des trajectoires de sujets humains et des trajectoires effectuées par le robot Tino	38
2.21. Courbes superposées des commandes envoyées et des courbes résultantes	39
2.22. Courbe présentant la relation entre commandes et positions finales à vide	40
2.23. Ensemble de courbes représentant la relation entre commandes et positions finales pour différents poids	41
2.24. Protocole de levée manuelle du bras du robot Tino	42
2.25. Bornes de stabilité min et max pour différents poids	42
2.26. Expériences conjointes au laboratoire SCALab de Lille 3	43
<hr/>	
3.1. Principe de l'apprentissage par renforcement	53
3.2. Résumé des zones cérébrales ayant un lien avec la commande motrice et des liens principaux entre ces zones	57
3.3. Boucle sensorimotrice à modéliser : concept	65
<hr/>	
4.1. Protocole d'association position → commande	69
4.2. Architecture naïve d'association position-commande	70
4.3. Erreur de reconstruction durant la progression de l'apprentissage	71
4.4. Erreur de reconstruction relative durant la progression de l'apprentissage	72
4.5. Superposition des courbes de positions stables et de la séquence de commande	73

4.6.	Erreur superposée avec les positions de stabilisations	74
4.7.	Analogie entre système musculaire et différents types de modèles	81
4.8.	Évolution de θ_{eq} et A^+ tout au long d'une convergence simulée	84
4.9.	Évolution de l'erreur de θ_{eq} et A^+ tout au long d'une convergence simulée	85
4.10.	Croissance de la force vis-à-vis de la croissance de θ_{eq}	86
4.11.	Simulation de la co-contraction soumise à un bruit blanc additif	87
4.12.	Projection visuelle dans un espace égocentrique	89
4.13.	Construction de la commande du cou et des yeux	90
4.14.	Architecture d'apprentissage	91
4.15.	Protocole d'exploration	95
4.16.	Évolution du nombre de neurones recrutés en fonction du temps sur les couches de catégorisations	96
4.17.	Erreur algébrique de reconstruction dans le temps pour un axe	97
4.18.	Évolution de la moyenne de modification des poids de la couche associative (pour un axe)	98
4.19.	Erreur de reconstruction moyenne et écart-type	98
4.20.	Résultat de l'imitation : photo	101
4.21.	Résultat de l'imitation du point de vue du robot	102
4.22.	Résultat de l'imitation au cours du temps du point de vue du robot et du point de vue extérieur	103
4.23.	Tino dans un contexte de geste de pointage émergent	103
4.24.	Pavage de l'erreur dans l'espace visuel du robot	104
4.25.	Force de co-contraction et position d'équilibre durant l'apprentissage	106
4.26.	Résultat de l'imitation du point de vue du robot pour 5 catégories apprises uniquement	107
<hr/>		
5.1.	Disposition générale des ganglions de la base	124
5.2.	Voie directe et indirecte dans les ganglions de la base	125
5.3.	Réseau d'apprentissage dans les ganglions de la base	130
5.4.	L'interprétation acteur-critique des ganglions de la base	132
<hr/>		
6.1.	Modèle acteur critique proposé	138
6.2.	Environnement simulé 2D	142
6.3.	Résultat de la propagation de récompense visualisé en deux dimensions	143
6.4.	Convergence vers la ligne droite	144
6.5.	Environnement modifié avec frottement	145
6.6.	Trajectoire « en cloche » en cas de punition due au frottement	146
6.7.	Résultat dans le cas d'un frottement sans punition	147
6.8.	Illustration du problème de la commande directe en force	147
6.9.	Résultat dans le cas d'un effecteur analogique non segmenté	148
6.10.	Différents types d'effecteurs analogiques pour la commande en force	149
6.11.	Réseau acteur pour un effecteur analogique par partie	150
6.12.	Résultat de trajectoire pour un effecteur analogique par partie	151
6.13.	Résultat des trajectoires pour un effecteur analogique et une zone de récompense	152
6.14.	Résultat des trajectoires pour un effecteur analogique et une catégorisation fine	153
6.15.	Résultat de trajectoire pour un effecteur analogique et un critère d'énergie	154
6.16.	Intégration comparative des forces au cours du temps avec et sans critères d'énergies	155
6.17.	Résultat de trajectoire dans un contexte « social »	157
6.18.	Comparaison des hauteurs relatives social/non-social avec les travaux de Becchio et collègues	158
6.19.	Schéma du pendule double et des valeurs utilisées dans les équations du bras	159
6.20.	Bilan des forces appliquées sur les segments du bras au niveau de leur centre de gravité	160
6.21.	La force de réaction du support appliquée par la table sur le bras, normale à la surface de la table	163
6.22.	La répartition des masses sur les deux segments de bras.	163
6.23.	Illustrations des équations	164

6.24. Amplitude de mouvement de l'épaule et du coude	165
6.25. Résultat du modèle appliqué à la simulation du bras	166
6.26. Résultat du modèle appliqué à la simulation du bras incluant toutes les variantes	167
6.27. Résultat du modèle appliqué à la simulation du bras dans un contexte social	167
<hr/>	
7.1. Modèle des effets de synchronisation dans les coordinations interpersonnelles	177
7.2. Principe de l'interaction bidirectionnelle	178
7.3. Photographie du protocole d'interaction	182
7.4. Protocole d'interaction	183
7.5. Paliers de fréquences mesurées	184
7.6. Résultat de la phase relative et de la fréquence instantanée	185
7.7. Résultat du taux de synchronisation selon les fréquences	186
7.8. Résultats pour les conditions ascendantes et descendantes.	186
7.9. Modèle oscillant simple	187
7.10. Réseau de contrôle robotique pour la synchronisation immédiate	188
7.11. Comparaison entre le système classique de synchronie par réseaux oscillants et le système employant des réflexes	191
7.12. Résultat synchronisation sans CPG	192
7.13. Évolution de la vitesse de déplacement de la différence de phase	193
<hr/>	
8.1. Problématique de l'emplacement dans le positionnement des CPG dans les modèles	197
8.2. Problématique de la connexion des interneurons face à un signal variable	198
8.3. Modèle oscillant inspiré des travaux de Wang	200
8.4. Modèle oscillant inspiré de Wang et simplifié	201
8.5. Apprentissage bilinéaire dans le modèle 1, partie 1	203
8.6. Apprentissage bilinéaire dans le modèle 1, partie 2	203
8.7. Explication du processus d'équilibre durant l'apprentissage	204
8.8. Convergence du modèle avec lien inhibiteur seul	205
8.9. Apprentissage dans le modèle 2 : mécanismes d'apprentissage sur le lien activateur	206
8.10. Poids synaptiques après convergence sur le modèle bilinéaire en fonction de l'énergie moyenne du signal injecté en entrée	208
8.11. Évolution des poids synaptiques en fonctions de la fréquence imposée	209
8.12. Exemples de formes de la fonction ϕ dans l'équation BCM	210
8.13. Exemple de fonction θ modifiée	210
8.14. Évolution de la fréquence et de la période du système en fonction de l'énergie moyenne entrante	212
8.15. Évolution des poids synaptiques en fonction de l'énergie entrante	213
8.16. Évolution du produit des poids synaptiques du duo de neurone en fonction de l'énergie moyenne	214
8.17. Évolution de la fréquence du doublet oscillant en fonction du seuil d'activation	215
8.18. Évolution du produit des poids synaptiques en fonction du seuil sur le neurone inhibiteur	215

Introduction

Problématique générale

La robotique voit aujourd'hui l'apparition de nouveaux usages, apportant de nouvelles contraintes et appelant donc à de nouvelles techniques : on cherche par exemple à pouvoir déployer des robots dans les espaces publics ou encore dans les foyers.

En parallèle, les usages traditionnels de la robotique ont tendance à se diversifier et leurs modes d'application actuels à être remis en question. Dans une usine par exemple, il est de plus en plus nécessaire de pouvoir disposer de robots capables d'agir dans un environnement partagé où les humains circulent et accomplissent leurs tâches. Dans ce contexte, certains besoins nouveaux se développent, comme celui de disposer d'outils robotiques de soutien et non plus uniquement de se restreindre à de lourdes chaînes de production automatisées.

Ces nouveaux besoins emmènent à se demander de quelles capacités pourrait-on doter un robot afin qu'il puisse interagir dans un environnement complexe, changeant, et potentiellement parcouru par d'autres robots ainsi que par des humains. On peut aussi se demander dans quelle mesure un robot pourrait alors interagir avec un humain lorsque les deux partagent la même tâche dans cet environnement complexe. Comment concevoir ou permettre au robot de développer de telles capacités d'interactions ?

La robotique traditionnelle a l'habitude de concevoir des robots spécialisés pour une tâche et qui l'exécutent de la façon la plus optimale possible. Souvent, dans un contexte d'interaction avec l'humain ce type de robot est source de frustration dans la tâche partagée. Un opérateur dans une chaîne de montage, par exemple, va devoir ralentir ou accélérer son travail en fonction du rythme des machines robotiques présentes sur cette chaîne. Les robots n'étant en effet pas capable de comprendre qu'une situation particulière nécessite un ajustement de leur allure pour permettre à l'opérateur de réaliser sa tâche dans de bonnes conditions. C'est alors l'incapacité du robot à saisir et à anticiper le comportement de l'opérateur qui est ici au cœur du problème.

L'humain, lui, est capable de comprendre et de prédire le comportement du robot, et il va donc pouvoir adapter son travail en fonction, toutefois cela lui demande des ressources mentales qui vont rendre l'interaction frustrante et peu intuitive, ce qui devrait être, dans l'idéal, évité.

On peut alors se demander quelle robotique concevoir pour répondre à ces problématiques, tant d'un point de vue du corps robotique en lui-même que du point de vue des algorithmes et modèles servant à contrôler ce corps.

Éléments de réponse pluridisciplinaires

Parallèlement à l'émergence de ces nouvelles problématiques, plusieurs champs de la science ont commencé à y apporter des éléments de réponses, notamment en étudiant les facultés de l'Homme et du vivant en général à s'adapter et à interagir avec l'autre ou avec un environnement complexe.

Le domaine de la psychologie a ainsi permis de dégager des pistes quant aux mécanismes de coordination permettant des interactions intuitives avec l'environnement ou entre humains. On parle notamment de coordination interpersonnelle pour décrire le phénomène permettant l'adaptation d'un comportement vis-à-vis d'un partenaire et vice-versa ([HARI et KUJALA, 2009](#)).

Certains effets étudiés par cette discipline peuvent être mis en avant. C'est le cas notamment des effets de synchronisation et des effets d'imitation (résonance motrice). La synchronisation est omniprésente entre humains ([OULLIER et al., 2008](#)). En plus d'aider à la réalisation de certaines tâches en interaction ([BERNIERI et ROSENTHAL, 1991](#)), celle-ci joue aussi un rôle clé en tant que signal permettant de renforcer l'interaction ([NADEL et al., 1999](#) ; [NADEL, 2002](#)). L'imitation, elle, que l'on peut concevoir comme innée

ou acquise, permet non seulement à certains phénomènes de communication sociale comme le bâillement de s'exprimer (CHARTRAND et BARGH, 1999) mais peut aussi être un facteur important de la capacité de l'humain à appréhender le monde, à communiquer par le biais de son corps et à développer ses capacités motrices (MARATOS, 1973; NADEL et BUTTERWORTH, 1996; BUTTERWORTH, 1999).

La psychologie nous renseigne également sur l'influence qu'a l'intention sur la forme des trajectoires des gestes et déplacements humains. Cela permettrait alors à une personne de prédire l'intention de l'autre à travers la simple observation de ses mouvements, ce qui facilite grandement l'interaction (GEORGIU et al., 2007; BECCHIO et al., 2008).

Toujours dans le même domaine, l'étude du développement de l'enfant a rendu possible la caractérisation fine des étapes par lesquelles celui-ci acquiert ses premières facultés d'interaction avec l'environnement et avec les autres, notamment avec sa mère. L'un des phénomènes clé analysé par la psychologie et que nous aborderons dans cette thèse est notamment le développement du phénomène d'imitation (PIAGET, 1945; MARATOS, 1973).

La biologie, et plus spécifiquement la neurobiologie, a quant à elle pu isoler de nombreux mécanismes permettant d'expliquer ces phénomènes. La découverte de « *neurones miroirs* » (RIZZOLATTI et al., 1996) a ainsi apporté un nouvel éclairage aux mécanismes de la résonance motrice (RIZZOLATTI et al., 1999) et notamment aux mécanismes d'imitation. Les études sur les CPG (Central Pattern Generator) et sur les réseaux oscillants en général ont apporté des débuts d'explications aux phénomènes de synchronisation comme la synchronisation non-intentionnelle, tout en affinant notre compréhension de la génération de mouvements complexes et rythmiques chez l'humain.

La neurobiologie, par l'étude des phénomènes d'apprentissage dans les réseaux de neurones du cerveau humain, a également permis d'isoler et de comprendre certains processus d'apprentissage chez l'Homme et leur implication dans le développement de certaines parties essentielles du cerveau. L'étude de ces différentes zones a ainsi permis de proposer diverses règles d'apprentissages générales tandis que l'analyse de certaines zones clés du cerveau humain a permis d'affiner plus précisément la connaissance des mécanismes d'apprentissage spécifique qui y ont lieu et leur implication dans le développement de l'humain, particulièrement dans son développement moteur. (on pense notamment aux zones du cervelet et des ganglions de la base)

Ainsi, la biologie apporte un éclairage indispensable aux mécanismes permettant à l'humain de se développer en interaction avec son environnement et aux phénomènes identifiés par la psychologie comme importants pour l'interaction.

Enfin, les domaines attenants à la mécanique ont mis en évidence que le corps à lui seul, dès lors qu'il possède certaines propriétés, pouvait contribuer à la résolution de certains problèmes d'interaction entre l'humain et son environnement que la robotique traditionnelle avait échoué à résoudre dans le cas d'environnement réel. En effet, des processus que l'on pensait autrefois principalement liés à des procédés complexes présents dans le cerveau, comme la marche ou la course, ont ainsi été résolus sans aucun contrôleur (marcheur passif (MCGEER, 1990)) puis avec des contrôleurs minimalistes capables de tirer parti au maximum des propriétés du corps (marcheur dynamique (COLLINS et al., 2005; WISSE et VAN FRANKENHUYZEN, 2006)). Ces avancées, que l'on pourrait regrouper sous le paradigme de « *l'intelligence du corps* », ont mis en lumière le fait que de nombreux aspects de l'interaction chez l'humain sont sans doute liés à des propriétés du corps et non uniquement à des phénomènes isolés dans le système nerveux.

Travaux en robotique

Il existe de nombreux travaux, en robotique et dans le domaine de l'intelligence artificielle (IA), qui, en exploitant les avancées des divers domaines scientifiques précédemment cités, ont tenté d'apporter des solutions à l'interaction d'un robot avec un environnement complexe et avec des partenaires .

Les travaux autour de l'intelligence du corps ont, par exemple, abouti au robot BigDog : un marcheur dynamique quadrupède capable pour la première fois en robotique de marcher dans un environnement complexe (RAIBERT et al., 2008).

Les différents travaux en neurobiologie et psychologie du développement forment la base d'un domaine complet de la robotique, la robotique développementale, qui postule l'importance du développement

d'un robot en interaction avec son environnement (LUNGARELLA et al., 2003) en tentant de mimer les étapes du développement de l'enfant (ou le développement chez les animaux).

De leur côté, les très nombreux travaux sur l'apprentissage en neurobiologie ont donné lieu à diverses applications en intelligence artificielle et en robotique. Les algorithmes d'apprentissage par renforcement, très utilisés, sont par exemple inspirés du modèle des ganglions de la base, un ensemble de noyaux dans le cerveau qui semble en lien avec le renforcement chez l'Homme. Ces algorithmes ont été largement adoptés par la suite dans le domaine des interactions Homme-machine (IHM) et plus particulièrement dans le domaine de l'interaction Homme-robot (IHR).

Les approches en psychologie, quant à elles, ont donné lieu à de nombreuses applications en robotique, non seulement autour de l'étude des phénomènes de résonance motrice comme l'imitation ou la synchronisation non-intentionnelle mais aussi autour de l'étude de la prédiction des intentions de l'autre au travers de ses actions (LEWKOWICZ et al., 2013).

Nous détaillerons plus précisément chacun des apports de la robotique vis-à-vis de ces différents domaines dans les sections concernées de ce manuscrit, mais retenons d'ores et déjà que la robotique a produit un ensemble de modèles appliqués et de robots capables de rendre compte d'une grande variété de phénomènes mis en avant par la neurobiologie, la psychologie et la mécanique dans le cadre de l'interaction de l'Homme avec un environnement complexe ou avec d'autres humains.

Justification scientifique

Le laboratoire ETIS a mené précédemment des études autour de l'intégration de plusieurs des phénomènes utiles à l'interaction au travers des trois prismes présentés précédemment : l'intelligence du corps, la psychologie développementale, et des outils inspirés des processus neurobiologiques.

Plus spécifiquement, les travaux de Philippe Gaussier puis de Pierre Andry ont montré que les phénomènes d'imitation étaient essentiels au développement moteur humain et à l'interaction de l'Homme avec un environnement complexe et des partenaires. En pratique, diverses implémentations développementales basées sur les connaissances acquises en neurobiologie les ont menés à des propositions de modèles montrant comment un robot est capable de former des liens entre sa perception et son action et comment ces liens peuvent ensuite être exploités dans un cadre d'interaction intuitive dans un environnement réel (GAUSSIER et ZREHEN, 1994; GAUSSIER et ZREHEN, 1995; GAUSSIER et al., 1998; ANDRY et al., 2005).

En parallèle, les travaux de Ken Prepin et d'Arnaud Revel (PREPIN et REVEL, 2007) puis d'Arnaud Revel et Pierre Andry (REVEL et ANDRY, 2009), d'Arnaud Blanchard et Jacqueline Nadel (BLANCHARD et NADEL, 2009) ou encore de Ken Prepin et Philippe Gaussier (PREPIN et GAUSSIER, 2010) ont permis, par l'introduction d'oscillateurs neuronaux dans les modèles, d'initier des effets de synchronisation et de montrer leurs apports dans un cadre d'interaction.

Plus récemment, les travaux de Khursheed Hasnain et Ghiles Mostafaoui (HASNAIN et al., 2013; HASNAIN et al., 2014) ont, via une collaboration expérimentale avec le laboratoire Euromov (anciennement M2H), étendu ce type de modèle et démontré l'importance de la synchronisation et son côté inévitable dans les interactions humaines et dans les interactions Homme-robot.

Les travaux de Caroline Lesueur-Grand et Ghiles Mostafaoui ont par la suite exploré l'usage de la synchronisation comme signal de renforcement dans l'interaction (GRAND et al., 2014) et l'ont appliqué en parallèle de modèles de navigation pour simuler la synchronisation durant la marche (LESUEUR-GRAND, 2017).

Parallèlement, des travaux préliminaires au laboratoire URECA ont donné lieu à des modèles permettant de prédire l'intention de l'autre au travers de ses trajectoires (LEWKOWICZ et al., 2013). Peu à peu, la question de l'influence des intentions sur le système moteur et dans quelle mesure cette influence peut permettre de lire l'intention de l'autre au travers de ses trajectoires motrices, est devenu une question centrale.

Le laboratoire ETIS a commencé à travailler sur l'importance du corps pour l'interaction au travers des travaux d'Alexandre Pitti qui a démontré l'intérêt du rythme et d'un corps bien conçu (PITTI et al., 2010).

Finalement, un partenariat entre ETIS et le LISV, un laboratoire de mécanique centré sur l'hydraulique, s'est développé. Il a abouti à la création d'un robot hydraulique humanoïde unique en France : Tino. Ce robot a été conçu dans l'esprit qu'un corps proche de celui de l'humain, à la fois pour agir dans le même domaine dynamique que lui, mais également pour amplifier les effets de résonance motrice entre l'humain et le robot, pourrait être pertinent dans un contexte d'interaction sociale.

Cette thèse est le premier travail utilisant le robot Tino. L'objectif de la thèse est de proposer un modèle de contrôle robotique adapté à un robot humanoïde hydraulique comme le robot Tino et s'inspirant des précédents travaux sur l'imitation, la synchronisation, et l'impact des intentions sur le mouvement afin de converger vers un modèle de contrôle développemental capable d'interaction dans un environnement complexe avec des humains ou d'autres agents.

Pour ce faire, nous analyserons les spécificités des modèles dont nous nous inspirons afin d'isoler les contraintes qui pourraient empêcher leur intégration sur robot ou leur combinaison entre elles dans un modèle unifié. Plusieurs solutions seront proposées pour répondre à ces contraintes, incluant parfois, quand c'était nécessaire, de nouveaux modèles. Puis nous appliquerons certains des modèles résultants sur le robot Tino, et d'autres, en simulation.

Du fait de ses objectifs et de ses sources d'inspiration, cette thèse se positionne dans le domaine de la robotique bio-inspirée développementale et du constructivisme. Nous essaierons de fait de respecter quelques principes généraux dans nos travaux :

- Les modèles proposés tenteront de s'inspirer au maximum de la biologie et de la neurobiologie.
- Les modèles proposés tenteront de rester les plus « écologiques » possible, c'est-à-dire de ne pas proposer de spécificités superflues et de faire au plus simple. De fait, nous tenterons de résumer au maximum les points inspirés de la neurobiologie et de n'en extraire que le nécessaire pour nos modèles afin de permettre des réutilisations et des implémentations simples de ces modèles dans un contexte expérimental en robotique.
- Chacun de ces modèles sera conçu dans l'idée d'une application sur robot réel déployée en environnement complexe dans un cadre d'interaction avec des humains. Nous analyserons à chaque fois dans ce cadre les contraintes que peuvent poser ce type d'environnement sur nos modèles.
- Pour rester dans un paradigme développemental, les modèles proposés tenteront de dépendre le moins possible de paramètres fixés valides uniquement sur un seul type de matériel ou dans un unique environnement. Nous chercherons à ce que nos modèles se développent, notamment par l'exploration et l'apprentissage, en lien avec le corps robotique, ses capacités et, dans l'idéal, en partant de zéro afin d'être déployable sur le maximum de matériel disponible. Cela pourra également permettre d'apporter, en parallèle, des éléments de compréhension au développement des capacités étudiées chez l'humain.

Plan de la thèse

Au vu de l'importance du robot Tino dans cette thèse, la première partie lui est consacrée. Dans un premier chapitre, nous détaillerons les raisons qui justifient l'emploi de ce robot dans un contexte d'interaction dans un environnement réel et avec l'humain. Nous analyserons également les raisons qui poussent à développer des algorithmes d'apprentissages directement sur robot réel et non en simulation, et nous verrons les contraintes qui s'appliquent au développement sur robot et qui peuvent justifier dans certains cas l'usage d'environnements simulés. Dans un second chapitre, au vu du caractère unique du prototype et du fait que cette thèse fut la première à l'utiliser, nous exposerons la mécanique du robot Tino ainsi que les travaux de caractérisation du robot qui ont été effectués en partenariat avec le laboratoire LISV. Nous présenterons ensuite certains phénomènes émergents du robot qui pourront être exploitables dans notre cadre afin de concevoir un contrôle tirant parti au maximum des propriétés de son corps. Nous concluons le chapitre par la présentation de résultats préliminaires issus d'une expérience menée en partenariat avec le laboratoire URECA et visant à juger de l'acceptabilité du robot dans l'interaction avec des sujets humains.

Dans la seconde partie de cette thèse nous présenterons la création d'un modèle de contrôle sensori-moteur bio-inspiré et développemental adapté au robot Tino. Dans le premier chapitre de cette partie, nous ferons un état de l'art sur les algorithmes de contrôles existants et sur ce qui nous pousse à sélectionner des modèles de contrôle issus des réseaux de neurones formels. Nous ferons ensuite une revue succincte des propriétés des diverses zones du cerveau humain qui nous paraissent importantes pour le contrôle moteur. Dans le second chapitre de cette partie, nous mettrons en place un premier modèle « naïf » tentant d'exploiter directement les propriétés émergentes du robot Tino isolées dans la partie précédente. Nous ferons ensuite la critique de ce modèle et des résultats obtenus pour ensuite introduire un modèle de co-contraction musculaire inspiré de la biologie dont nous analyserons les propriétés en simulation.

Enfin, nous présenterons un modèle capable d'établir des liens perceptions-actions, inspiré des précédents travaux du laboratoire. Il a été modifié dans l'idée de répondre aux principes énoncés dans la partie précédente et pour s'adapter au contrôle direct des vérins du robot Tino via le modèle de co-contraction musculaire. Nous présenterons les résultats obtenus sur le robot Tino, notamment la formation d'un espace de contrôle moteur correspondant à l'espace atteignable, dont le rôle nous semble analogue avec le rôle tenu par le cortex moteur chez l'humain. Nous analyserons ensuite ces résultats et nous mettrons en évidence les phénomènes qui émergent de ce type d'apprentissage, notamment l'imitation des gestes du partenaire et le pointage.

Dans une troisième partie, nous étudierons l'impact que peut avoir l'intention sur le mouvement, dans l'idée qu'un modèle capable de transcrire l'intention du robot dans ses trajectoires motrices permettra à l'humain de les lire simplement en observant les mouvements du robot. Pour ce faire, nous commencerons dans un premier chapitre par un état de l'art des mécanismes venant impacter les trajectoires motrices humaines selon le contexte, ce qui nous amènera à faire une revue du fonctionnement des ganglions de la base, un ensemble de noyaux dans le cerveau qui semble en lien avec le renforcement chez l'Homme. Dans le second chapitre de cette partie, nous proposerons un modèle inspiré d'une interprétation classique des ganglions de la base. Étant donné les limitations concernant l'implémentation de ce type de modèle en environnement réel, nous construirons une simulation simple pour le tester. Nous examinerons ses capacités à converger vers des trajectoires analogues aux trajectoires humaines en fonction de plusieurs critères et nous introduirons plusieurs variantes dues au modèle.

Nous placerons ensuite le modèle dans un contexte simplifié d'interaction sociale. Nous montrerons alors qu'il est capable de reproduire qualitativement et quantitativement certains résultats observés dans des expériences similaires de psychologie. Nous prouverons également que, pour une même tâche, le modèle est capable de converger vers des trajectoires différentes en fonction de la nature du contexte, social ou non. Par analogie avec les expériences de psychologie, nous concluons que ce type de modèles est capable d'introduire la notion d'intention dans les trajectoires motrices d'un agent. Nous illustrerons quelques points problématiques quant à l'adaptation de ces modèles pour un fonctionnement en accord avec l'architecture précédente, notamment la problématique du contrôle en force adapté à une commande musculaire.

Enfin, nous complexifierons l'environnement simulé en mettant en place une simulation 2D d'un bras humain soumis à quelques contraintes. Nous intégrerons l'ensemble des améliorations proposées sur le modèle précédent et nous montrerons que les résultats précédents peuvent se retrouver malgré un problème à résoudre plus contraint et un type de commande plus complexe. Nous concluons finalement par l'analyse des défauts du modèle et des points nécessaires à la combinaison de ce type d'architecture avec celle de la partie 2.

Dans la quatrième partie de cette thèse, nous étudierons la question de la synchronisation non-intentionnelle et du contrôle dynamique d'un robot pour l'interaction avec un environnement réel et le partenaire. En premier lieu, nous rappellerons les principes de la synchronisation non-intentionnelle, son utilité supposée chez l'humain et pourquoi il est intéressant de l'exploiter dans un contexte de robotique interactive. Nous introduirons nos travaux effectués avec l'équipe partenaire d'Euromov qui ont visé à préciser les limites de la synchronisation non-intentionnelle chez l'humain en mettant en interaction le

robot Nao (plus simple à utiliser dans ce contexte) face à plusieurs sujets. Nous présenterons ensuite brièvement le mode de fonctionnement des modèles classiques de synchronisation chez le robot tels qu'ils furent utilisés dans les travaux du laboratoire ETIS depuis plusieurs années.

Ensuite, nous formulerons une critique détaillée quant aux travaux récents mêlant synchronisation, réseaux de neurones oscillants et robotique. En nous appuyant sur des résultats issus du domaine des marcheurs passifs et dynamiques, nous mettrons l'accent sur le fait que certaines des hypothèses mêlant couples de neurones oscillants et synchronisation motrice ne nous paraissent pas solides. Nous proposerons une démonstration visant à justifier notre critique et nous la mettrons en place sur le robot Tino.

L'analyse des résultats nous amènera à décorrélérer l'étude du phénomène de synchronisation de l'étude des réseaux oscillants dans le mouvement humain et dans l'interaction humaine.

Dans un second temps, nous ferons un état de l'art succinct sur les réseaux oscillants et au vu des contraintes d'un environnement réel dans un cadre interactif ainsi que des contraintes permettant à ce type de réseau de fonctionner dans des modèles complexes, nous arriverons à la conclusion que l'étude de l'apprentissage dans les réseaux oscillants est nécessaire pour introduire ces outils dans nos modèles. Après une étude des modèles d'apprentissages existant sur les réseaux oscillants, nous introduirons un modèle inédit d'apprentissage qui se base notamment sur des découvertes récentes en neurobiologie quant à l'apprentissage dans les synapses GABAergique. Nous étudierons plusieurs variantes de ce modèle afin d'en éliminer itérativement les plus gros défauts et afin de converger vers un modèle de lien entre neurones excitateurs et neurones inhibiteurs qui n'est plus que dépendant des paramètres internes des neurones pour son fonctionnement, et qui est capable de s'adapter au contexte en se maintenant dans une plage utile. Nous ferons par la suite quelques propositions d'utilisation de ce modèle qui pourraient apporter des éléments de réponses à l'interaction dynamique entre un robot et son environnement ou entre un robot et un partenaire, dans un cadre d'environnement réel, complexe et changeant.

Enfin, nous concluons la thèse par l'analyse des travaux encore nécessaires à la combinaison de ces composantes entre elles : l'imitation, le biais des trajectoires selon le contexte et l'intention, la synchronisation par le biais du corps et les réseaux oscillants capables de s'adapter au contexte. Nous ferons ainsi un état des lieux des obstacles qui ont été levés, des points qui ont été identifiés dans la thèse et des points qu'il reste à éclaircir pour aboutir à une architecture unie composée de ces éléments, capable de s'adapter à un robot comme Tino et d'interagir ensuite dans un environnement réel.

I

Première partie :

Tino : Le robot humanoïde

Le choix d'un robot humanoïde hydraulique comme plateforme expérimentale

Sommaire du chapitre :

1.1. Introduction	9
1.2. Environnement simulé et plateforme robotique réelle : avantages et inconvénients	11
1.2.1. Les avantages d'un environnement simulé	11
1.2.2. L'apprentissage long sur robot	11
1.2.3. L'intelligence du corps	12
1.2.4. Le problème de la régularité en simulation	13
1.2.5. La physicalité de l'interaction	13
1.2.6. Conclusion	13
1.3. Pourquoi un robot humanoïde ?	15
1.3.1. La vallée de l'étrange	15
1.3.2. Des indices biomimétiques pointant la non-optimalité du choix d'une forme humanoïde	16
1.3.3. Les avantages d'un robot humanoïde	16
1.4. Pourquoi un contrôle hydraulique ?	18
1.4.1. Avantages et inconvénients des différents types de moteurs	18
1.4.2. La problématique de l'effort subit avec les moteurs électriques	19
1.4.3. La solution « compliant »	19
1.4.4. Hydraulique ou pneumatique : la question de la sécurité	20
1.4.5. Conclusion	20
1.5. Conclusion	21

1.1 | Introduction

Dans cette thèse nous soutenons l'hypothèse qu'une intelligence artificielle (IA) efficace est le fruit d'un développement en interaction avec le corps qu'elle contrôle. Nous défendons également le fait que des capacités spécifiques peuvent émerger, non seulement du détail d'implémentation des algorithmes, mais aussi des particularités du corps contrôlé par ces derniers. Il est donc nécessaire de décrire et caractériser le corps robotique pour lequel nos algorithmes ont été développés afin d'apporter une estimation de son impact sur les résultats observés.

Le robot Tino (et son successeur Tino 2) est la plateforme robotique principale utilisée dans cette thèse. Cette partie du manuscrit lui est dédiée afin d'étudier plus en détail les spécificités de ce corps robotique.

Ce robot est un prototype de robot humanoïde hydraulique constitué d'un buste humanoïde artificiel portant deux bras manipulateurs et d'une tête expressive équipée de deux caméras mobiles (voir la figure 1.1). Il est le fruit d'une collaboration entre le laboratoire LISV, le laboratoire ETIS et l'entreprise BIA qui s'occupa de sa construction. Sa conception est basée sur le projet HYDROÏD du LISV (ALFAYAD, 2009).

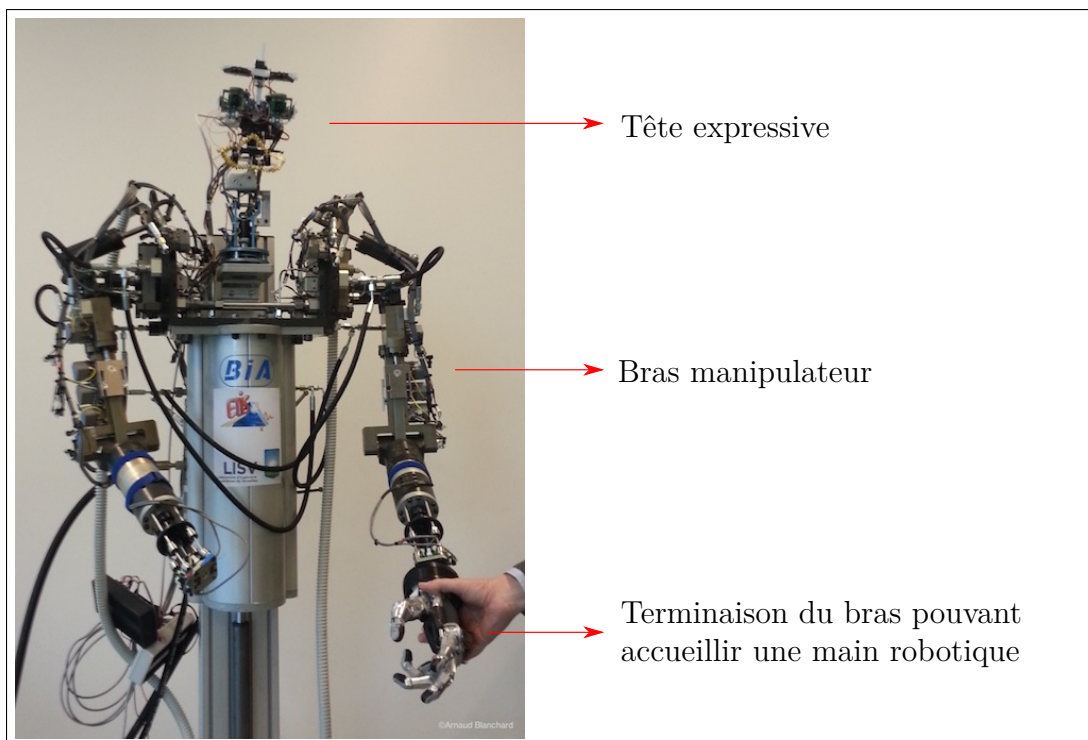


FIGURE 1.1. – *Le robot humanoïde Tino. On peut observer le robot et ses caractéristiques via une vidéo de présentation humoristique¹. Pour des informations plus complètes, on se référera à la page web dédiée².*

Cette thèse est la première à utiliser ce prototype, qui est d'ailleurs le premier et unique robot hydraulique humanoïde en France, ce qui explique également la nécessité des nombreux travaux de caractérisation et d'ingénierie qui ont été menés autour de cet outil robotique.

1. ↑ Présentation humoristique de Tino, <https://youtu.be/C5CE84qt59k>

2. ↑ <https://www.etis.ensea.fr//neurocyber/web/en/tino.php>

Nous décrivons donc dans cette partie tout ce qui a trait au robot humanoïde Tino, de la prise en main du prototype à sa caractérisation au cours de divers partenariats inter équipes. Nous illustrerons également les spécificités de ce robot, notamment ses capacités de contrôle en force, qui le rendent intéressant dans le cadre du développement d'algorithmes de contrôle moteur bas niveau. Dans un premier temps, nous discuterons dans ce premier chapitre des aspects théoriques qui orientent la méthodologie de notre travail autour de ce prototype et qui justifient notre approche. Nous verrons également les limites de celle-ci et préciserons dans quel cas il peut être utile de la compléter au travers d'environnements simulés.

Le chapitre suivant viendra ensuite décrire la mécanique du robot et les expériences réalisées afin d'en caractériser les propriétés.

1.2 | Environnement simulé et plateforme robotique réelle : avantages et inconvénients

Dans le contexte de la conception d'IA le choix d'un robot et l'importance du corps de ce robot sont tout sauf des évidences. Des questions telles que « Pourquoi ne pas se contenter d'un robot simulé ou de tout autre dispositif contrôlable simulé ? » ou encore « Pourquoi ne pas développer en premier lieu en simulation avant de transférer les résultats sur robot ? » sont des questionnements légitimes.

Dans cette section, nous avons pesé les avantages et inconvénients de chacune des deux approches. Nous essaierons ainsi d'apporter des réponses pragmatiques à ces questions qui expliciteront alors les directions et méthodes prises dans cette thèse.

1.2.1. Les avantages d'un environnement simulé

Une simulation ne tombe pas en panne. On sous-estime assez souvent le temps perdu en réparation de diverses sortes lors de la manipulation d'un outil robotique réel dans un cadre expérimental. Cet aspect, même s'il peut paraître anodin, est en pratique probablement la raison la plus forte qui ferait pencher la balance en faveur de la simulation lors du développement d'IA.

Une autre raison presque aussi importante est qu'une simulation informatique sera à même de réaliser, quand cela est nécessaire, un grand nombre d'expériences successives. Un corps robotique, lui, est limité, non seulement à cause de sa fragilité au court du temps et des expérimentations, mais aussi parce qu'on ne peut tout simplement pas accélérer la réalité et que le nombre d'expériences réalisables dans un temps donné s'en trouve donc contraint.

Si on prend l'exemple du développement du contrôle des bras d'un robot, celui-ci devra donc réaliser de nombreux mouvements avant que l'algorithme ne converge vers un résultat cohérent alors qu'à contrario certaines simulations peuvent par exemple prendre des mois ou des années virtuelles de calcul.

Cependant, la communauté du développement bio-inspiré incarné pointe du doigt que de telles simulations résolvent le problème posé de façon peu bio-inspiré, contraignant l'algorithme à, parfois, nécessiter un entraînement dépassant largement l'espérance de vie d'un humain. Nous pouvons citer ici le cas de l'IA AlphaStar, qui a nécessité un temps d'entraînement de plusieurs centaines d'années virtuelles pour battre un joueur professionnel sur le jeu-vidéo Starcraft 2 (VINYALS et al., 2019)). Cette échelle de temps de convergence est effectivement un non-sens si l'on souhaite s'inspirer du développement de l'être humain.

Si cette critique des environnements simulés est parfois justifiée, elle oublie toutefois que certains types d'algorithmes, qui cherchent à être un maximum bio-inspiré, pourront également mettre des semaines ou des mois virtuels à converger. Le reproche ne convient alors plus, car des mois ou des années de développement peuvent tout à fait être cohérents dans un contexte développemental. Il peut en effet être logique qu'un algorithme inspiré du cerveau de l'être humain demande à minima autant d'expérience qu'un bébé humain pour s'affiner. Certaines phases du développement de l'enfant durant effectivement des mois (CHANQUOY et NEGRO, 2004), on peut concevoir qu'il peut alors y avoir un avantage à privilégier un environnement simulé, accélérable à souhait, plutôt qu'un corps robotique avec lequel l'expérience durerait des mois réels, ce qui est à la fois incommode à mettre en œuvre, mais également difficile, voire impossible en pratique, au vu des pannes limitant l'usage intensif d'un corps robotique complexe sur le long terme.

1.2.2. L'apprentissage long sur robot

Certaines solutions existent pour pallier le problème commun d'apprentissage long sur robot. Des algorithmes à long temps de convergence ont ainsi parfois été testés, sur plusieurs robots en parallèle, afin d'accélérer virtuellement le nombre d'expériences possibles (GU et al., 2017). Mais outre la difficulté initiale à disposer d'un grand nombre de robots identiques en série et la démultiplication des pannes

en conséquence, cette solution empêche également l'algorithme de tirer profit de son adaptabilité à un corps spécifique.

Certains algorithmes, appliqués à un corps spécifique, peuvent en effet découvrir des solutions particulières non prévues en exploitant des caractéristiques uniques à ce corps. Des solutions qui pourront donc s'avérer invalides pour un autre corps robotique, même en considérant deux corps robotique conçus sur le même modèle.

Ce type de lien entre un corps spécifique et l'algorithme employé est le cœur de la philosophie de « *l'intelligence incarnée* ». Cette philosophie de travail en IA sur laquelle nous nous basons formule l'hypothèse que ces liens puissent aboutir à l'émergence de propriétés ou comportements spécifiques à un corps donné. Dans ce cadre de travail, il est donc dommage d'adopter des méthodes qui limiteraient ou empêcheraient cette émergence. La solution des corps parallèles est donc intéressante, mais présente toutefois certains défauts dans notre cas d'application.

D'autres réponses ont pu être développées en cherchant à optimiser certains paramètres des apprentissages dans les tests sur corps réels (voir par exemple : (MATIGNON et al., 2006)). Cependant, celles-ci sacrifient inévitablement à la richesse potentielle des comportements vers lesquels ces algorithmes convergent.

Enfin, une solution intermédiaire souvent appliquée consiste à faire l'apprentissage sur un corps robotique simulé informatiquement, puis soit à utiliser le corps simulé pour déduire les ordres à donner au corps réel dans le cadre d'une démonstration, soit à transférer l'apprentissage résultant directement sur le robot. Ces méthodes donnent globalement de bons résultats, mais présentent quelques défauts. Tout d'abord, il est nécessaire de disposer d'une simulation suffisamment précise du corps robotique utilisé pour que le transfert soit valable, ce qui est loin d'être évident à obtenir dans le cadre de prototypes. De plus, au fil des remplacements et réparations des pièces du robot, de tels modèles perdent en précision, ce qui rend difficile l'emploi de ces solutions lorsque l'on emploie des apprentissages très bas niveau qui cherchent à profiter des caractéristiques précises du matériel (ce qui est justement notre cas). Enfin, un autre défaut est que ce type de solution empêche l'utilisation par l'algorithme, comme pour la solution des multiples corps robotiques parallèles, de caractéristiques fines du robot potentiellement inconnues ou négligées par l'expérimentateur.

1.2.3. L'intelligence du corps

« L'Homme est intelligent parce qu'il a une main. »

Anaxagore (V^e siècle av. J.-C.)

Comme nous l'avons vu, les environnements simulés offrent des avantages indéniables et les solutions palliatives développées pour maintenir l'usage de robot réel malgré les défauts résultant d'apprentissages sur le long terme ont toutes des défauts notables.

Néanmoins, il existe de solides raisons qui amènent à privilégier un corps robotique réel à une simulation malgré les défauts de cette approche. La plupart de ces raisons sont inspirées d'un paradigme nommé « l'approche éactive » (VARELA et al., 1993) qui postule l'importance du développement de la cognition par l'interaction avec le corps et l'environnement. En s'inspirant de cette approche, nous donnons ici les avantages à travailler sur un robot réel en interaction avec l'environnement.

Si nous prenons l'exemple des travaux sur la marche, nous avons pu constater que les approches calculatoires classiques pouvaient très vite rencontrer des obstacles infranchissables (Exemple du robot ASIMO (KNIGHT et al., 2002)).

À l'inverse, des approches basées sur l'importance du corps et des phénomènes émergents qu'un corps bien conçu peut permettre d'exploiter, ont donné de bons résultats assez tôt dans le cadre de la marche, d'abord totalement passivement (sans dispositifs de contrôle) (MCGEER, 1990), puis dans le cadre de marcheurs dynamiques, dont le robot BigDog fut le succès le plus médiatisé (COLLINS et al., 2005 ; WISSE et VAN FRANKENHUYZEN, 2006). Tous ces travaux ont été réalisés dans un contexte d'avance par essai-erreur en interaction avec des corps réels ce qui simplifiait significativement les problèmes à résoudre

par les algorithmes de contrôle. Ce sont de bons exemples de ce que peut apporter l'expérimentation directe sur robot.

1.2.4. Le problème de la régularité en simulation

L'un des autres avantages d'un corps robotique est le test d'algorithme dans des situations bruitées, non prédictibles et irrégulières d'une expérience à l'autre. La régularité de la simulation peut être un important défaut lorsqu'il induit à ne pas voir qu'un algorithme s'est surspécialisé en profitant de la régularité du simulateur ou qu'il n'est stable que dans un ensemble de conditions précises correspondant au simulateur, l'amenant à ne pas pouvoir fonctionner ensuite sur un corps robotique réel. Les simulateurs ont donc le défaut de masquer à l'expérimentateur ce genre de cas.

À l'inverse, un algorithme commandant un corps réel permet de se rendre compte assez vite de ce type d'erreur, à la faveur des perturbations naturellement induites par l'environnement réel ou des changements de caractéristiques du robot au fil du temps (vieillesse des pièces et composants, etc.).

Les situations d'expérimentation en conditions réelles peuvent alors permettre de converger vers des algorithmes de contrôle robustes en toutes situations en évitant la plupart des écueils dus à un environnement trop parfait.

1.2.5. La physicalité de l'interaction

La physicalité est un anglicisme désignant, dans nos domaines, la présence physique réelle d'un agent, a contrario d'agent non présent physiquement comme des agents en simulation. Il s'agit d'un élément important de notre approche, autant du point de vue de l'expérimentateur que du point de vue des expériences qu'il est possible de mener avec le robot dans l'interaction.

L'importance de la présence physique d'un robot par rapport à un avatar virtuel est sujette à discussion dans notre communauté. Des travaux ont développé l'idée que la présence réelle d'un robot permet d'interagir avec lui de façon bien plus naturelle et en mettant en jeu plus de modalités sensorielles que ne le permet une simulation, par exemple via le toucher. De nombreuses études ont ainsi tenté de mesurer cette nuance (voir par exemple (KIM et al., 2017) ou (WAINER et al., 2006)). Cependant, bien que ces analyses apportent quelques indices sur l'importance de la physicalité dans l'interaction, aucune d'entre elles ne permet véritablement de répondre avec certitude à cette question, qui reste donc ouverte à la discussion.

Malgré cela, nous poserons comme hypothèse dans nos travaux que, particulièrement dans le cas d'interactions gestuelles, la présence réelle du robot est un point crucial dans le cadre d'expériences dédiées à l'interaction entre humains et robots.

Dans le même temps il est utile de noter que du point de vue de l'expérimentateur, un corps robotique réel, dès lors qu'il est conçu sur le modèle du vivant, permet une certaine empathie, ce qui facilite l'identification de comportements anormaux, qu'ils soient dus à l'algorithme en cours de test ou à une défaillance du robot. En profitant de cet aspect, il est ainsi aisé d'identifier si la tâche demandée au robot sera trop difficile, parce que par exemple mal pensée par l'expérimentateur. Cette physicalité du robot est donc importante d'un point de vue pragmatique dans le cas d'expériences répétées nécessitant de nombreux ajustements.

1.2.6. Conclusion

Les raisons évoquées dans les sections précédentes présentent les arguments principaux qui ont justifié, dans notre cas, l'utilisation d'une approche de développement d'algorithme au plus proche de l'interaction avec le robot réel. On essaiera ainsi dans ces travaux de ne compter sur des environnements simulés que lorsque nécessaire et de développer les algorithmes en gardant à l'esprit l'objectif de profiter au maximum des caractéristiques bas niveau du matériel.

Nous verrons dans les chapitres suivants que l'utilisation d'environnements simulés a parfois été nécessaire, notamment pour des raisons de rapidité de convergence dans le cas d'algorithmes développementaux à apprentissage lent.

1.3 | Pourquoi un robot humanoïde ?

Le choix d'un corps humanoïde lorsque l'on souhaite développer une IA de contrôle bio-inspirée basée sur un paradigme développemental n'est pas non plus évident au premier abord. Il existe en effet plusieurs arguments forts en défaveur du choix d'un corps humanoïde. Même si le robot humanoïde Tino a fait partie de cette thèse dès ses débuts, il peut être utile de s'arrêter un instant sur les inconvénients que peut avoir le développement d'algorithmes bio-inspirés et les expériences d'interactions sur robot humanoïde.

1.3.1. La vallée de l'étrange

L'un des problèmes les plus évidents d'un corps humanoïde est posé par la répulsion instinctive que semble éprouver l'être humain pour des êtres dont le physique s'approche particulièrement de l'Homme tout en en différant sur quelques détails. Cette répulsion a ainsi été théorisée par Masahiro Mori ([MORI, 1970](#)) : plus un corps ressemble à celui de l'humain, plus ses divergences nous sembleraient monstrueuses. Ce concept est théorisé sous le nom de « vallée de l'étrange ».

On notera que malgré la popularité actuelle de cette notion, qui fait presque consensus, son aspect scientifique ou non a souvent été débattu, de nombreuses études venant infirmer ou confirmer le phénomène sans qu'un consensus scientifique ne se dégage. Il est ainsi encore difficile aujourd'hui de trancher quant à sa réalité.

Néanmoins, en dehors du monde scientifique, certains aspects de ce phénomène ont eu de véritables répercussions dans l'industrie et sont pris tout à fait au sérieux. C'est notamment le cas de l'industrie de l'animation qui a longtemps été incapable de construire un processus de rendu en trois dimensions suffisamment maîtrisé pour mettre en scène des personnages humains sans que cela ne cause de répulsion. L'un des premiers films du studio Pixar, *Tin Toy*, mettant en scène un bébé humain avec des technologies d'animation encore embryonnaires, a notamment déclenché un débat à sa sortie quant à la répulsion ressentie en regardant ce personnage. Ce débat a par la suite amené le monde de l'animation à prendre en compte sérieusement ce phénomène dans les productions suivantes ([SIMPLYMAYA, 2016](#)).

Plus récemment dans cette industrie c'est le film *Polar Express* qui a donné lieu à un débat dans ce cadre, une partie du public et des enfants semblant, selon les critiques, trouver les personnages du film repoussant. Le débat qui a suivi sur le fait de savoir si oui ou non les techniques du film ont fait sombrer ses personnages dans la « vallée » a donné lieu à diverses études ([GELLER, 2008](#) ; [POLLICK, 2009](#) ; [TONDU, 2017](#)) encore une fois assez contradictoires et à diverses discussions internes au monde de l'animation quant au bien-fondé des techniques employées dans le film ([JENCKINS, 2004](#)).

De façon générale on retiendra donc que malgré le peu de preuves apportées par les approches scientifiques, l'industrie semble prendre en compte sérieusement la réalité du phénomène en pratique.

En dehors du monde de l'animation et malgré les preuves contradictoires quant à l'observation scientifique du phénomène, il existe d'autres indices qui pointent tout de même vers l'existence d'une sorte de vallée de l'étrange. Par exemple, il est aisé de faire l'expérience du phénomène soi-même devant un robot humanoïde ou une animation 3D mal réalisée. D'autre part, certaines théories semblent justifier l'existence de cette vallée de l'étrange. Par exemple, la vallée de l'étrange pourrait être une réponse défensive contre une probabilité de transmission de maladie ([RHODES et ZEBROWITZ, 2002](#) ; [MACDORMAN et al., 2009](#)) : un individu nous ressemblant suffisamment étant considéré comme humain, mais ses divergences vues comme le signe d'une anomalie qui doit être repoussée afin d'assurer la sécurité du groupe car potentiellement transmissible. D'autres théories voient le mécanisme comme une défense génétique de l'espèce (contre des malformations, etc.) ou à l'inverse comme un mécanisme de préférence génétique de l'espèce pour ceux qui, au contraire, cadreraient avec certaines normes ([RHODES et ZEBROWITZ, 2002](#) ; [GREEN et al., 2008](#)).

Pour finir, l'une des hypothèses inspirées de l'approche énaactive considère que l'apparence trop humaine d'un robot amène le cerveau humain à attendre du robot un comportement humain. C'est cette différence entre les attentes, la prédiction de comportement, et le comportement observé, qui pourrait ainsi causer la sensation de malaise. Cela a notamment été discuté dans le cadre du codage prédic-

tif (SAYGIN et al., 2011) et dans le cas d'un conflit de catégorie entre catégories attendues et réelles (FERREY et al., 2015). Notons que les deux approches peuvent tout à fait recouvrir le même phénomène.

Cet ensemble de faits amène à prendre en compte sérieusement le phénomène lorsque l'on doit considérer l'impact du corps d'un robot sur les expériences. Le choix d'un robot humanoïde, qui est le plus à même de causer une répulsion, peut donc être vu comme douteux lorsque l'objectif est de créer une plateforme robotique interagissant avec l'être humain. Néanmoins dans le cas de cette thèse il semble important de noter que le physique du robot Tino, assez rustique, semble être capable d'éloigner le robot de cette « vallée » comme nous le verrons brièvement dans la section 2.5 du chapitre suivant.

1.3.2. Des indices biomimétiques pointant la non-optimalité du choix d'une forme humanoïde

Le domaine de la réalisation d'IA bio-inspirée a ceci de particulier qu'il nous offre la facilité de toujours pouvoir estimer certains paramètres en observant le résultat obtenu par la nature dans les mêmes conditions. En effet, en prenant les mêmes voies qu'a empruntées la sélection génétique, et en supposant que les organismes existants sont plutôt bien optimisés, il est aisé d'obtenir des ordres de grandeur des mécanismes nécessaires à la réplication de telles ou telles de leurs propriétés.

Ainsi a-t-on une idée approximative, par exemple, de l'importance de la locomotion dans un cerveau, du poids calculatoire nécessaire pour le maintien de la posture en étudiant le cervelet humain, ou encore de la taille approximative d'un réseau de traitement visuel capable d'obtenir les mêmes performances que l'humain. Même si ces ordres de grandeur sont évidemment très approximatifs, ils fournissent néanmoins des renseignements précieux, une « vérité terrain ».

Prenons l'exemple des animaux se rapprochant le plus de l'Homme. On peut constater que la génétique les a pourvus de cerveaux très complexes, bien supérieurs à ce qu'on peut trouver ailleurs dans le règne animal. Si l'on souhaite travailler plus particulièrement sur le buste, comme c'est notre cas, en oubliant les spécificités de la locomotion, on peut observer que tous les animaux connus pour leur habileté à la manipulation d'objets sont considérés comme les plus intelligents du règne animal, et la plupart forment soit des structures complexes de ruches, dont la complexité de l'intelligence collective est débattue, soit sont pourvus de cerveaux parmi les plus complexes et les plus gros.

Cet ensemble d'observations soulève quelques questions. Si on prend ces analogies avec le vivant, il est alors prévisible que le système de contrôle nécessaire à un corps humanoïde, pour son développement et sa capacité à manipuler des objets, nous force à répliquer un cerveau « humain » ou du moins suffisamment complexe.

On peut alors se demander s'il est optimal de tenter de résoudre de front l'un des problèmes qui s'annonce comme étant parmi les plus complexes, en n'ayant, dans bien des cas, que des ébauches de systèmes plus simplistes. Est-il par exemple cohérent de résoudre le cas humain ou humanoïde avant d'avoir réussi à contrôler de façon crédible un « robot chat » ? Est-il cohérent de tenter le contrôle de robots manipulateurs avant d'avoir confirmé notre maîtrise dans d'autres domaines que l'on sait partagés bien plus largement dans le monde animal (vision, locomotion, etc.) ?

Cette question est d'autant plus importante, qu'a priori, rien n'interdirait de concevoir des robots capables d'interagir, mais non manipulateurs. Les Hommes interagissent en effet très largement avec des animaux quadrupèdes, et s'attaquer à cette problématique en prenant comme exemple ce type d'organisme semblerait, pour le domaine de l'interaction Homme-machine, a priori plus stratégique.

Ceci est peut-être la critique la plus solide contre le fait de développer des algorithmes d'interaction Homme-machine sur un corps robotique de forme humanoïde. Néanmoins nous allons voir que certains aspects positifs contrebalancent ces défauts.

1.3.3. Les avantages d'un robot humanoïde

Les avantages qu'un robot humanoïde apporte dans notre cadre peuvent être soulignés.

Avantages dans l'interaction

Dans l'interaction tout d'abord. Même si de fortes interactions entre un animal non humanoïde et un humain sont évidemment possibles (exemple du chat et du chien), une large part de l'interaction entre humains se fait via les expressions faciales et les mouvements des bras en plus du langage que nous n'aborderons pas dans cette thèse.

Un robot humanoïde nous permet donc de profiter de ces aspects (expressions faciales et mouvements des bras), et donc de les exploiter dans nos travaux, ce que d'autres types de robots ne permettraient pas sans paraître en décalage avec leur modèle d'inspiration.

On notera cependant que de nombreux travaux démontrent qu'il est possible de communiquer, du moins de transmettre émotions et ressentis, à travers des interfaces très minimalistes. Les travaux d'Heider et Simmel notamment peuvent être cités ([HEIDER et SIMMEL, 1944](#)).

Avantages physiques

Un corps humanoïde possède plusieurs avantages physiques dont il est intéressant de profiter en robotique. À titre d'exemple, certaines expériences précédentes du laboratoire ETIS ([RENVERGÉ, 2013](#)) nous ont permis de prendre conscience de l'importance du positionnement de la tête par rapport aux bras. Pour des raisons pratiques, ces expériences positionnaient les caméras à l'extérieur de la scène, pointées vers le bras et l'objet à manipuler. L'un des défauts majeurs de ce genre de positionnement est de dissimuler complètement de temps à autre l'objet par le bras. Il existe alors une gamme de mouvements difficile à réaliser faute d'observation, et certaines expériences peuvent alors être mises en échec par ce simple problème de positionnement. A contrario, une caméra placée sur une tête mobile avec des bras fixés à la position de l'épaule humaine permet d'observer la scène et l'objet manipulé sans que la capture d'image ne soit gênée par le bras lui-même. Le positionnement à l'épaule permet ainsi d'extraire au maximum les parties inutiles des bras de la scène à observer.

Un robot humanoïde avec une tête et des yeux capables de suivre l'objet à manipuler permet alors de profiter pleinement de ce type de simplifications implicites.

Avantages interdisciplinaires dans le cadre de la robotique humanoïde

Enfin, l'un des avantages les plus importants de la robotique humanoïde est que le corps humanoïde en robotique nous permet de communiquer de façon riche avec les différents domaines des « sciences cognitives ».

Les sciences cognitives forment une communauté pluridisciplinaire composée (sans ordre d'importance) de roboticiens, informaticiens, neurobiologistes, biologistes, médecins et psychologues. C'est cet ensemble de disciplines et la mise en commun des différents savoirs la constituant via des partenariats et des échanges qui constituent la clé du développement scientifique dans ce domaine.

Il faut alors remettre le choix d'un corps robotique dans ce contexte. Un corps robotique humanoïde nous permet en effet de profiter de collaborations riches avec les domaines scientifiques étudiant l'être humain, et notamment dans notre cadre avec les domaines de la psychologie ayant trait aux mouvements humains.

1.4 | Pourquoi un contrôle hydraulique ?

Le choix technique des types de moteurs équipant un robot peut paraître anodin face aux considérations précédentes, néanmoins, il n'est pas sans conséquence sur les capacités finales du robot et notamment sur ses capacités à interagir avec l'être humain.

Le besoin d'un contrôle dans les gammes dynamiques du mouvement chez l'humain

L'être humain fonctionne et interagit dans une gamme de dynamiques temporelles assez précise. Pour le choix d'un type de contrôle moteur au plus proche de l'humain, il est donc intéressant de prendre en compte la capacité moyenne d'un être humain dans diverses situations, notamment en situation de travail et de sport.

Dans le cadre restreint des mouvements du bras, l'être humain possède des vitesses moyennes, en situation de travail, de $1 \text{ m}\cdot\text{s}^{-1}$ en bout de bras à $25 \text{ m}\cdot\text{s}^{-1}$ en bout de bras en situation sportive. Les accélérations pour le bras chez l'humain peuvent quant à elles globalement aller de $5 \text{ m}\cdot\text{s}^{-2}$ à $7.5 \text{ m}\cdot\text{s}^{-2}$.

On retiendra donc pour la suite qu'il n'est pas possible d'utiliser n'importe quelles gammes de dynamiques de mouvements dans le cadre de la sélection d'un outil robotique interactif et qu'il faut respecter certains critères. D'une manière générale, au vu des autres contraintes à prendre en compte, la problématique se résumera à trouver un type de moteur suffisamment rapide pour que les mouvements possibles soient intéressants.

1.4.1. Avantages et inconvénients des différents types de moteurs

La difficulté du choix d'un moteur pour un robot, outre le fait de correspondre aux dynamiques du mouvement humain, est également de trouver un bon compromis entre ses capacités de vitesses, d'accélérations, mais surtout ses capacités à supporter des charges et des fluctuations de charges en cours de mouvement. Il faut en effet déjà être capable de porter l'architecture mécanique du bras robotique tout en maintenant une certaine dynamique.

Les moteurs électriques ont été largement utilisés dans le cadre de la robotique et dans le cadre de la robotique humanoïde. La simplicité de leur alimentation et leur taille aisément réductible les rendent adaptés à tout le domaine de la robotique embarquée. A contrario, les systèmes pneumatiques et hydrauliques ont le défaut de devoir nécessiter un compresseur, qui peut s'avérer assez volumineux. Même s'il existe divers travaux visant à réduire ou à intégrer les compresseurs dans l'architecture mécanique du robot ([ALFAYAD et al., 2011a](#); [ALFAYAD et al., 2011b](#)) ceux-ci n'existent encore qu'à l'état de prototype non déployable sur robot réel.

Les moteurs électriques ont toutefois beaucoup de mal à dépasser certaines contraintes. En général, la nécessité d'avoir un bras correspondant aux caractéristiques humaines pour l'interaction restreint le choix du modèle de moteurs électriques. En effet, si les moteurs électriques sont aisément réductibles, ils sacrifient dans l'opération leur capacité à maintenir de hautes vitesses et accélérations tout en portant des poids.

Ce constat a amené le milieu de la robotique à privilégier des systèmes de moteurs couplés à des réducteurs. Dans bien des cas le résultat est un compromis. Si on observe les modèles de bras robotiques électriques de tailles humaines existants, les réponses sont soit des bras robotiques capables de supporter une charge, mais en ne maintenant que des vitesses assez lentes, soit des modèles très rapides capables de ne porter que des charges assez faibles.

1.4.2. La problématique de l'effort subit avec les moteurs électriques

La capacité d'un bras robotique à subir un effort externe est primordiale dans le cas de la robotique interactive. On souhaite en effet obtenir un outil robotique capable de partager le même espace de travail qu'un être humain et donc, potentiellement, d'être heurté ou guidé par celui-ci. Pourtant, on souhaite également pouvoir porter des charges lourdes ou exécuter des mouvements rapides, ce qui, a priori, est incompatible. En réalité, ce que l'on souhaite pour combiner le meilleur de ces solutions, c'est de pouvoir contrôler la « dureté » du contrôle : souple dans les conditions dangereuses ou d'interaction donc facilement maîtrisable par un humain et plus dur dans des conditions sécurisées ou vitesse, précision et force sont nécessaires.

Un moteur électrique est théoriquement adapté à ce genre de cas. Il est normalement pleinement réversible et la force appliquée peut être contrôlée par la commande ou encore observée de l'extérieur (le courant passant dans le moteur électrique sera lié à l'effort en cours).

Néanmoins, en pratique, peu de moteurs électriques sont utilisés seuls sur un bras robotique, et la plupart passent par des réducteurs qui ne sont souvent tout simplement pas réversibles, ce qui annule ce lien entre l'effet observable externe et l'état du moteur.

Un autre problème majeur est que les moteurs électriques, en cas de blocage, en l'absence d'évacuation de leur énergie par un mouvement rotatif, se comportent alors comme un circuit fermé résistant, ce qui entraîne une surchauffe pouvant s'avérer destructrice.

Il existe quelques solutions à ce problème. La plus courante consistant à utiliser des moteurs standards tout en les dotant de mécanismes de détection d'effort (mécaniques ou en observant le courant). Ceux-ci agissent soit en compensant activement si un effort intense est détecté, soit en suivant le sens de l'effort demandé comme par exemple dans le cas d'un humain prenant le bras robotique pour le guider, soit en agissant comme coupe-circuit. D'autres solutions passent par des mécanismes de détections externes (optique par exemple) qui viennent couper le moteur en cas de problème.

Ces deux solutions ont l'inconvénient, dans notre cas, de rajouter des surcouches de système de contrôle, ce qui est problématique lorsque l'on souhaite pouvoir vérifier et observer chaque couche d'un algorithme de contrôle biomimétique et lorsque l'on veut, plus généralement, conserver la maîtrise des causes et conséquences du comportement que l'on observe. De plus, ces solutions sont asservies au fonctionnement et au degré de précision du mécanisme de contrôle externe ce qui pourrait être synonyme d'accident en cas de panne ou d'incident trop rapide.

1.4.3. La solution « compliant »

La compliance est un anglicisme désignant, en robotique, la capacité d'un matériel robotique à subir des efforts tout en étant capable d'être naturellement impacté par ces efforts sans phénomène destructif. Un bras robotique sera par exemple compliant si, en l'absence de commandes externes, il peut être déplacé sans mal d'un endroit à un autre par une personne, sans risque pour le bras.

En robotique interactive, cette propriété est évidemment particulièrement attractive et de nombreux travaux en robotique bio-inspirée ont visé à obtenir un système mécanique compliant ([VAN HAM et al., 2009](#)).

Quelques types d'actionneurs sont plus facilement compliant que le moteur électrique et n'ont pas ses défauts, c'est notamment le cas des actionneurs hydrauliques et pneumatiques.

En effet, contrairement aux moteurs électriques qui se chargent de l'intégralité de la transformation de la puissance en mouvement, ces actionneurs ont l'avantage d'être séparés en deux blocs : la pompe, qui sert à maintenir la pression du fluide, et le vérin, qui transforme cette pression en mouvement (généralement contrôlée par une électrovanne intermédiaire). Sur le robot final, on déportera la pompe en dehors du robot et il ne restera alors plus que le poids des vérins à porter.

L'avantage de ce système est que, pour la même place occupée sur le robot et le même poids, un fort gain en puissance est constaté. Bien que la nécessité de maintenir une pompe connectée au robot contraint énormément les utilisations qu'on peut en faire notamment dans le domaine de la robotique

mobile autonome, ce gain en puissance supprime un grand nombre des contraintes qui s'imposent aux moteurs électriques. Il n'est par exemple plus utile de mettre en place des systèmes de conversion vitesse-énergie comme les réducteurs et on est donc naturellement dans un système où l'actionneur est en contact direct avec l'environnement, sans effet d'irréversibilité.

Un autre avantage, qui profite de ce contact direct, est qu'il est facile dans certaines conditions d'arrêter le système hydraulique ou pneumatique et par un bon réglage de la pression et du débit il est possible d'être aussi « souple » ou « dur » qu'on le souhaite, ce qui était l'un des points essentiels désirés pour la robotique interactive.

Pour finir, l'avantage principal de ces deux types d'actionneurs est de pouvoir compter sur son système de pompe déportée. Ainsi, aucune action sur les vérins ne peut mener à une destruction de la pompe. La manipulation d'un bras robotique hydraulique ou pneumatique est donc ici sans danger pour le système lui-même.

1.4.4. Hydraulique ou pneumatique : la question de la sécurité

Les contrôles hydrauliques et pneumatiques sont tous deux compliants. Afin de choisir, d'autres éléments doivent être pris en considération.

Tout d'abord, il faut reconnaître qu'un système pneumatique a l'avantage de pouvoir évacuer ou se recharger en fluide à n'importe quel point du système, puisqu'il s'agit d'air. Une surpression peut donc être aisément expulsée au niveau du vérin lui-même et les fuites éventuelles n'ont que peu de conséquences. Un système hydraulique, lui, devra garder une réserve de fluide et rapatrier le fluide utilisé au réservoir dans tous les cas, sous peine d'obliger à un remplissage permanent du réservoir. Ces points influent sur la complexité du système et donc sur son coût. Ainsi, un système pneumatique est moins cher qu'un système hydraulique.

Cependant, un système pneumatique est bien plus dangereux qu'un système hydraulique. En effet, les systèmes pneumatiques recourent le plus souvent à l'air, qui est très compressible. Une rupture d'un tuyau véhiculant de l'air comprimé aboutira alors à une explosion ce qui peut s'avérer dangereux pour l'opérateur. À l'inverse, les fluides utilisés dans les systèmes hydrauliques (de l'huile dans notre cas) ont l'avantage d'être incompressibles. Ainsi, en cas de rupture d'un tuyau à forte pression le risque reste minime pour l'opérateur.

Dans le cadre de la robotique interactive, l'utilisation de prototypes de recherche fonctionnant à l'air comprimé est un risque trop grand à prendre en cas d'expérience impliquant des sujets humains.

1.4.5. Conclusion

Ainsi, c'est le trio compliance, faculté à faire varier la « souplesse » de l'actionneur selon le contexte, et sécurité, qui a justifié l'utilisation d'un robot hydraulique dans notre cas. On pourrait également ajouter la disponibilité, car d'autres dispositifs présentant également cet ensemble de propriétés existent, notamment dans le champ d'étude des muscles artificiels, un domaine intéressant qui se développe et qui est de plus beaucoup plus bio-inspiré. Cependant, ces dispositifs ne sont pas encore disponibles en série pour équiper des prototypes robotiques.

On tempérera dans le chapitre suivant les capacités réelles de notre prototype robotique Tino par rapport à ces avantages théoriques. En effet, nous verrons que certains critères tels que la compliance ne sont que partiellement respectés.

1.5 | Conclusion

Dans ce chapitre, nous avons abordé les principales justifications à certains des choix applicatifs faits durant cette thèse. Nous avons notamment vu quels sont les avantages et inconvénients du développement d'algorithmes sur robot réel et quelles sont les particularités justifiant d'utiliser, dans un contexte d'interaction Homme-machine, un robot hydraulique humanoïde comme base d'expérimentation pour des algorithmes de contrôle bio-inspirés.

Nous avons également évoqué les limites de ces choix, aussi bien du point de vue de l'expérimentateur et de la sécurité dans l'interaction qu'à des niveaux plus théoriques.

Dans le chapitre suivant, nous aborderons les détails et la caractérisation du prototype de robot hydraulique humanoïde, unique en France, qui sert de base de développement à certains des modèles étudiés durant cette thèse : le robot Tino.

Caractérisation du robot humanoïde hydraulique Tino

Sommaire du chapitre :

2.1. Introduction	23
2.2. Description des parties du robot Tino	24
2.2.1. Le torse	24
2.2.2. La tête expressive	26
2.2.3. Matériel de commande et de capture	28
2.3. Caractérisation expérimentale du robot Tino	30
2.3.1. Caractérisation de la tête du robot Tino	30
2.3.2. Précision des capteurs du robot Tino	33
2.3.3. Relation commande/mouvement	34
2.3.4. Capacité de mouvement du robot Tino	35
2.4. Phénomènes émergents de la mécanique et du contrôle	39
2.4.1. Validation du phénomène à vide	39
2.4.2. Variation du phénomène selon le poids	40
2.4.3. Validation du phénomène dans un cas plus général	40
2.4.4. Conclusion sur les phénomènes émergents du contrôle	41
2.5. Acceptabilité du robot Tino	43
2.5.1. Présentation résumée du protocole	43
2.5.2. Résultats et interprétations	43
2.6. Conclusion	45

2.1 | Introduction

Les robots Tino et Tino 2 sont des prototypes de robots humanoïdes hydrauliques non mobiles, constitués d'un torse et d'une tête (voir la page des robots Tino¹ ou la description en vidéo du robot²).

Le torse porte deux bras contrôlables indépendamment de 7 degrés de liberté chacun, 3 pour l'épaule, un pour le coude et 3 pour le poignet. Tino 2 quant à lui possède un degré de liberté supplémentaire pour les bras permettant une rotation vers l'intérieur.

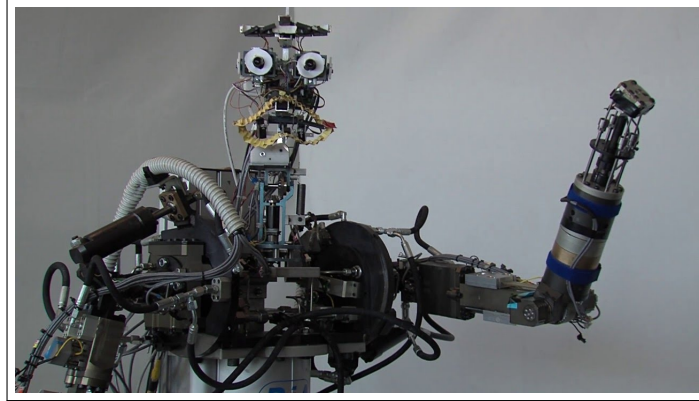


FIGURE 2.1. – *Le robot humanoïde hydraulique Tino.*

Pour des raisons de simplicité, la tête est entièrement électrique. Elle possède deux yeux (caméras) très réactifs capables de faire des saccades ou de suivre rapidement un objet en mouvement. Elle est également « expressive » et possède sourcils et bouche capables de réaliser des expressions faciales. Le tout est placé sur un cou mobile capable de mouvement de rotation sur un axe droite-gauche et haut-bas et qui est souvent contrôlé de la même manière que les yeux pour se diriger vers la cible du regard du robot.

Le robot Tino fut la plateforme principale utilisée durant cette thèse et constitue l'un des premiers prototypes du genre en France. Cette thèse également fut la première à utiliser et étudier ce robot, de fait nous allons détailler dans ce chapitre l'architecture mécanique du robot Tino, ses propriétés ainsi que les expériences de caractérisation que nous avons menées sur ce robot en partenariat avec le laboratoire LISV. Nous décrivons ensuite certaines propriétés émergentes que nous avons isolées et qui peuvent s'avérer utiles au contrôle. Enfin, nous concluons par la présentation d'un jeu d'expériences en partenariat avec le laboratoire SCALab qui nous a permis de jauger de l'acceptabilité du robot Tino dans un cadre d'interaction.

1. ↑ <https://www.etis.ensea.fr//neurocyber/web/en/tino.php>

2. ↑ Présentation humoristique de Tino, <https://youtu.be/C5CE84qt59k>

2.2 | Description des parties du robot Tino

2.2.1. Le torse

Le torse de Tino est la principale composante du robot (voir la photo présentée sur la figure 2.2). Il est constitué d'un torse manuellement relevable ou abaissable et de deux bras hydrauliques à 7 ou 8 degrés de liberté selon le modèle (Tino et Tino 2). Le torse et les bras ont été conçus par le laboratoire LISV sur la base de leur projet HYDROÏD, on retrouvera donc une abondante documentation sur ces bras et ce torse dans leurs travaux (ALFAYAD, 2009 ; ALFAYAD, 2011).

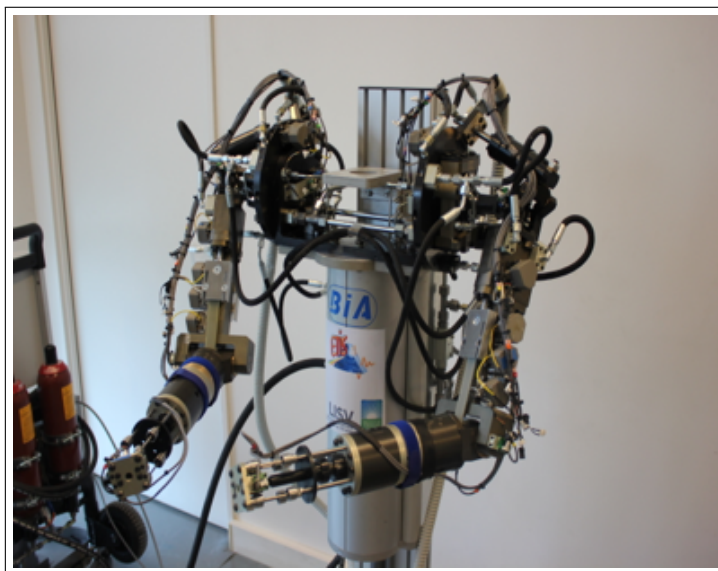


FIGURE 2.2. – Buste du robot Tino avant installation de la tête expressive.

On peut voir une vue synthétique du bras sur le schéma figure 2.3

On peut voir sur le schéma le positionnement important des deux vérins de l'épaule, qui assurent l'éloignement du bras au corps dans l'axe du corps et la levée du bras dans l'axe du corps. Le tout étant monté sur un vérin rotatif qui assure la rotation de l'ensemble de l'épaule autour de l'axe du corps.

Le coude, autre axe important, est actionné par un vérin hydraulique positionné sur le bras et permettant, via une courroie, un mouvement rotatif.

Pour finir, quatre vérins indépendants permettent le contrôle du poignet et un vérin rotatif asservit le tout afin d'obtenir toutes les positions désirées. Le schéma équivalent mécanique qui résulte de cet assemblage est présenté sur la figure 2.3 (cas de Tino 1).

Le robot Tino 2, quant à lui, dispose d'un vérin rotatif supplémentaire permettant de tourner les bras vers l'intérieur autour de l'axe du corps.

Compliance réelle du robot Tino

Nous avons vu que l'un des avantages théoriques d'un robot hydraulique par rapport à un robot basé sur des moteurs électriques était sa capacité à transmettre au système, de façon souple et sans casse, des mouvements ou des heurts imposés par l'extérieur. La capacité à déplacer ou arrêter un bras de robot est primordiale pour l'interaction humaine et la sécurité de l'interaction.

L'hydraulique apporte cet avantage théorique, mais qu'en est-il du robot Tino ? On peut noter qu'au moins l'un des axes du robot comporte un système intermédiaire entre le vérin et l'action puisque le

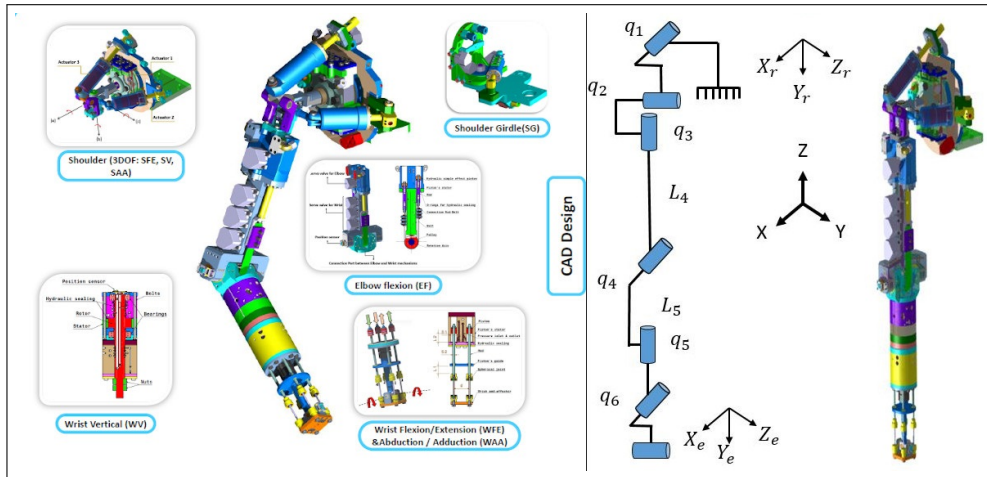


FIGURE 2.3. – À gauche : Détail d'un bras du robot Tino. Figure issue des travaux de l'équipe de Samer Alfayad au LISV. À droite : Schéma équivalent mécanique du bras de Tino 1. L'axe du corps est noté Y_r . Figure issue des travaux de l'équipe de Samer Alfayad au LISV.

coude se retrouve actionné via une courroie crantée. Cette courroie comporte en pratique les mêmes défauts que les systèmes réducteurs avec les moteurs électriques : elle isole les actions de l'extérieur du vérin et introduit des phénomènes non-linéaires (plissement de la courroie, saute de crans, etc.) qui empêchent de véritablement déplacer manuellement le coude du robot avec facilité. Il s'agit donc d'une importante réserve en pratique quant aux avantages théoriques que l'hydraulique pourrait ici nous apporter.

Fonctionnement des vérins et servovalves du robot Tino

Les vérins du robot Tino sont actionnés par un système assez classique de servovalves à buse (voir figure 2.4). Un champ électromagnétique vient déplacer une palette dans un sens ou un autre autour d'un pivot. Les extrémités de cette palette venant alors obstruer une sortie d'huile aux extrémités U ou V de deux buses. Selon la sortie obstruée, la pression augmente d'un côté ou de l'autre du système, poussant un tiroir vers A ou vers B. Le tiroir ainsi déplacé vient quant à lui obstruer ou libérer les canaux d'huile principaux qui pilotent le vérin.

Il est important de noter qu'il s'agit d'un système autorégulé car l'extrémité de la palette est reliée au tiroir. Un déplacement du tiroir vient donc recadrer la palette et réguler la pression. Ce qui donne donc une relation entre le champ électrique appliqué sur la buse et le déplacement du tiroir, permettant ainsi de contrôler le déplacement du tiroir.

Cas critiques découlant du fonctionnement des servovalves

Plusieurs problèmes apparaissent lorsque l'on se penche sur le fonctionnement des servovalves. Le premier est qu'il existe un système intermédiaire électromagnétique, puis hydraulique, de commande, avant même d'arriver au vérin. Il y a donc un risque pour que ce système intermédiaire comme dans le cas des réducteurs ou de la courroie pose problème vis-à-vis de la compliance désirée pour le robot.

Il est par exemple ici difficile d'observer l'impact d'une action appliquée au vérin sur le champ électrique de commande. Dans un sens, on est donc plus isolé ici des actions d'une personne en interaction avec le système qu'on ne le serait avec un moteur électrique simple, où l'observation du courant nous permettrait de déduire l'état de blocage ou non du moteur électrique.

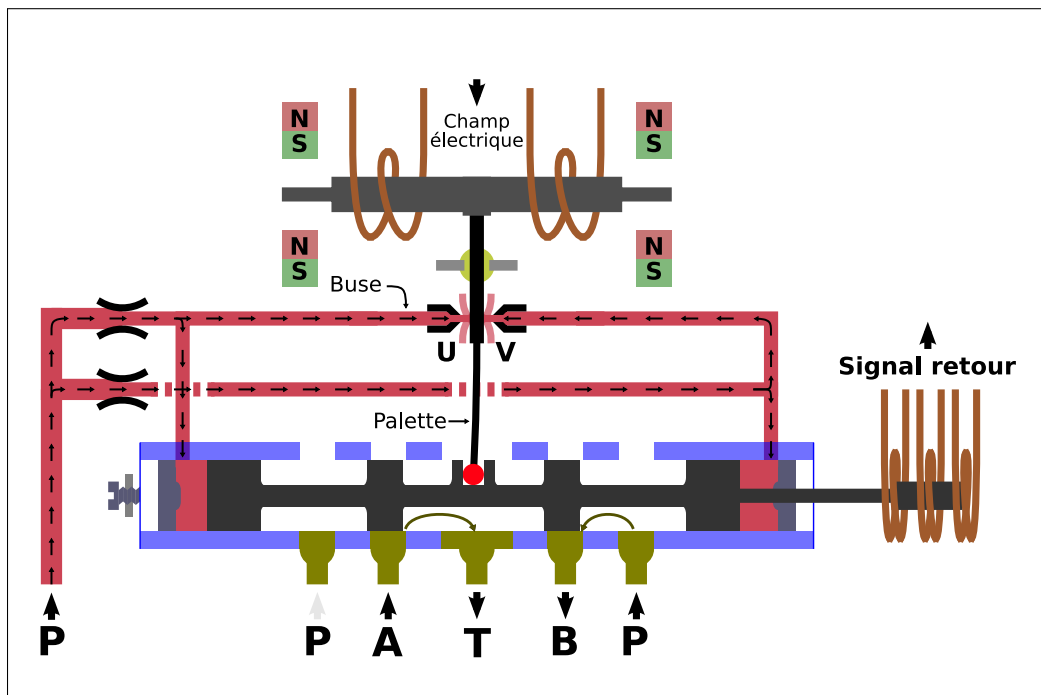


FIGURE 2.4. — Schéma du fonctionnement des servovalves : un champ déplace une palette (en noir) reliée à un tiroir (gris foncé) ceci permettant de boucher les orifices des buses U ou V ce qui fait monter la pression à gauche ou à droite. On ne représente pas l'écoulement depuis les buses, mais les buses s'écoulent en fait dans T, le retour d'huile. La pression augmentant à gauche ou à droite elle déplace le tiroir, ce qui rééquilibre la palette et rééquilibre donc la pression. Le tiroir va faire varier le débit en A ou B en changeant les flux d'huiles entre P, A, B et T. Le vérin est connecté à A et B.

Un autre défaut du système, plus problématique, est que le vérin se retrouve dans une situation de blocage complet si les deux canaux A et B se retrouvent fermés. Sur le schéma, 2.4 on observe que ce cas est obtenu lorsque l'on est presque au point d'équilibre du système.

On obtient un comportement particulièrement peu bio-inspiré puisque le point d'équilibre, qui correspond à une absence de force, est également le seul point où le vérin n'est pas mobile depuis l'extérieur car ses compartiments n'ont nulle part où déverser l'huile.

En pratique, il est possible de déplacer les bras du robot Tino manuellement sauf si l'on est sur ce point critique. Il s'agit donc là encore d'un bémol pratique important par rapport aux avantages théoriques de l'hydraulique pour nos applications.

Des solutions existent, mais elles nécessiteraient l'emploi conjointement de deux servovalves pour chaque vérin. Les servovalves étant ce qui coûte le plus cher sur le robot, cela reviendrait à doubler le coût du prototype, sans compter le poids supplémentaire sur chaque bras. Ce n'est donc pas cette solution qui a été retenue.

2.2.2. La tête expressive

La tête du robot Tino est une tête électrique dotée de deux caméras contrôlables indépendamment et d'un jeu de composantes mobiles (sourcils, bouche, etc.) capables de former des expressions faciales (voir figure 2.5).

La tête en elle-même, est principalement constituée d'un cou mobile capable de mouvement de rotation droite-gauche et haut-bas, relativement lents. On peut avoir un aperçu du mécanisme du cou sur la figure 2.7 et schématiquement sur le schéma figure 2.6

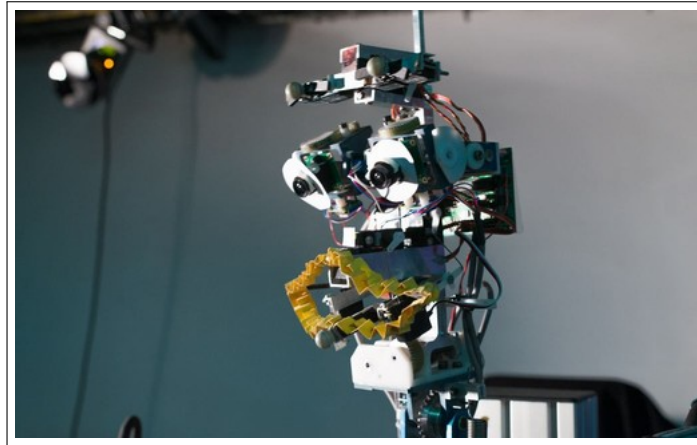


FIGURE 2.5. – Photographie de la tête expressive complète du robot Tino. Photographie prise durant la semaine de manipulation en partenariat au laboratoire SCALab.

Le cou est contrôlé par un jeu de roues crantées et de courroies. (voir figure 2.7 et figure 2.6) Les servomoteurs qui les commandent sont plutôt conçus pour être capables de contrebalancer l'inertie rotative de l'appareil plutôt que pour donner une vitesse forte. Les yeux sont également contrôlés par un jeu de servomoteurs, ils sont capables de mouvements particulièrement rapides (saccade oculaire).

Les diverses cartes de commandes (expression, récupération du flux vidéo des caméras, alimentations, etc.) sont fixées à l'arrière de la tête. En pratique cela a tendance à faire souffrir les moteurs du cou au vu du poids du matériel embarqué sur la tête. Diverses solutions ont été testées tout au long de cette thèse (abaisser les cartes de contrôle pour les rapprocher de l'axe de rotation par exemple), mais sans grand succès.

Les composants mobiles (sourcils, bouche, etc.) sont capables de former des expressions faciales grâce à de petits servomoteurs servant à contrôler la bouche et les sourcils (voir figure 2.5). Ce jeu d'expressions est directement inspiré des précédents travaux du laboratoire, notamment ceux de Sofiane Boucenna (BOUCENNA, 2011). Un jeu de 5 expressions primaires, qui peuvent être inter mêlées, est principalement utilisé.

Ce dispositif fut principalement utilisé pour renforcer l'interaction Homme-machine avec l'expérimentateur. Nous l'utilisons dans le cadre de l'interaction par le biais d'un objet que le robot tentait de suivre, le robot manifestait alors de la surprise via ses expressions faciales s'il se retrouvait dans une situation où il n'était plus capable de suivre l'objet puis malheureuse en cas de perte totale. Dans le cas contraire il manifestait une expression heureuse. Du point de vue de l'expérimentateur cela permet un retour très intuitif sur les éventuels accidents amenant le robot à perdre ou retrouver l'objet poursuivi et donc à des corrections rapides de l'expérience en conséquence. On peut voir un aperçu de ces propriétés expressives sur la vidéo de présentation du robot Tino (BLANCHARD, 2014)

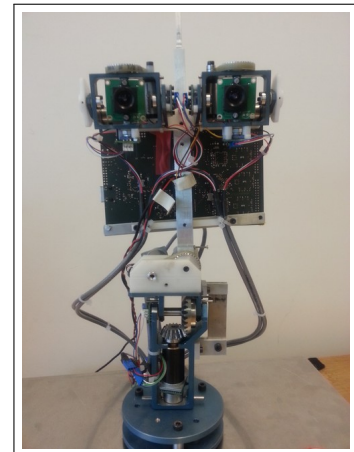


FIGURE 2.7. – Sans les expressions, on peut observer les mécaniques internes de la tête.

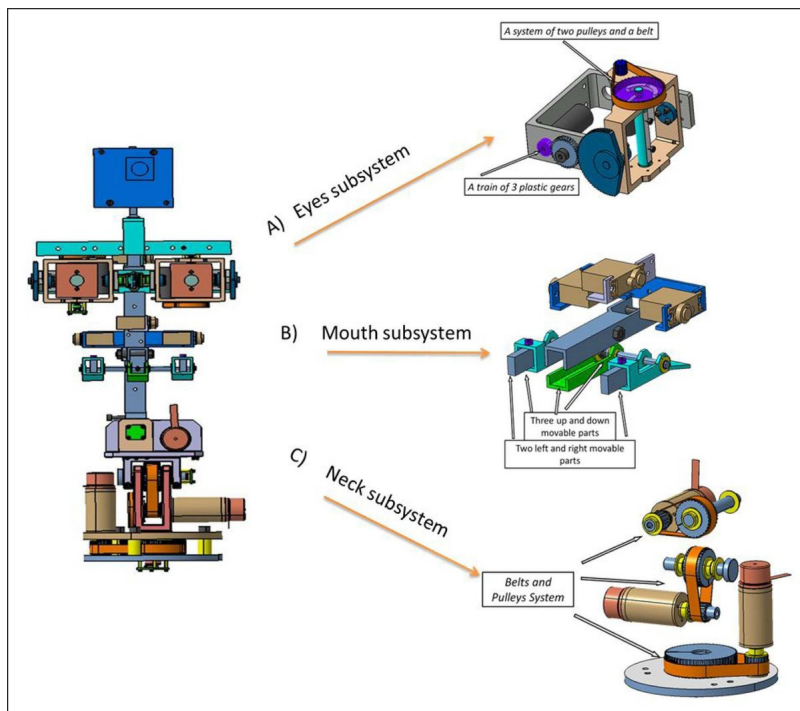


FIGURE 2.6. — Schéma de la tête du robot Tino, reproduit depuis la publication de nos travaux sur la tête en partenariat avec le laboratoire LISV (ALFAYAD *et al.*, 2016).

2.2.3. Matériel de commande et de capture

Le robot est équipé de divers senseurs angulaires permettant de récupérer en temps réel les angles de l'ensemble des articulations du robot. Notons qu'il s'agit toutefois de senseurs externes et non de senseurs placés sur les vérins ou de retour des commandes envoyées sur les servovalves. La caractérisation de ces senseurs est abordée plus en détail dans la section 2.3.2.

L'ensemble des signaux provenant de ces senseurs est capturé par une baie de contrôle qui centralise les ordres et les retours des senseurs du torse. Une seconde baie similaire est utilisée pour centraliser les ordres à destination de la tête expressive et les retours de ses senseurs.

Ces baies sont équipées d'un système temps réel de la société BIA qui fonctionne à 1000 Hz, permettant une acquisition des données et une prise en charge des ordres extrêmement fluide.

Les connexions externes communiquent avec ces baies par Ethernet et permettent d'écrire ou lire directement à des adresses spécifiques de la mémoire vive des baies ce qui assure une extrême fluidité dans la transmission.

Résumé des travaux effectués sur le système de commande et de capture

Le prototype Tino était fourni avec des routines de contrôle classiques passant par une couche de contrôleur PID directement intégrée dans la baie. Dans cette thèse, la nécessité de pouvoir travailler à très bas niveau nous a obligés à retravailler une partie des codes fonctionnant sur ces baies afin de nous permettre de modifier directement et sans intermédiaire les valeurs du courant des servovalves, ceci afin de pouvoir contrôler les forces exercées par les vérins sans aucun filtre.

Afin de conserver une certaine sécurité et pour éviter de forcer sur les vérins lorsque ceux-ci sont en bout de course, une sécurité a été implantée.

Il s'est avéré compliqué de concevoir une sécurité permettant d'arrêter véritablement un axe à cause

des effets d'élan dus à l'inertie du bras. De plus, une valeur nulle de commandes en courant ne correspond pas forcément à une ouverture nulle sur les servovalves. De fait, une valeur nulle sur la commande peut aboutir à une force exercée par les vérins.

La solution adoptée fut que, lorsque l'on dépasse une borne limite, la sécurité passe automatiquement sur un mode de contrôleur proportionnel pour ramener le bras à la borne limite. Le mode de contrôleur proportionnel permet en effet d'éviter ou de négliger une partie des problèmes de calibration de la commande si ses réglages sont suffisants.

Pour éviter de fausser les résultats de caractérisation, cette sécurité a été désactivée dans les expériences présentées dans la suite de ce chapitre.

2.3 | Caractérisation expérimentale du robot Tino

Lorsque cette thèse a débuté, il n’existait presque aucune documentation sur le robot Tino, le prototype venait d’être fabriqué par l’entreprise BIA et si les caractéristiques théoriques désirées étaient bien connues, les caractéristiques pratiques du prototype réel étaient quant à elle inconnues. Cela ne nous a pas empêchés de faire des expériences mêlant contrôle de robot réel et IA bio-inspirée dans la mesure où nous basons justement nos algorithmes sur la découverte par le robot de son propre corps de façon dynamique à travers l’apprentissage. Cependant, cela empêchait d’être rigoureux sur la fiabilité des données renvoyées par notre matériel embarqué, sur l’analyse des capacités de notre robot et sur leurs impacts sur les résultats observés.

Diverses collaborations inter équipes, notamment avec le laboratoire LISV de Versailles et le laboratoire SCALab de Lille 3, ont pu être mises à profit pour caractériser expérimentalement le robot Tino.

2.3.1. Caractérisation de la tête du robot Tino

Dans le cadre d’une collaboration avec le LISV il s’est avéré nécessaire de caractériser les capacités mécaniques de la tête du robot Tino.

Alors que le LISV s’occupait de l’aspect simulation, nous avons expérimenté les aspects pratiques de la caractérisation. En utilisant un algorithme de suivi (voir le schéma figure 2.8 nous avons pu tester la réactivité de la boucle complète par rapport au suivi d’objet. On peut retrouver l’ensemble des résultats de cette section à la référence : (ALFAYAD et al., 2016).

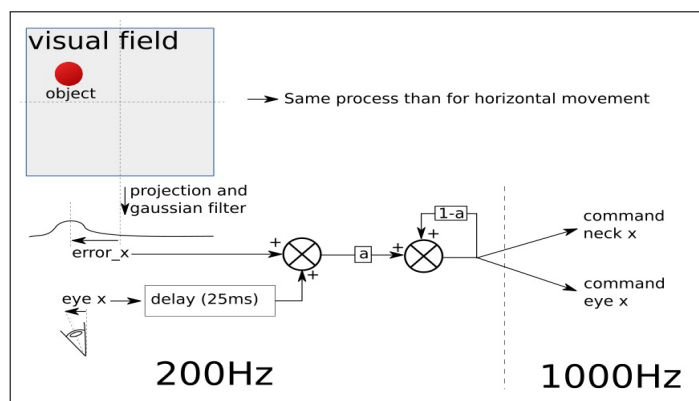


FIGURE 2.8. — *Présentation simplifiée de l’architecture de contrôle des yeux et du cou dans le cadre de la caractérisation de la tête. Graphique issu la publication collaborative avec le LISV sur ces travaux (ALFAYAD et al., 2016).*

Les premiers tests se sont fait simplement sur la détermination des capacités extrêmes de notre matériel et notamment des yeux du robot. On a ainsi pu éprouver les capacités de réaction des yeux à des commandes violentes en créneaux. Le changement d’état se fait en 1 ms. Le résultat est présenté sur le graphique figure 2.9 . On peut observer que selon le sens de variation (montant ou descendant), l’œil peut mettre de 0,2s à 0,5s à se stabiliser autour d’une position.

Ces résultats nous permettaient d’avoir une idée des capacités des yeux de Tino, mais il était tout de même difficile de combiner chaque caractéristique technique de la chaîne de traitement (capture des caméras, traitement électronique, capture de données côté ordinateur, algorithme de suivi, latence du contrôle, etc.) afin d’avoir une idée des capacités réelles de la tête du robot en situation expérimentale.

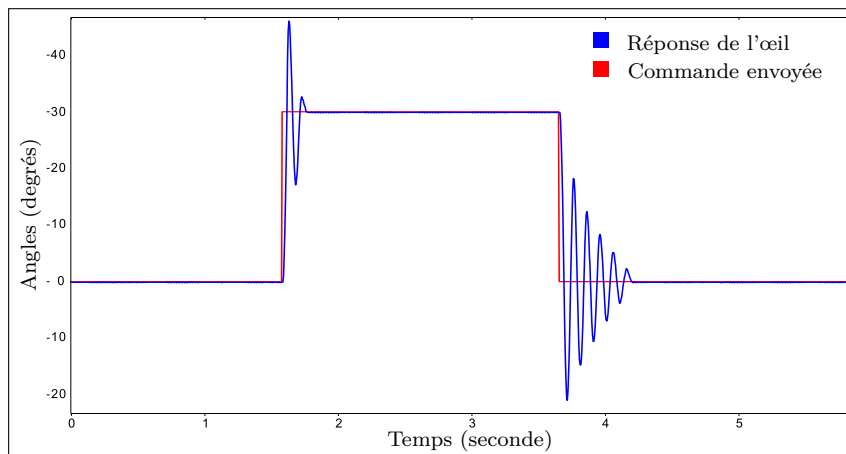


FIGURE 2.9. – Courbe de réponses des yeux du robot face à un créneau durant deux secondes. Le changement d'état se fait en 1 ms.

Pour pallier ce problème, l'idée de mettre en place un test complet de la boucle de réaction du robot a été développée. Nous avons donc effectué une gamme de tests dans lesquels le robot déplace un objet (ici une balle rose) pendant que la tête et les yeux tentent de poursuivre l'objet. Nous avons testé ce protocole sur des mouvements verticaux et horizontaux à diverses fréquences (voir figure 2.10).

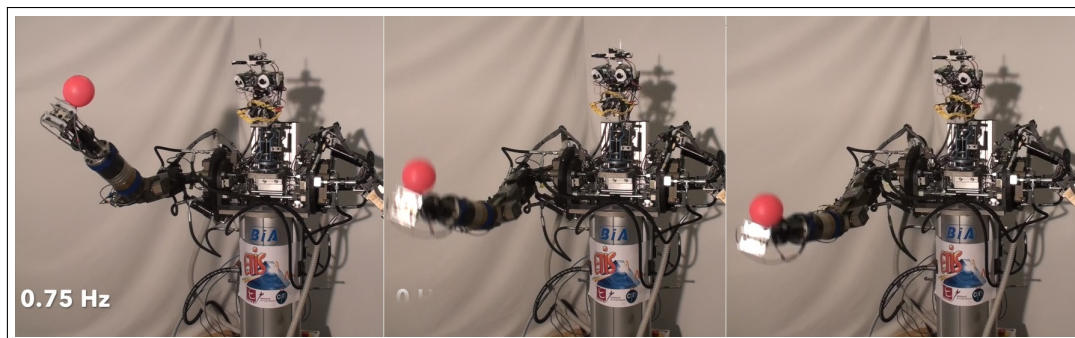


FIGURE 2.10. – Protocole illustré de caractérisation de la réactivité du cou et des yeux du robot Tino . Le robot déplace une balle rose verticalement grâce à ses bras hydrauliques pendant que sa tête et ses yeux tentent de poursuivre la balle. On mesure alors la différence entre la position de la balle dans le champ visuel et la position des yeux. Le protocole peut être observé en vidéo ¹.

Les résultats montrent une capacité de l'ensemble tête/œil à suivre des mouvements oscillants allant jusqu'à 1,5 Hz sur l'étendue du mouvement vertical du robot Tino avec une précision suffisante pour ne pas perdre la balle, ce qui était le point qui nous intéressait particulièrement. Les résultats des extremums sont présentés figure 2.11. Nous avons également jugé utile de caractériser la différence de participation relative des yeux et du cou. En effet, le cou étant plus lent que les yeux, des effets de compensation entrent en jeu et deviennent importants à certaines vitesses, notamment selon l'axe vertical. Les yeux arrivant plus vite sur la cible que le cou et devant ensuite compenser lorsque le cou se rapproche, cela donne lieu à quelques oscillations (voir figure 2.12).

1. [↑] <https://www.etis.ensea.fr//neurocyber/web/en/tino.php>

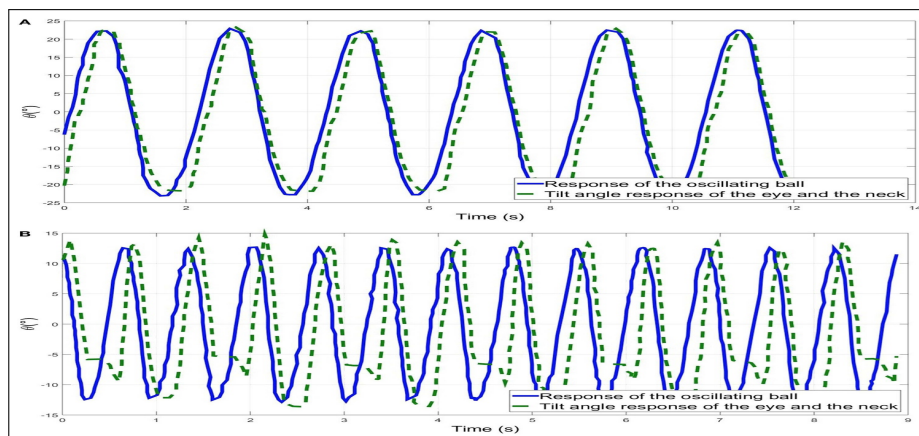


FIGURE 2.11. – Résultats extremum des expériences de caractérisation de la boucle complète de vision-action de la tête de Tino. Courbe du haut : 0,1 Hz, courbe du bas : 1,5 Hz. Le système accuse un retard dans cette dernière, mais il n’y a pas de décrochage, la largeur du champ de vision est suffisante pour éviter que le retard ne fasse perdre la cible.

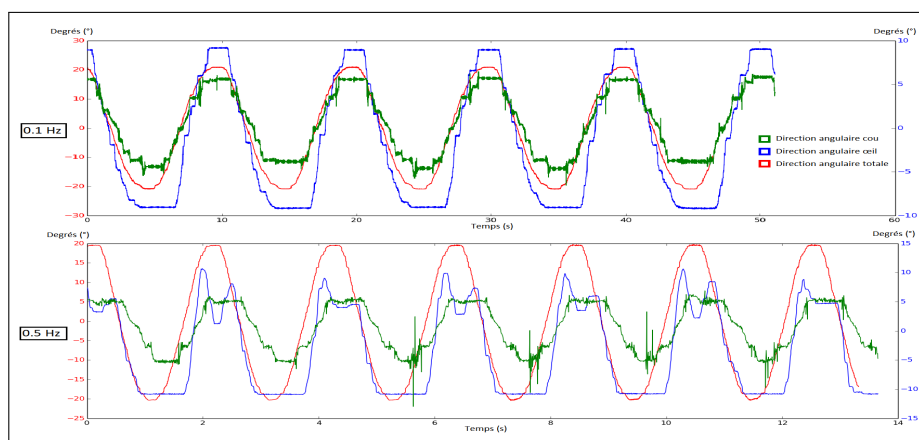


FIGURE 2.12. – Participation œil vs cou. En haut : suivi d’un objet oscillant vertical à 0,1 Hz, en bas : 0,5 Hz. La courbe rouge est la position dans le champ visuel (axe vertical gauche), la courbe bleue montre l’angle des yeux et la verte l’angle du cou, les deux sont dans le référentiel vertical de droite. À basse vitesse les participations sont équilibrées, à plus haute vitesse les yeux prennent de l’avance puis compensent.

2.3.2. Précision des capteurs du robot Tino

L'une des inconnues vis-à-vis du prototype Tino était de savoir à quel point nous pouvions nous fier à l'ensemble de la chaîne de capture d'information du robot Tino et à quel point elle correspondait à une image de la réalité. Afin d'évaluer ce point, nous avons profité d'une semaine d'expérimentation au laboratoire partenaire du LISV pour utiliser leur salle de capture de mouvement équipée de caméras permettant une reconstruction 3D. Nous avons testé dans divers cas la précision des capteurs face aux données capturées qui nous servaient alors de « vérité terrain ».

Protocole

Le robot Tino était placé au centre d'un environnement constitué de 7 caméras d'acquisition 3D de marque Vicon, fonctionnant via un jeu de réflecteurs infrarouges. Pour des raisons pratiques, nous avons supposé que les bras, dont la construction est rigoureusement symétrique, donnent des résultats identiques et nous nous sommes limités au seul bras droit durant ces séances d'acquisition. Nous avons ensuite transformé les données 3D vers le référentiel des capteurs du robot, un référentiel angulaire par rapport aux axes du bras considéré, afin de faire la comparaison. Nos tests en pratique nous ont permis de définir que chaque capteur angulaire réagit sous la forme d'une équation linéaire dont les paramètres peuvent diverger entre capteurs. Pour chaque capteur, on identifie donc au préalable ces paramètres avant d'en comparer la précision.

Résultats expérimentaux

On présentera ici la comparaison entre les données reconstruites et les capteurs, ainsi que la courbe d'erreur entre ces deux jeux de données. Le graphique présenté sur la figure 2.13 présente un des résultats dans le cas de l'épaule droite et le graphique présenté sur la figure 2.14 présente un des résultats dans le cas du coude droit.

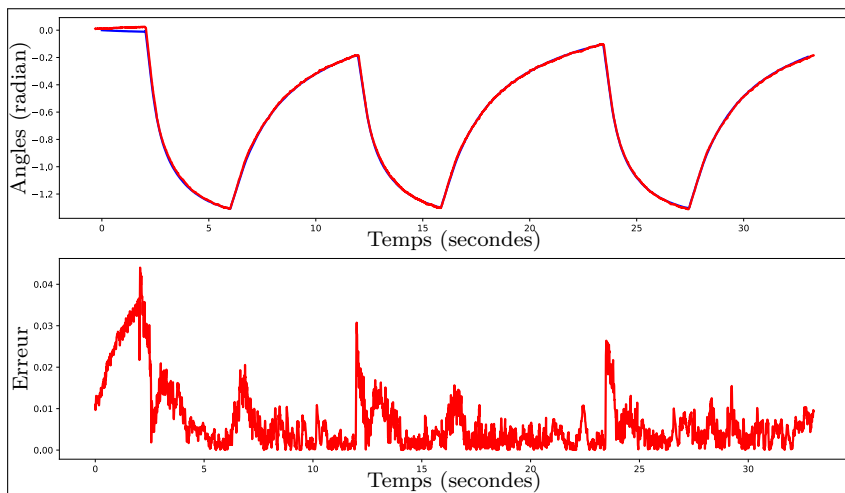


FIGURE 2.13. — Courbes comparant les données reconstruites depuis les caméras et les capteurs pour l'épaule droite. En bleu, le résultat de la capture 3D reconstruite et exprimée dans l'unité des capteurs. En rouge, le résultat capturé par les capteurs embarqués. La courbe en bas traduit l'erreur entre les deux.

Les résultats montrent que l'erreur maximale ne dépasse jamais 0,045 radian ($2,58^\circ$) entre la réalité telle qu'elle est mesurée par les caméras Vicon et la mesure effectuée par nos capteurs. Par ailleurs, le côté

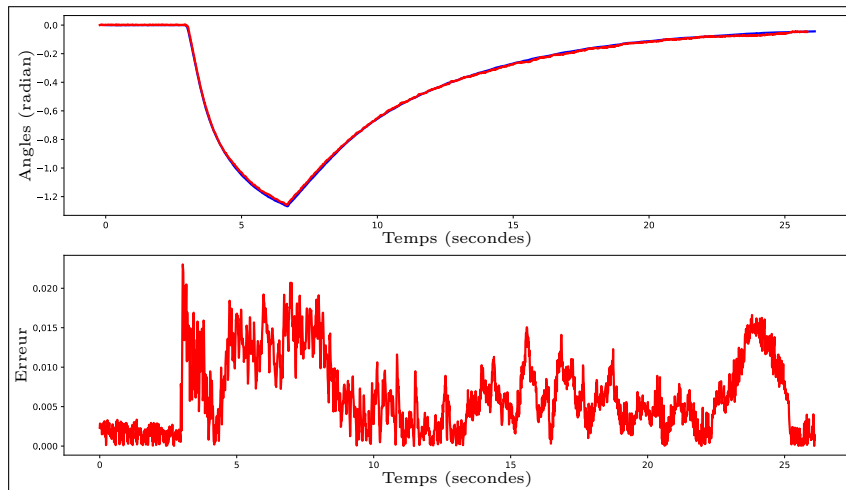


FIGURE 2.14. — Mêmes courbes que dans la figure, 2.13 mais pour le coude droit.

cyclique de la plus forte composante de l'erreur semble dû à un facteur mal calibré lors de la projection des données du Vicon dans l'espace angulaire des capteurs. Si l'on ne considère que l'erreur moyenne, celle-ci tombe à 0,0060 radian (soit $0,34^\circ$). On n'observe aucun décrochage dynamique particulier qui serait dû à la vitesse du mouvement.

Cette précision était plus grande que celle à laquelle nous nous attendions et constitue un bon socle car elle nous permet d'affirmer que les données recueillies grâce aux capteurs sont finalement très proches de ce que l'on pourrait obtenir en utilisant des reconstructions plus complètes à base de capture 3D.

2.3.3. Relation commande/mouvement

Savoir ce que l'on contrôle est un point crucial. En pratique, ce qui est contrôlé est finalement un ensemble assez complexe. En effet, nous manipulons principalement un champ électrique qui, à travers plusieurs systèmes intermédiaires physiquement autorégulés (voir section 2.2.1), pilote indirectement, non la pression, mais le débit circulant dans un vérin. Ce débit lui-même se traduisant par une modification de l'état du vérin dans un sens ou dans l'autre.

Même si en théorie il est possible de caractériser parfaitement ce système, il est plus difficile de savoir quel type de grandeur relie réellement le signal d'entrée à l'état du bras commandé après la prise en compte de tous les phénomènes présents (pertes de charge, frottement fluide à différents niveaux, frottements secs, etc.). Nous avons donc caractérisé les liens entre la réalité observée et notre signal de commande.

Résultats et interprétation

Nous présentons ici certains des résultats (pour l'épaule pour un mouvement selon un axe haut-bas). Les résultats sont illustrés sur les figures 2.15, 2.16 et 2.17.

Nous pouvons voir sur les courbes, figures 2.15, 2.16 et 2.17, une certaine correspondance entre les variations des courbes de vitesse et la forme du signal de commande. Pour préciser ce constat, on soumet l'axe à un créneau de commande tel que présenté sur la figure 2.18

On remarque alors que les axes, lorsqu'ils sont soumis à un créneau de commande de faible intensité, progressent avec une vitesse constante relative à l'intensité du créneau en question. On peut en conclure que notre commande permet de contrôler un analogue de la vitesse désirée. Ceci reste néanmoins à nuancer, en effet, comme l'ont montré les courbes précédentes cette relation semble assez approximative

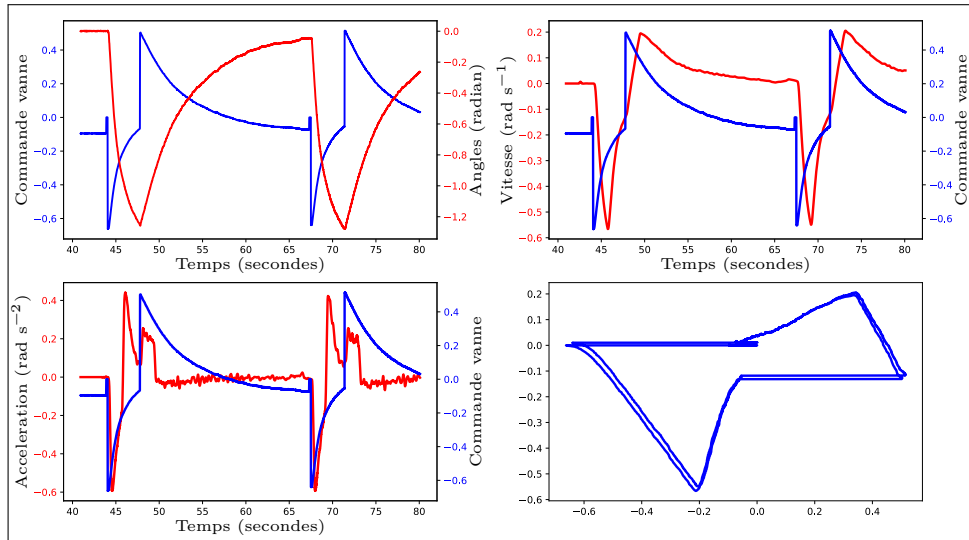


FIGURE 2.15. — Courbes illustrant le lien entre commande et résultat sur le bras de Tino. On met en parallèle la commande (courbes bleues) avec différentes dérivées angulaires de l'axe considéré afin d'observer des équivalences (courbes rouges). La courbe en bas à droite trace la commande par rapport à la vitesse sur plusieurs périodes.

lorsqu'il y a de forts changements de forme du signal. Le signal de vitesse dans ces cas accuse également des retards qui ne sont pas seulement le fruit d'un problème de facteurs multiplicatifs. Par ailleurs, on sait également que les « 0 » de commande ne correspondent pas forcément aux « 0 » réels des vitesses. Comme nous le verrons ensuite sur la figure 2.19 l'approximation linéaire n'est juste que localement ou pour des commandes faibles.

Néanmoins, nous nous servons dans la suite des travaux de cette thèse de l'approximation selon laquelle le signal de commande est approximativement lié à la vitesse angulaire dans certaines bornes :

$$\dot{\theta} \approx A * Commande + B \quad (2.1)$$

Pour finir, les réponses types pour plusieurs créneaux de commandes d'intensités différentes illustrées figure 2.19 peuvent être tracées. Une relation entre les vitesses montantes et descendantes et l'intensité des créneaux se dessine, mais, pour cet axe (l'épaule) et ce mouvement (mouvement haut-bas du bras), les formes tendent à s'aplatir et à ne plus être linéaires. Cela s'explique probablement par l'augmentation du moment angulaire de la gravité qui s'applique plus à une position horizontale qu'à la verticale.

De même, l'impact important de la gravité sur la dynamique du mouvement est visible avec des chutes du bras plus rapides que les montées.

2.3.4. Capacité de mouvement du robot Tino

Les capacités de mouvements maximales du robot Tino sont assez difficiles à estimer. En effet, les vérins servant à générer le mouvement sont surdimensionnés à la conception, ainsi que le compresseur servant à maintenir la pression de l'huile. Dans notre cas, celle-ci pouvait aller jusqu'à 90 bars, mais les vérins peuvent être alimentés avec des pressions supérieures. La conséquence de ceci est que les capacités du robot, en plus d'être variables avec la pression appliquée, sont également bien souvent abaissées artificiellement pour maintenir le robot dans une certaine zone de dynamique où il n'est pas dangereux pour l'expérimentateur, son matériel de capture et les personnes présentes dans la pièce. Par ailleurs, ses capacités sont également très variables selon les axes. L'inventaire exhaustif des capacités du

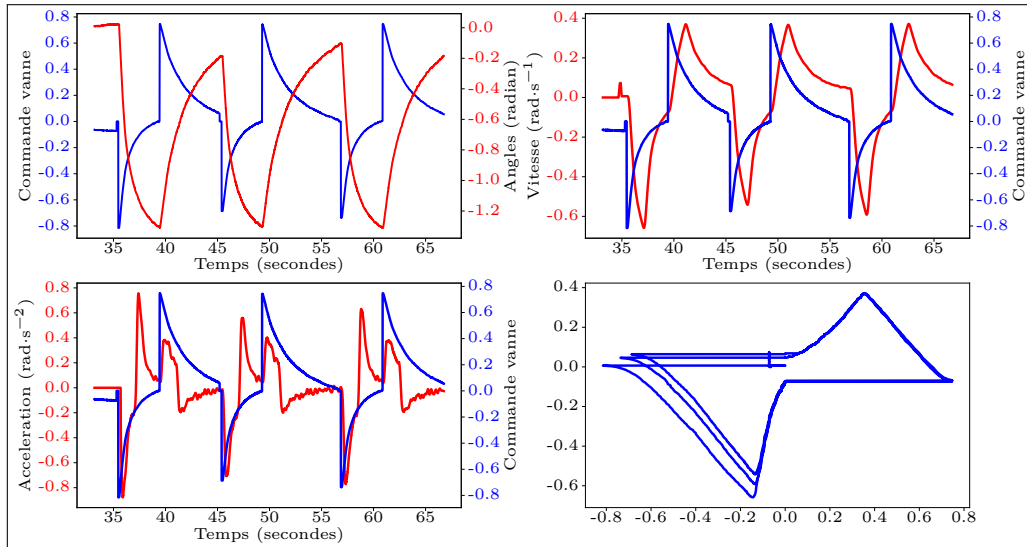


FIGURE 2.16. — Même protocole et représentation que pour la figure 2.15 mais sur des fréquences plus rapides

robot dans tous les cas de figure possibles n'a pas été détaillé ici, mais on peut en faire des estimations sur certains axes.

Ainsi, les capacités maximales du robot sur son axe principal haut-bas rotatif, dans des conditions sécurisées, sont exprimées par l'expérience illustrée figure 2.11 dans laquelle la tête suivait un bras qui effectuait des mouvements allant de 10° à -12° approximativement trois fois en une seconde. On peut estimer alors une vitesse rotative de $66^\circ \cdot s^{-1}$ soit une vitesse en bout de bras approximative de $1,15 \text{ m} \cdot s^{-1}$. L'accélération maximum dans ce cas est également estimable via ces résultats à $23 \text{ m} \cdot s^{-2}$.

Comparaison des résultats expérimentaux avec le domaine dynamique du mouvement humain

Comparer ces données avec le domaine de travail humain ne peut se faire qu'approximativement. Les vitesses de l'être humain sont, pour le bras, d'environ $1 \text{ m} \cdot s^{-1}$ en situation de travail et de $25 \text{ m} \cdot s^{-1}$ en situation sportive. Ces résultats sont à nuancer car les vitesses en bout de bras de l'être humain sont souvent un mouvement combiné du bras, de l'avant-bras et du poignet ce qui peut donner des vitesses plus élevées en bout de bras dans les situations sportives. Il est à noter que le robot Tino n'est pas loin des caractéristiques humaines sur son axe principal en situation de travail. Les accélérations humaines, pour un mouvement combiné bras/avant-bras, sont de $7,5 \text{ m} \cdot s^{-2}$ environ. A priori les accélérations maximales atteignables par le robot sont donc supérieures aux accélérations classiques du bras humain.

Pour compléter ces estimations, il faut rappeler que ces paramètres ne sont pas forcément les seuls points permettant de partager un domaine de travail. Les profils de vitesses, la forme des trajectoires et la possibilité ou non de partager un espace commun constituent des points importants. Nous n'avons pas, dans ces travaux, estimé chacun de ces points. Néanmoins, en partenariat avec le laboratoire SCALab, nous avons eu l'occasion de comparer pour une même consigne (mouvements verticaux et horizontaux) des trajectoires humaines par rapport aux trajectoires du robot Tino obtenues par des créneaux d'impulsion plus ou moins forts. Des résultats condensés sont présentés sur la figure 2.20.

Les résultats montrent que hormis une asymétrie due à la gravité (que l'on peut observer plus précisément sur la courbe 2.19) plus prononcée pour le robot, les formes et intensités de mouvement sont globalement les mêmes. Ces résultats nous permettent d'être assez confiants dans la capacité du robot Tino à évoluer dans un même espace dynamique qu'un être humain, à partager des tâches ou mener des actions conjointes.

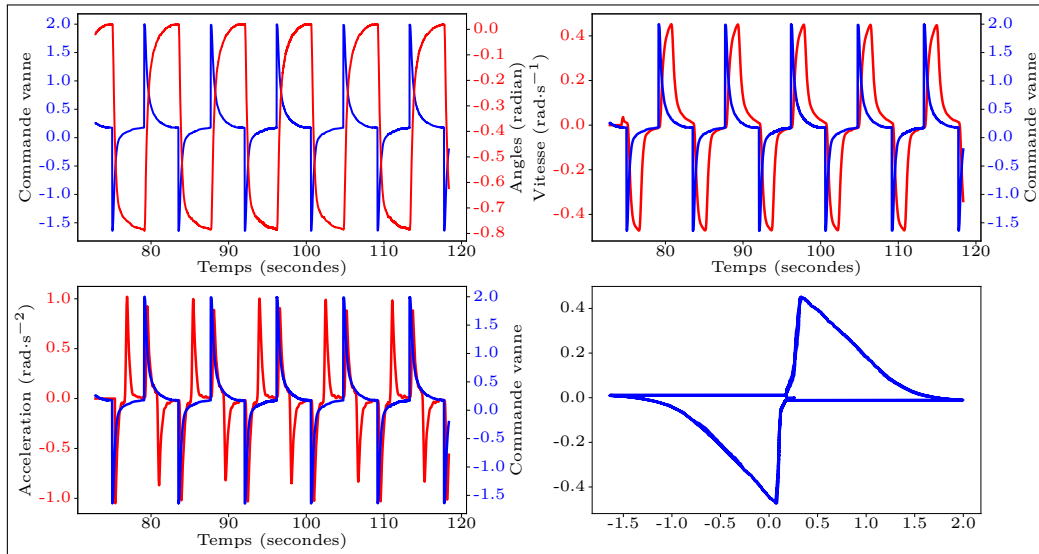


FIGURE 2.17. — Même protocole et représentation que pour la figure 2.15 et 2.16 mais sur des fréquences encore plus rapides

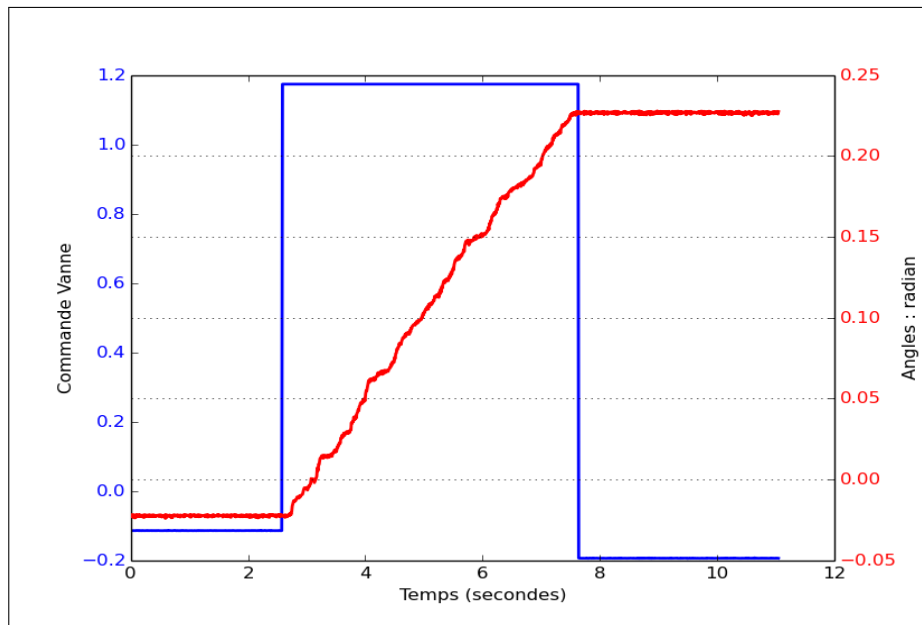


FIGURE 2.18. — Courbe de réponse angulaire d'un axe face à une commande en créneau. (La commande à atteindre est volontairement de faible intensité pour bien observer la progression.)

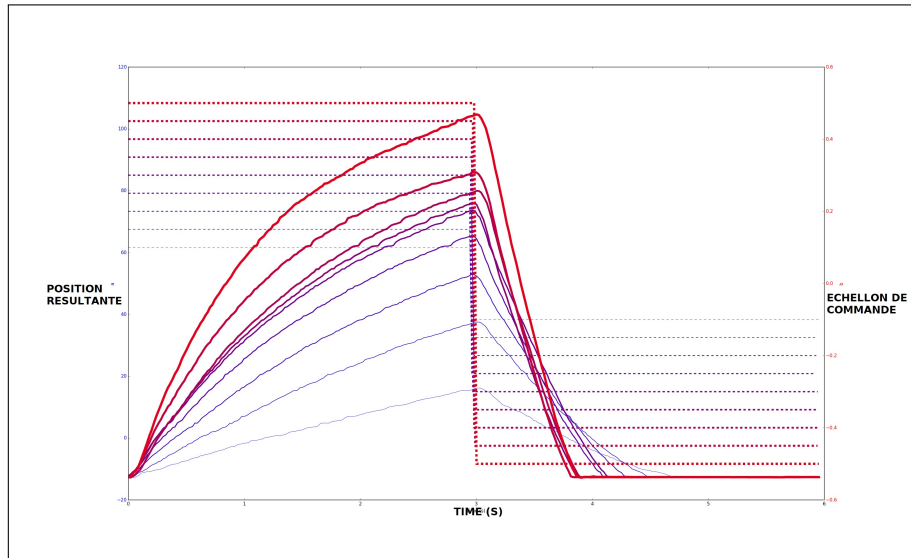


FIGURE 2.19. — Courbes de réponse dynamique angulaire à des créneaux d'intensités différentes.

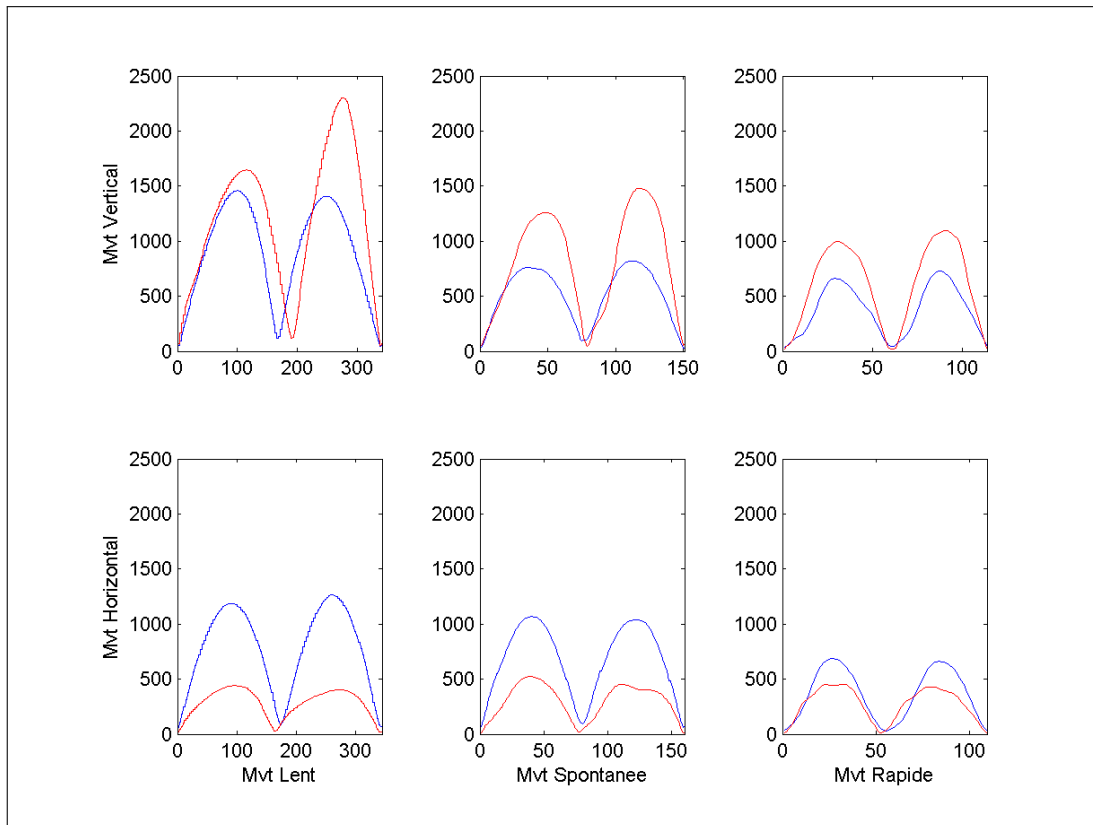


FIGURE 2.20. — Courbes de comparaison entre des trajectoires de sujets humains (en rouge) et des trajectoires effectuées par le robot Tino (en bleu). Le terme rapide correspond à un tempo de 300 ms et lent de 900 ms. Le protocole détaillé est disponible à l'annexe A. Courbe fournie issue des analyses de l'équipe du SCALab.

2.4 | Phénomènes émergents de la mécanique et du contrôle

Certains phénomènes utiles peuvent émerger des caractéristiques du robot vu ci-dessus. L'un des plus intéressants est le fait qu'à une commande spécifique, le bras du robot est soumis sur l'axe vertical à des forces contradictoires dépendantes de la commande, des frottements, mais également de la gravité (moment du poids du bras), des poids portés par le robot, etc. La courbe vue précédemment sur la figure 2.19 nous conforte dans l'hypothèse que ces forces, et notamment le moment angulaire dû au poids du bras, ont une action importante sur le résultat.

On fait alors la supposition qu'à une commande donnée sur un axe vertical, il est possible de profiter des effets de frottements et de moments angulaires pour que les bras s'arrêtent à une position donnée. Si c'est le cas, et si le phénomène varie en fonction du poids en bout de bras du robot, alors il peut être possible d'exploiter le duo commande/effet observé pour que le robot puisse déduire les poids portés durant la manipulation.

2.4.1. Validation du phénomène à vide

Pour valider le phénomène, on applique un protocole simple : on envoie une commande positive quelconque suffisamment faible sur le bras à vide partant de sa position la plus basse et on observe le résultat. Le bras est censé monter et on fait l'hypothèse que les frottements sont suffisants pour le stabiliser à une position angulaire dépendante de la commande.

Le résultat observé est que le bras s'arrête effectivement à une position fixe dépendante de la commande envoyée. Les courbes superposées des commandes envoyées et des courbes résultantes illustrant ce résultat sont visibles sur la figure 2.21.

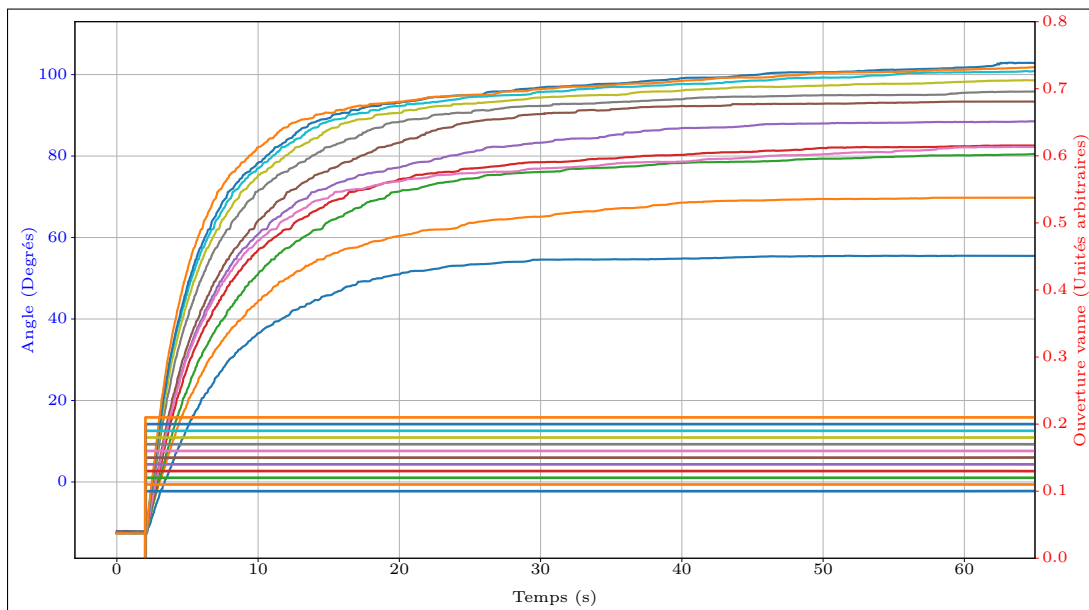


FIGURE 2.21. — Courbes superposées des commandes envoyées (créniaux de différentes tailles) et des courbes résultantes angulaires. On observe que chaque commande aboutit à une stabilisation sur une position d'équilibre différente. Une courbe de commande d'une certaine couleur est liée à une courbe de réponse de la même couleur.

Cette figure illustre bien le fait qu'il existe un type de relation entre la commande envoyée et la

position d'équilibre observée.

La relation entre ces commandes et la position d'équilibre finale peut être tracée. On obtient la courbe illustrée sur la figure 2.22. Une relation quasiment linéaire est bien visible sur cette courbe et pourra être utilisée dans le cadre d'algorithmes d'apprentissage.

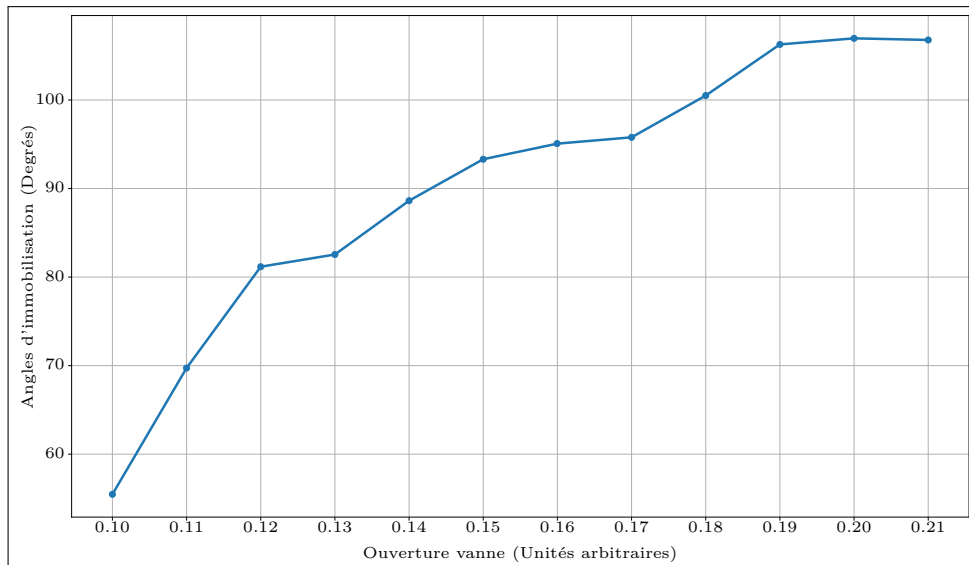


FIGURE 2.22. – Courbe présentant la relation entre commandes et positions finales à vide extraite de la courbe précédente (figure 2.21).

2.4.2. Variation du phénomène selon le poids

Le poids en bout de bras du robot permet normalement, selon nos hypothèses précédentes, de faire en sorte qu'à une commande identique les positions d'arrêts varient en fonction du poids. Un protocole proche du précédent a été préparé. Dans celui-ci le robot se met à soulever des haltères depuis la position la plus basse possible, toujours avec la même commande, et on note le résultat obtenu.

La courbe 2.23 illustre la compilation de ces résultats. Chaque courbe correspond à un poids testé et présente la relation entre position finale d'équilibre et commande. Il s'agit donc du même type de courbe que celle représentée sur la figure, 2.22 mais tracée selon différents poids.

On observe ici encore qu'au vu des formes des courbes, celles-ci forment des caractéristiques suffisamment cohérentes pour ne poser aucun problème à leur exploitation par des algorithmes d'apprentissages. Il devient donc théoriquement possible de profiter de cet effet émergent dans nos architectures de contrôle pour en déduire le poids appliqué en bout de bras et, à l'inverse, pour déduire la bonne commande à appliquer pour lever différents poids et les amener à une même position.

2.4.3. Validation du phénomène dans un cas plus général

La question de la stabilité du phénomène face aux perturbations se pose. En effet, est-il possible de partir de positions autres que le point le plus bas et de profiter, par exemple, de la gravité pour abaisser seule le bras du robot ? Est-il possible de partir de positions intermédiaires et, avec la même commande, d'atteindre le même point qu'en partant d'une position plus basse ? Dit autrement, est-ce que le point d'équilibre terminal est dépendant de la position de départ ou uniquement de la commande ?

Pour tester cette hypothèse, on soumet un bras du robot à un créneau de vanne, puis l'opérateur vient bouger manuellement ce bras de robot. Si on le bouge vers le bas le bras revient bien à la position

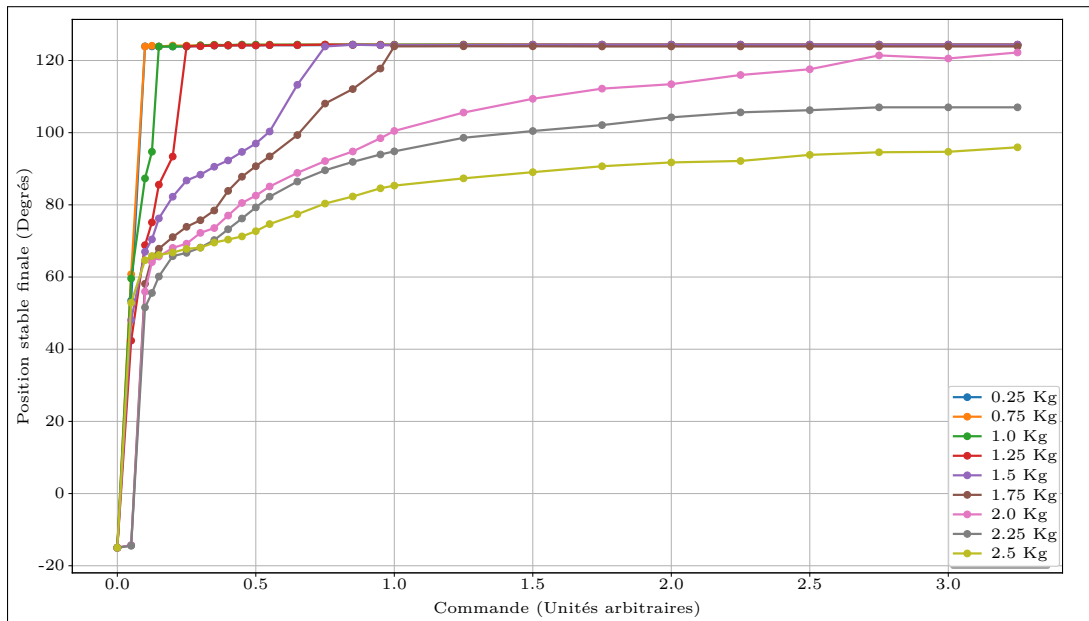


FIGURE 2.23. – Ensemble de courbes présentant la relation entre commandes et positions finales pour différents poids.

d'équilibre. Néanmoins si on le bouge vers le haut celui-ci se stabilise tout de même à une nouvelle position sans redescendre à sa position initiale.

Pour caractériser ce phénomène, on répète l'opération de plus en plus haut pour observer s'il existe une borne haute stable où le bras de robot retombe grâce à la gravité à une nouvelle position d'équilibre. Le résultat de l'expérience est illustré sur la figure 2.24.

Le résultat montre bien que le point d'équilibre n'est pas indépendant de la position de départ. Durant toute l'expérience, la même commande reste appliquée et pourtant le bras reste plutôt stable (grâce aux frottements secs, aux frottements fluides, etc.) sur une large plage de valeur. Finalement à partir d'un certain angle la gravité finit par triompher des diverses forces de frottements et de la force des vérins pour retomber à une même position d'équilibre.

La zone de stabilité pour une même commande et un même poids ne forme donc pas une position unique, mais une large zone entre deux positions dont la borne basse est confondue avec la position d'équilibre en partant du point le plus bas du bras, et la borne haute confondue avec une position d'équilibre partant du point le plus haut du bras.

On peut tracer ces deux bornes pour différents poids afin d'observer la variation de ces bornes plutôt que la variation d'un point fixe. Le résultat est présenté sur la courbe 2.25 et montre bien que les poids ont un impact sur le positionnement de ces bornes.

2.4.4. Conclusion sur les phénomènes émergents du contrôle

Certains des phénomènes que nous pensions voir émerger des caractéristiques du robot se sont vu confirmer et présentent des caractéristiques qui étaient a priori exploitables par des algorithmes classiques d'apprentissage. Néanmoins ces phénomènes présentent également des caractéristiques problématiques pour une exploitation telle quelle, notamment des zones de stabilité larges, empêchant d'exploiter précisément ces phénomènes seuls pour le contrôle d'un bras robotique à vide ou en charge. Nous verrons dans le chapitre suivant une démonstration d'un apprentissage employant ces phénomènes émergents dans un cas restreint.

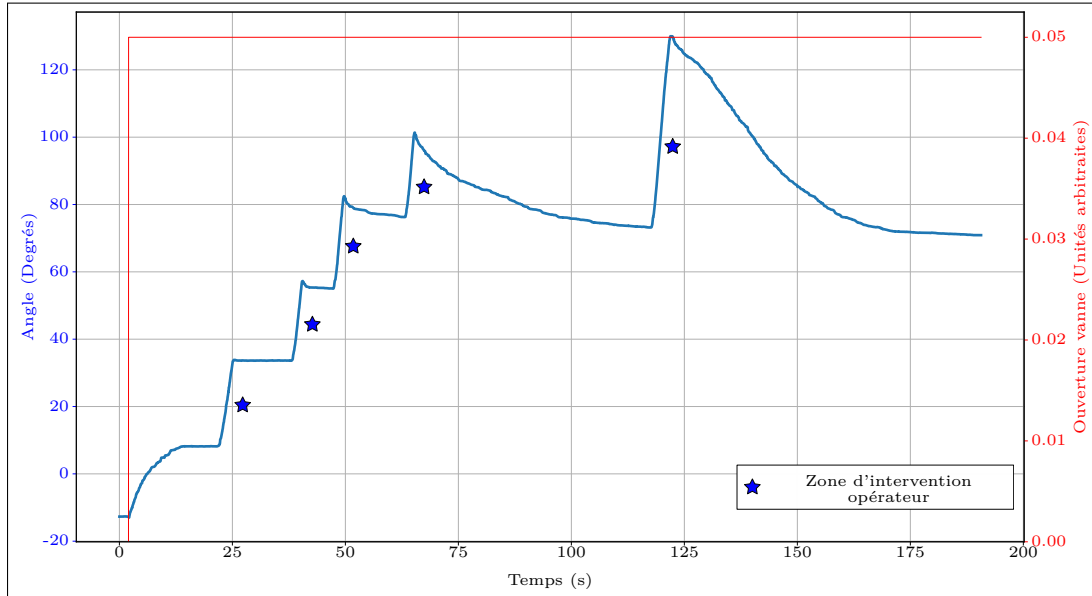


FIGURE 2.24. — *Protocole de levée manuelle du bras du robot Tino. La première montée du bras (courbe bleue) est due aux moteurs du robot réagissant face au créneau. Le bras se stabilise à une position d'équilibre. Par la suite c'est l'opérateur qui vient bouger à la main le bras du robot pour induire un déséquilibre et observer le résultat.*

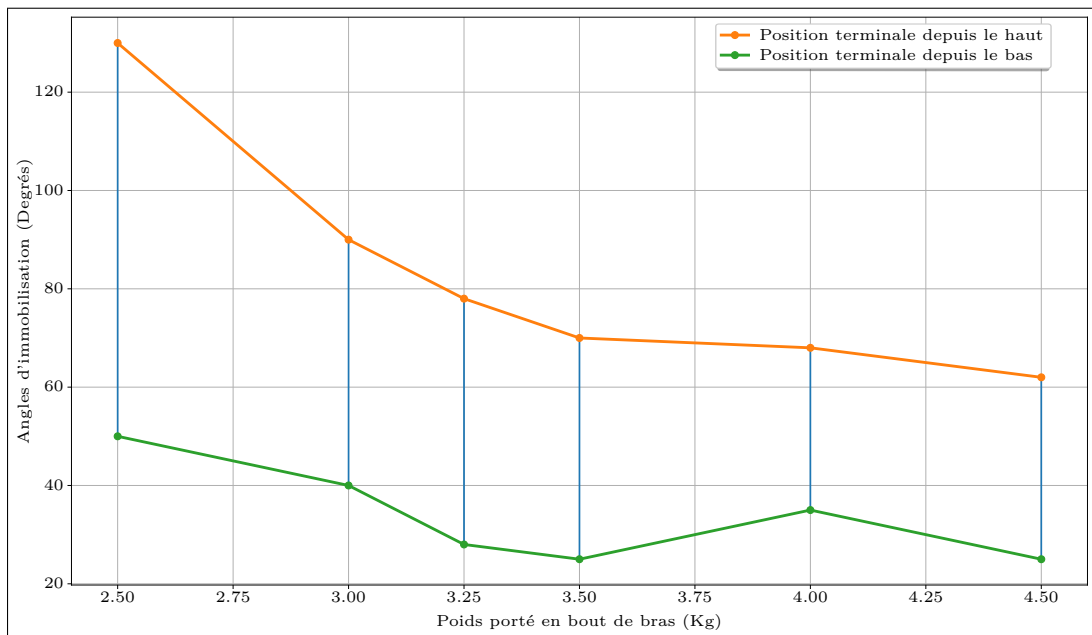


FIGURE 2.25. — *Bornes de stabilité min et max pour différents poids. L'écart important entre min et max indique qu'il pourra être difficile d'exploiter un protocole simple d'association « commande-position atteinte ».*

2.5 | Acceptabilité du robot Tino

En partenariat avec l'équipe SCALab de Lille 3 nous avons mené une expérience conjointe dont l'objectif était de comparer les cinématiques de mouvement du robot Tino, d'un robot électrique et d'humains, puis de jauger de leurs influences sur la qualité ressentie et observée d'une interaction avec un agent humain dans des tâches nécessitant des mouvements conjoints.

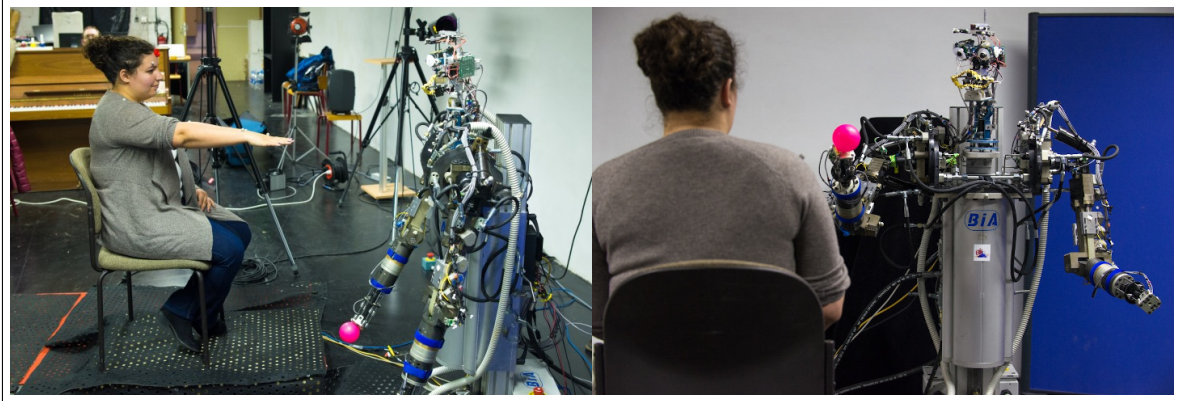


FIGURE 2.26. – *Expériences conjointes Tino-humain au laboratoire SCALab de Lille 3 visant à comparer les dynamiques de mouvement et à juger de l'impact du robot sur l'appréciation de l'humain. Protocole détaillé disponible à l'annexe A.*

Nous présenterons ici une comparaison des mouvements du robot Tino avec ceux de l'humain pour des mouvements primaires ainsi que les résultats des questionnaires remplis par les sujets jugeant de l'acceptabilité du robot.

2.5.1. Présentation résumée du protocole

Le protocole détaillé étant disponible à l'annexe A nous nous contenterons ici de présenter un résumé du protocole. Dans une pièce deux acteurs se font face assis : un participant humain non entraîné et un agent qui pourra être un expérimentateur (humain entraîné) ou un robot. Le protocole prévoit de tester deux cas principaux : l'un où le participant doit suivre l'expérimentateur et l'un, identique, à la seule différence que l'agent tente également de suivre le participant et ce, à différents tempos imposés par un signal sonore. L'objectif était alors de jauger par l'analyse des cinématiques des participants et des questionnaires rendus ensuite (questionnaires de type « Godspeed » spécifiquement conçus pour jauger l'interaction d'un robot avec un participant, voir (BARTNECK et al., 2009) et (WEISS et BARTNECK, 2015)) de l'influence du tempo, du type d'acteur et du fait que le participant suive un agent « aveugle » ou soit plutôt dans une mécanique de mouvement conjointe avec l'agent qui tente également de se synchroniser sur ses mouvements.

2.5.2. Résultats et interprétations

Les résultats de cette expérimentation furent partiels. Tout d'abord pour des raisons de temps manquant dans l'expérience : nous avons en effet sous-estimé le temps nécessaire au passage des sujets. Mais également parce que certaines branches du protocole étaient incomplètes à la base : il n'y avait par exemple aucun test avec des sujets « naïfs », la plupart d'entre eux étant au courant des tenants et aboutissants de l'expérience. Autre défaut de l'expérience, la nécessité de contrôler finement le robot

dans la partie où celui-ci doit suivre l'agent amène automatiquement l'usage d'algorithmes intermédiaires (de type PID ici) qui viennent entacher la fiabilité des conclusions quant aux capacités brutes du robot.

Les résultats des données cinématiques de mouvement sont déjà présentés section 2.3.4. Ici, nous présenterons donc les résultats extraits des questionnaires. Les résultats sont détaillés dans le tableau 2.1.

God-Speed questionnaire results										
	Lateralité	Anthropom*	Animation	Appréciation	Intelligence	Suret�	Interaction	Total	Technocity	Technophily
Sujet1	0,9	1,8	2,8	3,0	3,2	2,3	2,3	2,6	31,0	8,0
Sujet2	0,9	3,8	3,0	3,0	2,8	2,7	2,3	2,9	39,0	10,0
Sujet3	1,0	4,0	3,8	4,8	3,6	3,3	2,2	3,6	42,0	14,0
Sujet4	0,9	3,5	3,6	3,6	3,6	2,3	2,5	3,2	26,0	6,0
Sujet5	1,0	3,8	3,6	4,0	4,0	3,0	2,5	3,5	27,0	8,0
Sujet6	1,0	1,0	3,4	3,6	3,4	2,7	2,0	2,7	27,0	5,0
Sujet7	1,0	2,5	3,4	4,0	3,4	3,0	1,5	2,9	23,0	5,0
Sujet8	1,0	2,0	2,0	3,2	2,2	2,3	3,2	2,5	28,0	5,0
Sujet9	0,9	2,0	3,8	4,2	3,2	2,7	2,2	3,0	32,0	8,0
Sujet10	1,0	3,0	4,2	4,4	3,0	3,3	2,2	3,3	40,0	13,0
Sujet11	1,0	1,3	2,8	3,2	3,0	3,0	3,3	2,8	21,0	3,0
Sujet12	0,9	2,8	2,8	3,8	2,4	3,3	2,2	2,8	33,0	8,0
Sujet13	1,0	3,3	2,8	4,0	2,2	3,0	2,5	2,9	35,0	11,0
Sujet14	1,0	1,8	2,0	2,6	1,6	3,0	3,3	2,4	32,0	10,0
Sujet15	1,0	2,0	2,6	2,8	1,8	3,0	2,8	2,5	30,0	6,0
Sujet16	0,9	5,0	5,0	5,0	5,0	3,7	2,3	4,3	40,0	15,0
Moyenne	1,0	2,7	3,2	3,7	3,0	2,9	2,5	3,0	31,6	8,4

Tableau 2.1. – *Tableau r sumant les r sultats des questionnaires Godspeed pos s aux participants de l'exp rimentation conjointe au laboratoire SCALab, hormis la lat ralit , les 5 premi res colonnes sont not es sur 5, les colonnes suivantes sont des indices compos s*

De fa on g n rale, il est difficile d'interpr ter ces r sultats seuls sans les comparer   d'autres robots, ce qui est l'un des points probl matiques de cette  tude. N anmoins, on peut noter que le score d'appr ciation, qui  tait l'un des plus importants   observer, est particuli rement haut (3,7/5) en moyenne, ce qui montre qu'a priori le robot Tino ne d clenche pas de rejet particulier vis- -vis des sujets humains.

2.6 | Conclusion

Dans ce chapitre, nous avons étudié l'outil principal de cette thèse et expliqué pourquoi dans notre cadre les caractéristiques de cet outil robotique sont importantes. Nous avons également exposé les caractéristiques théoriques et les travaux de caractérisation technique qui ont eu lieu dans le cadre de cette thèse et qui permettent d'estimer ses caractéristiques et leurs potentiels impacts sur le résultat observé lors du développement d'IA sur ce robot.

Nous avons exposé les quelques problématiques théoriques qui émergent du développement d'IA sur un robot hydraulique humanoïde et montré quelques résultats partiels qui nous permettent d'affirmer que ce robot est bien capable techniquement d'interagir avec un humain (comparaison des domaines de travail et trajectoires, acceptabilité du robot qui semble prédire un faible risque de tomber dans la vallée de l'étrange). Nous garderons néanmoins à l'esprit pour la suite de cette thèse les problématiques de temps de convergence en environnement réel, qui n'ont pas de réelles solutions et amèneront pour certains algorithmes l'obligation de passer par des simulations.

Pour finir nous avons exposé certains phénomènes émergents de l'architecture robotique, principalement dus à la combinaison du mode de contrôle particulier avec des phénomènes de frottements secs et fluides et nous avons montré en quoi, et avec quelles réserves, ces phénomènes pouvaient être exploités pour le contrôle du robot.

Nous verrons dans la partie suivante comment tirer parti des caractéristiques de ce robot, de sa capacité à interagir dans des gammes dynamiques « humaines » et, enfin, de ces phénomènes émergents, ceci afin de concevoir une IA de contrôle bio-inspirée et développementale adaptée aux spécificités de cette plateforme robotique.

II

Deuxième partie :

**Architecture bio-inspirée pour le
contrôle d'un robot**

État de l'art sur le contrôle sensorimoteur et positionnement

Sommaire du chapitre :

3.1. Introduction	48
3.2. Types de modèles existants	49
3.2.1. Contrôle optimal	49
3.2.2. Apprentissage et adaptabilité : existence de nombreux modèles	52
3.3. Propriétés du système nerveux central et contrôle moteur chez l'humain	57
3.3.1. Cortex visuel	57
3.3.2. Thalamus	58
3.3.3. Cortex associatif pariétal	58
3.3.4. Cortex moteur et prémoteur	58
3.3.5. Ganglions de la base	58
3.3.6. Hypothalamus	59
3.3.7. Hippocampe	59
3.3.8. Cervelet	60
3.3.9. Développement du contrôle moteur chez l'enfant	63
3.3.10. Positionnement et choix des boucles sensorimotrices à modéliser	64
3.4. Conclusion	66

3.1 | Introduction

Nous cherchons à réaliser un modèle pour le contrôle bio-inspiré d'un robot dans le cas d'une interaction dans un environnement réel ou avec un partenaire, tout en restant positionné dans un cadre développemental et constructiviste. Nous souhaitons également permettre au modèle l'exploitation des phénomènes émergents identifiés sur le robot Tino dans le chapitre 2. L'idéal serait que l'exploitation de ces propriétés soit faite de façon implicite dans l'apprentissage, sans nécessiter de composante supplémentaire. Avant d'aborder le type de modèle que nous pensons idéal dans notre cadre, nous allons évaluer dans ce chapitre les raisons qui nous poussent à délaissier certaines approches plus classiques de contrôle moteur en robotique.

Le chapitre commencera ainsi par une revue des algorithmes de contrôle robotique classiques, notamment du contrôle optimal, puis nous introduirons succinctement les algorithmes d'apprentissages exploités en robotique. Pour finir, nous ferons un état de l'art des connaissances actuelles concernant le fonctionnement de certains noyaux et parties importantes du cerveau humain, notamment celles des plus impliquées dans le cadre du contrôle moteur.

3.2 | Types de modèles existants

De très nombreux modèles existent pour contrôler un robot et l'amener à réaliser des mouvements. On peut citer l'exemple de modèles très classiques de l'automatisme qui consistent à présenter simplement des suites de points à une architecture bas niveau de type contrôleur proportionnel afin d'exécuter une séquence d'actions. Ces modèles ont le désavantage d'être complètement déconnectés de l'environnement dans lequel ils évoluent, forçant l'utilisateur à mettre en place un environnement parfaitement délimité et prédictible.

Même dans un tel environnement ils ne réalisent pas non plus forcément un travail parfait à tout point de vue. En effet, la nécessité de mettre en place un environnement parfaitement contrôlé demande du temps et des ressources et la moindre erreur apparaissant dans l'environnement peut vite déclencher un effet domino. Dans une chaîne de montage entièrement automatisé par exemple, une erreur implique souvent l'arrêt de l'ensemble de la chaîne par sécurité.

Plus important ces algorithmes, lorsqu'ils sont déployés dans un environnement où des humains sont présents, ont tendance à augmenter la charge de travail des opérateurs. Par exemple face à des machines incapables de s'adapter au rythme de travail ou de prédire les actions des humains et de s'y adapter en conséquence, il devient alors nécessaire d'effectuer un travail de surveillance, un travail d'adaptation au rythme de la machine (qui ne correspond pas forcément à celui de l'être humain) et pour finir un travail de formation et d'expertise nécessaire à l'interaction avec ces machines.

Finalement, la machine automatique dans ce cadre ne fait pas parfaitement ce pour quoi elle fut conçue : décharger les humains de certaines tâches. C'est l'humain qui au contraire doit se former à ce nouvel outil « non intelligent » et s'adapter à lui. Pour avoir une discussion plus complète sur cet aspect on pourra se référer au très classique « *Ironies of Automation* » (BAINBRIDGE, 1983).

Dans cette thèse, l'un des objectifs étant de proposer des modèles permettant d'interagir de façon intuitive avec l'être humain dans un environnement réel il est alors évident que les défauts inhérents à ces types d'approches les rendent peu intéressantes pour nos applications.

Comme nous l'avons vu dans l'introduction, l'un des objectifs de cette thèse est de proposer une alternative à ce problème avec des architectures souples, écologiques et adaptatives, capables d'interagir dans un environnement réel où évoluent d'autres agents et humains et à terme de partager certaines tâches. De fait, nous préférons nous orienter vers d'autres types d'approches.

3.2.1. Contrôle optimal

Le contrôle optimal est une théorie dérivée du milieu de l'automatisme et encore très utilisée aujourd'hui. Contrairement à l'automatisme classique certaines applications du contrôle optimal tentent de reproduire certaines caractéristiques du mouvement humain ou d'apporter des interprétations au fonctionnement du système contrôle moteur chez l'humain. (IVALDI et al., 2012).

Le contrôle optimal repose sur deux principes mathématiques : le principe d'optimalité de Bellman (BELLMAN, 1957) et le principe du *maximum de Pontryagin* (PONTRYAGIN et al., 1962). Ces principes aident à la minimisation d'un critère mathématique, qui est à définir. Cette minimisation donnera alors une trajectoire de contrôle sous contrainte. Todorov décrit dans « *Optimal Control Theory* » comment lier ces deux méthodes pour divers critères (TODOROV, 2006).

Selon le critère on peut alors tenter d'obtenir une trajectoire optimale pour de nombreux cas particuliers : pour éviter de trop trembler par exemple ou pour minimiser les variations brusques dans les virages, ou encore pour optimiser la dépense d'énergie, mais celles-ci doivent être parfaitement modélisable en avance.

Par exemple, certains des critères les plus célèbres visent à minimiser des paramètres de la cinématique :

- La loi du « Jerk minimal » (HOGAN, 1984a ; FLASH et HOGAN, 1985) vise à minimiser la variation de l'accélération sur un mouvement afin de reproduire des trajectoires humaines.
- La loi du « Torque minimal » vise à minimiser la dérivée du couple (UNO et al., 1989)(NAKANO et al., 1999).

- La loi de la « variance minimale » (HARRIS et WOLPERT, 1998) ou de l'intervention minimale (TODOROV et JORDAN, 2003) vise à minimiser l'erreur de position en bout de bras. Les erreurs de trajectoire n'aboutissant pas à une erreur de position en bout de bras ne seront ainsi pas corrigées, d'où le concept d'intervention minimale.

D'autres vont viser à réduire l'énergie comme le ferait un humain pour optimiser sa dépense d'énergie musculaire (HATZE et BUYS, 1977). Certains vont aller plus loin dans la prise en compte de la biologie dans l'objectif d'affiner cette minimisation d'énergie et prendront même en compte les aspects de co-contraction et d'énergie de co-contraction dépensée (BERRET et al., 2008 ; GAUTHIER et al., 2010).

La plupart de ces modèles donnent de bons résultats quant à la reproduction de trajectoires humaines sur des sections de trajectoire dans chacun des cas particuliers étudiés, néanmoins le concept même sur lequel ils sont basés porte en lui de gros défauts qui empêchent de l'employer dans des cas plus généraux (environnement complexe) ou dans des cas interactifs.

Critique du contrôle optimal

Les modèles de contrôle optimal ont plusieurs défauts. Tout d'abord, ils se basent sur une approche qui consiste à imaginer, dans beaucoup de cas, un objectif au système et à l'intégrer dans le contrôle (minimiser l'énergie, le jerk, le torque, etc.), bien souvent ces critères dépendent de l'interprétation du concepteur et donc finalement cela reste une interprétation « humaine » du fonctionnement du système. Comme nous l'avons illustré dans le Chapitre 1 sur l'exemple d'Asimo et des marcheurs passifs, de telles approches, faites par l'observation d'un système depuis l'extérieur en vue de reproduire ses caractéristiques (la marche dans l'exemple), peuvent aboutir à des erreurs d'appréciations sur le fonctionnement du système dont on s'inspire et peuvent donc finalement donner de mauvais résultats.

Un second point négatif est que tous les critères ne sont pas minimisables aisément. Les équations de base permettant de construire ces modèles sont en effet assez complexes et certains critères sont difficiles à utiliser. De façon générale, ces équations demandent énormément de prérequis, notamment une modélisation fine de l'outil robotique, ou encore la connaissance des types de perturbations rentrant en jeu. Appliqué à un robot, cet ensemble de contraintes impose donc la nécessité d'une connaissance préalable de l'environnement et du robot, ce qui est problématique dans bien des applications et ce qui limite le procédé à un environnement contraint.

Un autre aspect problématique de ce type de contrôleur est la difficulté à prendre en compte des perturbations non prédictibles. En effet, la plupart des modèles de contrôleurs optimaux se contentent de fournir des trajectoires en simulation et de les réappliquer ensuite sur l'outil robotique, ce sont des contrôleurs que l'on dit en boucle ouverte : aucune information venant de la réalité ne vient modifier dynamiquement la réaction du robot. On retrouve en cela l'héritage de l'approche de l'automatique classique.

Néanmoins, certains modèles de contrôleurs optimaux (TODOROV et JORDAN, 2002 ; WADA et KAWATO, 2004) viennent commander un système tiers qui lui est en boucle fermée (contrôleur en impédance de type masse-ressort amorti par exemple (HOGAN, 1984b)) afin de permettre d'ajouter une adaptation en temps réels. Toutefois, ces systèmes s'appuient sur des équations encore plus complexes à calculer a priori et nécessitant donc d'avantages de prérequis (correspondances et lois sensorimotrices du système doivent notamment être connues a priori).

Cet ensemble de contraintes rend le contrôle optimal difficile à appliquer dans notre cadre. Nous tentons en effet de contrôler un robot dans des environnements diversifiés dans lesquels l'humain vient modifier l'environnement régulièrement et peut également interagir avec le robot. La nature non adaptable de ces modèles les rend alors selon nous peu pertinents pour ce type d'application. De même, quand l'objectif est plutôt d'obtenir des modèles qui s'adaptent aux caractéristiques de l'outil robotique sans avoir a priori de connaissances étendues quant à cet outil, alors les approches de ce type deviennent désavantageuses, car elles nécessitent trop de connaissances préalables sur l'outil contrôlé.

Le contrôle optimal et le concept de critères à minimiser dans le cas de l'interaction sociale

Dans notre cadre, nous partons de l'hypothèse que les mouvements chez l'humain présentent des caractéristiques intéressantes qui aident à l'interaction et à la communication avec d'autres êtres humains. Il peut alors être intéressant d'obtenir un algorithme capable de reproduire certaines de ces caractéristiques.

Les méthodes dépendant de critères ou fonction de coût définissant une trajectoire ainsi que d'autres méthodes aboutissant à des trajectoires aux caractéristiques prédéfinies a priori (automatiques, lois prédéfinies de mouvement, etc.) ont du mal à reproduire certaines de ces caractéristiques.

En effet, les trajectoires réalisées par les humains dans de nombreuses tâches ne sont pas uniquement définies par des critères énergétiques, mais sont également très variables, non seulement en fonction du contexte environnemental de la tâche réalisée (porter des poids, risques, etc.), mais également en fonction des partenaires d'interactions lors des interactions sociales.

Les mouvements humains à un certain instant sont par exemple fortement dépendants des tâches qui seront à réaliser *après* la tâche en cours ou des variations qui auront lieu plus tard dans la tâche en cours. Ils sont donc très variables à l'ordre de réalisation des mouvements dans une tâche ou à l'ordre de succession de plusieurs tâches. (ANSUINI et al., 2006 ; ROSENBAUM et JORGENSEN, 1992 ; SHORT et CAURAUGH, 1997). Ceci montre que de façon générale si l'on devait établir un critère de mouvement, en oubliant même la problématique de l'interaction sociale et en se concentrant sur un ensemble de tâches variables dans un environnement simple, alors ce critère, pour être fonctionnelle devrait être extrêmement complexe, ce qui en ferait un critère difficile, voire impossible, à résoudre et implémenter dans le cadre du contrôle optimal.

Au-delà de cet aspect, le problème devient plus marquant quand on analyse le contexte d'une interaction sociale. Il a en effet été démontré que les mouvements humains sont fortement dépendants, dans leurs trajectoires et dynamiques, du contexte social de la tâche : est-on en interaction avec quelqu'un ou non ? Quel est le comportement de cette personne ? Sont autant de questions dont les réponses vont fortement influencer la trajectoire du mouvement chez l'humain.

L'expérience de pensée la plus célèbre qui illustre cette problématique fut posée par Pierre Jacob et Marc Jeannerod (JACOB et JEANNEROD, 2005) en imaginant un protocole de type « Dr Jeckyl et Mr Hyde » dans lequel un patient se fait opérer par un médecin dont les intentions sont douteuses. La question est de savoir à la simple vue du médecin portant son scalpel et s'apprêtant à opérer le patient si le médecin va soigner le patient ou plutôt le tuer. La question posée est celle de l'intention du médecin : est-on capable de percevoir par l'observation cette intention juste en prenant en compte la dynamique du mouvement du médecin manipulant son scalpel ?

Les auteurs répondent dans cette analyse que si l'intention motrice (ici : faire un mouvement avec le scalpel) est la même alors l'intention du praticien sera imperceptible dans ces mouvements. Interdisant donc de connaître a priori les intentions du praticien par une simple observation.

Néanmoins, malgré la position de ces auteurs sur le sujet, les expérimentations pratiques visant à éclaircir cette question ont plutôt démontré l'inverse. Ainsi, l'expérience d'Ioanna Georgiou et collègues (GEORGIOU et al., 2007) puis celle de Cristina Becchio et collègues (BECCHIO et al., 2008) ont démontré qu'une tâche ayant le même objectif final aboutit à des résultats différents selon qu'un humain est impliqué ou non dans celle-ci (poser un objet dans la main d'un humain ou sur la table au même endroit ne donne pas les mêmes trajectoires) et selon le contexte social de l'interaction (compétition ou coopération). L'expérience a depuis été reproduite avec les mêmes résultats (LEWKOWICZ et al., 2013).

Pour finir, le principe a également été reproduit dans le cadre de l'interaction avec des robots, montrant que si le contexte « social » établi par le partenaire robotique est différent (semble-t-il attentif ou est-il plutôt passif?), alors les dynamiques des mouvements observés chez l'humain, pour une même tâche sont bel et bien différentes (NAGAI et al., 2010 ; THOMAZ et CAKMAK, 2009).

Cet ensemble d'expériences achève de contredire la position de Jacob et Jeannerod non seulement dans le cas humain, mais également dans des cas bien plus généraux.

De manière générale, cet ensemble d'expériences démontre donc bien que si on devait résumer les dynamiques du mouvement humain à un critère à optimiser alors ce critère devrait prendre en compte

l'agencement des tâches, la nature de ces tâches, les variations à l'intérieur de celle-ci, la nature du partenaire d'interaction, le but de la tâche en cours, la dynamique du partenaire d'interaction, etc. donc un grand nombre d'informations venant à la fois de l'environnement et à la fois du contexte social.

La nature de ces algorithmes rend l'élaboration de ce critère global peu crédible et la résolution mathématique qui en découlerait difficile, voire impossible.

De notre point de vue ceci rend ce large ensemble d'algorithmes peu crédible pour la réalisation d'interaction motrice avec un partenaire et plus généralement pour l'interaction avec un environnement complexe, si on considère, ce qui est notre hypothèse, que les caractéristiques du mouvement humain présentent un intérêt pour l'interaction entre humains et donc pour l'interaction Homme-machine.

3.2.2. Apprentissage et adaptabilité : existence de nombreux modèles

Les algorithmes hérités de l'automatisme et les lois décrivant le mouvement de l'Homme ont donc de nombreux défauts dans notre cadre. Pour répondre à une partie de ces défauts, toute une gamme de modèles hérités de l'apprentissage automatique a été proposée. Leur point commun est de tenter d'apporter des solutions à la réalisation de mouvements artificiels en utilisant des outils capables d'apprentissage. Ces outils ont l'avantage de résoudre implicitement, au moins partiellement, certains des défauts des modèles que nous avons étudiés précédemment, notamment leur manque d'adaptabilité selon l'environnement, ce qui les rend intéressants dans notre cadre.

Il existe une très grande variété de modèles et en faire un état de l'art complet irait bien au-delà de l'ambition de cette thèse, nous résumerons donc ici succinctement les types de modèles existants et critiquerons certaines de leurs propriétés.

Apprentissage par démonstration

L'apprentissage par démonstration n'est pas un type d'algorithme en soi, mais regroupe tous les types d'algorithmes fonctionnant par démonstration : il s'agit d'un mode d'apprentissage ou un « expert » de la tâche à apprendre vient guider le robot (ou l'agent simulé) sur les trajectoires ou les successions de tâches désirées. L'apprentissage par démonstration peut-être réalisé à base de nombreux algorithmes, mais a en général l'avantage de demander peu de complexité algorithmique. En effet, l'algorithme se contente d'associer un état à l'exécution d'une commande pertinente fournie par l'expérimentateur.

L'apprentissage par démonstration est également assez riche, car il repose sur une base de connaissances héritée de la psychologie du développement de l'enfant décrivant le développement du bébé et notamment l'acquisition de ses compétences motrices. L'exemple de l'imitation est ainsi souvent donné par les auteurs du domaine comme un exemple dans lequel, via un phénomène de résonance moteur, l'imitation, il devient possible de guider un agent pour apprendre des tâches complexes.

Ce type d'apprentissage possède malgré tout quelques défauts si on le considère seul, car il suppose bien souvent que les boucles motrices nécessaires à cette résonance ou au fonctionnement moteur bas niveau de l'agent considéré sont déjà construites, en ne se posant généralement pas la question de cette construction. Dans notre cadre, la question de la construction de ces boucles nous apparaît pourtant primordiale et nous partons de l'hypothèse que ce sont ces boucles qui sont responsables de certains phénomènes émergents du comportement utile à l'interaction. À ce titre, l'apprentissage par démonstration seul nous paraît donc peu approprié pour répondre à notre problématique.

Une autre limitation intrinsèque de l'apprentissage par démonstration tient dans le fait de nécessiter un « démonstrateur », qui devra guider l'algorithme à chaque étape de l'apprentissage, ce qui n'est pas très pratique pragmatiquement pour l'expérimentateur en premier lieu, mais ce qui est aussi problématique si l'on souhaite étudier le développement d'un agent sur un temps long. D'autres stratégies doivent alors être considérées afin de permettre une certaine autonomie dans l'apprentissage.

Algorithme de régression sur les noyaux : RFWR, LWPR, GMM, etc.

Il existe toute une classe d'algorithmes capables, par régression, d'apprendre des lois complexes en venant modifier les noyaux de catégorisations des données. (ATKESON et al., 1997). C'est notamment le cas des algorithmes RFWR (SCHAAL et ATKESON, 1998) puis LWPR (VIJAYAKUMAR et al., 2005). On retrouve également ce type d'algorithme dans de nombreux exemples dérivés des algorithmes EM (Expectation - Maximisation), par exemple les modèles GMM (CALINON et al., 2007).

Ces algorithmes forment un ensemble assez vaste. Ils peuvent être utilisés comme outil pour traiter les données dans un modèle plus vaste, mais ils sont aussi employés par certains auteurs pour apprendre directement des lois motrices, ce qui explique que nous en parlons ici. En général, leur défaut principal pour notre cadre applicatif est selon nous d'adopter un point de vue très inspiré du point de vue du contrôle optimal. Ainsi ils sont souvent employés dans l'optique d'un apprentissage de trajectoires précises pré-fournies ou fournies via démonstration. C'est donc finalement plus les modes d'applications que les algorithmes en eux-mêmes qui s'avèrent problématiques pour nos approches.

D'autres aspects font que nous n'emploierons pas directement ces classes d'algorithmes dans nos travaux. Par exemple, certains de ces algorithmes demandent des étapes supplémentaires de post-traitement (pour déduire les fonctions inverses par exemple), ce qui fait qu'ils ne sont pas parfaitement autonomes.

Nous verrons néanmoins que les algorithmes que nous employons, même s'ils sont différents par l'approche, et parfois par la mise en application, ont tout de même de nombreux points communs avec cette classe d'algorithme.

Apprentissage par renforcement et programmation dynamique : Q-learning et TD-Lambda

Les algorithmes du contrôle optimal ont enfanté d'un nouveau domaine grâce aux travaux de Sutton et Barto dans les années 80 (SUTTON et BARTO, 1988) : l'apprentissage par renforcement.

Le principe est que l'algorithme vient agir sur l'environnement, et en fonction du résultat de l'action, un observateur (qui peut être un expérimentateur ou un autre algorithme) vient punir ou récompenser l'algorithme. La prochaine fois que l'algorithme se retrouvera dans la même situation, il améliorera son choix en tenant compte des punitions passées, c'est le renforcement.

Deux algorithmes majeurs sont à signaler : le TD-Lambda et le Q-Learning.

Le TD-Lambda fut introduit par Richard Sutton en 1988 (SUTTON, 1988), il s'agit d'un modèle fonctionnant dans un espace discret (l'environnement est discrétisé, on est soit dans un état soit dans un autre) : à chaque passage d'une portion de l'espace (A) à une autre (B), une récompense est calculée et vient se propager sur l'algorithme. La prochaine fois que l'algorithme se retrouvera dans l'état A il aura tendance à aller un peu plus vers B (ou l'inverse si la récompense est négative). L'idée intéressante dans cet algorithme est qu'à chaque passage une portion de la récompense de l'état B vient s'ajouter à la récompense désignant l'état A. Le résultat étant que la prochaine fois que l'on passera d'un état antérieur (C) vers l'état B alors la portion de récompense désormais sur B permettra de récompenser l'algorithme en C pour le guider, ou l'éloigner, de B. Ainsi la récompense de A se propage et guide, itération après itération l'algorithme depuis les états alentour. On résout donc un problème global : aller vers des récompenses ou les éviter en passant d'un état à un autre, en appliquant une politique uniquement locale (l'équation n'a lieu qu'entre deux états successifs), on parle de « programmation dynamique » (BELLMAN, 1957).

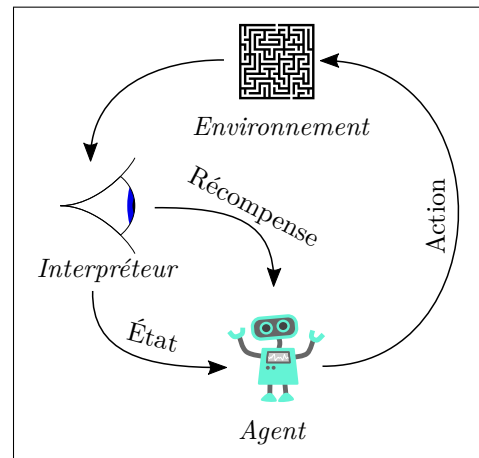


FIGURE 3.1. – Principe de l'apprentissage par renforcement, Image Creative Commons CC0 1.0

Le Q-learning est très proche du TD-Lambda, mais les implications entraînées par ses différences sont importantes. Introduit par Christopher J.C.H. Watkins dans sa thèse en 1989 (WATKINS, 1989) l'algorithme divise les états du TD-Lambda en autant d'états qu'il y a de possibilité d'action pour l'algorithme. L'avantage de cette séparation est que si un état A est adjacent à un état B et à un état C, que l'état B porte une récompense positive et l'état C une récompense négative, alors le passage de A vers C n'impacte que l'état « A + action vers C » alors que le passage de A à B impacte l'état « A + action vers B ».

Cette modification implique quelques étapes supplémentaires dans l'algorithme, il faut par exemple trouver le maximum des états-actions pour choisir l'action à réaliser et il faut rétropropager uniquement le maximum des récompenses des états-actions suivants sur l'état-action précédent.

L'un des avantages de cette méthode par rapport au TD-Lambda est d'éviter le moyennage des récompenses et punitions entourant un état, qui pourrait conduire l'algorithme à éviter les états proches de récompense négative, ce qui peut être non pertinent si l'algorithme a parfaitement appris à éviter l'action problématique.

Le Q-learning a également la particularité d'avoir été démontré mathématiquement et rigoureusement comme un algorithme convergeant bien vers une politique optimale (WATKINS et DAYAN, 1992) contrairement au TD-Lambda.

Les deux algorithmes sont également connus comme ayant été mis en relation avec le fonctionnement de la dopamine dans le cerveau des mammifères (HOUK et ADAMS, 1995)). Ce qui leur donne une légitimité certaine dans le domaine des algorithmes bio-inspirés.

Q-learning/TD-Lambda et variantes : cas du contrôle sensorimoteur Ces deux algorithmes ont l'avantage de résoudre le problème existant dans le contrôle optimal vis-à-vis de la définition et du calcul d'un critère à minimiser. Il devient ainsi possible de définir de multiples types de récompenses et punitions dans un grand nombre de contextes puis de laisser l'algorithme converger itérativement vers une trajectoire optimale sans qu'il ne soit nécessaire de formuler un critère global à minimiser.

Ce type d'algorithme prend également nativement en compte les problématiques d'ordre des tâches ou de variation du critère dans la tâche elle-même dès lors que toute la combinatoire est explorée.

Un autre avantage important est qu'il devient possible d'imaginer des récompenses et punitions qui correspondraient implicitement à plusieurs « critères » sans qu'il ne soit nécessaire de les séparer ou même de les formaliser. Il est ainsi possible de donner des récompenses globales et de laisser l'algorithme se charger de la convergence sans qu'il ne soit nécessaire d'interpréter les trajectoires comme étant le respect de tel ou tel sous-critère.

Ce dernier point est particulièrement intéressant, car il permet en théorie à l'algorithme de prendre en compte certaines interactions sociales dans ses trajectoires en ne postulant que le fait que les interactions vont avoir un impact sur le résultat final. Par exemple : donner un objet à un humain ou le poser sur le sol, si cela est fait de la même manière, peut être sous-optimal dans un cas (inutile de prendre des précautions en déposant un objet sur le sol) ou problématique dans l'autre (l'humain pouvant faire des erreurs, on risque de rater la main et donc d'augmenter le taux d'erreur final si on ne fait que lâcher l'objet toujours au même emplacement). Il est donc possible de concevoir des récompenses globales de réussite ou d'échec qui contiennent implicitement ces informations.

Ces avantages se retrouvent sur les variantes plus modernes de ces algorithmes qui ne sont pas forcément appliquées en environnement discret tels que les mécanismes acteur-critique et l'apprentissage profond par renforcement.

De très nombreuses architectures exploitants ces concepts existent dans le contrôle moteur afin de faire apprendre à un robot ou à un agent certaines tâches, citons par exemple : pointages d'objet (CHAO et al., 2014), comportements d'imitations (KOBBER et PETERS, 2010) ou encore la saisie d'objet chez l'enfant (BERTHIER, 1996).

Retenons donc ici que ces algorithmes sont parfaitement capables de prendre en compte l'ensemble des points critiques présentés dans la partie 3.2.1 et 3.2.1 de ce chapitre et qu'ils en ont fait la démonstration dans de nombreux cas particuliers.

Critique des algorithmes d'apprentissage par renforcement dans le cadre de l'apprentissage sensorimoteur Cette catégorie d'algorithme semble tout indiquée pour résoudre de nombreux types de problèmes, dont certains posés dans le cadre de cette thèse. Toutefois, ces algorithmes ne sont pas exempts de défauts dans leur application en pratique.

Le défaut principal de ces algorithmes dans notre cadre est que leur capacité à résoudre des problèmes très généraux est compensée par un temps de convergence extrêmement long. Ce qui peut être très problématique, voir inapplicable sur robot réel comme nous en avons discuté dans la partie 1.2 du chapitre 1.

Prenons l'exemple d'un ensemble de 10 tâches simples et imaginons que l'on souhaite profiter des bonnes propriétés de l'algorithme qui permettrait de montrer l'adaptation du mouvement dans une tâche en fonction des tâches à venir. Alors il serait nécessaire, non seulement d'apprendre avec l'algorithme les dix tâches en question, mais également de parcourir l'ensemble des combinaisons possibles de ces dix tâches, ce qui aboutit à un environnement à parcourir extrêmement grand et à une expérience quasiment impossible en pratique sans recourir à un environnement simulé.

Même sur des tâches relativement simples, si l'expérience doit se dérouler en pratique en impliquant des sujets humains alors le nombre d'expériences à effectuer rend la chose impossible en pratique.

Pour pallier ces problèmes, plusieurs solutions ont été proposées. La plus commune étant d'apprendre en environnement simulé pour ensuite calquer le résultat sur un robot réel. Comme nous en discutons dans le chapitre 1, ceci a le désavantage important de dépendre de la justesse de la correspondance entre le robot simulé et le réel.

Une autre possibilité, décrite par Sutton et Barto (SUTTON, BARTO et al., 1998), et largement utilisée dans cette communauté, est de mettre en place un *échantillonnage préférentiel* en réutilisant les données d'une expérience réelle et en les réorganisant puis en les piochant aléatoirement de manière à offrir à l'algorithme, à partir d'un ensemble de données restreint, l'impression d'une diversité dans les expériences tout en augmentant le nombre d'expériences présentées. La convergence de l'algorithme a tout de même lieu, ici encore, en simulation avant portage sur robot réel.

Cette méthode introduit de plus un biais important, car l'échantillonnage préférentiel n'invente évidemment aucune donnée : si l'espace exploré est trop réduit par rapport à la taille réelle de l'environnement alors le résultat sera mauvais par endroit. Si le concept s'adapte donc bien aux tâches simples, il est par contre évident que dès qu'il s'agit de capturer des détails d'interaction la faible quantité de données entrantes pose problème.

L'incapacité du système, dans les deux cas, à être véritablement utilisé en direct pose également problème si l'objectif est de simuler une interaction tout au long de l'apprentissage ou d'observer l'évolution d'un apprentissage sur certaines tâches.

Apprentissage par renforcement et approches bio-inspirées L'apprentissage par renforcement est comme nous l'avons vu un type d'algorithme très généraliste et capable de résoudre de nombreux problèmes efficacement, dont certaines des problématiques auxquelles nous nous intéressons.

Néanmoins, l'usage de cet apprentissage pour tous les cas de figure soulève quelques questions étant donné les défauts cités précédemment. On peut alors se demander comment la biologie utilise l'apprentissage par renforcement et surtout dans quel cas elle ne l'utilise pas et pourquoi.

Si on prend l'exemple du système musculaire, on remarque que la biologie a convergé au fur et à mesure de l'évolution vers des systèmes dont les propriétés tendent à simplifier, voire résoudre implicitement, les problèmes qui pourraient se poser au Système Nerveux Central (SNC). Le fait que le SNC utilise des types d'apprentissages par renforcement ne fait pas débat dans la littérature, mais on peut donc imaginer que, tout comme pour les muscles, de très nombreuses parties du système nerveux humain sont dotées de propriétés spécialisées et efficaces pour résoudre les tâches qui se posent. Ainsi l'apprentissage par renforcement viendrait, en appui d'autres systèmes, résoudre un problème qui lui a déjà été fortement simplifié.

On peut alors se poser la question dans le cadre du contrôle : à quel point les autres parties du cerveau fournissent-elles « clé en main » des composantes du contrôle moteur ?

De cette question, que nous allons aborder dans les parties suivantes, découle la principale critique

des usages des algorithmes de renforcement classique que l'on retrouve dans la littérature : la majorité de ces algorithmes est en effet utilisée comme une « boîte noire » et si leur utilisation est illustrée dans diverses applications, bien souvent la question du découpage du problème reste non abordée. Cette boîte noire fort pratique pousse parfois à résoudre des problèmes en utilisant l'apprentissage par renforcement dans des applications dans lesquelles l'humain n'en fait pourtant pas usage (par l'intelligence du corps ou par des propriétés de sous-parties spécialisées du cerveau, ou encore par des apprentissages locaux bien plus simplistes). Le fait qu'à chaque fois la boîte noire donne un résultat efficace ne pousse en effet pas particulièrement à la remise en question a posteriori de la pertinence du modèle employé.

Dans les parties suivantes nous aborderons le contrôle moteur humain sous l'angle de cette question : quelle partie du contrôle moteur le cerveau et le corps humain peuvent réaliser sans l'intervention d'un système généraliste d'apprentissage par renforcement et comment modéliser ces propriétés dans notre cadre ?

3.3 | Propriétés du système nerveux central et contrôle moteur chez l'humain

Dans cette partie, nous allons tenter de donner un aperçu concis du fonctionnement du contrôle moteur chez l'humain, tel qu'on le connaît dans l'état de l'art. Plus précisément, nous nous focaliserons sur deux uniques entrées du système nerveux : la vision et la proprioception et leurs liens avec le système moteur afin de générer un signal d'action cohérent. Nous aborderons également une description sommaire de quelques-uns des noyaux intermédiaires dont les rôles sont importants dans cette boucle.

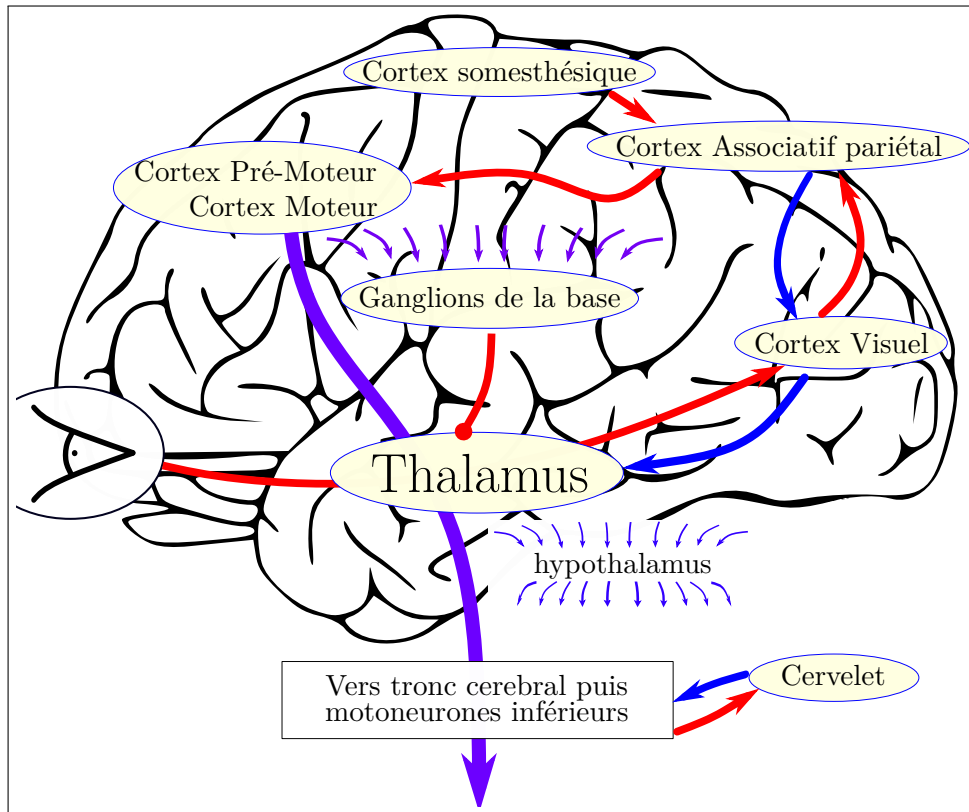


FIGURE 3.2. – Résumé des zones cérébrales ayant un lien avec la commande motrice et des liens principaux entre ces zones.

3.3.1. Cortex visuel

Le cortex visuel constitue la couche de traitement corticale des signaux visuels. Ces derniers partent de la rétine pour ensuite traverser le *Corps Géniculé Latéral* (CGL), abrité dans le thalamus, qui sert de relai entre la rétine et le SNC.

Ce cortex est habituellement divisé en plusieurs régions : le cortex visuel primaire est le premier ensemble qui reçoit les entrées rétiniennes, il va assurer un traitement rudimentaire de l'information à un niveau très local. Cet ensemble, souvent désigné par « V1 », envoie ensuite ces informations traitées à V2, puis V3, etc. à chaque nouvelle couche les informations sont de plus en plus intégrées et de moins en moins locales (on peut y rencontrer des neurones réagissant pour un certain motif positionné n'importe où sur la rétine). On appelle cette transformation de l'information la voie du « Quoi » (« *What* » en anglais) qui correspond à une indication sur la nature de l'objet. En parallèle existe une voie du

« Où » (« *Where* » en anglais) avec des neurones qui sont plutôt spécialisés dans la détection rapide du mouvement et de caractéristiques simplistes, mais localisés (les deux voies partagent évidemment des informations et se complètent). Ainsi le système nerveux central est informé sur la nature de l'objet via des primitives visuelles complexes et en même temps sur sa position via des primitives moins complexes, mais rapides à donner un résultat et localisées spatialement.

Certains des algorithmes les plus connus de catégorisations sont inspirés directement des jeux de connexions organisées qui furent observés dans ces ensembles, c'est notamment le cas de la proposition de carte auto adaptative de Kohonen (souvent abrégée « Carte de Kohonen ») (KOHONEN, 1982).

Retenons pour nos applications que le cortex visuel est un ensemble qui classe l'information visuelle en primitives exploitables par le système nerveux central.

Notons également qu'il existe en réalité énormément de connexion en sens inverse dans ce cortex (partant de V2 pour revenir vers V1 par exemple), parfois plus que de connexions dans le sens classique. La littérature fait souvent l'hypothèse que cette propriété pourrait aider à la localisation de caractéristiques d'objets dans le champ visuel. Mais il existe également de nombreuses hypothèses démontrant qu'il pourrait s'agir d'un moyen d'améliorer l'auto-organisation et d'autres propriétés de ces couches (notamment la stabilité dans l'apprentissage), c'est le cas par exemple du modèle ART de Stephen Grossberg qui se base sur ces liens récursifs entre couches (GROSSBERG, 1987).

3.3.2. Thalamus

Le Thalamus est un ensemble de noyaux servant de liens entre le SNC et le reste du système nerveux. On y retrouve donc énormément de noyaux relais pour diverses boucles (par exemple le Corps Géniculé Latéral). Le Thalamus est ainsi considéré comme la « porte de sortie » du système nerveux. C'est notamment via le thalamus que sont véhiculés les ordres moteurs ou retours sensoriels. Certains noyaux projettent ainsi sur le Thalamus afin d'apporter un dernier contrôle sur les sorties motrices ou afin de filtrer les entrées sensorielles.

3.3.3. Cortex associatif pariétal

Le cortex associatif présent dans l'aire pariétale est un cortex qui reçoit des afférences des centres de traitement sensoriels. Ici, nous nous intéressons surtout aux afférences qu'il reçoit du cortex visuel et du cortex somesthésique (ce dernier, que nous n'aborderons pas en détail, tient le même rôle pour les sens du toucher et de la proprioception que le rôle tenu par le cortex visuel pour la vision). Le cortex associatif mêle ces différentes données pour construire des ensembles de données correspondant à des événements existant sur plusieurs modalités. Il est souvent vu dans la littérature comme fonctionnant également sur le concept de cartes auto-organisées.

3.3.4. Cortex moteur et prémoteur

Le cortex moteur est l'emplacement des motoneurones supérieurs qui convoient les ordres vers les motoneurones inférieurs. C'est le centre moteur du cerveau. Ce cortex est organisé en carte topologique de façon à coder des directions de mouvements pour chaque binôme de muscles (GEORGOPOULOS et al., 1986). Il est donc suspecté de posséder également des propriétés d'auto-organisation. Il reçoit des afférences de nombreuses régions du cerveau, notamment des cortex sensoriels et associatifs. Le cortex prémoteur disposé non loin du cortex moteur est souvent vu comme étant un concepteur de séquences complexes qui seront ensuite activées sur le cortex moteur.

3.3.5. Ganglions de la base

Les ganglions de la base (ou noyaux gris centraux) sont un ensemble de noyaux complexes disposés au centre du cerveau. Nous détaillerons plus précisément ces noyaux dans le chapitre 4, mais résumons cet ensemble en deux structures. L'une de ces structures (dont fait partie le Striatum (BEISER

et al., 1997)) est souvent vue comme un sélecteur (REDGRAVE et al., 1999). C'est-à-dire qu'il reçoit des entrées d'une zone du cerveau puis va être capable de sélectionner les entrées les plus pertinentes parmi celles-ci selon le contexte en fonction d'un apprentissage par renforcement préalable guidé par des récompenses et punitions. Enfin, un ensemble de noyaux de cette structure s'occupera de sélectionner avec précision la sortie gagnante (ces noyaux sont souvent modélisés très simplement comme un algorithme de type « Winner Take All »), cette sortie viendra impacter ensuite le thalamus qui sert de relai à de nombreuses entrées et sorties du SNC, permettant ainsi aux ganglions de la base d'avoir un rôle de superviseur généraliste. Il est notamment impliqué dans la modulation de la sortie motrice (par inhibition et désinhibition (CHEVALIER et DENIAU, 1990)) au niveau du thalamus et a donc un lien fort avec la sélection de l'action motrice (AOSAKI et al., 1995 ; GURNEY et al., 2001 ; DOLLÉ et al., 2010).

Le second ensemble de noyaux des ganglions de la base est principalement constitué de la SNC (Substance noire pars Compacta) qui s'occupe de gérer la récompense qui viendra moduler le premier ensemble sous forme d'émission de dopamine. Cette dopamine de la SNC a été mise en lien avec la récompense des algorithmes de type TD-Lambda (MONTAGUE et al., 1993 ; SCHULTZ et al., 1997 ; SUTTON, 1988 ; DOYA, 1999). Depuis cette découverte, il existe un certain consensus sur le fait que la SNC viendrait moduler le premier ensemble de noyaux afin de sélectionner une action de façon pertinente et donc que l'ensemble exécuterait une forme de TD-Lambda neuronal aux équations équivalentes.

La réalité est un peu plus complexe, tout d'abord parce que les noyaux simplifiés en *Winner Take All* dans le premier ensemble ont un fonctionnement un peu plus subtil que cette simplification (GIRARD et al., 2002 ; GIRARD et al., 2003). Et ensuite parce que la dopamine est propagée de façon très étendue dans le cerveau, elle pourrait donc non seulement moduler le système de sélection d'action présent dans les ganglions de la base, mais également moduler tous les autres systèmes d'apprentissage présents dans les différentes aires du cerveau.

Notons aussi que les ganglions de la base ont une disposition en colonnes parallèles, ils peuvent donc être utilisés par de nombreuses zones du cerveau. Ils reçoivent de fait des afférences provenant de l'ensemble du cerveau, d'où leur position centrale qui favorise cet état de fait. Il est alors intéressant de considérer les ganglions de la base comme un outil auquel viendrait se raccorder les différentes parties du cerveau selon leurs besoins, ce qui s'accorde à la théorie du « neural reuse » de réutilisation des signaux disponibles par chaque ensemble du cerveau (ANDERSON, 2010) et à la vision que Kenji Doya avait explicitée dans le cadre de son analyse du cervelet (DOYA, 2000). Cette vision en tant qu'outil, c'est-à-dire sans chercher un « rôle global » à ces parties, mais plutôt en analysant le type de calculs qu'elles réalisent, évite ainsi de se perdre dans les multiples utilisations parallèles dont ces ensembles sont l'objet par d'autres parties du cerveau.

3.3.6. Hypothalamus

L'Hypothalamus est un noyau un peu à part, car il n'a un lien a priori que très indirect avec le système de contrôle moteur. Sa fonction est de tenir lieu de relai entre le système nerveux central et le système endocrinien. Il sert ainsi à traduire des signaux tels que la fatigue, le sommeil, etc. en signaux compréhensibles par le système nerveux et vice-versa. Il est ainsi le cœur d'émission de nombreux neuromédiateurs qui viennent changer l'état de fonctionnement de zones du cerveau.

3.3.7. Hippocampe

L'Hippocampe est une structure à la géométrie particulière qui fut mise en relation avec la création de « cellules de lieu » aidant au repérage dans un espace. Plus tardivement, des hypothèses plus générales ont émergé sur l'Hippocampe, le considérant plutôt comme un système de création de « transition » entre un état passé et un état futur, permettant ainsi de représenter une large combinatoire qui servirait alors d'entrée à d'autres structures qui pourraient exploiter ces données, comme les ganglions de la base. De fait, cette hypothèse explique également pourquoi on peut observer des « cellules de lieu » ou « de grille » dans l'Hippocampe, car une transition entre deux états observés peut représenter un marqueur du « lieu ».

3.3.8. Cervelet

Le cervelet est une structure un peu à l'extérieur du cerveau, proche de ce que l'on appelle le pont cérébral, le tronc cérébral ou plus anciennement le bulbe rachidien : cet ensemble de noyaux hétéroclites servant de relais et de centre de traitement réflexe entre le thalamus et les ordres moteurs. Le cervelet se situe juste à côté de cet ensemble et prend la forme d'une unité autonome qui forme un « petit cerveau », d'où son nom. D'un point de vue phylogénétique, il s'agirait d'un ancien reste de ce qui était le système nerveux central auparavant. Puis, le cortex tel qu'on le connaît s'est développé et le cervelet se serait retrouvé à jouer un rôle d'appoint.

Le cervelet reste néanmoins l'une des plus grosses structures du cerveau, on estime ainsi le nombre de neurones total dans le cerveau à approximativement 120 milliards (PELVIG et al., 2008) et ceux dans le cervelet à 101 milliards (ANDERSEN et al., 1992). Le cervelet constitue donc plus de 80% des neurones du cerveau humain (HERCULANO-HOUZEL, 2009).

L'anatomie du cervelet est, par rapport à d'autres structures du cerveau, très bien connue. En effet, la structure particulière du cervelet et la présence de neurones de tailles importantes ont rendu l'étude du cervelet aisée assez tôt dans l'histoire des sciences. Les « cellules de Purkinje » ont ainsi été les premiers neurones à être décrits en 1839 par l'anatomiste tchèque Jan Evangelista Purkyně (VOOGD, 2000).

L'anatomie du cervelet a donné lieu à une intense et abondante littérature. Comme dans le cadre des ganglions de la base, la littérature au départ s'est concentrée sur l'étude de chaque boucle qu'entretient le cervelet avec les entités voisines. On a ainsi trouvé des liens entre le cervelet et les récompenses, entre le cervelet et le traitement visuel, entre le cervelet et le mouvement, etc. La plupart du temps, ceci a surtout abouti à une certaine confusion, un peu comme dans le cadre des ganglions de la base, chacun voyant dans le cervelet un organe destiné à tenir des rôles très différents en fonction des boucles étudiées.

Ces modes d'études, même s'ils sont encore très présents dans la littérature, ont peu à peu été mis de côté par un autre point de vue. Kenji Doya a ainsi appelé à voir le cervelet plutôt comme un outil qui exécuterait une fonction calculatoire donnée et auquel viendrait se connecter les autres parties du cerveau lorsque cette fonction leur est nécessaire (DOYA, 2000). En le voyant ainsi il devient assez logique que le cervelet soit mis en jeu dans de nombreuses boucles différentes et parallèles.

D'un point de vue phylogénétique cette manière de concevoir le cervelet fait également sens, car son état de structure assez ancienne dans le cerveau plaide en faveur de connexions opportunistes ensuite faites par des structures voisines lorsque nécessaire.

Le cervelet est constitué de plusieurs couches. La première couche est la couche des *cellules granulaires*. Ces cellules sont extrêmement nombreuses et forment la large majorité des 100 milliards de neurones du cervelet, de fait sont les neurones les plus nombreux du cerveau humain. Il y aurait à peu près 200 de ces cellules, chez l'humain, pour une seule fibre entrante.

Ceci fait que les cellules granulaires sont suspectées de servir à la *décomposition des signaux*, soit grâce à des différences d'excitabilité entre-elles, soit à des différences de longueurs entre les fibres sortantes qui s'y connectent (on parle d'hypothèse des lignes à retard (BRAITENBERG et ATWOOD, 1958)) soit en profitant de connexions différentes avec leurs interneurons inhibiteurs (cellule de Golgi). Il s'agit en réalité probablement d'un peu des trois, car la littérature abonde de preuves neurobiologiques semblant indiquer que les trois aient une influence.

Notons que de façon assez paradoxale il y a peu de capture de signaux et d'études sur ces cellules elles-mêmes, car il s'agit de neurones excessivement petits et proches de leurs voisins, ce qui rend les analyses de signaux difficiles. Ce point est notable quand on sait que les cellules de Purkinje, qui forment la couche suivante, sont au contraire parmi les premières à avoir été étudiées du fait de leur grosseur.

Les fibres entrantes sur les cellules granulaires sont nommées « fibres moussues », elles proviennent d'endroit très divers du système nerveux, ce qui n'a rien d'étonnant si l'on considère l'hypothèse de « l'outil » de Kenji Doya. Un point qui appuie encore plus cette hypothèse est que les fibres d'une zone ont tendance à rester proches et à exciter des cellules granulaires proches. Cette proximité restera valide tout au long du passage dans les couches suivantes, formant des « bandes » parallèles qui chacune desserviront des emplacements différents du cerveau. Les sorties finales seront elles aussi très souvent

liées à l'emplacement des entrées.

Cette décomposition de signaux va servir à nourrir plusieurs couches de ce que l'on appelle les cellules de Purkinje. Si l'on ne considère que la première couche, le rôle de ces cellules apparaît alors nettement comme celui d'intégrateur de signaux. Les cellules de Purkinje sont en effet connectées à de très nombreuses cellules granulaires via des fibres nommées *fibres parallèles*, leur organisation nécessaire pour connecter les très nombreuses cellules granulaires leur donnant un aspect de ligne parallèle. L'arbre dendritique nécessaire à la cellule de Purkinje pour se connecter à autant de lignes parallèles en fait l'un des plus larges arbres dendritiques du cerveau, ce qui a facilité son étude. Une cellule de Purkinje serait connectée à environ 500 cellules granulaires, mais ces 500 fibres parallèles feraient néanmoins plus de 200 000 connexions avec l'arbre dendritique de la cellule de Purkinje, ce qui multiplie le potentiel excitateur des entrées.

Les cellules de Purkinje ont un type d'apprentissage assez particulier lié à un organe extérieur au cervelet : *l'olive*. L'olive est un organe extérieur au cervelet qui émet des fibres vers les cellules de Purkinje, ces fibres sont appelées fibres grimpanes et chaque cellule de Purkinje n'est connectée qu'avec une unique fibre grimpanne. On pourrait se dire qu'une unique fibre n'aurait que peu d'impact sur un arbre dendritique qui reçoit des centaines de milliers d'excitations. Mais c'est là qu'une particularité étonnante rentre en jeu : chaque fibre grimpanne vient s'entourer autour de l'arbre dendritique de la cellule de Purkinje, formant de très nombreuses connexions synaptiques. Ce qui fait qu'une unique émission de potentiel d'action depuis cette fibre surpasse largement les émissions des autres émetteurs, même si l'émission est faible.

Cette particularité a, très vite, fait suspecter que le cervelet jouerait le rôle d'un gigantesque réseau d'association. Notamment à partir des travaux de David Marr et Thomas W Thach en 1969 (MARR et THACH, 1991) puis de James S. Albus (ALBUS, 1971). En effet, ce signal unique prenant le pas sur tous les autres ressemble fortement au « signal désiré » des divers modèles de neurones associatifs. Il existe néanmoins une petite subtilité, deux écoles s'affrontant sur le sujet : une école considère que, comme dans le cas des réseaux de neurones formels, il s'agirait d'un signal « professeur » (école de Marr), et une autre école considère qu'il s'agirait au contraire d'un signal d'erreur (école d'Albus) s'appuyant sur certains travaux montrant que les cellules de Purkinje, après un fort potentiel d'action, subiraient plutôt une LTD (Long-Term Depression), soit un désapprentissage, plutôt qu'une LTP (Long-Term Potentiation : un renforcement). Dans ce cas, chaque cellule de Purkinje apprendrait en permanence et verrait ses « erreurs » se faire sanctionner. C'est cette seconde école qui semble aujourd'hui faire consensus.

Retenons donc que cette couche semble fonctionner dans tous les cas comme un gigantesque centre d'association. Ce qui va dans le sens de l'hypothèse « outil » de Kenji Doya. Un tel système étant probablement utile et « générique » dans de très nombreuses boucles.

Ensuite vient un point un peu moins tranché. Les noyaux profonds du cervelet qui constituent la couche suivante sont également constitués de cellules de Purkinje et forment les mêmes connexions avec les fibres grimpanes (en partageant donc une fibre grimpanne à chaque fois avec la couche du dessus) et avec les fibres parallèles. Mais ils reçoivent aussi directement des connexions des fibres moussues. Jusqu'ici rien qui ne les distingue de la couche supérieure à ceci près qu'ils sont moins nombreux et qu'ils sont donc de plus gros centres intégrateurs et qu'ils reçoivent également les fibres moussues en entrée donc qu'ils ont accès à la donnée d'origine pour leur intégration et l'apprentissage résultant. Néanmoins, la connexion est particulière avec la couche précédente de cellules de Purkinje. Cette dernière envoie en effet des efférences inhibitrices sur les noyaux profonds.

De nombreuses hypothèses existent ici : l'une d'elles voudrait que cette dernière couche ne serve qu'à calculer une erreur (erreur qui serait alors la sortie du cervelet) en soustrayant les calculs précédents avec les informations brutes. Néanmoins, ceci ne rend pas compte de l'apprentissage qui a également lieu dans cette couche de la même manière que dans la couche précédente, avec les mêmes données et le même signal de récompense.

Une autre hypothèse, plus simple, est de considérer que la couche précédente servirait en quelques sortes de « brouillon » de l'apprentissage avec lequel les noyaux profonds viendraient composer, par la liaison inhibitrice, un signal affiné en utilisant les données brutes d'entrées (fibres moussues) et les

données décomposées des fibres parallèles. D'une certaine manière, on peut voir cela comme une capacité d'apprentissage utilisant la couche intermédiaire comme réservoir de signaux inhibiteurs. Un peu de la même façon qu'un perceptron ou LMS est capable de faire converger certains de ses poids vers des poids négatifs pour affiner son calcul. Ici, la nature aurait donc décomposé le système de cette manière parce que le système biologique humain doit composer avec deux types de liens différents pour porter un signal inhibiteur et excitateur. Deux types de liens qui ne peuvent, en outre, être émis que par un neurone à la fois (un neurone ne peut pas émettre un signal inhibiteur et excitateur). Ce dédoublement pourrait donc trouver son explication ici.

Les noyaux profonds ont ensuite une sortie excitatrice qui innerve de nombreuses parties du cerveau. Des interneurons inhibiteurs vont quant à eux renvoyer un signal négatif vers l'olive. Ce qui va dans le sens d'une olive qui serait là pour construire une erreur en soustrayant le signal calculé (S) à un signal entrant ($S_{\text{désiré}}$?).

Ce dernier point va encore plus dans le sens d'un cervelet vu comme un gigantesque outil d'association. Notons que le cervelet peut recevoir des afférences dopaminergiques, ce qui le transformerait alors en centre d'apprentissage par renforcement. Ceci n'étant pas incompatible avec son rôle supposé précédemment (les deux fonctions pouvant avoir lieu via le même substrat tant que la dopamine irrigue une zone et pas l'autre).

Rôles du cervelet

Le cervelet fait partie, comme nous l'avons vu, de très nombreuses boucles, mais ce qui nous intéresse surtout c'est son rôle dans le cadre du contrôle moteur. Dans ce cadre, le cervelet est soupçonné de tenir un rôle majeur dans certains réflexes (réflexes posturaux) qui demandent des ajustements rapides et fins (pour l'équilibre par exemple). Mais il est également vu comme un centre d'automatisme moteur qui viendrait construire la précision du mouvement (MARR et THACH, 1991). Ce point est assez logique si l'on voit le cervelet comme un gigantesque outil de décomposition-association. Cela le rend alors capable d'apprendre un mouvement par rapport à son passé et donc d'apprendre des trajectoires réalisées finement, rapidement, et automatiquement. On peut ainsi imaginer que tous les mouvements humains qui viendraient se répéter suffisamment longtemps seraient alors appris comme des automatismes dans le cervelet, automatismes qui viendraient alors prendre le pas peu à peu sur les ordres du système moteur central, qui n'aurait plus besoin que de donner certains ordres de départ pour entraîner le déclenchement de l'automatisme.

La façon dont ce passage se fait est sujette à débat.

Le cervelet peut aussi être vu comme un compensateur en cas d'absence sensoriel. Il est ainsi capable de faire des liens complexes entre les différents senseurs et de venir compenser un senseur lorsque celui-ci viendrait à faire défaut, en utilisant les données du passé et celles d'autres senseurs.

Cette hypothèse, malgré son élégance, n'a pas forcément de bases fortes neurobiologiques. Les sorties du cervelet sont en effet bien moins nombreuses que ses entrées. S'il est ainsi possible d'imaginer l'entrée, sans le cervelet, de l'ensemble des senseurs du système nerveux, au vu du très grand nombre de cellules granulaires, le sujet devient par contre plus délicat si l'on considère la possibilité que ce grand nombre de signaux soit également présent dans les sorties plus limitées du cervelet.

Néanmoins, retenons ces quelques rôles du cervelet dans notre cadre :

- Centre de calcul pour les réflexes
- Centre d'apprentissage d'automatisme
- Centre d'association sensoriel

Dans notre cadre, on peut toutefois noter après cette analyse que si le cervelet remplit des fonctions qui semblent essentielles, nombre d'entre elles en dehors de certains réflexes, semblent plutôt lier à l'acquisition d'automatisme et à l'affinage d'apprentissage sur le temps long et ne semblent pas forcément nécessaires à la compréhension des boucles motrices initiales du système nerveux humain dans le cadre des tout premiers stades du développement.

3.3.9. Développement du contrôle moteur chez l'enfant

« L'enfant explique l'Homme autant, et souvent plus, que l'Homme n'explique l'enfant. »

Jean Piaget (1966)

On se pose la question de la sélection d'une boucle impliquant les parties du cerveau précédemment présentées qui permettrait de rendre compte des premières étapes du développement. Notre hypothèse est en effet que le développement de fonctions plus complexes, qu'elles soient motrices ou non, se base forcément sur le développement préalable du système bas niveau moteur. On postule également que c'est ce développement moteur qui permet l'émergence de certaines propriétés essentielles à l'interaction avec l'environnement ou avec un partenaire.

Les idées qui postulent l'importance du développement moteur comme base nécessaire au développement des autres capacités, imaginant ainsi un développement par strate de l'être humain, nous viennent principalement de Jean Piaget.

Piaget, par l'observation du développement des enfants, postule ainsi un ordre de développement qui sera toujours le même en faisant l'hypothèse que chaque strate de développement est naturellement dépendante de la précédente, les capacités d'une strate dépendant en effet, selon son hypothèse, des capacités de la strate précédente (PIAGET, 1936; PIAGET, 1948; PIAGET, 1937). Piaget présente le développement de ces strates ainsi :

- Stade sensorimoteur (0-2 an), dans lequel l'enfant explore l'environnement, compte sur ses réflexes, montre des phénomènes d'imitation immédiate et différé et mature sa représentation mentale des objets.
- Stade préopératoire (2-7 ans), dans lequel l'enfant maîtrise le langage, le dessin, et les premiers raisonnements.
- Stade des opérations concrètes, dans lequel l'enfant maîtrise la conceptualisation de raisonnement logique concret, commence à raisonner avec des nombres, etc., mais toujours en relation avec des phénomènes observables.
- Stade opérations formelles : maîtrise des logiques formelles et abstraites.

Piaget postule ainsi la primauté des réflexes sur le développement, puis des capacités sensorimotrices basiques, et seulement ensuite viendrait le langage, etc. Dans les raisonnements de Piaget, celui-ci accorde une grande importance à l'interaction avec l'environnement qui viendra changer radicalement les résultats de chacune de ces étapes, sans remettre toutefois en question leur ordre respectif. Cette théorie, dont Piaget est l'un des pères, est nommée le « *constructivisme* ». Il s'oppose à son époque aux « *béhavioristes* » (tel John Broadus Watson et Burrhus Frederic Skinner) qui considèrent l'enfant comme une « ardoise vierge » sur lequel l'apprentissage aura un pouvoir presque absolu et aux « *innéistes* » (tel Arnold Gesell) qui à l'inverse considèrent que le comportement humain est avant tout inné.

Pour Piaget, seul l'ordre de la maturation cérébrale, qui donnera l'ordre des comportements observés, est contraint, car certaines compétences doivent obligatoirement se baser sur des compétences de plus bas niveau. Mais en dehors de cet ordre, c'est l'environnement et l'interaction de l'enfant avec l'environnement qui dictera le résultat de chacun de ces stades qui peuvent donc aboutir à un résultat très variable.

Piaget formalise sur ces bases la théorie de « l'équilibration » (PIAGET, 1975) qui postule l'adaptation entre l'organisme et le milieu, la continuité entre le biologique et le cognitif (très proche de ce que l'on entend aujourd'hui par « l'intelligence du corps ») et qui intègre deux mécanismes : l'assimilation (intégration du monde extérieur) et l'accommodation (modification des schémas de comportement face au monde extérieur). Ces points font également partie des fondations de notre approche des problèmes posés en robotique interactive.

Les théories de Piaget sont encore aujourd'hui l'une des clés essentielles à la compréhension de certaines pathologies chez l'enfant. Il a ainsi été démontré que les enfants présentant des retards dans le développement de certains stades présenteraient un risque pour les stades suivants (WANN et al., 1998 ;

MELILLO et LEISMAN, 2010). À l'inverse il a été démontré que certains facteurs neuronaux importants dans les parties du cerveau en lien avec les premières étapes du développement pouvaient aboutir à de meilleures performances dans le développement de ces stades et ainsi par la suite à des fonctions cognitives plus abouties et saines. (MURRAY et al., 2006 ; RIDLER et al., 2006). Graham K Murray et ses collègues ont ainsi analysé ces questions dans des domaines plus larges et démontrés que les premières étapes du développement sont bien primordiales pour la réalisation avec succès des étapes suivantes (MURRAY et al., 2007).

À un point donné d'une étape du développement, ces aspects se démontrent également en observant une corrélation entre le développement moteur d'un enfant et ses autres capacités (BURNS et al., 2004) et à l'inverse en montrant la corrélation entre certaines absences de capacités motrices et certaines pathologies comme l'autisme (TEITELBAUM et al., 2002).

La pensée de Piaget se généralise même en amont des stades sur lesquels il a travaillé. Il a en effet été démontré que des retards dans le développement des premiers réflexes chez des enfants prématurés pouvaient poser ensuite des problèmes dans la première étape du développement sensorimoteur (MARQUIS et al., 1984).

De manière générale, la théorie de Piaget repose donc aujourd'hui sur une solide base d'observations diverses dans le développement de l'enfant (pour plus de bibliographies sur les atteintes aux développements de l'enfant, se référer aux excellentes bibliographies de Gerry Leisman et collègues (LEISMAN et al., 2014) et de Robert Melillo (MELILLO, 2011)) et d'observations neurobiologiques générales démontrant le développement successif des aires cérébrales qui semblent effectivement aller dans le sens des propositions de Jean Piaget (CASEY et al., 2005).

Cet ensemble nous permet de poser comme hypothèse solide que pour obtenir des architectures présentant des caractéristiques proches de celles d'un humain, notamment dans la capacité d'interagir, il faut donc avant tout nous concentrer sur les premières boucles motrices se développant chez l'enfant.

Comme nous l'avons vu, le développement de l'humain vers un être humain sain et n'ayant aucun désordre mental, donc capable d'interaction avec ses semblables, passe par le développement des premières boucles motrices et des réflexes chez l'enfant. On se pose alors la question de la modélisation des éléments primordiaux de ces boucles.

3.3.10. Positionnement et choix des boucles sensorimotrices à modéliser

En s'inspirant de la littérature présentée précédemment, trois aspects principaux nous paraissent importants à considérer :

- Le développement des réflexes. Comme en fait l'hypothèse Piaget et comme l'illustre (MELILLO, 2011) ils sont à la base du développement. Le SNC s'appuie en effet sur les boucles réflexes constituées dans les noyaux du pont, dans le cervelet et sur les réflexes au niveau des muscles pour se développer. Comme il a été démontré sur les enfants prématurés, un dysfonctionnement de ces boucles réflexes entraîne ensuite en cascade des problèmes aux stades suivants. On pourrait parler de stade de développement « 0 », ce qui en fait un aspect important à considérer.
- La boucle sensorimotrice. Premier ensemble à se développer après les réflexes (comme le théorise Piaget et comme l'illustre la biologie), cette boucle est constituée, pour nous, des cortex sensoriels (dans notre cas visuel et somesthésique, car on ne considère que ces deux modalités dans cette thèse) et du cortex moteur. C'est dans cet ensemble que se forment les primitives motrices cohérentes vis-à-vis des possibilités motrices de l'enfant ainsi que les primitives visuelles et sensorielles en amont.
- L'intervention des ganglions de la base. Les ganglions de la base et leur activité de sélection et d'émission de dopamine jouent un rôle primordial dans le développement de l'enfant et dans la résolution de problèmes moteurs plus complexe (LEISMAN et al., 2014). On suppose que cette boucle comprend non seulement les ganglions eux-mêmes, mais fait également jouer un rôle important à l'hippocampe (comme entrée de données complexes assimilable par les ganglions

de la base), au thalamus (comme intermédiaire à moduler, sortie des ganglions de la base) et à l'hypothalamus comme modulateur selon diverses contraintes que les ganglions doivent prendre en compte (faim, fatigues, sommeil, etc.).

- L'aide à la réalisation de séquences de mouvements complexes. Il existe de nombreux sous-systèmes qui aident à la réalisation de mouvements complexes, c'est notamment le cas du cortex prémoteur, mais on suppose également que le cervelet avec ses capacités d'association importantes permet également d'apprendre des automatismes et donc des mouvements complexes et fins par répétition. De même que le cervelet, des boucles du pont cérébral sont supposées être le cœur de mouvements rythmiques complexes (Central Pattern Generator).

Sélection de la boucle sensorimotrice, justifications et hypothèses.

Nous nous focaliserons principalement sur le développement de la boucle sensorimotrice. Nous considérerons ici que tous les aspects réflexes sont en effet déjà développés et calibrés (en gardant de plus à l'esprit que beaucoup de ces réflexes n'existent que pour des raisons de contraintes biologiques qui peuvent ne pas être pertinentes en robotique comme nous l'aborderons dans la section 4.2).

De même, on considérera que la gestion fine et complexe du mouvement selon des récompenses et punitions est un aspect qui vient après le développement basique des primitives motrices, non seulement par analogie avec les hypothèses de Piaget et les vérifications ultérieures de ses hypothèses en biologie, mais aussi par analogie entre les ganglions de la base et les algorithmes TD-lambda et Q-learning qui ne convergent que sur des temps longs. Nous aborderons néanmoins les aspects liés aux ganglions de la base et à l'apprentissage par renforcement dans le chapitre 5 en considérant que les étapes précédentes, notamment la boucle sensorimotrice, sont déjà développées.

De même, la réalisation de séquences de mouvements fins ou de répétitions complexes nous semble nécessiter, toujours dans le même esprit, la présence de primitives motrices déjà développées. Nous aborderons les aspects de répétitions et de structure à la base du rythme dans les mouvements dans la partie IV.

Après cette réduction de notre problème, en considérant que les réflexes de base nécessaires aux fonctionnements des muscles sont préexistants, on se retrouve avec une boucle sensorimotrice restreinte présentée sur le schéma 3.3.

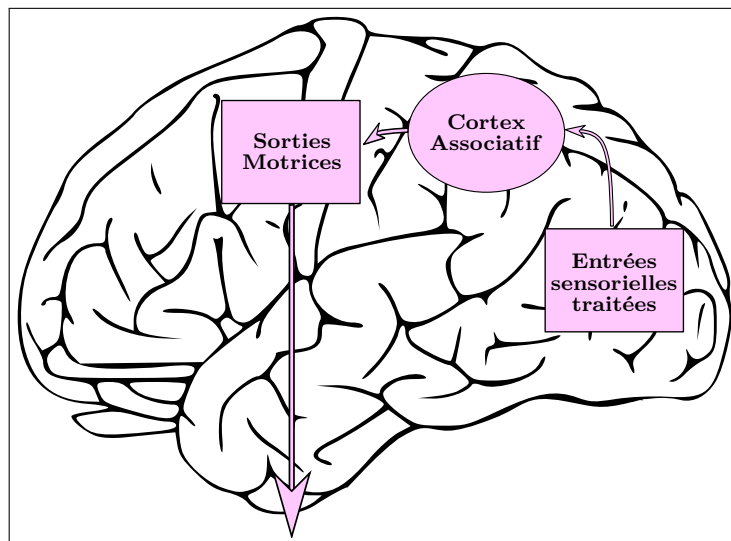


FIGURE 3.3. — Concept simplifié une fois pris en compte les hypothèses développées dans la partie 3.3.10.

3.4 | Conclusion

Nous avons abordé dans ce chapitre les différentes classes de modèles existants et nous avons discuté de leurs avantages et inconvénients respectifs. Ceci nous a permis d'isoler une classe de modèle qui semble être la plus pertinente dans notre cadre : les algorithmes basés sur l'apprentissage. Les algorithmes existants ne nous ont pas paru répondre à l'ensemble des problématiques soulevées dans le cadre du développement, sur le robot Tino, d'un modèle de contrôle capable d'interaction avec un environnement réel et des partenaires.

Nous avons ensuite résumé le fonctionnement de quelques structures importantes du système nerveux chez l'humain impliquées dans le contrôle moteur. Ceci afin de mettre en avant une boucle sensorimotrice à modéliser.

En nous aidant des travaux de Piaget sur les stades du développement et sur les démonstrations de ces stades apportées ultérieurement par les domaines de la biologie et de la psychologie, nous en avons conclu qu'un robot interagissant dans un environnement réel ou avec un partenaire devrait, si on s'inspire du cas humain, baser ses compétences d'interaction sur les premiers stades du développement moteur, notamment sur les stades de conception des réflexes puis sur la boucle sensorimotrice. En prenant comme hypothèse que les réflexes sont déjà partiellement construits nous en avons conclu que la boucle à modéliser en priorité dans nos modèles serait la boucle sensorimotrice et plus particulièrement son développement dans l'interaction avec l'environnement.

Dans les parties suivantes, nous aborderons un premier modèle naïf qui tente d'utiliser les particularités émergentes du robot Tino dans une boucle sensorimotrice simple. Nous étudierons ensuite les défauts de ce modèle avant de proposer et d'étudier un modèle intermédiaire de contrôle musculaire qui servira à une architecture plus développée.

Modèle développemental d'un contrôleur sensorimoteur

Sommaire du chapitre :

4.1. Première architecture naïve	68
4.1.1. Concept général	68
4.1.2. Protocole expérimental	68
4.1.3. Architecture	68
4.1.4. Résultats expérimentaux	71
4.2. Contrôle musculaire et modèle de contrôle bas niveau pour robot	75
4.2.1. Le muscle seul et ses systèmes de régulations	75
4.2.2. Le jeu des muscles agonistes-antagonistes : co-contraction (ou co-activation musculaire)	79
4.2.3. Point de vue du roboticien sur les modèles de muscles	80
4.2.4. Modèle de co-contraction musculaire pour robot	80
4.2.5. Conclusion	87
4.3. Modèle développemental pour l'apprentissage de liens perceptions-actions	88
4.3.1. Description du modèle	88
4.3.2. Résultats et convergence de l'algorithme	96
4.3.3. Imitation immédiate : effet émergent du modèle en situation d'ambiguïté perceptive	99
4.3.4. Analyse et critiques, de l'algorithme et perspective pour une évolution du modèle	105
4.4. Conclusion	119

4.1 | Première architecture naïve

Dans la partie précédente, nous avons illustré un phénomène émergent de la mécanique du robot Tino. Comme nous en avons discuté dans l'introduction de cette thèse, nous nous positionnons dans une approche développementale et constructiviste, nous mettons donc l'accent sur le développement des capacités du robot à travers ses possibilités d'actions et la perception de la réalité que lui permet son corps. Dans ce cadre, il nous a paru judicieux de nous intéresser aux caractéristiques présentées dans le chapitre précédent afin de savoir s'il était possible de les exploiter dans le cadre du contrôle du robot via une approche développementale.

Nous nous basons également sur la boucle sensorimotrice décrite précédemment et fruit de notre réflexion sur les parties intéressantes à modéliser dans le cadre de cette thèse.

Nous mettons en premier lieu en place une architecture « naïve » basée sur des réseaux de neurones artificiels et qui a pour but de démontrer la possibilité de profiter des phénomènes émergents de la mécanique de Tino.

4.1.1. Concept général

Le phénomène émergent présenté à la fin du chapitre 2 se résume ainsi : à un point de départ « A », le bras immobile portant une charge « C » est soumis à t_0 à une force due à son vérin que l'on suppose constante, image de la commande électrique « I ». Le bras étant soumis à un ensemble de frottements et de forces dues à son poids propre et à celui de la charge portée il finit par ralentir et s'arrêter totalement à une position finale « F ». Ce phénomène peut être exploité à la fois pour lier une force à une position finale et à la fois pour lier un poids porté à une modification dans la commande nécessaire pour atteindre une position donnée.

Dans cette expérience nous nous contenterons de lier sans charge, autre que le poids propre du bras, la commande « I » avec la position finale « F » afin de jauger s'il est possible ou non d'exploiter ce phénomène émergent directement dans le contrôle du robot sans études supplémentaires.

4.1.2. Protocole expérimental

Le protocole est simplement la mise en forme du concept ci-dessus : on démarre d'une position minimale pour un axe du bras (ici l'expérience est faite sur l'axe haut-bas de l'épaule du robot, qui permet le plus d'étendue de mouvement) puis on applique une commande, en commençant par les commandes les plus faibles, et en progressant par incrément. On laisse ensuite l'architecture (détaillée plus loin) apprendre le lien entre la position obtenue après stabilisation et la commande envoyée. L'objectif étant d'exploiter le résultat de l'apprentissage afin de pouvoir ensuite utiliser l'architecture pour atteindre une position désirée.

Ici, on fait l'approximation que le temps de stabilisation après une commande est très supérieur à celui du trajet pour atteindre la position finale, alors que la commande, elle, est toujours émise de façon constante. Une simple association *position* \rightarrow *commande* sera donc en moyenne équivalente, même en laissant l'apprentissage fonctionner de façon continue, à une association *position finale* \rightarrow *commande* à partir du moment où le temps d'intervalle entre deux itérations de l'expérience est suffisamment long pour que le temps durant lequel le bras est stable soit supérieur au temps passé en mouvement.

4.1.3. Architecture

L'architecture est une association simple présentée sur la figure 4.2.

Les deux signaux sont remis dans une échelle commune avant d'être catégorisés via un réseau classique de type K-moyenne (« *K-mean* » en anglais) (STEINHAUS, 1956) capable, contrairement aux K-means classiques, de créer dynamiquement des noyaux en plus de les déplacer (ceci afin d'accélérer l'apprentissage, au détriment de l'utilisation en mémoire).

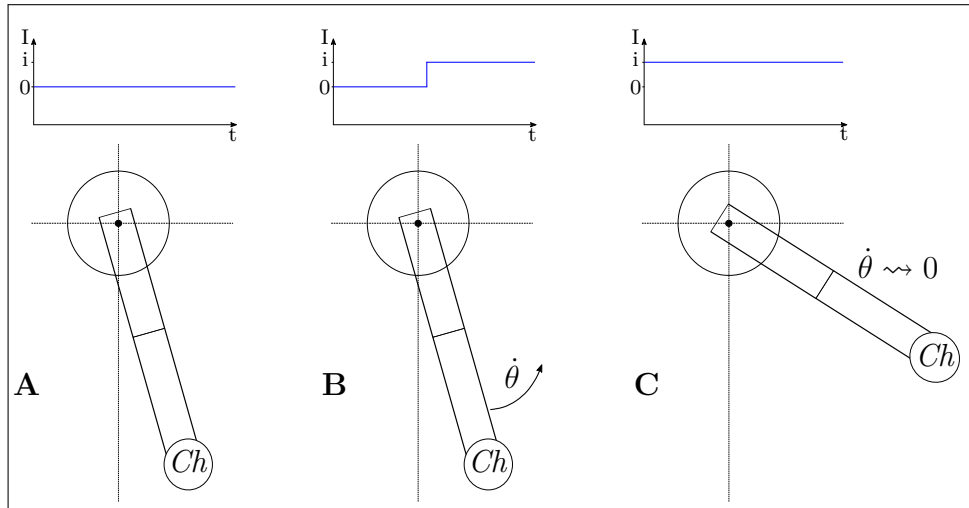


FIGURE 4.1. — On représente le protocole destiné à l'association entre la position et la commande. Une commande est appliquée à t_0 (A) au bras du robot portant une charge (Ch) (ici charge constante équivalente au poids de sa main) puis le bras se met en mouvement (B) pour se stabiliser (C) lorsque la vitesse angulaire ($\dot{\theta}$) devient nulle. On exécute cette séquence en boucle pour différentes commandes. Une étape n'est pas représentée ici : entre l'étape C et l'étape A de la séquence suivante on exécute une étape « D » en envoyant une légère commande négative qui ramène l'axe considéré à sa position minimale.

L'équation de réactions des noyaux est définie comme suit :

$$Acx_i = 1.0 - e^{-\frac{1}{2} \cdot \sum_j \frac{(W_{ij} - Ecx_j)^2}{\sigma^2}} \quad (4.1)$$

Où i est un itérateur pour la couche actuelle et j pour la couche d'entrée. Acx_i est l'activation du noyau i , Ecx_j la sortie du neurone j de la couche d'entrée. W_{ij} est le poids synaptique reliant i à j et représente la position du noyau dans l'espace d'entrée. σ est un facteur définissant la largeur de réaction du neurone en fonction de la distance.

L'activation des noyaux est ensuite remise sous forme neuronale afin de pouvoir donner un ensemble de données cohérent pour un apprentissage neuronal ensuite. La sortie du neurone associé au noyau est donc définie par :

$$Scx_i = f(Acx_i) \quad (4.2)$$

Où f est une fonction d'activation non linéaire dont la sortie est définie telle que :

$$f(Acx_i) = \begin{cases} 1 & Acx_i \geq 1 \\ Acx_i & 0 \leq Acx_i \leq 1 \\ 0 & Acx_i \leq 0 \end{cases} \quad (4.3)$$

Le système de création dynamique de nouveaux noyaux, quant à lui, fonctionne comme ceci :

$$\begin{array}{l} \text{si } Scx_i < \text{Vigilance} \quad \forall i \text{ alors} \\ | \quad W_{(i_{max}+1)j} = Ecx_j \quad \forall j \end{array}$$

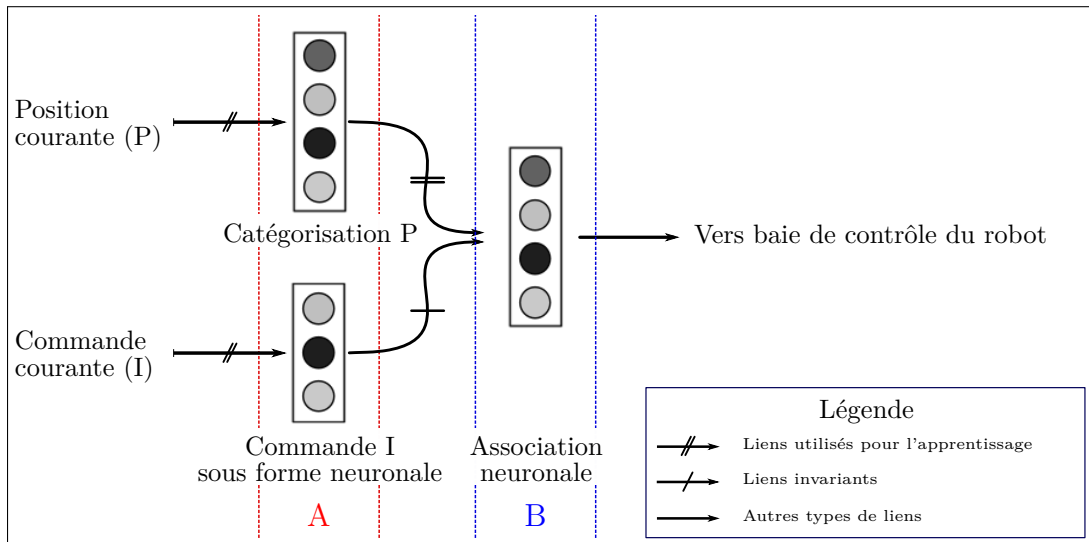


FIGURE 4.2. – L'architecture récupère en premier lieu la position courante (P) et la commande du robot (I) qui sont ensuite projetées sur la couche A . La commande est projetée sous forme neuronale (un neurone par axe considéré) tandis que la position est catégorisée grâce à un algorithme de type K -mean (détaillé à partir de l'équation (4.1)) Le résultat de ces catégorisations est ensuite utilisé pour associer la position à la commande actuelle dans la couche B grâce à un algorithme de type LMS (détaillé à partir de l'équation (4.4)).

Où i_{max} est le dernier noyau créé et $i_{max} + 1$ le suivant.

Le recrutement est conçu afin de s'assurer que les équations (4.1) et (4.2) fassent que la sortie soit maximale (ici, le maximum est défini à 1,0) quand le recrutement a lieu, ceci afin de créer un nouveau noyau exactement là où il est le plus nécessaire vis-à-vis de l'espace d'entrée pour représenter la nouvelle donnée au mieux. Chaque noyau peut en plus se déplacer comme dans le cadre du K -mean classique (mêmes équations).

La couche d'association, pour finir, est un algorithme classique d'association de type moindre carré (« LMS » pour « *Least Mean Square* ») (WIDROW et HOFF, 1960) proche d'un perceptron monocouche (ROSENBLATT, 1958). L'activité de chaque neurone est simplement définie par :

$$A_i = \sum_{\forall j} W_{ij} * Am_j \quad (4.4)$$

et les poids modifiés en fonction de la sortie désirée tels que :

$$\Delta W_{ij} = \epsilon * (Sd_i - S_i) \quad (4.5)$$

Où Sd_i est la sortie désirée et S_i la sortie actuelle du neurone i .

La sortie du neurone i est, elle, définie par :

$$S_i = f(A_i) \quad (4.6)$$

Où f est la même fonction non-linéaire classique définie précédemment dans l'équation (4.3)

On détaillera plus précisément les subtilités de ce type d'architecture dans la section 4.3.

4.1.4. Résultats expérimentaux

Dans un premier temps, on cherche à observer la progression de l'erreur au cours de l'apprentissage, ceci afin de valider que l'architecture est capable d'apprendre dans ces conditions. On peut observer le résultat sur la figure 4.3.

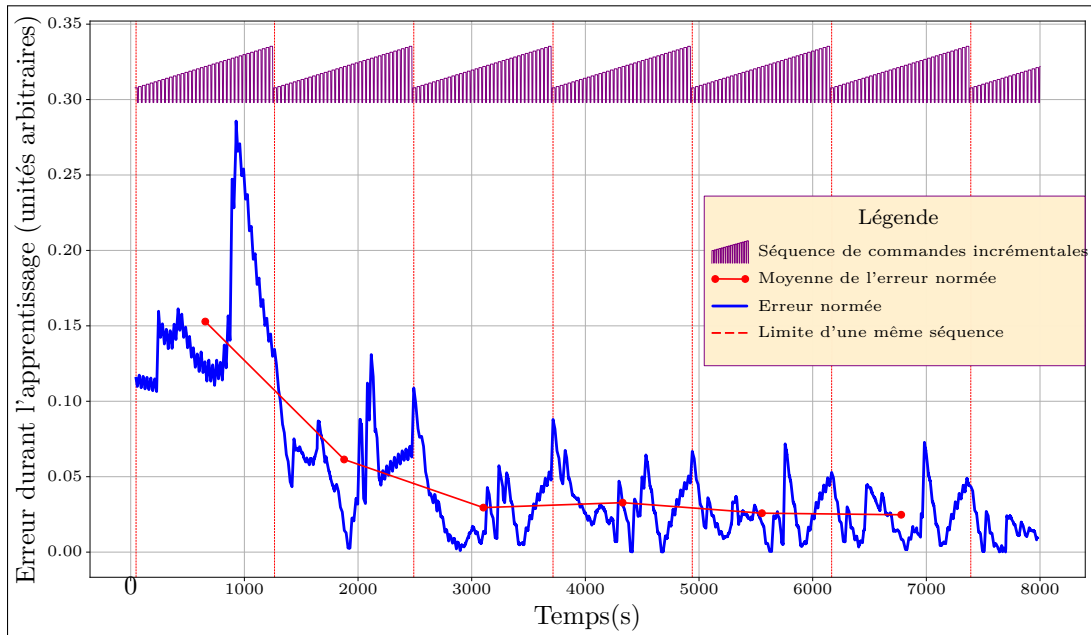


FIGURE 4.3. – *Erreur de reconstruction absolue normée durant la progression de l'apprentissage, on peut observer sur cette figure la courbe d'erreur en fonction du temps (en bleu), la courbe représentant les séquences de commandes (en violet : unités arbitraires, la courbe n'est pas à l'échelle ; elle n'est ici que pour donner une idée du lien entre commande et erreur) et la courbe d'erreur moyenne par séquence.*

Comme on peut le voir, l'erreur diminue très vite puis finit par stagner. La stagnation éventuelle due au pas de l'apprentissage n'a plus d'influence au-delà de celui utilisé pour cette courbe, c'est donc bien une limite due à la nature des données apprises qui se manifeste par cette stagnation et non une limite due à des paramètres que l'on pourrait encore optimiser.

La courbe se déroule sur plusieurs séquences successives d'expériences composées de séquences de commandes qui augmentent par incrément (en violet au-dessus de la courbe), ce qui explique le motif répétitif que l'on distingue sur la courbe d'erreur.

L'erreur absolue est pertinente pour avoir une idée du décalage dans la commande du robot sur le résultat qualitatif observé, mais pour juger de la progression de l'apprentissage et de ses défauts il peut être plus pertinent de tracer l'erreur relative. On peut observer celle-ci sur la figure 4.4. On y observe bien une chute importante de l'erreur de 50% en moyenne au départ à une moyenne d'environ 10% en fin d'apprentissage.

La forme particulière de ces courbes montre une répétition de motif à chaque nouvelle expérience, motif corrélé à la progression des pas de commande, on observe particulièrement ce motif sur la courbe relative, qui montre les différences importantes d'erreur entre les différentes parties de la courbe durant une séquence.

Pour mieux comprendre les raisons d'un tel motif, on peut s'intéresser aux positions qui sont effectivement atteintes après stabilisation sur l'une des séquences de commande. On observe une séquence superposée à la position atteinte après stabilisation sur la courbe 4.5.

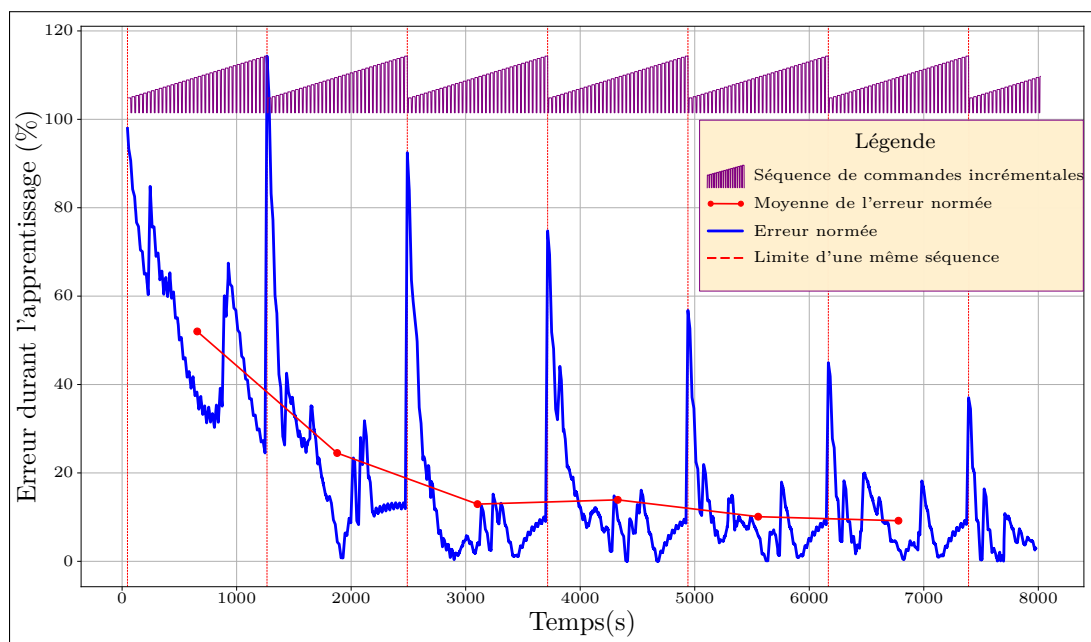


FIGURE 4.4. – *Erreur de reconstruction relative (normée) durant la progression de l'apprentissage, on peut observer sur cette figure la courbe d'erreur en fonction du temps (en bleu), la courbe représentant les séquences de commandes (en violet : unités arbitraires, la courbe n'est pas à l'échelle; elle n'est ici que pour donner une idée du lien entre commande et erreur) et la courbe d'erreur moyenne par séquence.*

Comme on l'observe sur cette courbe, la progression n'est pas parfaitement linéaire tout du long et possède des non-linéarités fortes au début et à la fin de la séquence. Au début de la séquence, car les commandes sont alors trop faibles pour vaincre les forces de frottement et d'inertie du système, ce qui aboutit à un décollage vif de la courbe lorsque la commande atteint un certain palier. Et à l'inverse une fois la position mécanique maximale atteinte les commandes supérieures ne donnent alors pas d'autres résultats et aboutissent logiquement au même palier.

Ces deux non-linéarités ont des implications importantes sur le motif d'erreur relatif. On peut tracer pour s'en rendre compte le motif d'erreur, sur une séquence, superposé à la progression de ces positions stables pour chaque commande durant la séquence. On peut voir cette superposition sur la figure 4.6.

On peut observer que plusieurs des pics principaux sont corrélés avec les non-linéarités de la courbe de progression des positions finales. Ce qui est logique. Prenons l'exemple des non-linéarités en début d'expérience. La commande est alors trop faible pour vaincre les forces de frottement et l'inertie (entre 6 150 s et 6 300 s sur la figure). En effet, le problème que l'algorithme d'apprentissage tente de résoudre n'est alors pas soluble, car il existe une infinité de commandes correspondant à la même position. Cet algorithme seul ne pouvant choisir entre ces possibilités il en découle une erreur qui reste irréductible. Le problème est le même au-delà d'une certaine commande envoyée (vers la fin de l'expérience, à partir de 7 000 s sur la courbe). Dans ces deux zones, on obtient donc un moyennage de l'apprentissage avec un unique emplacement « juste » et le reste présentant des erreurs d'où cette formation de pics, par exemple, entre 7 000 s et 7 400 s.

Les pics d'erreurs restant entre ces deux zones « mortes » sont quant à eux probablement dus à la variabilité des données et pourraient sans doute être encore optimisés.

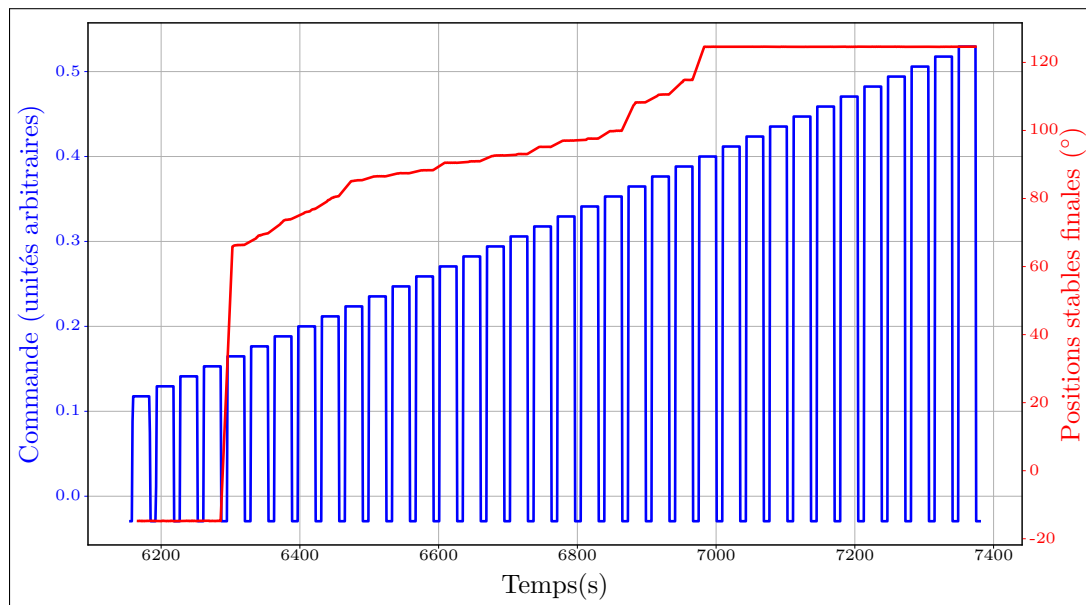


FIGURE 4.5. – *Superposition de la courbe représentant la position atteinte après stabilisation, à chaque palier du protocole (courbe rouge), avec la courbe représentant les paliers de commandes envoyées (courbe bleue). On observe bien la progression de la position stable en fonction de la progression incrémentale des commandes. On ne représente ici les positions stables que pour l'expérience en cours, pour plus de conditions représentées par rapport à la commande, au poids, etc. : voir les courbes de la partie 2.4. On ne représente ici que les positions stables terminales et non l'ensemble de la courbe du mouvement pour ces créneaux, pour une idée du motif durant un créneau voir la courbe 2.19.*

Analyse qualitative

Si le résultat d'un point de vue quantitatif est acceptable et montre que l'on est capable d'apprendre la relation entre commande et position résultante avec une erreur décroissante et une erreur moyenne en fin d'apprentissage assez basse, en réalité le résultat observé lorsque l'on tente d'utiliser cette prédiction comme commande de contrôle sur le robot est, lui, assez mauvais. Le robot est alors incapable d'atteindre une position désirée avec ce seul algorithme.

Plusieurs causes se recoupent pour expliquer ceci. Tout d'abord, même en mettant de côté les phénomènes non linéaires, une commande pour aller d'un point A à un point B, si elle est appliquée à un troisième point C ne ramène évidemment pas en B. Il est donc difficile d'utiliser ce système avec des associations simples qui viendraient guider le mouvement, car le résultat est alors très dépendant de la phase.

Ensuite, comme on l'a vu dans le chapitre précédent, notamment sur la figure 2.25, une position résultante peut être associée à plusieurs commandes dès lors que le contexte change (et vice-versa : une commande peut être associée à plusieurs positions résultantes selon le contexte). En limitant le robot au même protocole que celui décrit sur la figure 4.1, c'est-à-dire en le limitant à une unique position de départ, on peut avoir un début de résultat qualitatif, mais dès lors que celui-ci dérive un peu le résultat observé peut s'avérer assez différent et incohérent.

On pourrait pallier à une partie de ce problème en apprenant la relation entre une image de l'état passé du robot (en plus de la position actuelle) avec la commande en cours, ce qui permettrait de mieux généraliser. Néanmoins, les premiers essais utilisant un tel dispositif n'ont pas été concluants, la

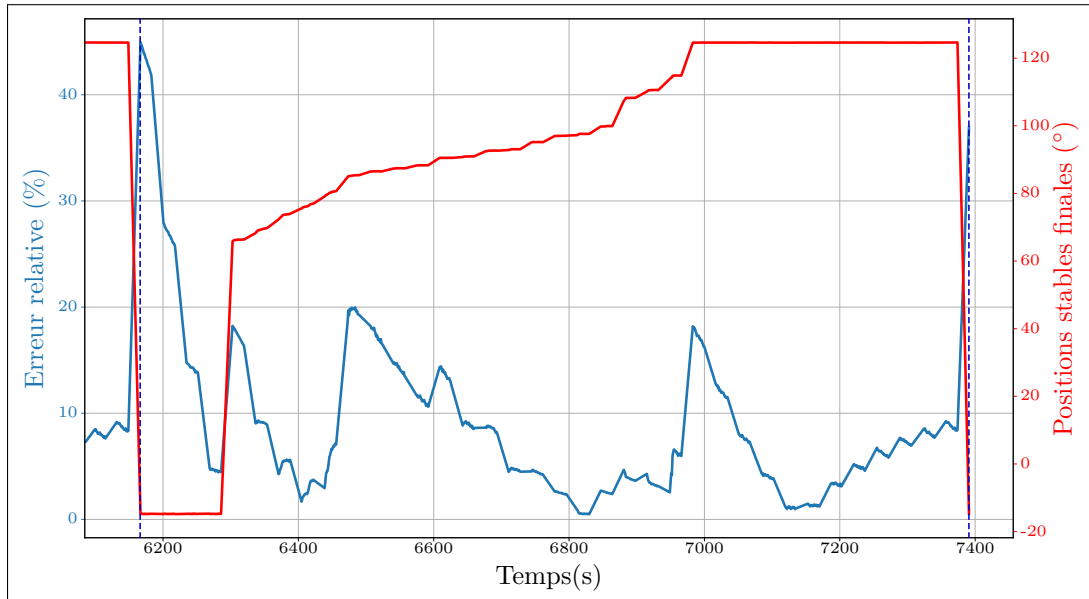


FIGURE 4.6. – Variation de l'erreur d'apprentissage au cours du temps (courbe bleue) en fin d'apprentissage et des positions stables finales pour chaque palier de commande (courbe rouge), les limites de la séquence sont matérialisées par des pointillés violets.

combinatoire de tous les états possibles des mouvements passés étant probablement trop élevée pour donner un premier résultat intéressant.

Notons que même si cela donnait de bons résultats alors la gestion de la mémoire du passé dans ce type de dispositif s'avérerait primordiale et aurait donc un fort impact sur le résultat. De plus, les phénomènes de frottement feraient également qu'une grande plage de commande pourrait être compatible avec le fait, par exemple, de rester immobile à une certaine position.

Le fait que le résultat qualitatif soit mauvais et qu'aucune voie simple n'a semblé se dégager pour le corriger nous a poussés à remettre en question notre approche du problème.

On va alors essayer d'analyser la problématique du point de vue bio-inspiré, en se ramenant au fonctionnement du vivant, plutôt qu'en tentant de profiter directement des caractéristiques émergentes de notre outil robotique. Dans les sections suivantes, nous allons donc construire le contrôle de notre robot par apprentissage en nous appuyant sur les apports de la biologie et neurobiologie, de la psychologie du développement humain et en faisant un parallèle avec certaines théories du contrôle en robotique. Néanmoins, nous n'avons pas abordé dans les sections précédentes un aspect du contrôle chez l'humain qui nous semble maintenant avoir un lien fort avec la façon de présenter les signaux au contrôleur haut niveau, et dont l'absence pourrait expliquer les problèmes rencontrés ici : c'est l'impact et l'importance des parties musculaires et de certaines boucles réflexes les contrôlant chez l'humain.

Avant donc d'aborder un nouveau modèle de contrôle sensorimoteur, nous allons donc analyser les avantages du contrôle des muscles chez l'humain puis proposer et analyser un modèle qui sera capable de mimer certaines de ces propriétés.

4.2 | Contrôle musculaire et modèle de contrôle bas niveau pour robot

Dans cette partie nous ferons un résumé des propriétés bas niveau du contrôle moteur chez l'humain, notamment des muscles et des quelques boucles de régulations principales mises en jeu dans leur contrôle. Ceci afin de faire un parallèle avec les propriétés de notre robot et les algorithmes de contrôles classiques. Nous commenterons régulièrement tout au long de cette partie l'utilisation de chacune de ces propriétés dans notre contexte. On proposera ensuite un modèle simple afin de reproduire certaines de ces propriétés et on le critiquera.

4.2.1. Le muscle seul et ses systèmes de régulations

Fibres musculaires (ou fibres extrafusales)

Les muscles sont des ensembles de fibres musculaires (ou fibres extrafusales) disposées en parallèle. Chaque fibre peut se contracter ou se relâcher (mais elles ne peuvent pas « s'étirer », c'est la contraction du muscle antagoniste qui indirectement étire les fibres du muscle agoniste) et chaque fibre a des propriétés différentes face à la fatigue et l'effort demandé. On distingue 3 types de fibres musculaires chez l'humain :

- Les fibres de type I
- Les fibres de type IIa
- Les fibres de type IIb

Les fibres de type I sont responsables du maintien de l'effort dans le muscle, elles se fatiguent lentement, tandis que les muscles de type IIa et IIb sont plutôt mis en jeu lors d'efforts intenses, rapides et brefs, mais se fatiguent plus vite. C'est l'usage de ces types de fibres différents de manière coordonnée qui permet d'obtenir des mouvements variés répondant aux impératifs de l'humain en interaction avec son environnement. Chacun de ces types de fibres est donc recruté selon des taux différents selon le mouvement ([WILMORE et COSTILL, 1998](#)), notons également que les fibres varient et adaptent leurs nombres (l'effort musculaire détruit et reconstruit régulièrement des fibres, ce qui est l'une des raisons à ces changements) durant la vie du muscle pour s'adapter au mieux au type d'effort régulièrement demandé, ce qui a été mis en évidence par l'analyse des muscles d'athlètes sportifs ([COSTILL et al., 1976](#)).

De façon encore plus surprenante il a été démontré en croisant artificiellement les connexions entre deux types de fibres, que si les signaux envoyés à une fibre sont plutôt proches de la nature d'un autre type de fibre alors la fibre qui reçoit cette demande erronée de la part du système nerveux central (SNC) change de phénotype et se transforme peu à peu dans le type de fibre approprié pour répondre au signal envoyé ([BULLER et al., 1960](#)), ainsi le phénotype de ces fibres, même individuellement et en dehors de la dynamique de leur renouvellement, n'est pas fixe et dépend du SNC.

Chacune de ces fibres est connectée à un seul motoneurone excitateur, qui fait le lien entre le système nerveux et les muscles en envoyant un signal à ces fibres musculaires en fonction des ordres arrivant sur le motoneurone. Chaque motoneurone excitateur peut par contre être connecté à plusieurs fibres musculaires. L'ensemble du groupe de fibres connectées au motoneurone plus leur motoneurone est appelé *unité motrice*. Il est à noter qu'un motoneurone est spécifique à un unique type de fibres. On a donc des motoneurones spécialisés dans les fibres I, IIa ou IIb. Pour des informations plus précises sur les unités motrices durant toute cette partie, se reporter si nécessaire à la revue des travaux du domaine : ([HECKMAN et ENOKA, 2012](#)).

Interprétation dans le cadre de la robotique Dans notre approche constructiviste bio-inspirée, nous tentons d'atteindre, comme il fut noté dans l'introduction, un certain minimalisme

« écologique ». C'est-à-dire que nous tentons de séparer ce que les processus vivants ont de spécifique, conduisant à de bonnes propriétés ou à des effets émergents intéressants du point de vue de la robotique, de ce que le vivant construit par nécessité comme processus au regard des limitations de la biologie. On peut ici se demander dans le cas des fibres musculaires si leurs différences ne sont pas de simples conséquences du fait que la nature des muscles interdit toute autre solution efficace pour à la fois maintenir un effort et à la fois produire des efforts intenses rapidement. Dans nos applications, ce type de limitation n'apparaît pas forcément, les vérins hydrauliques de notre robot étant suffisants pour combiner les deux types de propriétés.

Néanmoins, il est intéressant de noter les conséquences que ceci a sur le système de contrôle, car celui-ci a à disposition deux types d'effecteurs différents, l'un pour le maintien du mouvement et l'autre pour son lancement. Cette coupure du contrôle peut peut-être aider le SNC à construire des mouvements complexes. Notons également que la réadaptation dynamique du type des fibres musculaires selon la nature des ordres dans le temps peut être intéressante à exploiter.

Les fuseaux neuromusculaires (ou fibres intrafusales)

Les fuseaux neuromusculaires (ou « fibres intrafusales ») sont responsables, entre autres, de la proprioception musculaire chez l'humain et fournissent une information relative à leur élongation. Il s'agit en réalité biologiquement de fibres proches des fibres musculaires classiques, mais modifiées pour servir de récepteur. Ces fibres sont disposées en parallèle des fibres musculaires et apportent ainsi une information lorsqu'elles sont étirées ou raccourcies. Information qui peut ensuite être employée pour avertir le système nerveux central d'un événement imprévu (« *quelqu'un déplace mon bras* », par exemple, ou des vibrations du sol) ou pour alimenter des réflexes.

Ces fibres étant proches des fibres musculaires classiques elles ne sont actives que tendues et deviennent inactives lorsqu'elles sont totalement distendues, ce qui survient par exemple si le muscle auquel elles sont connectées est trop contracté. Ceci pose un problème pour fournir une information précise sur une grande plage de mouvement. Pour pallier à ce problème leurs extrémités, connectées à certains types de motoneurons spécifiques, peuvent se contracter ou se décontracter comme des fibres musculaires. Elles peuvent ainsi « suivre » le muscle pour rester dans tous les cas à leur maximum de sensibilité. C'est la co-activation de deux types de motoneurons, qui commandent aux fibres musculaires et aux fuseaux neuromusculaires, qui permet de faire en sorte que lorsqu'un ordre est envoyé par le système nerveux alors les fibres musculaires s'activent en même temps que le fuseau neuromusculaire, permettant de toujours conserver le fuseau sous tension et prêt à émettre une information.

Réflexe ostéotendineux (ou réflexe d'extension, ou réflexe myotatique) Un réflexe important connecte les fuseaux aux fibres musculaires, c'est *le réflexe ostéotendineux* (ou « réflexe d'extension », ou « réflexe myotatique ») qui permet, lors d'une perturbation, de compenser celle-ci en agissant directement sur le muscle sans attendre d'ordre du SNC. En clair s'il y a contraction involontaire due à un événement extérieur alors le fuseau neuromusculaire détectant cette perturbation non seulement envoie l'information de la perturbation au SNC, mais en plus demande à la fibre musculaire connectée par l'arc réflexe de compenser cette perturbation. Ceci permet une compensation très rapide des perturbations dues à l'environnement.

Ce réflexe agit également sur le muscle antagoniste en effectuant l'action inverse. Une perturbation détectée dans un sens ou dans l'autre par le muscle agoniste ou antagoniste agira ainsi de façon cohérente sur les deux muscles. Il s'agit du réflexe le plus rapide de tous les réflexes du corps humain.

Notons pour finir qu'un réflexe inverse empruntant les mêmes chemins (réflexe de flexion) existe et est connecté aux récepteurs de chaleurs, permettant le retrait rapide des membres en cas de brûlure. Nous n'aborderons pas ce réflexe en détail ici, car il sort largement du cadre de cette thèse, notre robot n'étant pas équipé de capteurs capables de mimer les propriétés tactiles et de réception de la douleur de la peau humaine.

Types de fibres intrafusales Les fibres intrafusales, comme les fibres extrafusales, sont classifiées en différentes catégories qui donnent lieu à différents types d'innervations :

- À sac nucléaire statique
- À sac nucléaire dynamique
- À chaîne nucléaire

Chacun de ces types de fibres intrafusales du fuseau neuromusculaire a des propriétés différentes, est contractile pour éviter de perdre en sensibilité et est connecté à un ensemble de nerf qui mélange les signaux de ces types de fibres. En sortie, on ne distingue dans le mélange plus que deux types de signaux, les deux types de fibres nerveuses portant ces signaux peuvent alors être classifiés comme suit :

- Les connexions de Type Ia (ou $A\alpha$) qui sont connectés à l'ensemble des types de fibre du fuseau.
- Les connexions de Type II (ou $A\beta$) qui sont connectés à l'ensemble sauf au sac nucléaire dynamique.

On considère en général que la mixture de signaux résultante fait que le type Ia porte une information à la fois statique et dynamique, c'est-à-dire qu'elle réagit à la fois temporairement à l'impulsion d'élongation de façon proportionnelle à cette impulsion et à la fois conserve un signal de base statique relatif à l'élongation courante alors que le Type II ne porte que l'information sur l'élongation courante sans prise en compte des phénomènes dynamiques (BUSER et IMBERT, 1982). La connexion spécifique entre ces deux types et les motoneurons pour le réflexe ostéotendineux n'est pas clair dans la littérature : une partie de la littérature ne considérant que le Type Ia et une autre considérant que les deux types sont connectés au neurone alpha pour le fonctionnement du réflexe, sans préciser explicitement la distribution de ces connexions selon le type.

Les motoneurons qui se connectent en entrée des fibres du fuseau neuromusculaire afin de réguler la sensibilité de ce fuseau sont également connectés différemment selon les types de ces fibres. Ils ne suivent par contre pas exactement le motif de connexion des nerfs sortants des types Ia et II. Par exemple les motoneurons « statiques » sont connectés à tous sauf aux fibres de type sac nucléaire dynamique, alors que les motoneurons « dynamiques » ne sont connectés qu'aux sacs nucléaires dynamiques. Ce qui est légèrement différent de la connexion des nerfs qui propagent le signal de sortie (il n'y a pas de motoneurone connecté à l'ensemble), mais reste porteur du même sens : une séparation statique/dynamique.

On retiendra ainsi de cet ensemble que comme pour le contrôle des fibres musculaires classiques, il y a ici une séparation entre les mécanorécepteurs dynamiques et statiques et que leurs systèmes d'adaptation respectifs, afin que les fibres du fuseau restent actives, sont également séparés afin d'ajuster différemment les parties statiques et dynamiques.

Interprétation dans le cadre de la robotique Ce réflexe s'apparente à un système de contrôle proportionnel de l'élongation des muscles, il permet au système de s'assurer que l'élongation effective est bien celle donnée en ordre par le SNC. Indirectement, ce réflexe aboutit à un genre de contrôle proportionnel de type PID (contrôle Proportionnel Intégral Dérivé) avec des composantes proportionnelles (réactions statiques) et des composantes réagissant plutôt à l'intensité de la variation temporaire (composantes dérivées).

Notons également que pour assurer une commande optimale le système nerveux doit réussir à envoyer deux valeurs aux effets identiques sur les motoneurons gérant l'élongation du fuseau et sur celles gérant les fibres musculaires afin que le fuseau soit toujours tendu et prêt à émettre des informations, quelle que soit l'élongation des muscles. La façon dont le SNC s'y prend pour gérer cette co-activation est encore sujette à débat dans la littérature biologique.

Notons toutefois que du point de vue de l'apprentissage dans les réseaux de neurones artificiels et du roboticien ceci semble peu problématique au premier abord, car le réflexe ostéotendineux rentre automatiquement en jeu afin de corriger l'erreur si le SNC échoue à fournir deux signaux aux effets adéquats, ce qui donne automatiquement une information sur la correction que le SNC doit apporter dans un sens ou l'autre afin de fournir des signaux plus adéquats à la prochaine co-activation.

Dans notre cadre on pourrait faire une analogie avec ce système et le système mécanique d'autorégulation de l'ouverture des vérins, illustré au paragraphe 2.2.1 du chapitre précédent, les deux ont en effet pour objectif d'autoréguler la force en sortie du dispositif actif sur le bras (robotique ou humain) afin qu'il y ait une correspondance avec la commande envoyée par le contrôleur central.

Réflexe myotatique inverse (ou réflexe tendineux de Golgi)

Il existe un réflexe particulier qui a son importance dans cette analyse : le réflexe myotatique inverse ou réflexe tendineux de Golgi. Il est constitué d'un capteur de force positionné sur les tendons nommé « *Organe tendineux de Golgi* » et qui détecte pour chaque fibre du tendon indépendante la force qui est appliquée sur cette fibre. Si une force trop intense est détectée, les organes tendineux de Golgi libèrent un potentiel d'action qui va inhiber la contraction de la fibre musculaire associée (via un interneurone inhibiteur). Le résultat est *une protection des fibres musculaires contre leur destruction* (GUYTON et HALL, 2006)

Il est fait l'hypothèse que cela servirait, au-delà de la fonction de protection pure, à bien répartir la charge à travers les fibres musculaires. En effet, comme les fibres sont recrutées peu à peu de façon incrémentale la charge augmente déjà de manière bien répartie entre celles-ci. Néanmoins s'il y a un signal trop fort ou mal calibré qui se retrouve envoyé à certaines fibres alors ce système de régulation supplémentaire s'assure que cela ne mène pas à la destruction de la fibre.

Comme pour les autres réflexes le système est en miroir vis-à-vis du muscle antagoniste et active donc des fibres musculaires antagonistes.

Motoneurones

Les motoneurones sont les neurones de sorties qui commandent les fibres musculaires et les fuseaux neuromusculaires vus précédemment. Les motoneurones sont nommés « *motoneurones alpha* » pour ceux qui commandent aux fibres musculaires et « *motoneurones gamma* » pour ceux qui commandent aux fuseaux neuromusculaires, la co-activation des deux en vue de donner une commande qui évite d'induire un déséquilibre et d'activer le réflexe ostéotendineux est donc nommée « *co-activation alpha-gamma* ». Il existe des motoneurones réalisant les deux types de commandes en même temps et nommés motoneurones bêta, mais ils sont le sujet de moins de travaux dans la littérature.

Les motoneurones en plus de leurs séparations alpha-gamma sont également différenciés en fonction de leurs tendances à plutôt commander aux versants statiques ou dynamiques des fibres musculaires et des fuseaux neuromusculaires. On pourrait donc parler de motoneurones « *alpha statique* », « *alpha dynamique* », « *gamma statique* » et « *gamma dynamique* » (même s'il est rare de croiser ces dénominations sous cette forme dans la littérature).

Les motoneurones « *statiques* » auront tendance à être excitables facilement et les motoneurones dynamiques plus difficilement, favorisant le recrutement en premier lieu des motoneurones statiques résistants à la fatigue, mais peu capables de faire des efforts pour ensuite recruter les motoneurones plus difficiles à activer « *dynamique* » qui serviront pour les efforts physiques (WALMSLEY et al., 1978).

À cette excitabilité de ces neurones, souvent relative à leur taille, s'ajoute leurs connexions sortantes : en général, comme on l'a noté précédemment, un motoneurone va être connecté à plusieurs fibres musculaires ou éléments du fuseau neuromusculaire, formant une unité motrice (UM) qui peut alors être de taille très variable.

Principe de la taille de Henneman Cet ensemble de conditions s'assemble et réagit à une règle qui fut énoncée par Henneman : *le principe de taille de Henneman* qui stipule que les unités motrices sont recrutées dans l'ordre en fonction de leur diamètre et de leur excitabilité afin de respecter une gradation dans l'effort.

Ce principe a été énoncé par Elwood Henneman en 1965 (HENNEMAN et al., 1965 ; MCPHEDRAN et al., 1965 ; WUERKER et al., 1965) puis largement vérifié expérimentalement (WALMSLEY et al., 1978 ; KAMEN et DE LUCA, 1989 ; DOUTRELOUX, 1992 ; CONWIT et al., 1999).

Notons qu'en plus de recruter plus d'unités motrices au fur et à mesure de la croissance de la commande, les motoneurones eux-mêmes peuvent augmenter leurs fréquences de décharge peu à peu, permettant d'augmenter la contribution de leurs fibres musculaires au fur et à mesure de la décharge (KUKULKA et CLAMANN, 1981). Ainsi, même si toutes les unités motrices sont recrutées cela permet de continuer à augmenter la force délivrée par le muscle.

Interprétation dans le cadre de la robotique Le principe de taille est intéressant, car il signifie que le système bas niveau tente de « masquer », en quelque sorte, l'aspect granulaire du contrôle bas niveau au système nerveux central. Ainsi un simple signal « analogique » croissant lancé sur l'ensemble des motoneurones recrutera de façon incrémentale les plus intéressantes en matière de résultat vis-à-vis de l'effort demandé et de leur fatigabilité afin de créer un lien linéaire entre le signal envoyé et la force effectivement déployée tout en faisant en sorte de ne pas épuiser les fibres musculaires inutilement.

De même, les capacités de réadaptation des fibres et de changements de phénotypes en fonction de la commande vont dans le même sens : camoufler la nature des détails du système bas niveau au SNC.

Du point de vue de la robotique cet ensemble de faits est pratique, car il signifie qu'il est pertinent, dans un modèle de contrôle inspiré du SNC, de faire l'approximation d'une sortie unique analogique par muscle. Ce qui nous simplifie grandement l'analogie avec la commande électrique de nos vérins.

Notons tout de même que dans le détail l'analogie est strictement fautive, car si des ensembles de motoneurones sont bien commandés indistinctement par le SNC il reste que pour un même muscle plusieurs neurones du SNC (nommés *motoneurones « supérieurs »*) commandent à plusieurs ensembles de *motoneurones « inférieurs »* (ceux connectés aux muscles : Alpha, Bêta, Gamma). Donc le masquage de la nature granulaire du signal n'est que partiel, il s'agit plutôt d'un changement de dimension. Il est difficile de trouver une réponse claire dans la littérature sur l'échelle de cette réduction de dimension. La connectivité détaillée entre les motoneurones supérieurs et inférieurs étant complexe. Le trajet entre le SNC et le muscle passe en effet par plusieurs intermédiaires, ce qui complique l'analyse.

Dans ce chapitre nous ferons l'hypothèse d'une réduction maximale de dimension, c'est-à-dire d'un signal analogique commandant l'ensemble d'un même muscle, mais nous verrons dans la partie III que dans certains cas la nature granulaire incrémentale du contrôle peut-être utile comme couche de sortie d'un algorithme d'apprentissage.

4.2.2. Le jeu des muscles agonistes-antagonistes : co-contraction (ou co-activation musculaire)

Ce que nous avons vu précédemment est le détail d'un seul muscle. Néanmoins, le contrôle d'un membre chez l'humain dans l'un des axes de l'un de ses membres passe par deux muscles qui agissent conjointement afin de fournir une force résultante. Ces muscles s'opposent l'un à l'autre. On les nomme *agoniste* et *antagoniste*. En effet, comme chaque muscle ne peut que se contracter il ne peut alors, seul, fournir une force que dans un unique sens. Il a donc besoin d'un muscle disposé en miroir sur le membre à contrôler qui générera donc la force opposée par sa propre contraction.

Ainsi pour mouvoir un membre, le SNC contracte un muscle tout en relâchant l'antagoniste, le membre se déplace alors jusqu'à un nouveau point d'équilibre.

Le point important de ce système est qu'un duo de commandes sur ces muscles donne alors un duo de contractions désirées qui finit par aboutir à un état d'équilibre dans une certaine position. Le système peut-être vu par l'analogie de deux ressorts en opposition dont on commanderait l'élongation au repos.

C'est souvent, dans la littérature, cette analogie des ressorts qui est utilisée pour proposer des modèles du muscle, notamment le plus connu et classique d'entre eux, le modèle proposé par Archibald Vivian Hill (HILL, 1938; HILL, 1949; HILL, 1950; HILL, 1970) (appelé simplement « Modèle de Hill ») qui propose la modélisation du muscle comme un élément contractile en série avec un élément passif, le tout en parallèle avec un autre élément passif par analogie avec les fibres musculaires, leurs propriétés et les tendons.

Notons un point important : un couple de forces maintenues sur l'agoniste et l'antagoniste aboutit à une position d'équilibre, néanmoins il existe plusieurs couples de force différents aboutissant à cette

même position d'équilibre. Ainsi, il est possible pour le système de contrôle de se contracter plus ou moins fortement en se maintenant à une même position d'équilibre.

Ceci aboutit donc à un système qui se maintient ou se déplace vers sa position de façon plus ou moins dure selon l'amplitude de la co-contraction. La co-contraction porte donc implicitement la notion de force et de dureté dans le mouvement.

4.2.3. Point de vue du roboticien sur les modèles de muscles

Du point de vue du modélisateur et du roboticien le jeu de co-contraction signifie que le système musculaire peut être vu comme un système de commande proportionnel classique. En effet, si l'on déplace un bras alors qu'il est soumis à une co-contraction alors celui-ci va revenir à sa position initiale par le jeu des deux éléments contractiles opposés dont les intensités de contraction sont maintenues. Notons qu'en plus, chaque muscle est soumis à une boucle interne pour conserver l'élongation imposée via le réflexe ostéotendineux, ce qui assure la stabilité de l'ensemble.

Il est donc difficile de trouver le bon niveau de modélisation, doit-on s'en tenir à une boucle de type PID pour le contrôle des bras du robot au vu des propriétés des muscles ou au contraire tenter de modéliser finement les propriétés des deux muscles, simuler leur existence physique sur le bras du robot et en déduire la commande résultante à envoyer sur les vérins du robot ?

Le débat est courant dans la littérature et même le modèle de Hill que l'on pourrait qualifier d'intermédiaire est souvent vu soit comme trop simple (incapable de transcrire réellement les propriétés intéressantes du muscle) soit comme trop complexe (inutile par rapport à un système de commande classique ou un ressort seul, car aboutissant au même comportement macroscopique dans beaucoup d'application et donc limitant et trop complexe pour modéliser de très nombreux muscles en parallèle) (WINTERS, 1990).

De notre point de vue, au vu de la difficulté d'adapter la simulation d'un modèle de muscle complexe à un système à vérins (pas forcément de correspondance a priori entre la notion de force en sortie du modèle et la commande des vérins), il est par exemple tentant de simplifier le contrôleur grâce à un modèle de type PID.

Toutefois, il existe un point qui nous semble marquant et important dans l'idée de co-contraction : c'est qu'elle porte à la fois, via deux grandeurs de mêmes unités, la position désirée et la raideur pour s'y maintenir (ou y aller) (qui correspond dans l'analogie du ressort à la constante de raideur de l'équation d'un ressort). Dans le cadre de cette thèse, le contrôle de cette raideur est en effet primordial pour l'interaction Homme-robot, car c'est elle qui permet de passer d'une interaction « souple, sécurisée, mais imprécise et peu rapide » à une interaction « dure, précise et potentiellement dangereuse » en s'adaptant selon le contexte. Si l'on utilisait une boucle de contrôle de type PID, les variables à commander pour obtenir cet effet seraient la position désirée et le « P » (terme proportionnel souvent nommé k dans les formules, similaire à la raideur du ressort), néanmoins ces variables n'évolueraient pas du tout dans le même ordre de grandeur ce qui peut poser problème en sortie d'un algorithme d'apprentissage.

Le système en co-contraction a donc cela d'élégant qu'il mêle ces deux variables dans deux grandeurs exprimant la même chose (une élongation au repos désirée du muscle), donc du même ordre et dont l'apprentissage se fera en conséquence dans les mêmes gammes de vitesse.

Dans la section suivante, on présente donc un modèle de co-contraction qui conserve l'aspect « doublet de commandes » tout en étant plus simple que celui de Hill et nous détaillerons les propriétés intéressantes de ce modèle dans notre cadre.

4.2.4. Modèle de co-contraction musculaire pour robot

Nous nous sommes inspirés des travaux de (RENVERGÉ, 2013) pour la mise en place de ce modèle. Le modèle fait une analogie entre le système de co-contraction musculaire et un système à deux ressorts en torsion. On peut voir l'évolution de cette analogie entre le système musculaire humain, un système à double ressort tel que le modèle de Hill et un système à double ressort en torsion sur la figure 4.7.

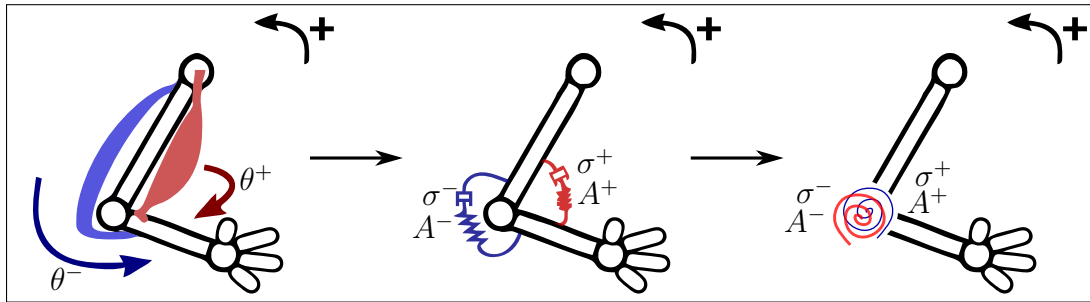


FIGURE 4.7. – Analogie entre système musculaire et différents types de modèles.

Le défaut principal de cette simplification est que les équations de deux ressorts classiques (ou du modèle de Hill) appliquées à un bras sont dépendantes de la géométrie du système alors que des ressorts en torsions appliquent directement un couple qui n'est pas dépendant de la géométrie du système, ou plutôt qui n'est dépendant que d'un unique facteur d'inertie qui résume le système. On perd donc en sens du point de vue de la modélisation puisque l'impact de la géométrie du bras disparaît. Mais c'est à la fois un inconvénient et un avantage, car cela permet également de grandement simplifier les équations du système et de ne pas avoir besoin d'y introduire les dimensions du bras.

Le modèle mis en place pour le robot est donc régi par les équations de deux ressorts en torsion mis en opposition. Une fois simplifiée (en considérant notamment que les angles sont normés entre 0 et 1) l'équation se résume par :

$$\ddot{\theta} = \frac{K}{I} \cdot (\theta_{eq} - \theta) - \frac{\sigma}{I} \cdot \dot{\theta} \quad (4.7)$$

Où θ est l'angle de l'articulation commandée et $\dot{\theta}$ la vitesse angulaire de l'articulation. $\ddot{\theta}$ quant à lui représente l'accélération angulaire. Notons aussi que $\theta_{eq} = \frac{A^+}{A^+ + A^-}$ et $K = A^+ + A^-$. Où A^+ et A^- sont les valeurs désirées de contractions pour les muscles agonistes et antagonistes.

I quant à lui est le moment d'inertie du système considéré.

On peut observer la relation évidente ici entre cette équation et un contrôleur proportionnel, à ceci près que la commande est un couple de valeur : A^+ et A^- . Notons que contrairement aux travaux dont nous nous inspirons pour ce modèle nous allons ici contrôler totalement indépendamment A^+ et A^- . Les travaux dont on s'inspire ne contrôlaient en effet que l'un des deux membres du binôme et déduisaient le second membre en faisant la supposition que la co-contraction était toujours de même force, ce qui réduit substantiellement l'intérêt d'un tel système.

Se pose ensuite la question de l'adaptation de cette équation au robot. En effet, il n'y a pas de relation évidente entre l'accélération et la commande de nos vérins comme on a pu le montrer dans la partie 2.3.3 au chapitre précédent.

Toutefois dans cette même partie on a pu montrer que sous certaines conditions il y avait une relation entre la commande et la vitesse angulaire de l'axe considéré. Pour profiter de cette approximation, on ramène donc l'équation (4.7) à :

$$\ddot{\theta} = \frac{1}{I} \cdot [(A^+ \cdot (1 - \theta) - A^- \cdot \theta) - \sigma \cdot \dot{\theta}] \quad (4.8)$$

Où l'on peut ensuite déduire $\ddot{\theta}$, $\dot{\theta}$ et θ . On applique ensuite l'approximation :

$$A \approx \alpha \cdot \dot{\theta} + B \quad (4.9)$$

Où A est la valeur de contrôle de la commande électrique sur l'axe considéré (relative à une ouverture de vanne) et α le facteur de proportionnalité. B est un facteur additif que nous fixerons à 0 par simplicité. Nous essaierons de garder le système calibré au niveau des capteurs pour qu'une commande nulle aboutisse à une vitesse nulle, ce qui nous permet d'ignorer ce facteur ici.

Ensuite A peut être déduit en utilisant les équations (4.8) et (4.9).

$$\dot{A} = \frac{1}{I} \cdot [(A^+ \cdot (1 - \theta) - A^- \cdot \theta) - \sigma \cdot A] \quad (4.10)$$

Transformable itérativement en :

$$A(t) = \frac{\Delta t}{I} \cdot [(A^+ \cdot (1 - \theta(t-1)) - A^- \cdot \theta(t-1)) - \sigma \cdot A(t-1)] + A(t-1) \quad (4.11)$$

Où Δt peut être vue comme une constante ici puisqu'en pratique l'équation est exécutée sur un système dont la période d'échantillonnage est contrôlée.

Propriétés du système de co-contraction

Le système de co-contraction a l'avantage de faire en sorte qu'à un couple de commandes donné le système se stabilise naturellement sur une position donnée, mais tout en conservant une même grandeur pour les deux variables, contrairement à un système de type PID classique où la position désirée serait dans une unité différente du facteur proportionnel.

Toutefois, le système peut également contrôler des « simili-vecteurs vitesse » puisqu'il est possible d'utiliser des couples de co-contraction dont les positions d'équilibre sont au-delà des bornes mécaniques du membre considéré, en utilisant ce type de couple le système de contrôle en amont peut ainsi déclencher un mouvement dans un sens ou dans l'autre sans se soucier de la position d'équilibre visée.

Notons une propriété très importante de ce système : comme les deux commandes sont de même nature, il est assez simple de combiner plusieurs commandes via une règle additive tout en conservant l'ensemble cohérent. Prenons quelques exemples : deux co-contractions identiques que l'on sommerait donneraient alors en résultat une co-contraction maintenant la même position d'équilibre, mais avec une force double.

De même, deux co-contractions maintenant deux positions différentes aboutissent après sommation à une position stable intermédiaire avec une force plus élevée. Et finalement, deux co-contractions identiques dont les positions stables sont éloignées des bornes mécaniques, qui vont correspondre à des « simili-vecteurs vitesse », vont donner après sommation un résultat correspondant à une dynamique de mouvement doublée.

On ne retrouve par exemple pas cette propriété sur les contrôleurs proportionnels classiques. Un couple de positions désirées associé à leurs facteurs proportionnels désirés, que l'on additionnerait à un couple identique, donnera une force certes doublée, mais également une position stable finale qui est deux fois plus éloignée que celle d'origine, car les positions désirées s'additionnent.

Cette propriété fait que la co-contraction est une grandeur très pratique à manipuler en général et, plus spécifiquement : sa faculté à produire naturellement des interpolations pertinentes lorsque l'on somme plusieurs commandes fait que dans le cadre de contrôleurs qui vont effectuer des sommations de sous-ensembles de commandes, comme c'est le cas de contrôleurs neuronaux, le fonctionnement de ces contrôleurs pourra s'en trouver grandement simplifié.

D'autres propriétés de la co-contraction sont intéressantes dans le cadre spécifique d'un apprentissage incrémental. Nous en discutons dans la section suivante.

Le système de co-contraction dans un contexte d'apprentissage

Comme nous allons utiliser ce système dans un cadre d'algorithmes bio-inspirés développementaux, on s'intéresse aux propriétés théoriques de ce système dans le cas de son interaction avec un modèle capable d'apprentissage sur les doublets de co-contractions. Dans cette section, on discutera et on illustrera donc les propriétés de ce type de système dans un cadre d'apprentissage en prenant comme hypothèse que l'on emploie des apprentissages incrémentaux classiques.

Stabilisation précoce L'une des propriétés intéressantes de ce système de co-contraction est que les deux variables commandées sont dans le même espace, ce qui favorise un apprentissage uniforme des deux variables. Il est intéressant de noter un aspect : comme on est sur un système qui cherche une position d'équilibre entre deux forces, même si l'on n'utilise dans la commande qu'une portion de ces deux forces, alors la position d'équilibre reste juste.

La conséquence de ceci est que si l'on considère un apprentissage incrémental où, pour une certaine situation, les commandes A^+ et A^- sont en cours d'apprentissage alors le résultat sur la position est très vite juste dès lors que A^+ représente une portion du A^+ final à atteindre et que A^- représente également une portion équivalente du A^- final. Ce qui est le cas ici, car l'apprentissage est uniforme et de même vitesse sur ces deux valeurs, qui représentent le même type de donnée.

Cela peut se mettre en évidence par le fait que les apprentissages utilisés classiquement emploient des équations de progression du type :

$$W = W + \lambda \cdot (S_{\text{desiré}} - S_{\text{estimé}}) \quad (4.12)$$

Si l'on ne considère qu'un seul poids pour la démonstration alors le système peut se simplifier :

$$S_{\text{estimé}} = S_{\text{estimé}} + \lambda \cdot (S_{\text{desiré}} - S_{\text{estimé}}) \quad (4.13)$$

Et comme le système commence dans notre cas à $S(0) = 0$ il devient alors évident que le premier terme de la suite s'écrit comme $S_{\text{estimé}} = \lambda \cdot S_{\text{desiré}}$ puis pour le second terme, $S = \lambda \cdot S_{\text{desiré}} + \lambda \cdot (S_{\text{desiré}} - \lambda \cdot S_{\text{desiré}}) = S_{\text{desiré}} \cdot \lambda \cdot (2 - \lambda)$ etc. Itérativement, l'ensemble des termes s'écrit ainsi de façon évidente à chaque fois comme un produit d'un facteur (croissant, représentant l'évolution de la suite et tendant vers 1) avec $S_{\text{desiré}}$. L'important ici étant que l'équation garde cette forme à chaque itération et que ce même facteur s'applique à un $S_{\text{desiré}}$ comme à une seconde opération parallèle convergeant vers $T_{\text{desiré}}$ (par exemple). On peut généraliser cette déduction à plusieurs poids et des points de départ non nuls (à ceci près que cette conclusion ne sera pas totalement juste dans les premières itérations dans le cas d'un point de départ non nul).

Dans le cas des équations en co-contraction, ce fait devient particulièrement intéressant, car la position d'équilibre atteinte par l'équation s'écrit sous la forme $\theta_{eq} = \frac{A^+}{A^+ + A^-}$. En appliquant les observations sur les équations d'apprentissages on peut alors noter qu'à chaque pas de la convergence on peut écrire que $A^+_{\text{estimé}}$ qui est en cours d'apprentissage, répond à l'équation : $A^+_{\text{estimé}} = \beta \cdot A^+_{\text{desiré}}$ et de même pour A^- , avec le même facteur bêta : $A^-_{\text{estimé}} = \beta \cdot A^-_{\text{desiré}}$. Si on souhaite observer la position d'équilibre atteinte à chaque pas de l'apprentissage, on peut alors noter :

$$\left| \begin{aligned} \theta_{\text{eq_estimé}} &= \frac{A^+_{\text{estimé}}}{A^+_{\text{estimé}} + A^-_{\text{estimé}}} \\ &= \frac{\beta \cdot A^+_{\text{desiré}}}{\beta \cdot A^+_{\text{desiré}} + \beta \cdot A^-_{\text{desiré}}} \\ &= \frac{A^+_{\text{desiré}}}{A^+_{\text{desiré}} + A^-_{\text{desiré}}} \end{aligned} \right. \quad (4.14)$$

On note donc que l'équation de cette position d'équilibre se simplifie et qu'elle est ainsi égale à la position d'équilibre désirée dès la première itération de l'apprentissage. Dans le cas d'un point de départ non nul pour A^+ ou A^- l'équation n'est plus juste dès le départ et son temps de convergence dépendra de ce point de départ. En pratique cela reste toutefois assez rapide et surtout plus rapide que l'évolution de A^+ et A^- seul. On peut l'observer en simulation pour différents couples de A^+ et A^- en début d'apprentissage sur la figure 4.8 et on trace l'évolution de l'erreur sur la figure 4.9.

Ce phénomène de stabilisation malgré un apprentissage très partiel est particulièrement positif dans le cadre d'apprentissages développementaux. En effet si on imaginait a contrario un système de type contrôle proportionnel classique avec un apprentissage fait sur la position désirée, alors celle-ci commencerait très bas pour croître au cours du temps, le résultat serait alors que les mouvements du robot commenceraient à être justes sur les positions apprises qu'à la toute fin de l'apprentissage et serait bien

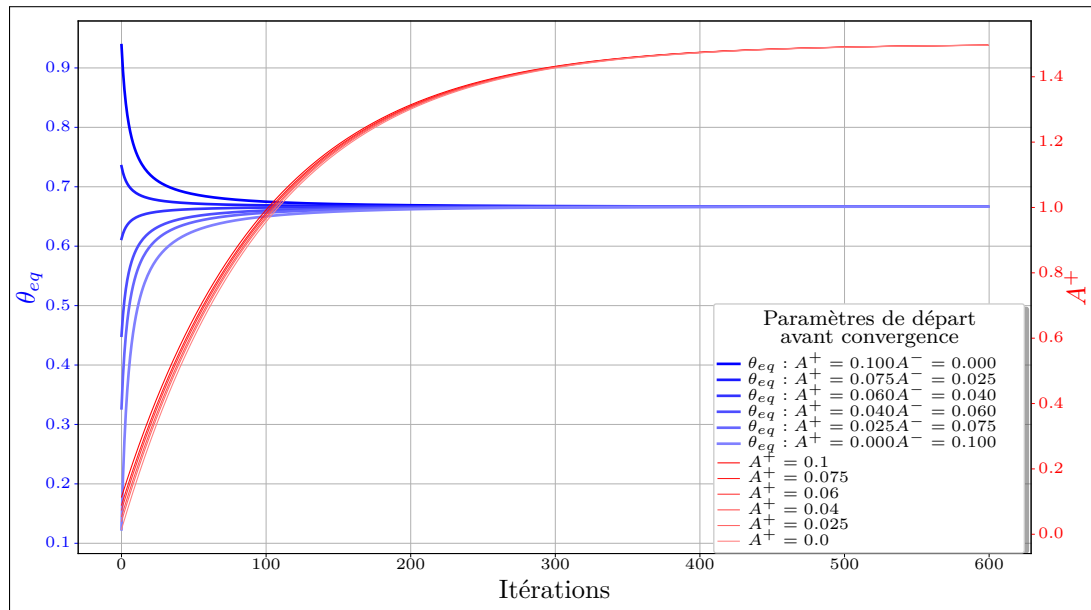


FIGURE 4.8. – Évolution de θ_{eq} et A^+ tout au long d'une convergence simulée : on peut observer l'évolution de θ_{eq} pour divers couples A^+/A^- en début d'apprentissage. Des A^+ et A^- différents de 0 en début d'apprentissage augmentent légèrement le temps de convergence de θ_{eq} , mais ce temps reste très inférieur à la convergence de A^+ et A^- seul (on ne trace que A^+). On ne trace pas le cas d'un A^+ et A^- nul car alors θ_{eq} atteint sa valeur finale en une seule itération. On trace les erreurs sur la figure 4.9.

en deçà de ces positions dans tous les autres cas. Ce point est problématique dans le cas d'un apprentissage développemental interactif où la nécessité de pouvoir interagir durant l'apprentissage implique que des comportements cohérents (même si peu précis) émergent rapidement afin de soutenir l'interaction.

Ce système en co-contraction nous permet donc d'éviter le problème en convergeant très tôt dans l'apprentissage vers un bon résultat quant aux points d'équilibres atteints.

Croissance régulière de la force La force de maintien de la co-contraction augmente logiquement de façon régulière au cours de l'apprentissage, car la force de co-contraction considérée est le fruit de l'opposition de A^+ et A^- , même si cette opposition se stabilise, comme nous l'avons vu, sur une position adéquate assez tôt elle reste néanmoins de faible force au début de l'apprentissage (par analogie avec un PID, de faible facteur proportionnel). Ceci est très avantageux, car cela signifie que dans le cas où l'on souhaite utiliser un mécanisme d'apprentissage incrémental avec ce système de co-contraction pour apprendre à un robot un mouvement « dur » qui se termine par une position maintenue avec force, alors au début de l'apprentissage on atteint bel et bien la position à cause de la stabilisation précoce, mais de façon « souple », avec une force moindre. Donc le robot agit implicitement de façon lente et sécurisée si son apprentissage est balbutiant, ce qui est une propriété intéressante du système dans un cadre d'interaction. On illustre ceci en simulation sur la figure 4.10.

Stabilisation qui conserve les bonnes propriétés de dépendances aux forces extérieures Nous avons donc vu que la co-contraction permet de se stabiliser à une position donnée avec une force donnée. Mais n'y a-t-il pas un risque de masquage excessif des phénomènes émergents intéressants vu en fin de chapitre précédent ? Par exemple un contrôleur proportionnel classique dont chaque facteur P, I et D est réglé de façon appropriée aurait tendance à masquer l'impact des frottements et forces extérieures sur le dispositif, du moins dans la position finale de stabilisation puisque le concept de

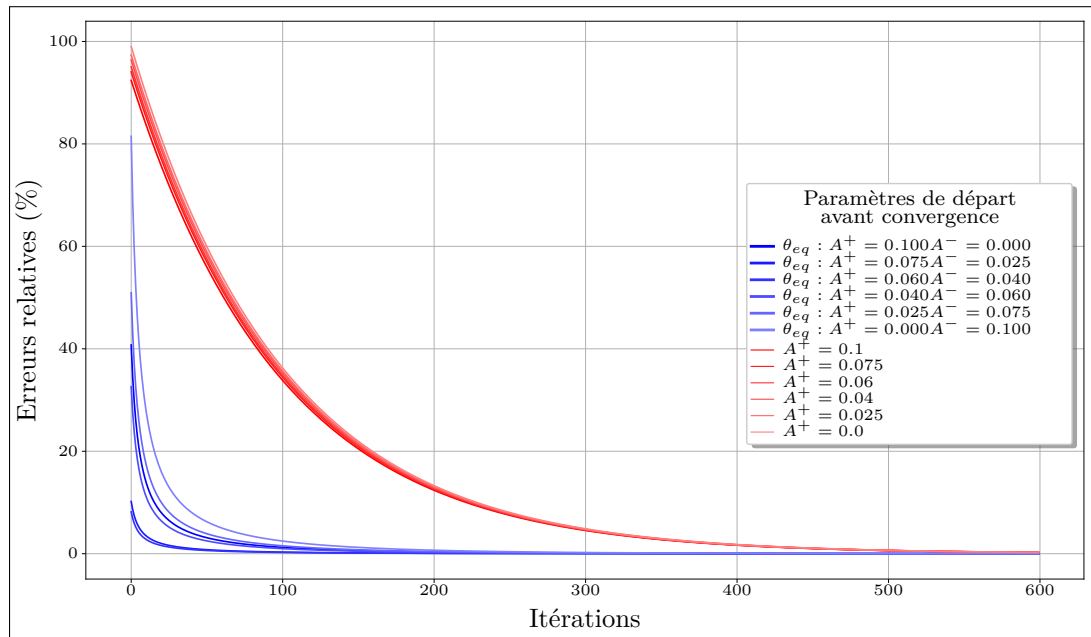


FIGURE 4.9. – Évolution de l'erreur de θ_{eq} et A^+ tout au long d'une convergence simulée. On trace les mêmes variables dans les mêmes conditions que celles présentées sur la figure 4.8. Cette fois si on présente leur erreur par rapport à la valeur vers laquelle elles convergent.

ce type de système de contrôle est justement de venir à bout de ces contraintes dans un contexte de robotique classique. Mais dans notre contexte, il s'agirait d'un défaut puisque cela interdirait aux couches hautes du système de contrôle de tirer des conclusions sur la situation au regard des nuances d'effets obtenues pour une même commande. L'idéal serait donc de conserver un effet dynamique et statique de l'environnement (poids, frottements) sur le comportement pour une même commande envoyée.

Ici, ce type de système en co-contraction s'avère être capable de conserver les effets que l'on a illustrés au chapitre précédent. L'absence de boucle d'augmentation automatique de la force en cas d'erreur (boucle d'intégration, le 'I' dans un modèle de type PID) permet de laisser les phénomènes illustrés dans la partie précédente avoir un effet observable sur le mouvement. Il reste donc possible d'utiliser ce modèle dans un apprentissage liant un contexte (poids différents portés par exemple) à une action et à l'inverse de déduire d'un mouvement et d'une commande le poids porté à ce moment-là.

Notons également que lors de l'apprentissage il y a croissance régulière de la force de co-contraction ce qui signifie que les forces au début de l'apprentissage sont faibles et donc que les effets en question sont forts quand l'apprentissage est partiel. Ce qui peut donner des propriétés exploitables intéressantes puisque cela signifie que pour un certain contexte : plus celui est nouveau plus les effets dus à l'environnement seront forts et donc plus ils pourront être pris en compte facilement durant les débuts de l'apprentissage. Ce qui est pratique si l'objectif est justement de forger un apprentissage capable de tenir compte de ces effets.

Exploration progressive implicite La stabilisation précoce vue précédemment est intéressante, mais une autre propriété importante découle de celle-ci. En effet, comme la stabilisation précoce tire parti du fait que les forces initiales s'opposant convergent vers une position d'équilibre juste, même lorsqu'elles sont faibles en début d'apprentissage alors il apparaît qu'en début d'apprentissage la position d'équilibre est juste, mais qu'elle peut être très sensible à un bruit additionnel.

En effet, si on considère que les commandes A^+ ou A^- ne sont, en début d'apprentissage, que

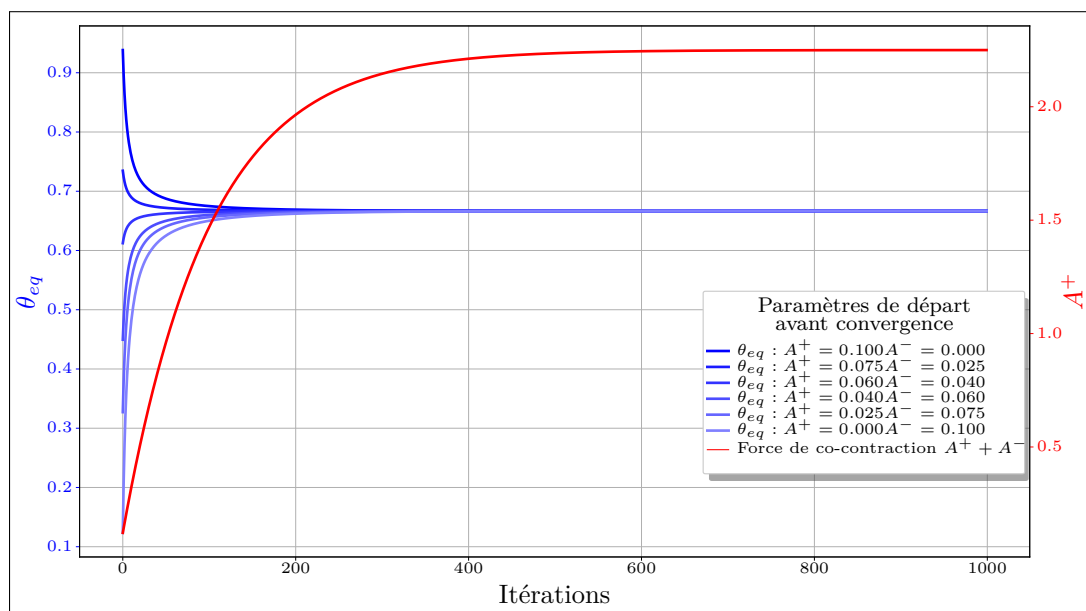


FIGURE 4.10. – Croissance de la force vis-à-vis de la croissance de θ_{eq} en fonction de A^+ et A^- différents en début d'apprentissage.

très inférieures aux valeurs désirées après la convergence, alors même un bruit faible par rapport aux valeurs désirées sera important relativement aux valeurs de A^+ ou A^- en début d'apprentissage. Ceci aboutit à une force résultante de cette erreur pouvant s'avérer importante devant la force maintenant la stabilisation vers la position souhaitée.

Ainsi, si un bruit d'amplitude constante est appliqué sur A^+ et A^- alors l'erreur moyenne commise sur la position d'équilibre au début de l'apprentissage sera grande. Par contre, et c'est là le point important, pour ce même bruit constant l'écart sur la position d'équilibre deviendra très faible en fin d'apprentissage.

On illustre ce phénomène en simulation sur la figure 4.11

On obtient alors un système dont les propriétés émergentes sont, lorsqu'il est soumis à un bruit blanc, une *exploration graduelle de l'environnement de façon variable là où l'apprentissage est faible*.

Dans la littérature moderne la plupart du temps le système nerveux central est considéré comme le responsable direct de l'exploration et de la variabilité de cette exploration, c'est l'explication qui est notamment apportée à un phénomène important que l'on peut constater chez l'humain dans lequel l'erreur durant un mouvement/une tâche est relative à ce qu'il reste à apprendre dans la zone parcourue ou dans la tâche.

Les hypothèses modernes impliquent ainsi que le SNC guide explicitement l'exploration là où elle est nécessaire c'est-à-dire là où il y a de la nouveauté ou des récompenses pour le système (IZAWA et SHADMEHR, 2011 ; HERZFELD et SHADMEHR, 2014 ; PEKONY et al., 2015 ; DHAWALE et al., 2017).

Sans remettre en question le fait que des systèmes au niveau du SNC existent et puissent effectivement tenir ce rôle, on peut tout de même constater que ce phénomène chez l'humain peut tout à fait être, comme dans ce modèle, un phénomène émergent de la progression de l'apprentissage et d'un simple bruit constant. Bien sûr, comme l'apprentissage dans ce modèle peut être vu comme ayant lieu dans le SNC alors c'est tout de même un apprentissage lié au SNC qui resterait guide de l'exploration, néanmoins cette hypothèse permet d'éviter l'ajout d'un système supplémentaire qui viendrait moduler un bruit afin d'optimiser l'exploration dans les zones les plus intéressantes. Ici le résultat observé est le même, mais est implicite au système de contrôle musculaire en co-contraction.

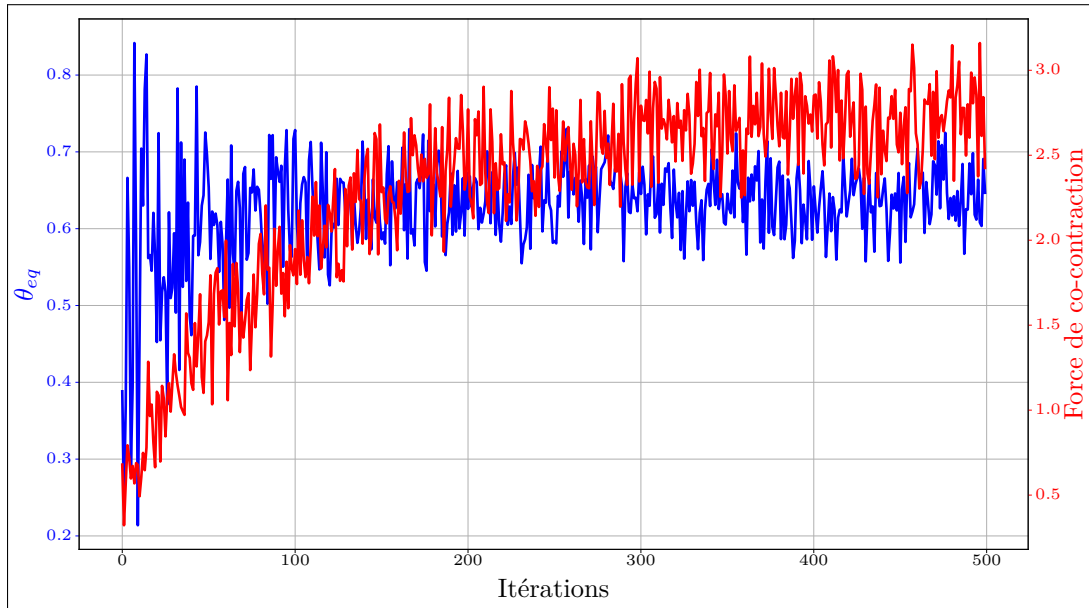


FIGURE 4.11. – *On peut simuler le comportement de l'équation (4.7) dont les deux composantes d'entrée A^+ et A^- sont soumises à un bruit blanc constant. A^+ et A^- étant croissant tout au long de l'apprentissage le bruit blanc a un fort impact sur la variation relative de ces deux termes au début de l'apprentissage (forte erreur au début), mais peu d'impact à la fin. On l'observe sur la courbe de force de co-contraction (en rouge) dont la valeur moyenne augmente, mais avec un bruit blanc d'amplitude constante tout au long de l'apprentissage. Mais du fait de la forme particulière de l'équation de θ_{eq} le résultat sur θ_{eq} est très différent avec une amplitude très importante de l'erreur au départ puis faible ensuite, notons que du fait de cette forme d'équation, même des erreurs bien plus faibles que celles représentées ici aboutissent à des erreurs importantes sur θ_{eq} en début d'apprentissage.*

4.2.5. Conclusion

Dans cette section, nous avons vu quelques bases bibliographiques du contrôle musculaire chez l'humain et les quelques boucles qui lui sont liées. Nous avons fait un parallèle entre certaines des propriétés de ces boucles et les propriétés existantes implicitement sur le robot, nous avons finalement déduit des propriétés n'existant pas nativement sur notre système lesquelles de leurs caractéristiques sont intéressantes à reproduire dans notre cas. Nous avons conclu sur une proposition de modèle musculaire simple permettant de reproduire l'une des caractéristiques manquantes principales identifiées : la co-contraction musculaire et ses propriétés. Nous avons discuté et illustré en simulations quelques-unes des propriétés intéressantes du système dans notre cadre.

Dans la section suivante, nous verrons comment exploiter ce modèle pour le contrôle robotique dans une architecture d'apprentissage robotique formant une boucle entre les perceptions du robot et l'action.

4.3 | Modèle développemental pour l'apprentissage de liens perceptions-actions

Dans le chapitre précédent nous avons vu les justifications qui ont permis de sélectionner une boucle sensorimotrice particulière pour la modélisation dans un cas de modèle visant à être bio-inspiré, développemental et applicable sur robot pour l'interaction avec un environnement réel ou un partenaire. Nous avons également proposé un modèle de muscle capable de tenir un rôle d'intermédiaire et de simplifier le problème du contrôle moteur tout en reproduisant certaines bonnes propriétés du système de contrôle bas niveau moteur de l'être humain. Dans cette section, nous proposerons une modélisation de la boucle sensorimotrice sélectionnée en interaction avec le modèle musculaire.

4.3.1. Description du modèle

En pratique, comme on l'illustre sur le schéma 3.3, l'idée générale derrière le concept est de supposer que les premières étapes du développement moteur peuvent s'effectuer par la simple construction de liens entre la perception et l'action, cela sans aucune intervention particulière de sous-systèmes supplémentaires tels que les ganglions de la base.

Ce type de liens simples sensorimoteurs a déjà donné lieu à de nombreux travaux dans le domaine qui ont su démontrer l'existence de nombreuses propriétés émergentes que ce soit sans apprentissage (BRAITENBERG, 1986) ou avec des apprentissages associatifs simples. (BACH-Y-RITA et al., 1969 ; BEER, 1995 ; CHIEL et BEER, 1997 ; GAUSSIER et ZREHEN, 1994 ; TRULLIER et al., 1997 ; GAUSSIER et ZREHEN, 1995 ; ANDRY et al., 2001 ; RAZAKARIVONY et al., 2010 ; GAUSSIER et al., 1997 ; BOUCENNA et al., 2010 ; RENGÉ et al., 2010 ; RENVERGÉ, 2013). Nous nous inspirons ici de cet ensemble de travaux et tentons de raffiner et généraliser le concept dans le cadre de la commande du robot hydraulique Tino via le modèle musculaire vu précédemment. On essaie également de rester sur une modélisation « écologique » et donc la plus minimaliste possible.

L'objectif est également de permettre à ce modèle d'être appliqué à terme sur un large ensemble de systèmes robotique en dépendant le moins possible d'informations a priori sur ces systèmes.

Traitement préalable des entrées visuelles et contrôle du cou et de la tête du robot

Pour concevoir les entrées visuelles, on base notre hypothèse sur le fait que les jeunes bébés humains ont des capacités visuelles très peu développées. Les bébés ne peuvent en effet se focaliser que sur des objets proches et dont les caractéristiques visuelles sont simples ou sur des objets en mouvement proches (ATKINSON et al., 1974 ; BANKS et SALAPATEK, 1981 ; BROWN et YAMAMOTO, 1986 ; GWIAZDA et al., 1980).

Dans le cadre de cette expérience, nous avons artificiellement simplifié ce système à une unique primitive visuelle : une couleur dont seul le mouvement/la position dans l'espace visuel est perçu. On essaie ainsi de se ramener au cas simple d'un bébé dont la catégorisation d'objet est peu développée, mais qui est tout de même capable de suivre des objets en mouvement (visage, mains de la mère, etc.). Nous discuterons de ce choix dans la partie critique finale.

L'entrée visuelle se résume à un point bougeant dans l'espace 2D du champ de vision du robot. Comme nous l'illustrerons dans la partie détaillant le protocole d'apprentissage le marqueur de couleur sera ici positionné sur l'extrémité du bras du robot en suivant l'hypothèse qu'au vu des mouvements possibles chez l'enfant et au vu de sa vision lacunaire alors l'extrémité de sa main et plus généralement toute l'extrémité de son bras constituent les principaux facteurs de mouvements induits par son propre corps dans son champ visuel (voir travaux de (PIAGET, 1945) qui se basent sur les mêmes hypothèses).

Le modèle qui résume la capture visuelle du point et sa transformation dans un espace égocentrique est résumé sur la figure 4.12. Ces informations sont ensuite utilisées pour commander le cou et les yeux du robot. Notons qu'il est inutile dans notre cadre d'utiliser la projection dans l'espace égocentrique pour

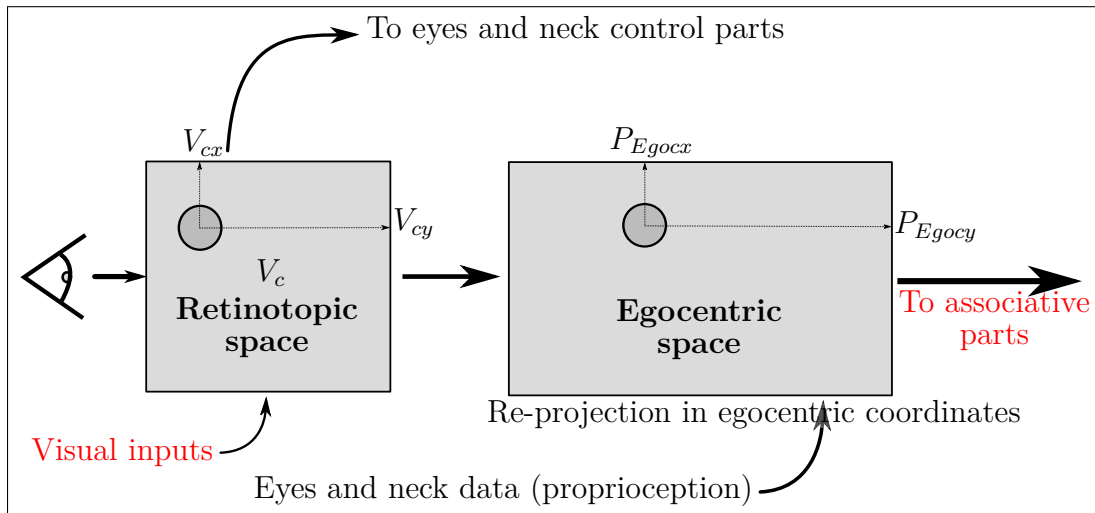


FIGURE 4.12. – *Modèle simple de traitement de la vision pour en extraire une primitive visuelle simpliste (ici couleur, mais peut-être réalisé avec une détection de mouvement). Les données initialement dans un espace 2D rétinotopiques sont utilisées pour le contrôle du cou et de la tête du robot afin de suivre l'objet. Ensuite, elles sont reprojctées dans un espace égocentrique pour simplifier les associations. L'information sur l'espace égocentrique est ensuite envoyée sur la couche associative suivante.*

contrôler le cou et les yeux. Cette dernière projection n'étant utile que pour les traitements associatifs suivants.

L'hypothèse principale ici est donc une reprojction du champ visuel rétinotopique dans un référentiel égocentré.

Les équations qui définissent cette partie du modèle sont :

$$V_c = F(V) \quad (4.15)$$

Où V est le champ visuel initial (sous forme matricielle) et V_c le champ filtré pour récupérer le marqueur de couleur. F est la fonction de filtrage.

On peut ensuite extraire V_{cx} , la position x du stimulus (V_{cx}) et V_{cy} , sa position y dans l'espace rétinotopique. On utilise ensuite P_{ex} et P_{ey} , respectivement la position en x des yeux et la position en y des yeux pour obtenir une position dans un référentiel centré sur la tête (voir l'équation (4.16)) .

$$\begin{aligned} P_{hcx} &= P_{ex} + V_{cx} \\ P_{hcy} &= P_{ey} + V_{cy} \end{aligned} \quad (4.16)$$

Le résultat est ensuite directement utilisé pour contrôler l'orientation des yeux. (voir figure 4.13)

On peut ensuite utiliser P_{nx} et P_{ny} , la position du cou en x et y , combinées avec P_{hcx} et P_{hcy} (voir l'équation (4.17)) pour obtenir une position dans un espace égocentré.

$$\begin{aligned} P_{Egocx} &= P_{nx} + P_{hcx} \\ P_{Egocy} &= P_{ny} + P_{hcy} \end{aligned} \quad (4.17)$$

Le résultat rétinotopique est ensuite utilisé pour contrôler directement l'orientation du cou (voir figure 4.13) et l'information égocentrique est, ensuite, utilisée par les parties associatives suivantes (comme on peut le voir sur la figure 4.12).

La reprojction dans un espace égocentré est soupçonnée de se produire dans le cerveau humain (BOTTINI et al., 2001 ; VALLAR et al., 1999) (voir également une revue du domaine des « head direction

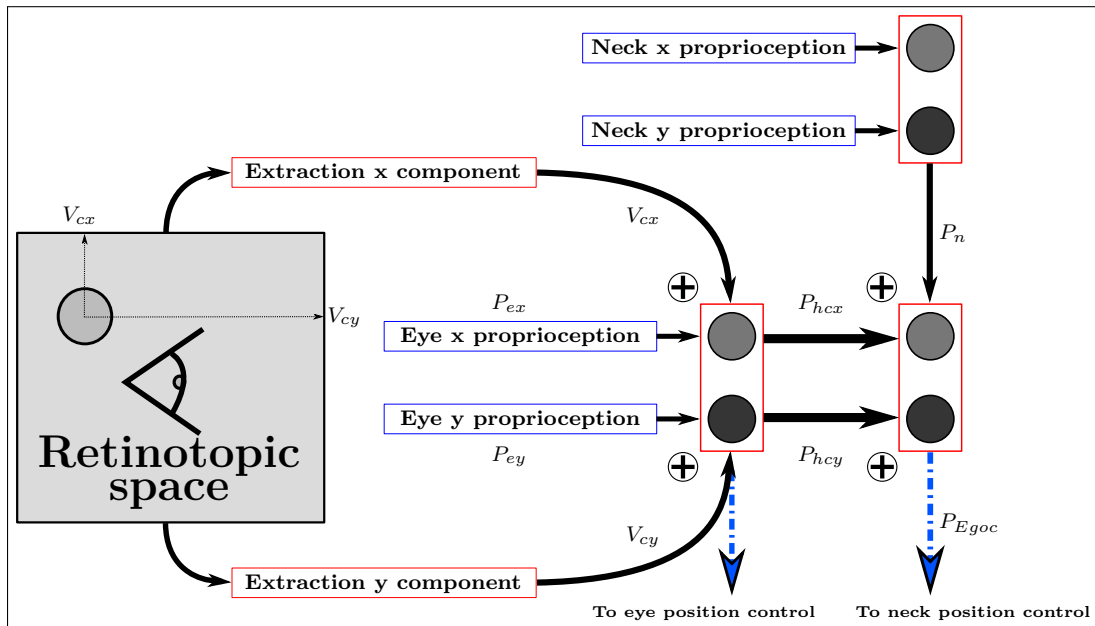


FIGURE 4.13. – Schéma du contrôle des yeux et du cou. Il n'est pas nécessaire ici d'utiliser la projection dans l'espace égocentrique. Par contre, l'espace égocentrique sera ensuite formé en s'aidant des retours des yeux et du cou. Notons qu'a priori ceci est une prise de liberté sur l'aspect bio-inspiré, car les retours sensoriels des yeux et du cou ne semblent pas impliqués, dans la littérature, avec les changements de référentiels.

cell » dans (TAUBE et al., 1996) et de travaux plus récents dans (DUDCHENKO, 2015)). Cette reconstruction pourrait aussi être réalisée directement par la catégorisation de donnée rétinienne dans un espace partageant des entrées avec des informations proprioceptives de la tête et des yeux (voir travaux de Kohonen (KOHONEN, 1982)).

Notons que nous n'avons pas détaillé ici la construction de la commande pour la tête et les yeux au-delà de ce qui a été résumé sur le schéma 4.13 car nous l'avons déjà évoqué dans le chapitre 2. En résumé, les yeux et le cou convergent vers un point unique donné en entrée dans l'espace rétinotopique, principalement en utilisant des mécanismes de type contrôleur proportionnel. Le décalage dans l'espace rétinotopique par rapport au centre de la vision est ainsi vu comme une erreur. Ce mécanisme est assez proche de la circuiterie connue dans la littérature du colliculus supérieur qui tient ce rôle (entre autres) dans la biologie humaine.

La latence des yeux et du cou combinés est faible, comme vu dans (ALFAYAD et al., 2016), et nous la considérerons dans la suite de ce chapitre, avec les autres phénomènes dynamiques du contrôle des yeux et du cou, comme négligeable.

Parties associatives et apprentissages

Il s'agit de la section principale de notre architecture perception-action. Le modèle (Figure 4.14) est séparé en quatre parties principales. Les entrées (visuelles dans un espace égocentrique et sensorielles) sont catégorisées via un système simple inspiré du K-mean.

On utilise une mesure de catégorisation pour déduire la distance des entrées aux prototypes de manière similaire au processus utilisé sur les classiques cartes de Kohonen ou sur les K-means (KOHONEN, 1982). Cette formulation en « distance » dans l'espace des entrées a déjà été démontrée comme neurobiologiquement plausible en faisant l'hypothèse d'une normalisation de la surface occupée par les synapses d'entrées d'un neurone, une mécanique déjà utilisée dans les réseaux de neurones artificiels (ARBIB,

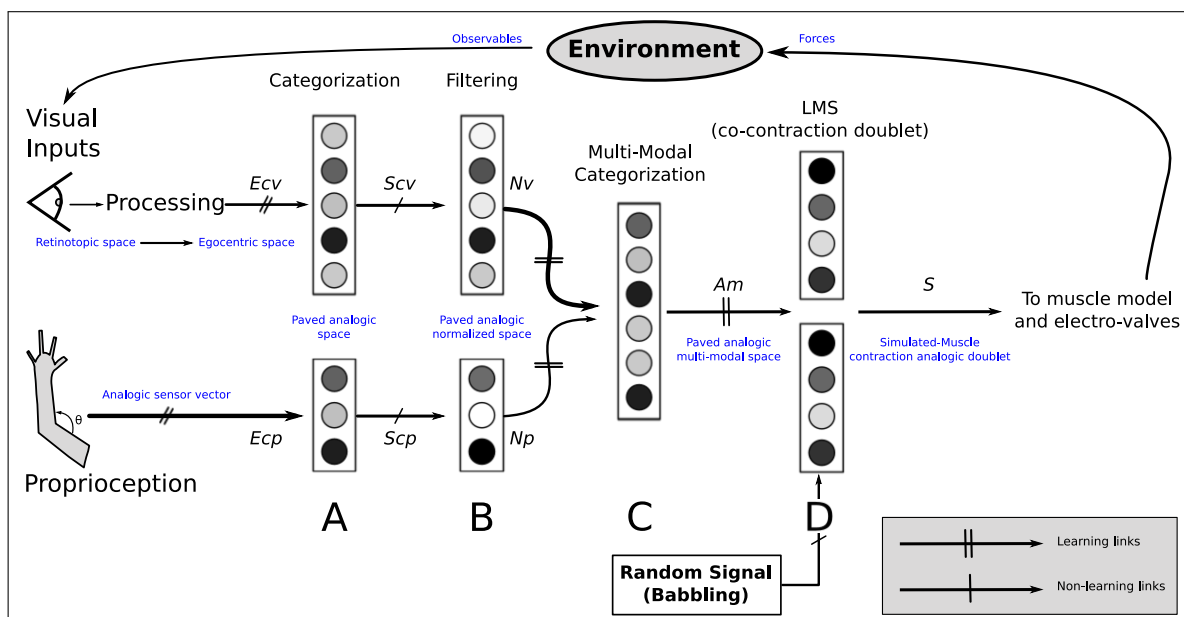


FIGURE 4.14. – Architecture de la partie apprentissage du modèle. Les couches antérieures transmettent les informations visuelles et proprioceptives et sont ensuite catégorisées indépendamment dans une couche de catégorisation (A). Cette couche fonctionne à partir d'un mécanisme de recrutement, les neurones sont recrutés quand la distance aux entrées devient trop importante. Le fonctionnement ensuite est assez similaire à un K-mean dont chaque noyau serait une sortie neuronale. On filtre ensuite les données dans la couche (B) basée sur le modèle « Instar » de Grossberg (GROSSBERG et LEVINE, 1975) qui permet une normalisation « douce » et un rehaussement de contraste à partir d'un modèle bio-inspiré. On crée une catégorisation multimodale en (C) puis les catégories sont utilisées en (D) en associant cette représentation des états au doublet de co-contractions utilisé durant l'exploration motrice (babillage moteur ou « babbling »). Dans notre cas, il y a deux neurones par articulation et comme on commande principalement 4 axes dans le cas de notre bras cela fait une sortie de 8 neurones.

2003) et observée en neurobiologie (ROYER et PARÉ, 2003).

Cette première couche de catégorisation que l'on nomme traditionnellement « SAW » (Selective Adaptive Winner) dans certains travaux précédents (BRAUD et al., 2014; GRAND et al., 2013) peut être vue comme une forme très simplifiée d'un réseau ART (CARPENTER et GROSSBERG, 1987) ou comme un réseau de type K-mean dont les sorties d'activations des noyaux seraient présentées sous la forme de neurones et dont il serait possible d'augmenter dynamiquement le nombre de noyaux aux endroits peu pavés au lieu d'attendre la convergence des noyaux adjacents.

Chaque neurone de ce réseau est, de la même manière qu'un K-mean, et contrairement aux travaux de réseaux de type Kohonen, complètement indépendant des précédents et peut être « recruté » quand aucun autre neurone n'a de réponse appropriée à une nouvelle forme présentée en entrée. Ceci permet au réseau de toujours avoir des neurones répondant suffisamment aux entrées durant une exploration de l'espace. Le mécanisme de recrutement permet de converger très rapidement vers un ensemble de noyaux représentant l'espace. Cet aspect nous permet, de façon pragmatique, une utilisation simple du réseau dans nos applications robotiques (alors qu'un réseau de type Kohonen demande souvent un grand temps de convergence). Notons qu'ici contrairement aux algorithmes de type GMM, il n'y a pas d'adaptation des noyaux. Ce qui peut être vu comme un inconvénient (les noyaux représentent moins

précisément les données) ou un avantage (les parties des noyaux qui seraient écrasées par l'adaptation et qui sont ici « non pertinentes » sont encore disponibles rapidement pour les algorithmes des couches suivantes s'il y a changement de contexte). On peut par contre empiler des couches utilisant différentes largeurs de noyaux pour fournir un large échantillon de décompositions aux couches suivantes. Dans les quelques tests effectués, cela n'a pas changé le résultat qualitatif dans notre cas d'application.

Les équations de cette couche sont implémentées ainsi :

$$Acx_i = 1.0 - e^{-\frac{1}{2} \cdot \sum_j \frac{(W_{ij} - Ecx_j)^2}{\sigma^2}} \quad (4.18)$$

Où i est l'itérateur pour la couche courante et j pour la couche d'entrée. Acx_i est l'activation du neurone i et Ecx_j est le signal de sortie du neurone j de la couche d'entrée. W_{ij} est le poids de la synapse entre le neurone j de la couche d'entrée et le neurone courant i . σ est le facteur qui détermine la largeur de la réaction du neurone comme dans une fonction de type gaussienne. x peut-être « v » ou « p » selon que l'on considère la branche de l'architecture traitant de la vision ou de la proprioception.

La sortie du neurone i est calculée ainsi :

$$Scx_i = f(Acx_i) \quad (4.19)$$

Où f est une fonction d'activation non linéaire classique définie comme suit :

$$f(Acx_i) = \begin{cases} 1 & Acx_i \geq 1 \\ Acx_i & 0 \leq Acx_i \leq 1 \\ 0 & Acx_i \leq 0 \end{cases} \quad (4.20)$$

Le processus de recrutement qui s'active quand les neurones ne répondent pas suffisamment est défini ainsi :

$$\begin{aligned} &\text{si } Scx_i < \textit{Vigilance} \quad \forall i \text{ alors} \\ &| \quad W_{(i_{max}+1)j} = Ecx_j \quad \forall j \end{aligned}$$

Où i_{max} est le dernier neurone recruté et $i_{max} + 1$ le suivant. Le recrutement est construit afin d'assurer que les équations (4.18) et (4.19) aboutissent à une valeur maximale (ici 1,0 dans notre cas) quand le recrutement a lieu pour le motif présenté à ce moment-là.

La principale différence dans cette version de l'architecture en comparaison des précédents travaux sur le sujet est que nous n'utilisons aucun filtrage binaire des sorties. Nous pensons effectivement que pour des apprentissages associatifs simples (en dehors des processus décisionnels) il est plus logique de laisser l'apprentissage s'effectuer sur l'ensemble des données.

À la place d'un système de type « Winner Take All » nous utilisons donc un système de renforcement neuronal du contraste : le système « Instar » de Stephen Grossberg. (GROSSBERG et LEVINE, 1975) (Couche B de la Figure 4.14). L'équation de fonctionnement de ce système est définie ainsi :

$$\dot{N}x_i = -A * Nx_i + (B_i - Nx_i)f(Nx_i) - Nx_i * \sum_{k \neq i} f(Nx_k) + Scx_i \quad (4.21)$$

Avec ici Nx_i la sortie du réseau de renforcement de contraste.

Ce réseau a certaines propriétés intéressantes : il permet d'aboutir sous certaines conditions à une normalisation « douce » du signal d'entrée. Mais qui conserve la signification de l'absence de signal (0), le réseau donnant alors une sortie nulle. Ceci est très important pour nous dans la couche suivante.

La couche suivante de catégorisation a ceci de particulier qu'elle va mêler deux catégorisations précédentes du système, ce qui peut, si on utilise un système classique, donner d'assez mauvais résultats, notamment avec des traitements classiques de type « distance d'entrée » qui ont tendance à se moyennner

lorsque l'on travaille avec des entrées très nombreuses pour donner un contraste faible en sortie dans ce type de cas.

L'idée ici est alors d'utiliser le fait que les précédentes couches ont été normalisées pour se ramener aux travaux vus précédemment qui montrent l'équivalence entre distance et systèmes neuronaux simples à la condition que les surfaces synaptiques d'entrées soient normées (ce qui est neurobiologiquement plausible : (ROYER et PARÉ, 2003)) et que l'entrée elle-même porte une énergie normée. On utilise ainsi une combinaison de liens neuronaux classiques dans cette couche qui vont aboutir à une réaction finale proche de la « distance », mais tout en conservant l'importance du signal « nul » et plus généralement de signaux très faibles, qui mettront en échec le système de normalisation. Ce qui nous permet de conserver la propagation naturelle d'un signal nul ou faible tout au long de notre architecture, nous allons voir par la suite que cela à son importance dans le contrôle. De façon générale, ceci permet également de s'affranchir des contraintes du faible contraste face à un très grand nombre d'entrées. Ce type d'équation donnant alors un excellent contraste sur les données.

Les équations sont définies ainsi :

$$Am_i = \sum_{\forall j} W_{ij} * Nx_j \quad (4.22)$$

Puis $Sm_i = f(Am_i)$, où f est la fonction décrite dans (4.20).

Les poids sont normalisés ainsi par analogie avec la surface constante synaptique :

$$\sum_{\forall j} W_{ij}^2 = N \quad (4.23)$$

N est ici une constante conçue pour être égale au nombre d'entrées dont la norme est sensée faire « 1 » en condition optimale. Ce type de mécanisme de catégorisation basé sur la compétition entre synapses et sur des modèles neurobiologiquement plausibles neuronaux est très débattu dans la littérature (voir (BEDNAR et CHRISTOPHER WILLIAMS, 2016) pour une revue du sujet).

Notons ici que N est un méta-paramètre venant complexifier la couche et qui est peu crédible biologiquement. Néanmoins, il est aisé d'imaginer des mécanismes capables d'aboutir au même résultat dès lors que l'on emploierait des modèles de neurones légèrement plus complexes (tels que les équations de type BCM). En effet, un mauvais N donnerait ici une sur-activation importante ou une sous-activation de certains neurones. Des modèles neuronaux capables de réguler leurs entrées en fonction de ces sur-activations ou sous-activations pour se maintenir dans une plage de fonctionnement, associées à la normalisation des surfaces synaptiques, donneraient automatiquement le même résultat sans nécessiter donc de méta-paramètres.

La notion de propagation du signal nul, qui nous a fait introduire ces mécanismes, est très utile dans nos travaux, car nous travaillons avec des concepts de forces. Ainsi on préserve le lien intuitif qui existe entre l'intensité du signal et l'intensité de la force qui sera déployée en fin d'architecture. Comme nous utilisons en fin de boucle un système en co-contraction, cela permet également de se servir du phénomène de variabilité du contrôle en fonction de la force du signal appliqué et de l'évolution de l'apprentissage, phénomène que nous avons illustré dans la partie 4.2.4. Ici un signal en entrée faible d'un point de vue des catégories, donc peu connu du point de vue du système aboutira alors à un contrôle de force faible, lent et hésitant en présence de bruit.

La couche intermédiaire, par son fonctionnement particulier, ne peut plus être soumise au même recrutement simple qui est effectué sur les couches fonctionnant en « distance ». Un recrutement en distance est en effet simple puisqu'il s'agit de retenir les valeurs du motif à conserver à un endroit (généralement sur les liens synaptiques simulés pour que ce soit imagé) puis d'en faire la comparaison ensuite. Ici, le fonctionnement de cette couche intermédiaire n'est pas relatif à une distance directement, mais fonctionne comme un neurone formel classique. Pour recruter, il s'agit donc de découvrir la combinaison de poids capable d'activer le neurone au maximum pour les entrées considérées.

Pour ce faire, on exécute une équation de type « LMS » en boucle à chaque recrutement en appliquant

à chaque fois le système de normalisation des poids vu dans l'équation (4.23) :

tant que $Sm_i \neq 1,0$ **faire**

$$\Delta W_{ij} = \epsilon * (Sd_i - S_i)$$

S'assurer que :

$$\sum_{\forall j} W_{ij}^2 = N$$

Puis calculer les sorties selon :

$$Am_i = \sum_{\forall j} W_{ij} * Nx_j$$

$$Sm_i = f(Acx_i)$$

(4.24)

Enfin, la dernière couche est une simple couche associative de type (LMS) :

$$Aa_i = \sum_{\forall j} W_{ij} * Am_j$$

(4.25)

Dont les poids synaptiques sont modifiés tels que :

$$\Delta W_{ij} = \epsilon * (Sd_i - S_i)$$

(4.26)

Où Sd_i est la sortie désirée pour le neurone i et S_i la sortie actuelle pour le neurone i . Sd est ici la paire de co-contractions appliquée durant l'exploration de l'environnement.

La sortie du neurone i est définie ainsi

$$S_i = f(Aa_i)$$

(4.27)

Où f est la même fonction d'activation non linéaire classique telle que définie dans l'équation (4.20)

L'idée ici est de considérer que les motoneurones au départ sont simplement soumis à des impulsions aléatoires, et que celles-ci aboutissent alors à un nouvel état observable qui sera catégorisé puis finalement associé en fin de boucle avec l'ordre émis. Permettant ainsi de construire peu à peu le lien entre les conséquences de sa propre action et la commande à appliquer sur le doublet de co-contractions qui est à l'origine de l'état observé.

Ici, les deux éléments de co-contractions sont explorés en même temps de façon aléatoire sans faire de supposition sur la force résultante ou ses effets, qui, combinés avec les effets émergents dus à l'architecture du robot, illustrés en fin de partie I, aboutiront à un mouvement. On espère également profiter des bonnes propriétés du contrôle en co-contraction qui aboutissent à des interpolations pertinentes dans le cas de la somme de plusieurs commandes. En effet, ici, dans le cas de la rencontre d'un état observable qui se résume à une combinaison de deux états déjà appris (appelons les A et B), alors ce sont les deux ensembles de catégories correspondants à ces états déjà appris qui vont s'activer pour représenter cet état intermédiaire. Les propriétés du système de co-contraction dont nous avons discuté dans la section 4.2.4 assurent que la somme de ces deux états donne une commande pertinente dont l'état d'équilibre sera une interpolation entre les états d'équilibre de A et B. De fait, le travail de l'apprentissage sur cette dernière couche s'en trouve grandement simplifié.

Protocole d'apprentissage

Comme nous l'avons vu, cette architecture est donc capable de construire les liens entre perceptions et actions. Pour construire ces liens de façon stable sur la longueur, le protocole est construit comme suit :

Un signal aléatoire est envoyé sur le doublet de co-contractions (chaque élément étant indépendant) de la simulation musculaire et ce signal est maintenu jusqu'à ce que la position se stabilise. On reprend ici les hypothèses de l'expérience naïve décrites dans la partie 4.1.2, c'est-à-dire que l'on fait la supposition que le temps mis à se stabiliser est négligeable devant le temps une fois stabilisé, ceci afin de privilégier une association entre points statiques. Notons toutefois qu'il est possible d'imaginer la généralisation de cette architecture à des associations dynamiques, mais ce ne sera pas le cas dans ce chapitre.

Notons également que dans les tests qualitatifs le résultat final change assez peu pour des temps de stabilisation proches du temps passé à la position stable (apprentissage rapide) ce qui tendrait à montrer que l'hypothèse vue précédemment n'a pas nécessité à être respectée de façon stricte pour obtenir un résultat qualitativement correct. Ceci peut sans doute s'expliquer par le fait que lors de l'exploration les associations entre les positions observées sur la trajectoire ne sont pas pertinentes, car de nombreuses trajectoires peuvent correspondre à un même doublet de co-contractions (tout dépend du point de départ), de fait les erreurs induites par ces points ont tendance à être moyennées pour disparaître.

Le résultat des signaux de ces signaux d'exploration, ou, pour employer un terme de la psychologie du développement de l'enfant, de « babillage moteur » (« *babbling* » en anglais), est donc un robot qui explore aléatoirement son environnement et les bornes de cet environnement. On illustre une session d'exploration du robot Tino sur la figure 4.15 ou en vidéo ¹.

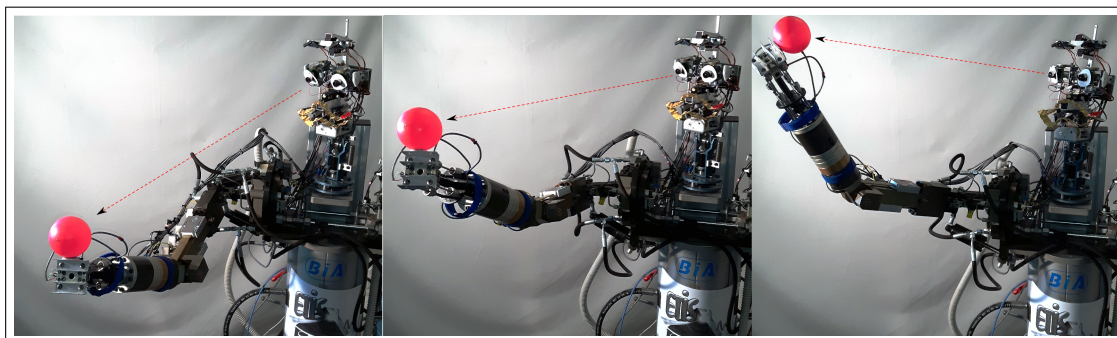


FIGURE 4.15. – *Le robot Tino explore aléatoirement l'environnement puis se stabilise quelques secondes, la tête et les yeux du robot suivent le marqueur de couleur (une balle rose ici) positionné en bout de bras du robot, l'apprentissage fonctionne pendant ce temps de façon constante sans contrôle de l'opérateur, les associations sont alors faites entre ce que le robot observe (position de sa main dans l'espace égocentrique + proprioception) et l'action réalisée, le protocole d'apprentissage peut également être observé en vidéo ¹.*

Dans ce modèle, les méta-paramètres d'apprentissage tels que le taux d'apprentissage pour la dernière couche sont définis afin de s'assurer d'un apprentissage lent capable d'accumuler de nombreuses expériences. C'est un parti pris qui nous distingue ici fortement des travaux précédents qui employaient ce genre d'algorithme (par exemple dans les travaux de (RENVERGÉ, 2013)) qui privilégiaient les apprentissages rapides « en un coup » et qui nécessitaient alors d'autres sous-systèmes pour compenser les défauts induits par des stratégies d'apprentissages rapides.

Ici, nous faisons donc l'hypothèse que la formation des primitives motrices dans le cortex moteur humain est au contraire guidée par un mécanisme lent permettant de forger des primitives fiables capables d'être utilisées par les autres sous-systèmes du cerveau.

Après apprentissage, l'exploration est arrêtée et le comportement observé est alors uniquement le résultat de l'apprentissage.

1. [↑] <http://perso-etis.ensea.fr/neurocyber/Videos/imitation/tino/>

4.3.2. Résultats et convergence de l'algorithme

Comme expliqué dans les parties précédentes le robot explore son environnement aléatoirement avec son bras. Durant cette exploration, il suit l'extrémité de son bras avec ses yeux et son cou. La position visuelle de l'objet suivi est ensuite re-projetée dans un espace égocentrique qui va permettre de faire un apprentissage visuo-moteur entre les conséquences de la commande et les ordres envoyés. Un autre point à retenir est que l'on se place dans un contexte favorable à l'association d'une position stable à un ordre stable.

Durant les premières étapes de l'expérience, ce sont les premières couches de catégorisations visuelles et sensorimotrices (notées comme « A » sur la figure 4.14) puis multimodales qui vont se développer afin de représenter l'espace d'entrée.

On peut observer la convergence de l'algorithme de cette première couche sur la courbe 4.16 pour plusieurs paramètres de vigilance (paramètre dans les équations de la partie 4.3.1).

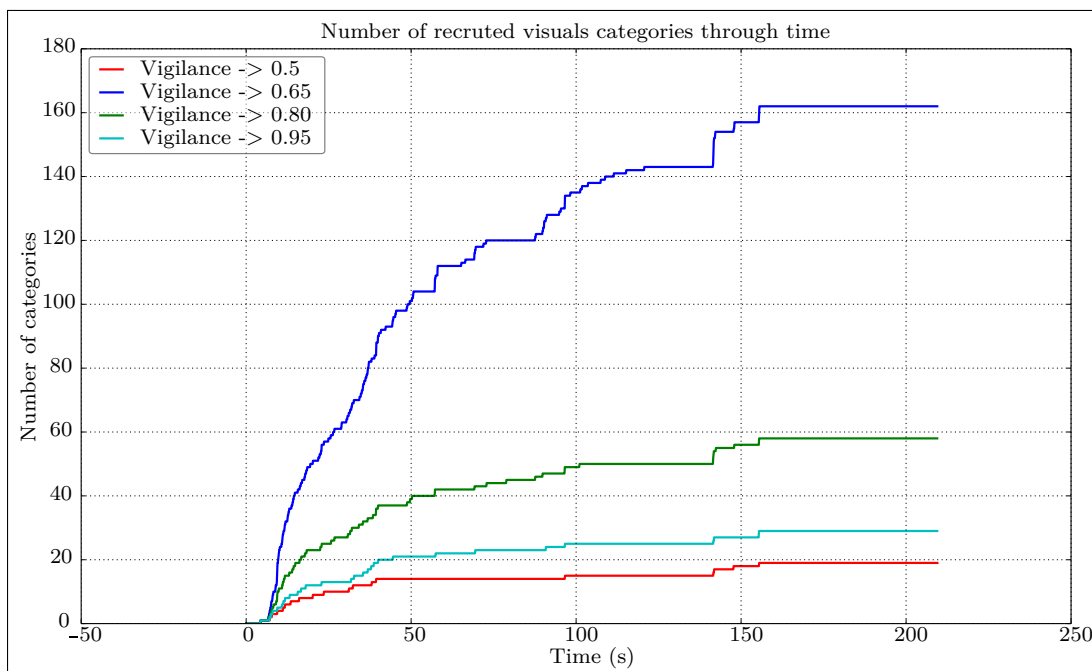


FIGURE 4.16. — Évolution du nombre de neurones recrutés en fonction du temps sur les couches de catégorisations. Durant une session d'apprentissage, on peut observer qu'après un certain temps le nombre de neurones recrutés se stabilise, ils sont alors capables de représenter l'ensemble des états proprioceptif et visuel. Les couleurs ici représentent les différentes « vigilances » utilisées pour chaque courbe (voir équations (4.5)). On remarque que des vigilances plus élevées aboutissent à de plus grands nombres de neurones recrutés, mais ne demandent pas forcément plus de temps de convergence. La vigilance peut être vue comme étant liée avec le nombre de catégories prédéfinies au lancement d'un algorithme de type K-mean.

L'importance du facteur de vigilance est discutée dans la partie « analyse et critiques ». La couche C (association multimodale) se comporte de la même façon.

L'association entre la couche multimodale et l'action à réaliser est ensuite faite sur la couche finale du réseau (noté « D » sur la figure 4.14). Ici on utilise un long apprentissage à bas taux d'apprentissage ce qui nous permet une convergence uniforme et une bonne généralisation des lois de commandes qui

vont utiliser les bonnes propriétés des noyaux de la couche C pour interpoler les lois de commandes entre différents pôles appris.

On peut observer la progression de l'erreur de reconstruction à travers le temps sur cette dernière couche sur la figure 4.17 qui est la moyenne de reconstruction des deux muscles simulés (agoniste et antagoniste) sur un unique axe (ici l'épaule dans un axe haut-bas). On peut également observer l'amplitude moyenne normée de modification des poids tout au long de l'apprentissage sur un autre axe du robot sur la figure 4.18.

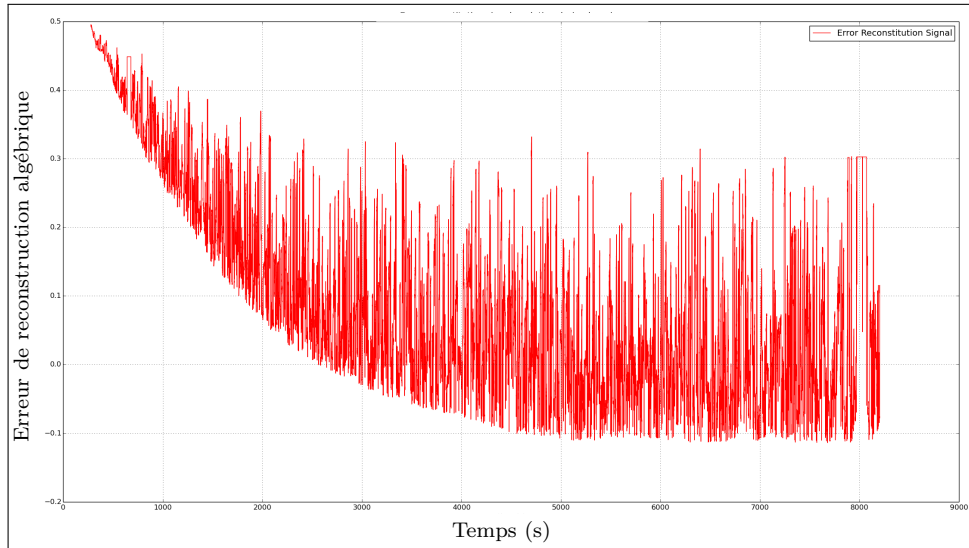


FIGURE 4.17. – *Erreur de reconstruction algébrique (il ne s'agit pas de la norme, mais de ce qui est utilisé pour l'apprentissage donc l'erreur possède un sens et donc un signe) pour un axe dans le temps (sans filtrage).*

Il est intéressant de noter la structure particulière de l'erreur sur la première figure, car si celle-ci diminue, elle reste néanmoins variable au cours de l'expérience. En réalité c'est normal, car l'apprentissage favorise ici les associations entre positions statiques et commandes statiques. De plus il n'y a ici pas suffisamment d'information pour discriminer les parties dynamiques de la courbe puisqu'à une commande correspond un ensemble de positions (la trajectoire durant le mouvement) puis la position finale. Comme l'apprentissage favorise la position finale (car le robot y reste stable plus longtemps) et que les positions intermédiaires ne sont pas pertinentes pour le lien direct avec la co-contraction (car très variable en fonction du point de départ), alors les autres cas rencontrés durant le trajet forment des erreurs fortes de reconstructions. En réalité si on devait ne considérer que les erreurs statiques il faudrait ne regarder que la ligne de base de l'erreur moyenne normée. Nous verrons que ces erreurs intermédiaires fortes n'ont que peu d'influence sur le résultat qualitatif.

On peut retracer le même type de courbe en séparant les variations et l'erreur moyenne normée, ceci pour l'ensemble des axes en moyenne. On illustre le résultat sur la figure 4.19.

On observe ainsi que toutes les couches de l'algorithme convergent au fur et à mesure de l'apprentissage. Notons toutefois que certains axes comme l'épaule convergent de façon plus optimale que d'autres. Ceci est principalement dû au fait que le robot ne dispose pas d'informations en trois dimensions sur l'espace autour de lui, nous n'utilisons ici en effet qu'un seul « œil » du robot comme sources de données visuelles. Ceci ne pose en effet pas de problème quand les mouvements des axes correspondent à l'espace visuel perçu, comme les mouvements haut-bas de l'épaule par exemple, mais le problème apparaît dans les cas où des axes jouent sur la profondeur (notamment le coude). Dans ces cas-là, le robot est incapable de distinguer deux états qui se superposent sur sa vision 2D malgré des ordres moteurs différents ce qui aboutit au fait que l'erreur reste alors élevée en fin d'apprentissage sur les axes considérés.

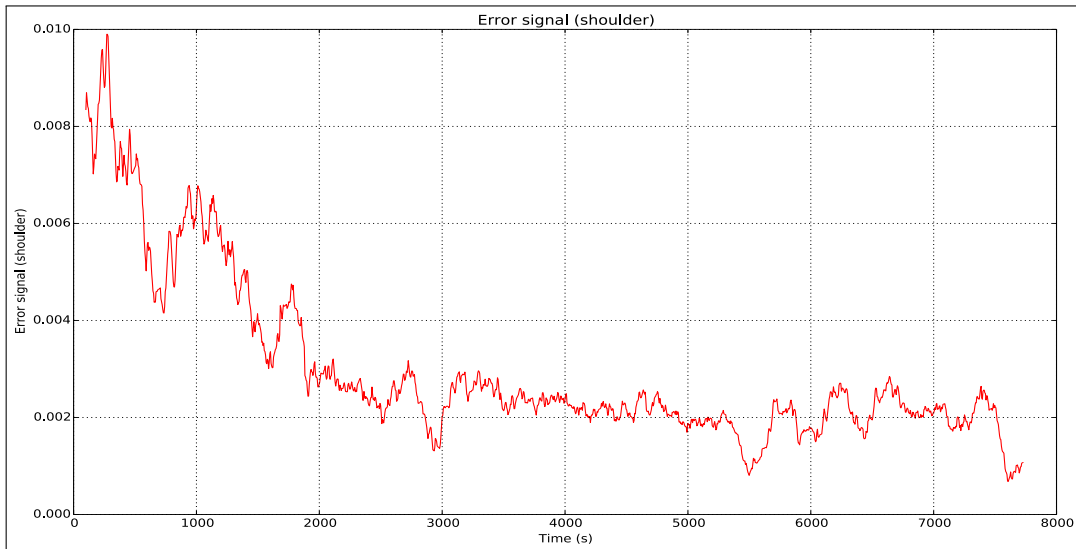


FIGURE 4.18. – *Évolution moyenne de la modification des poids de la couche associative (pour l'épaule). On observe la convergence au cours du temps. Comme l'épaule est facilement liée à l'espace visuel, on observe peu d'erreurs en fin d'apprentissage.*

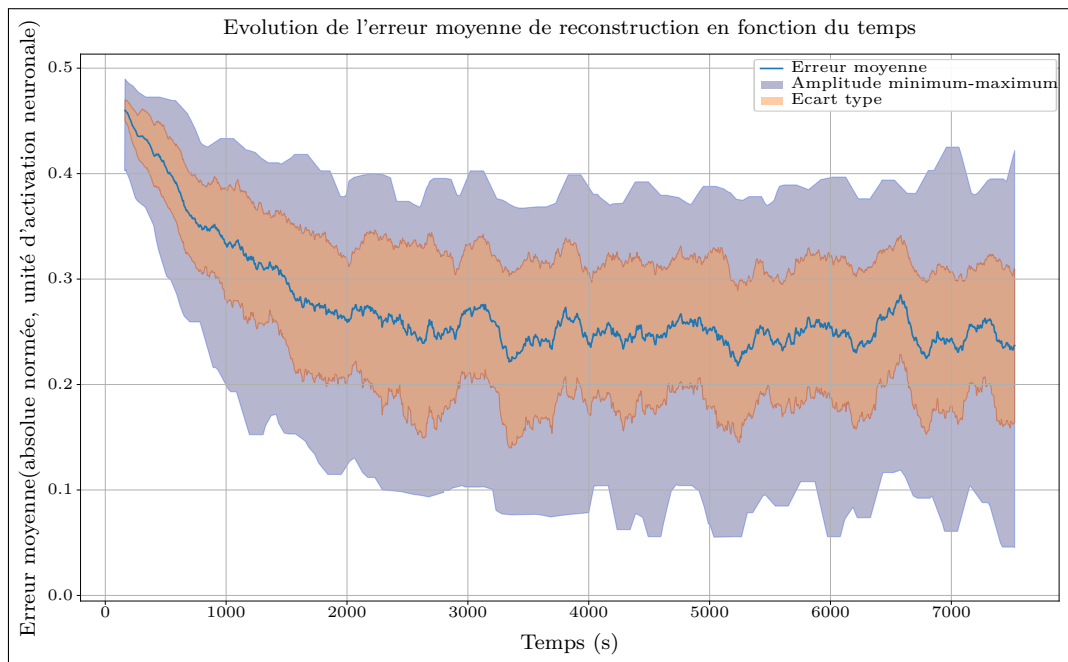


FIGURE 4.19. – *Erreur de reconstruction moyenne au cours du temps. Durant les mouvements l'erreur est élevée car le système ne peut lier des situations dynamiques (l'erreur est responsable du mouvement dans notre cas). On observe un motif particulier d'erreur doté d'une forte amplitude en fin d'apprentissage, synonyme de large amplitude de mouvement, tandis que l'erreur moyenne diminue. La borne basse représente l'erreur sur les positions stables. Les poids sont identiques au départ et ne permettent aucun mouvement, l'erreur moyenne y est donc élevée, mais l'amplitude autour de cette erreur est logiquement très faible.*

4.3.3. Imitation immédiate : effet émergent du modèle en situation d'ambiguïté perceptive

Jusqu'ici, nous avons démontré que nous sommes capables avec ce modèle de lier la position finale observée à la suite d'une action à cette action en profitant des propriétés de stabilisation pour une commande maintenue qui découlent du modèle de co-contraction musculaire et des autres effets de la mécanique du robot. Néanmoins du point de vue du comportement du robot l'algorithme se contente pour le moment *d'apprendre à rester immobile*. En effet, puisqu'à une position observée on est capable de lier la commande permettant de se maintenir à cette position alors si on laisse l'algorithme fonctionner en boucle fermée sans introduire volontairement de perturbation pour illustrer artificiellement son fonctionnement : celui-ci se contente alors de maintenir la position actuelle puisqu'il a appris, justement, la commande qui permet de s'y maintenir.

Donc on obtient finalement un robot qui n'agit pas même si l'on peut dire que ça lui a permis de construire effectivement un ensemble de commandes motrices pertinentes et réutilisables.

Mais en réalité il existe une situation où le robot agit effectivement, c'est la situation *d'ambiguïté perceptive*.

L'ambiguïté perceptive

« L'intelligence ce n'est pas ce que l'on sait,
mais ce que l'on fait quand on ne sait pas »

Jean Piaget

L'ambiguïté perceptive est le nom d'un phénomène dans lequel l'Homme ou l'architecture robotique est en situation de se tromper entre ce qu'il perçoit et ce qui est la réalité. Par exemple dans notre cas le robot a appris à associer un stimulus (ici un marqueur de couleur, mais ça pourrait être une position d'objet en mouvement) avec une commande. Ce qu'il a appris finalement c'est le modèle de commande associé à son bras et plus précisément à l'extrémité de son bras. Et comme il est dans une situation où il n'a jamais vu d'autres marqueurs du même type, mais ne représentant pas la même chose il est incapable d'introduire une nuance dans son lien entre perception et action.

On peut imaginer la chose en la formulant ainsi : le robot « pense » que tout marqueur de ce type, tout mouvement par exemple, représente son bras.

Ce qui signifie que l'on est aussi capable de le « tromper » en introduisant ensuite un second marqueur du même type. Le robot percevant le nouveau marqueur va alors « croire » que son bras est à une autre position et compensera automatiquement au vu de son apprentissage en emmenant son bras à cette position erronée perçue.

Le résultat est très intéressant, car si le marqueur erroné est un autre objet, un phénomène quelconque, une personne dans le champ visuel, on assiste alors automatiquement à un geste de *pointage* et si au contraire il s'agit de la main d'une personne en face alors on assiste à l'émergence d'un phénomène *d'imitation immédiate* (si on fait l'hypothèse que finalement, si le robot a une main proche, visuellement, de la main de l'expérimentateur, alors avec une catégorisation visuelle supplémentaire il aura plus tendance à se tromper avec la main de l'expérimentateur qu'avec un autre stimulus très différent).

Imitation et pointage chez l'enfant, deux phénomènes importants et bien étudiés du développement

L'imitation immédiate (et l'imitation motrice en général) ainsi que le pointage sont deux phénomènes connus comme apparaissant à certaines étapes du développement et comme étant des marqueurs importants de ces étapes. Le pointage est par exemple vu comme un moyen de communication implicite (RASHDAN, 2010) puisqu'il implique qu'un observateur extérieur est capable de déchiffrer ce qui constitue un point d'intérêt pour l'enfant à cet instant.

L'imitation est un phénomène qui fut aussi postulé comme majeur dans les sociétés humaines pour la transmission de compétences et dans la structure des sociétés. C'est notamment l'œuvre de Gabriel Tarde en 1890 « *Les lois de l'imitation* » (TARDE, 1890) et les travaux de Charles Darwin (1872) qui emmena la communauté scientifique à considérer l'importance de l'imitation. Ce n'est que bien plus tard que l'imitation fut analysée comme un point majeur du développement humain dans l'enfance. C'est principalement l'œuvre de Piaget (PIAGET, 1945) qui permit de caractériser nombre de ces phénomènes, de l'imitation immédiate à l'imitation différée.

Dans la vision de l'imitation de Piaget, l'imitation est un moyen de transmettre des compétences motrices à un partenaire en interaction. En effet, une fois que l'enfant (ou le robot dans notre cas) est dans une boucle d'imitation alors le partenaire peut lui « montrer » des successions de mouvements à réaliser, par exemple, ou lui transmettre des mouvements typiques utiles que l'enfant pourra alors réutiliser : c'est l'imitation « différée ». En accord avec les hypothèses de Piaget sur ce sujet on peut alors faire l'hypothèse qu'un déficit, dans ces stades, de l'imitation, peut alors aboutir à un écroulement de toute la chaîne de transmission de compétences. Il a ainsi été prouvé que des retards mentaux et certaines pathologies mentales peuvent être corrélés avec des troubles dans le stade imitation (NADEL et POTIER, 2002).

En pratique, l'imitation peut avoir de nombreuses formes selon le moment du développement et avoir lieu sur tous les effecteurs finaux de l'être humain. Les plus connus sont l'imitation au niveau des gestes des bras et l'imitation au niveau des muscles du visage (expression) ou de la langue (MELTZOFF, MOORE et al., 1977; FIELD et al., 1982; MELTZOFF et MOORE, 1983; KUGIUMUTZAKIS, 1999). L'imitation est également mis en lien avec l'empathie et plus récemment avec la présence de « neurones miroirs » (concept que nous détaillerons plus en détail dans les discussions) qui ont été eux-mêmes proposés comme étant en lien avec l'empathie (GALLESE et GOLDMAN, 1998; PRESTON et DE WAAL, 2002; DECETY, 2002). Elle apparait à partir de 1 mois et se développe jusqu'à 9 mois pour ses formes les plus simples, avant de prendre des formes plus complexes (imitation différée, réutilisation des séquences de mouvements, etc.). On passe ainsi d'un enfant « réactif » qui n'agit qu'en réaction à un partenaire, à un enfant actif capable de réutiliser les mouvements et compétences même en l'absence du partenaire. On trouvera d'amples travaux bibliographiques supplémentaires sur l'imitation immédiate, différée, et leurs conséquences dans (NADEL et BUTTERWORTH, 1996), (BUTTERWORTH, 1999) et (MARATOS, 1973).

Applications en robotique L'imitation et le pointage ont, pour ces raisons, fait l'objet de nombreux travaux en robotique et dans les applications attenantes (agents artificiels simulés). Il serait vain ici de faire une bibliographie complète de ces applications, mais notons tout de même que quelques groupes se distinguent dans cet ensemble :

- Ceux qui se basent sur un apprentissage par démonstration de l'imitation ou du pointage, généralement en lien avec des algorithmes de renforcement (par exemple (KOBBER et PETERS, 2010) pour l'imitation ou (CHAO et al., 2014) pour le pointage)
- Et ceux qui exploitent l'hypothèse de l'ambiguïté perceptive comme moteur de l'imitation sur des apprentissages associatifs simples, comme dans notre cas (GAUSSIER et ZREHEN, 1995; ANDRY et al., 2002; ANDRY et al., 2004; RENGÉRVÉ et al., 2010; BOUCENNA et al., 2010).

Nous ferons une rapide comparaison de nos travaux avec ces groupes dans la partie analyses et critiques.

Ambiguïté perceptive appliquée à l'algorithme et émergence de l'imitation immédiate et du pointage

Notre algorithme peut donc être mis dans un cadre d'ambiguïté perceptive, ce qui aboutit automatiquement à l'émergence de gestes de pointages ou d'imitations.

Ce phénomène peut être illustré sur la photo 4.20. On peut également observer le phénomène du point de vue du robot (dans un espace égocentrique) sur la figure 4.21 disponible en animation ¹ et sur

1. ↑ <https://perso-etis.ensea.fr/neurocyber/web/fr/imitation.php>

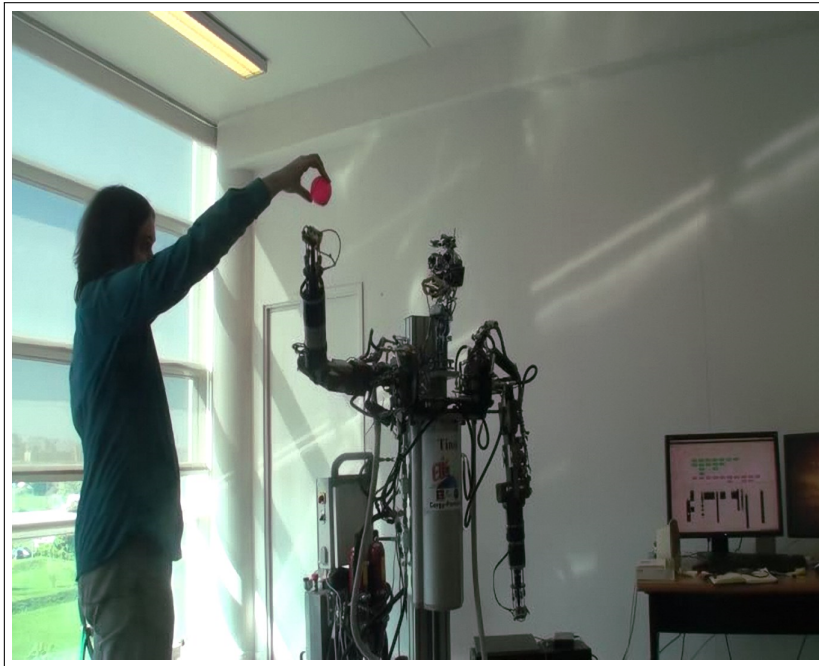


FIGURE 4.20. – *Émergence de l'imitation : en enlevant le marqueur du bras, on crée un contexte d'ambiguïté perceptive, le robot tente alors de corriger l'erreur en repositionnant son bras, on assiste à l'émergence d'un phénomène d'imitation*¹.

la figure 4.22 ou en vidéo¹. On peut alors utiliser l'imitation pour créer des situations de déséquilibres permettant de jauger des capacités de suivi du robot et donc de la justesse en pratique des primitives motrices apprises. On peut ainsi observer plus précisément la capacité du système à interpoler entre les différentes catégories sensorielles pour obtenir une commande efficace sur la figure 4.24 où l'erreur d'atteinte entre la « cible » et la position réelle atteinte est tracée (espace égocentrique du robot). Les points sur cette carte d'erreur sont les positions des noyaux de la première couche de catégorisation (A) et on se limite, comme pour la courbe 4.21 à l'affichage de ceux représentant les catégories visuelles, facilement représentables (le robot ayant 7 degrés de liberté, les points de l'espace proprioceptif sont en 7 dimensions et donc difficile à représenter simplement).

On remarque que l'erreur n'est pas particulièrement corrélée à la densité ou à la proximité des points ce qui illustre les capacités d'interpolation et de généralisation de l'algorithme. Par contre, on remarque une plus forte erreur dans les positions moyennes que dans les positions extrêmes. Ceci est dû au fait que les positions à l'extrémum de l'espace atteignable sont généralement liées à un unique couplet de commande (par exemple : pour toucher une position en haut à droite de la tête du robot, celui-ci devra tendre le bras, ce qui aboutit à un unique couplet de commande liée à cette position) alors que les positions intermédiaires peuvent être liées à plusieurs couplets de commande (différentes postures) qui ne sont pas faciles à discriminer en l'absence d'informations perceptives supplémentaires (vision 3D ou vision du bras complet par exemple) ainsi le robot ne peut que se « tromper » sur cet ensemble de position. Notons toutefois que d'un point de vue qualitatif (voir les vidéos), cette erreur reste relativement faible du point de vue de la personne en interaction. Notons également que les propriétés d'interpolation sont ici grandement facilitées grâce au système musculaire en co-contraction qui permet nativement le mélange de commande en aboutissant à une combinaison résultante qui a du

1. ↑ <http://perso-etis.ensea.fr/neurocyber/Videos/imitation/tino/>

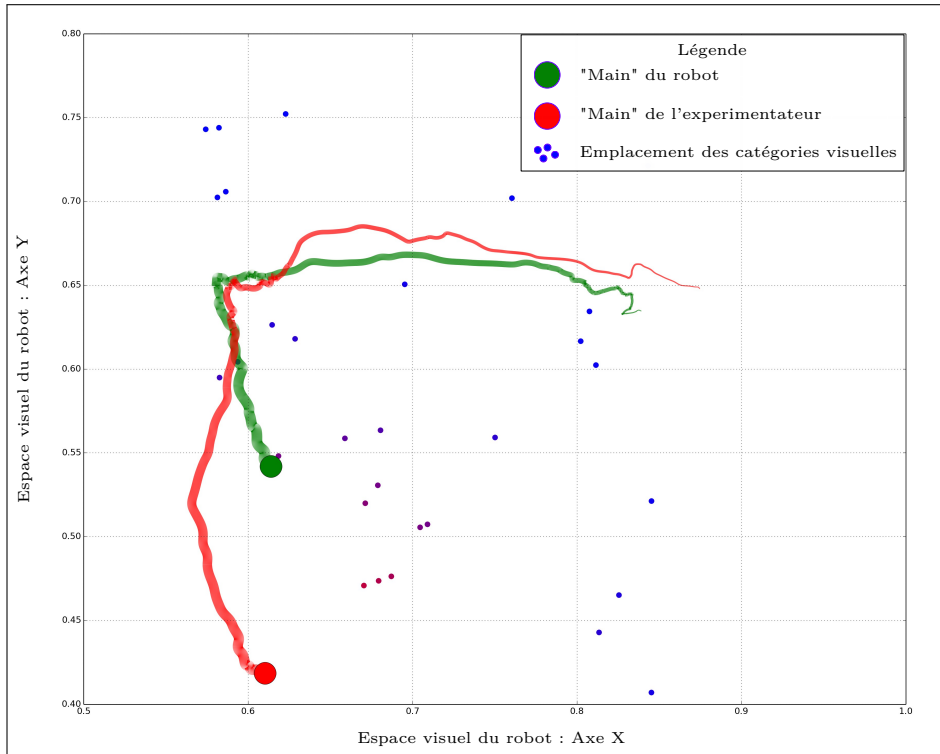


FIGURE 4.21. – *Résultat de l'imitation du point de vue du robot. Les axes représentent son espace visuel, le marqueur rouge est ce qu'il observe, sa traîne représente le déroulement temporel, le marqueur est déplacé pour introduire une ambiguïté perceptuelle. Le marqueur vert est la position réelle du bras du point de vue du robot. Il n'est utilisé que pour l'illustration et non dans l'algorithme. Les points représentent les catégories (visuelles) et leur couleur leurs activations respectives. Chacun tire la commande vers lui, ce qui aboutit à la propriété d'interpolation¹.*

sens sur le résultat final, comme nous l'avons abordé dans la section 4.2.4 ou en vidéo¹. Le résultat obtenu montre bien que nous sommes capables d'obtenir des comportements émergents d'imitations (ou de pointages) avec notre architecture de lien entre perception et action, en exploitant la notion d'ambiguïté perceptuelle.

En dehors d'un contexte d'imitation, le robot peut être mis dans un contexte où il produit des gestes de pointage. Il suffit pour cela que l'ambiguïté perceptuelle ne soit plus causée par une erreur concernant la main d'une personne en face du robot, mais par un objet quelconque, par exemple un objet en mouvement. On peut observer ce type de mise en contexte sur la suite de figure 4.23 ou en vidéo².

1. ↑ <https://perso-etis.ensea.fr/neurocyber/web/fr/imitation.php>

2. ↑ <https://www.etis.ensea.fr//neurocyber/web/en/tino.php>

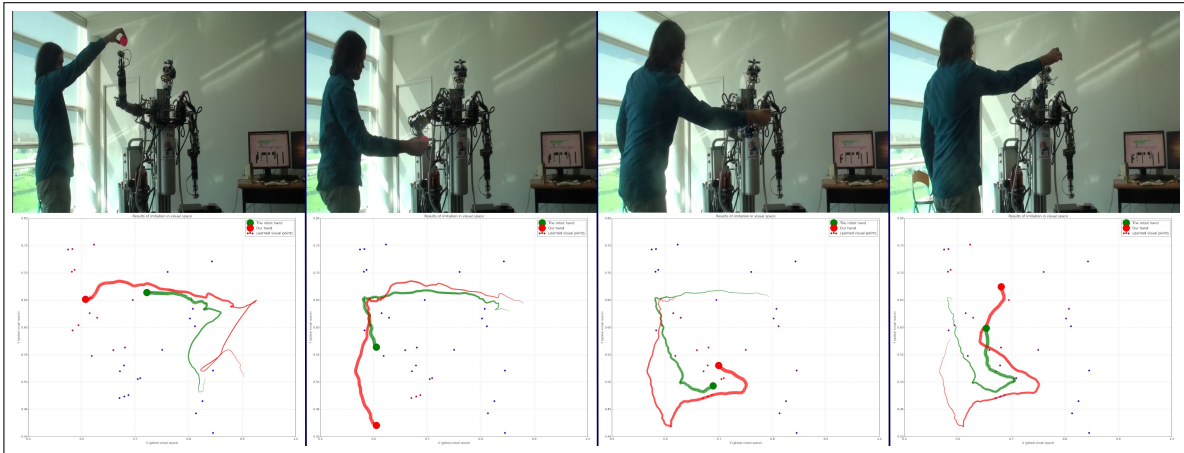


FIGURE 4.22. — On peut observer ici l'évolution de la situation du point de vue du robot et du point de vue extérieur, que nous avons abordé plus en détail sur les figures 4.21 et 4.20.



FIGURE 4.23. — Si l'ambiguïté découle d'un objet on observe l'émergence du pointage (le public pense que le robot « montre l'objet » et qu'il dispose d'un programme conçu pour). Notons l'interaction riche qui émerge du jeu d'expression et du pointage¹.

1. ↑ voir vidéo de l'expérience : <https://www.etis.ensea.fr//neurocyber/web/en/tino.php>

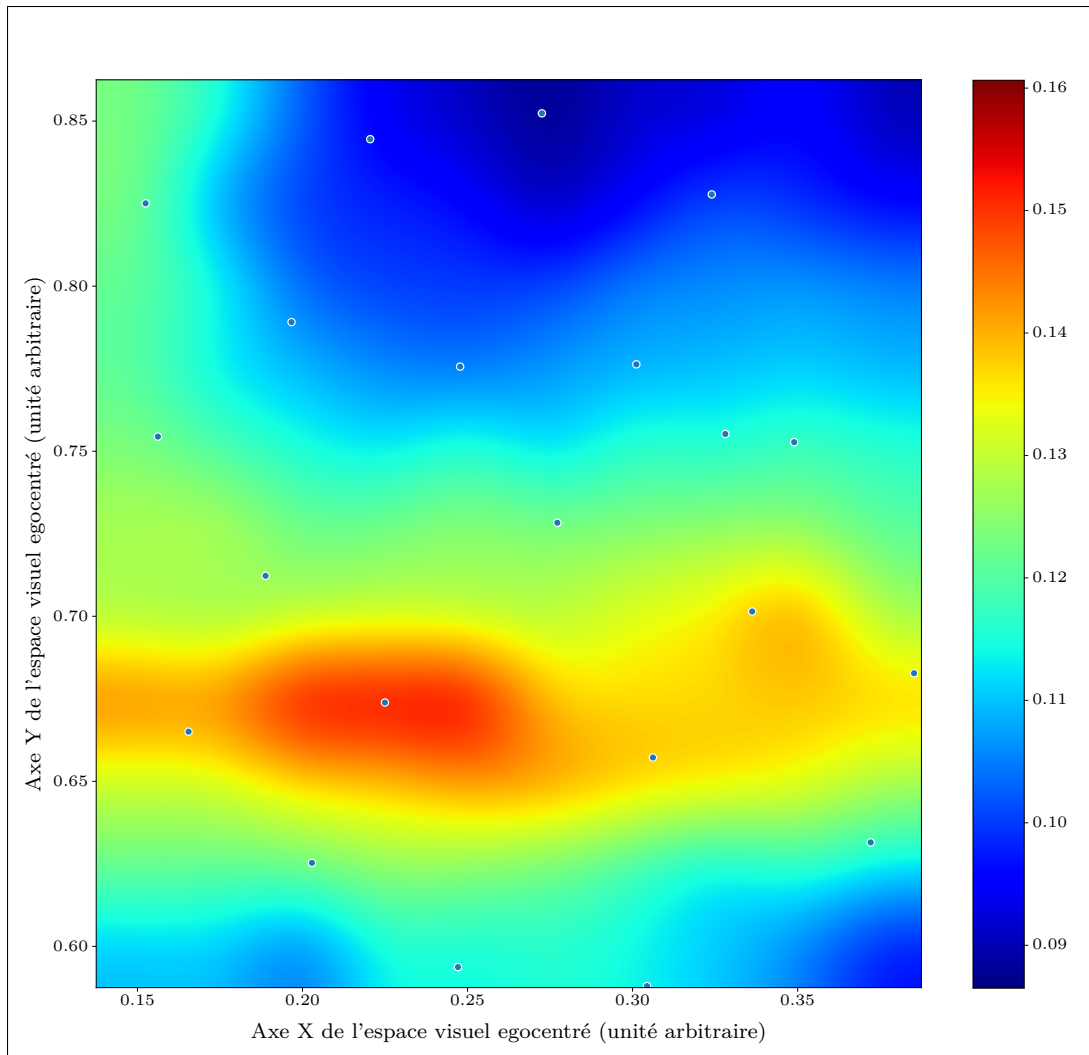


FIGURE 4.24. – *Après apprentissage, voici un pavage de l'erreur sur une portion de l'espace visuel du robot (unités arbitraires de l'espace visuel du robot). On remarque que l'erreur est surtout faible dans les zones extremums. L'erreur n'est pas relative à la position des catégories visuelles (ronds bleus sur la figure), ce qui montre que les lois de commandes ont bien été généralisées et que l'algorithme est capable d'interpolation. Ce n'est pas ici le meilleur apprentissage possible, dans cette thèse les aspects de performances pour atteindre un point précis nous intéressent assez peu nous n'avons pas cherché à optimiser les résultats dès lors que les phénomènes émergents étaient bien présents.*

4.3.4. Analyse et critiques, de l'algorithme et perspective pour une évolution du modèle

Dans cette section, nous discuterons des apports du modèle, le comparerons avec d'autres approches, nous critiquerons et discuterons de divers aspects du modèle et verrons également des pistes pour ses améliorations potentielles.

Les propriétés du modèle

Une certaine analogie avec le rôle du cortex moteur L'un des points essentiels de ce modèle dont nous avons peu discuté dans la partie résultat est que la couche intermédiaire de catégorisation tient un peu le même rôle que celui du cortex moteur chez l'humain.

Dans notre modèle, après convergence l'activation d'un neurone de cette couche intermédiaire vient activer sur la dernière couche (le LMS) une combinaison d'activation musculaire représentant une commande.

Mais du fait de l'apprentissage, toutes les catégories présentes dans cette couche se retrouvent liées à des commandes pertinentes vis-à-vis de l'espace exploré par le robot. Ce qui revient à dire que les catégories de cette couche représentent un ensemble d'actions *pertinentes pour agir dans l'espace atteignable du robot*.

Même si ce n'est pas le cas dans l'architecture présente pour des raisons de rapidité de convergence, si l'on remplace cette couche avec une architecture prenant en compte la proximité des neurones les uns envers les autres (comme un réseau de Kohonen) alors on obtient une carte organisée et formée dont les neurones ont un sens, car ne représentant que des commandes utiles à l'espace atteignable.

Finalement, cette couche tient ainsi un peu le rôle du cortex moteur chez l'humain : offrir un espace d'action utilisable par les autres parties du cerveau, et pertinent.

Ce « simili-cortex moteur » est alors réutilisable par d'autres algorithmes une fois qu'il est formé. On pourrait par exemple l'imaginer en entrée d'algorithmes de sélection de l'action comme dans la partie suivante.

Sécurité de l'interaction : croissance de la force relative à l'apprentissage et faible erreur de position en début d'apprentissage

Comme nous l'avons évoqué précédemment dans la section 4.2 le modèle de co-contraction musculaire intermédiaire a plusieurs avantages, il a notamment la propriété de donner, pour une croissance des activations durant l'apprentissage, une force faible en début d'apprentissage tout en atteignant pourtant parfaitement la position demandée très tôt dans l'apprentissage (voir, en théorie, comme nous l'avons démontré, dès l'instant 0 si les poids initiaux sont à 0). Nous avons simulé ces propriétés, mais on peut maintenant les observer sur l'architecture en pratique durant l'apprentissage. Pour simplifier la visualisation, on effectue ici un apprentissage où l'ensemble des forces de co-contraction apprises est uniforme pour chaque position. On peut observer le résultat sur la figure 4.25.

On confirme donc ici en pratique l'apport de l'algorithme de co-contraction que nous avons démontré en simulation. Cet algorithme permet en effet une « souplesse » dans le mouvement initial (et donc également une vitesse faible de mouvement) grâce à un déploiement de force faible au début de l'apprentissage, ce qui est une propriété émergente importante pour un algorithme visant à l'interaction avec des humains, car cela implique que même sans sécurité particulière l'algorithme est « souple », et donc sécurisé, dans les zones où son apprentissage est déficient. Car c'est un point que nous n'illustrons pas ici, mais qui a son importance : dans l'hypothèse d'un apprentissage non uniforme selon la zone, alors cette propriété sera variable en fonction de la zone ou, plus généralement, en fonction du contexte observé. Un nouveau contexte peu connu donnera donc automatiquement lieu à des actions « souples » dans ce contexte. Ce qui assure que l'algorithme n'est précis et « dur » que dans des contextes qu'il reconnaît parfaitement.

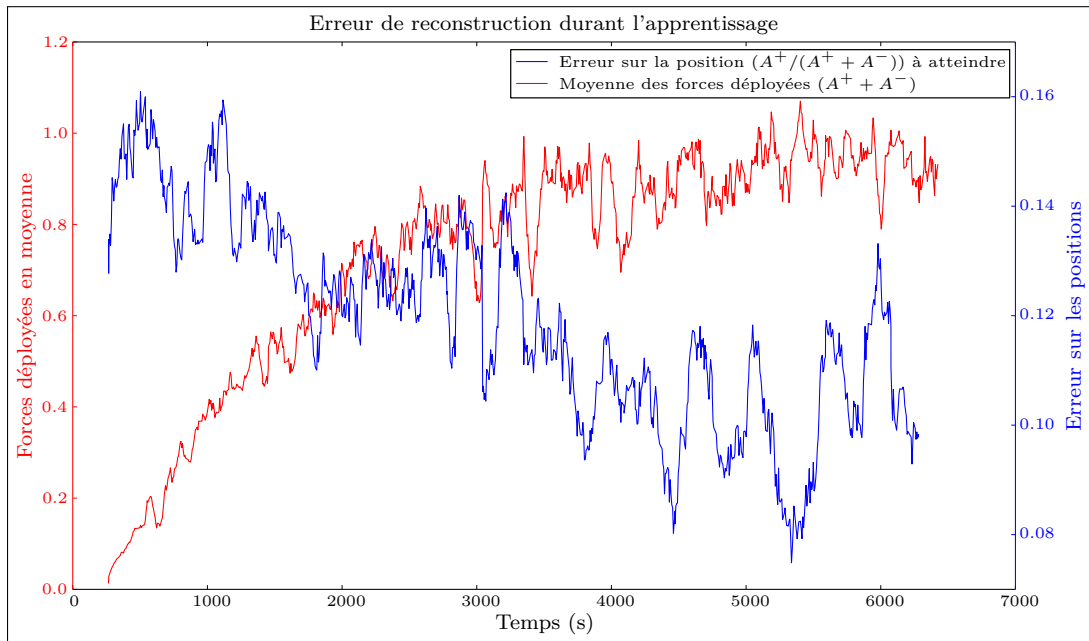


FIGURE 4.25. — On décompose ici la force et la position d'équilibre du système depuis les A^+ et A^- qui sont appris par l'architecture de liens perception-actions au cours du temps. On observe bien une erreur de position faible dès le départ (et décroissante ensuite), mais surtout une force nulle en début d'apprentissage qui tend au fur et à mesure vers la force finale (ici 1,0).

Apprentissage de positions et forces liées Un autre point important du système est de ne pas faire de distinction entre la notion de forces et de positions puisque la position maintenue et une simple simulation de deux forces s'opposant. Il est donc tout à fait possible de voir le système comme un système d'apprentissage de position ou de voir le système comme un système capable d'apprendre des mouvements liés à des forces de maintien. Même sur un robot hydraulique comme Tino ceci a des conséquences, car, comme nous l'avons détaillé dans la partie I des forces de commandes différentes donnent des résultats différents en réaction à des forces extérieures différentes, comme des poids.

L'architecture qui a été présentée ici est donc un substrat idéal à l'apprentissage de mouvements liés à des forces. Il est par exemple possible de l'employer pour lier un contexte « objet porté/position » à un doublet de commande et ainsi faire émerger du système une approximation implicite sur la force à utiliser pour porter un objet à une certaine position.

Les applications possibles sont très nombreuses, surtout dans un contexte d'interaction Homme-machine où il est important que la notion de forces mises dans l'interaction soit centrale au système.

Propriétés d'interpolations Comme nous l'avons vu sur la figure 4.24 l'erreur n'est pas particulièrement liée aux positions des catégories précédemment apprises. En effet, toute la chaîne de traitement de l'algorithme fonctionne selon un paradigme de codage en population. Dans cette chaîne de traitement, les informations analogiques dues à la vision dans un premier temps, puis aux catégories proprioceptives et visuelles sont fournies sans filtrage de type « tout ou rien » ou « *Winner take all* », les rehaussements de contraste utilisés à la place sont basés sur des modèles bio-inspirés qui conservent l'information sans la détruire tout en la rehaussant et en la filtrant dynamiquement (résistance au bruit, etc.). C'est cette information riche qui arrive ensuite au système de catégorisation multimodale, qui, par ses équations spécifiques ne fait que former des combinaisons linéaires de cette information sans la détruire ou la filtrer trop intensément.

Le résultat de ces choix dans la chaîne de traitement est que la dernière couche d'association va utiliser de nombreuses informations pour converger et va alors associer ces ensembles à des commandes. On peut imaginer le résultat comme une succession de ressorts associés à chaque « point » représentés par les catégories, qui chacune vont alors tirer l'effecteur vers leur position. Dans cette analogie, une position visuelle entre plusieurs catégories donnera alors une commande qui est *une interpolation des commandes associées à ces catégories*.

Notons qu'en plus, comme nous l'avions évoqué précédemment, le système de co-contraction qui reçoit cet ensemble de commandes mélangées possède la propriété de donner un résultat pertinent pour la somme de deux commandes (voir 4.2.4). Ce qui simplifie grandement la tâche à la dernière couche associative pour converger vers des commandes aboutissant à un comportement cohérent.

Il y a plusieurs avantages à tout ceci : Le premier étant que l'apprentissage dessine alors une carte de commande continue sans ruptures liée à la position des catégories. Ceci est bien illustré par la figure 4.24

Le second avantage est qu'un faible nombre de catégories est suffisant pour borner et paver l'espace afin de fournir une commande acceptable. On peut observer ce nombre assez faible sur la figure 4.21 et sa vidéo associée¹ mais on peut aussi pousser l'idée à l'extrême pour illustrer le comportement avec très peu de catégories sur la figure 4.26 et la vidéo disponible à la même adresse¹.

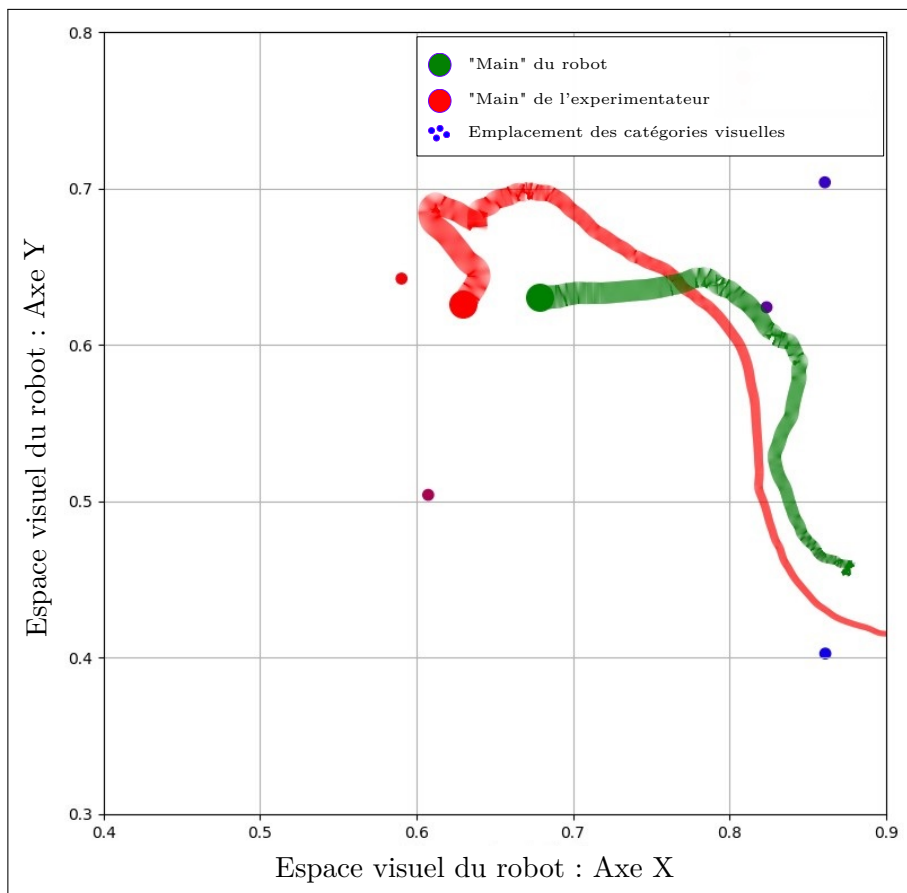


FIGURE 4.26. — On reproduit l'expérience en limitant le nombre maximum de catégories visuelles à 5 afin de tester les limites des capacités d'interpolation de l'algorithme. On peut observer le résultat en vidéo¹.

1. ↑ <https://perso-etis.ensea.fr/neurocyber/web/fr/imitation.php>

Discussion autour de quelques points critiquables

Émergence des phénomènes d'imitation et pointage et étapes du développement

Comme nous l'avons montré, le modèle permet d'apporter un début d'explication assez simple à la formation des phénomènes d'imitation et de pointage chez le jeune bébé. Le temps d'apparition assez court chez l'enfant des premiers phénomènes d'imitation moteur (au bout de 1 mois) semble de plus compatible avec la simplicité de l'apprentissage que nous avons constaté dans notre modèle simplifié. Par contre, on pourrait se poser la question de la disparition de ces effets émergents. En effet : ces effets chez l'enfant disparaissent au cours du temps pour laisser place à un contrôle sur ses actions qui peut certes encore employer l'imitation, mais dans un contexte volontaire et non plus seulement en réagissant.

Mais qu'en est-il de notre modèle ? Est-il possible d'expliquer la disparition de ces phénomènes dans le cadre de notre modèle sans ajout supplémentaire ?

En réalité, un début d'explication peut être apporté par le fait que, dans notre hypothèse, ces phénomènes existent grâce à un facteur : l'ambiguïté perceptive. Si on installait par exemple, un système visuel plus complexe capable d'affiner ses catégories visuelles au cours du temps alors ce que l'on pourrait obtenir est bien un système qui se « trompe » dans une première phase de son développement, car il est incapable de fournir des primitives distinguant son bras en mouvement d'un autre objet en mouvement, ou parce qu'il est incapable de faire la distinction entre sa perception d'une main, vague et floue, d'une autre main qui n'est pas la sienne.

Mais peu à peu, deux phénomènes vont entrer en jeu :

- Tout d'abord, l'affinage de ses caractéristiques visuelles va lui permettre de différencier l'autre de soi-même.
- Ensuite, l'affinage de l'apprentissage va lui permettre d'utiliser les informations de proprioception au fur et à mesure de ses actions dans un environnement bruyant pour filtrer le bruit et en évitant de prendre en compte des stimuli qui ne sont pas liés à la position perçue du bras. Notons tout de même que ce phénomène n'est probablement pas assez fort chez l'humain pour expliquer la totalité de la perte d'ambiguïté, car la proprioception humaine est notoirement imprécise. Il est ainsi tout à fait possible de tromper un humain adulte en lui faisant croire que sa main est à un autre endroit (des jeux de miroirs ou des positions des bras peu communes peuvent être employés à ces fins) ce qui montre que la proprioception seule peut ne pas suffire pas à lever l'ambiguïté.

Ces deux phénomènes ensemble peuvent être suffisants pour démontrer avec cet algorithme le passage d'une situation d'ambiguïté perceptive à celle de « non-ambiguïté », et donc le fonctionnement du passage d'une étape du développement à l'autre.

On peut toutefois avancer que ces phénomènes de perte d'ambiguïté par le gain de précision de l'apprentissage ne sont que partiellement mis en jeu et que d'autres phénomènes peuvent être tout aussi importants. Trois phénomènes majeurs sont à retenir :

- Une inhibition venant d'autres parties du cortex (cortex préfrontal par exemple) peut entrer en jeu pour éviter que les stimuli extérieurs ne déclenchent des actions injustifiées.
- Des phénomènes d'inhibition guidée par la récompense peuvent rentrer en jeu via les ganglions de la base au fur et à mesure que l'enfant apprend à restreindre ses mouvements face à son environnement.
- L'imitation peut aussi disparaître par l'inhibition des phénomènes connus par apprentissage. Certaines parties du cerveau, comme le cervelet, sont parfois supposées être au cœur de phénomènes de prédiction visant à inhiber les phénomènes sensoriels connus afin de laisser le système nerveux ne traiter que les informations non connues.

La proprioception dans le modèle : phénomène ralentissant, perte d'ambiguïté définitive ou « désambiguïté » favorable à l'action ? On a assez peu discuté de l'impact de la proprioception et de sa précision dans le modèle alors que celui-ci n'est pas anodin. À première vue la proprioception n'apporte en effet que des effets négatifs défavorables à l'action dans ce modèle et cela de plusieurs manières.

La première est qu'une proprioception précise implique une information précise pour faire le lien entre les sens et l'action, ce qui supprime alors toute ambiguïté, le modèle se met dans les tests à converger vers un lien parfait entre proprioception et action en négligeant les données visuelles. Cependant, ce lien parfait est insensible à toute notion d'ambiguïté perceptive et est donc impropre à l'action. Le modèle se stabilise alors sur une simple boucle de maintien en position.

Ce premier problème peut être résolu en s'inspirant de la biologie : la proprioception chez l'humain est en effet peu précise, et donc pour introduire cette mauvaise précision on introduit à dessein un flou sur les catégories proprioceptives et un facteur qui fait que la proprioception est moins prise en compte que la vision du point de vue de la couche d'intégration multimodale (comme on peut le voir schématiquement sur la figure 4.14 via la différence de largeur des liens).

On peut d'ailleurs se poser la question : la proprioception de mauvaise qualité chez l'humain ne l'est-elle pas « à dessein » d'un point de vue évolutionnaire pour justement favoriser la création de liens entre les stimuli extérieurs et les actions dans les premières étapes du développement et ainsi favoriser un développement orienté vers l'interaction avec l'environnement en négligeant les informations venant du corps sans liens avec l'extérieur ? Nous laissons cette question sans réponse, car elle dépasse largement le cadre de cette thèse, mais nous pensons qu'elle vaut sans doute la peine d'être étudiée dans des travaux supplémentaires.

Cette proprioception peu précise dans notre modèle peut être vue comme un facteur négatif, même si elle permet maintenant le mouvement. En effet, si on imagine les catégories comme des ressorts qui tirent, en quelque sorte, l'action vers ce qu'elles représentent alors la proprioception dans notre modèle agit comme un ralentisseur, car elle porte finalement en elle une absence d'ambiguïté, puisque l'ambiguïté n'est que visuelle. Dans nos tests qualitatifs, il est en effet apparu qu'une absence complète de proprioception donnait des mouvements plus rapides et que la largeur du lien entre cette proprioception et l'espace multimodal était directement liée à la lenteur du mouvement.

On peut alors se poser la question de l'utilité de la proprioception dans l'effet émergent d'imitation et de pointage.

En réalité la proprioception porte un autre phénomène, bien plus positif celui-là et qui vient compenser ce défaut : c'est qu'elle permet de préciser certaines situations d'ambiguïté et de rendre le mouvement plus juste en l'absence d'informations visuelles adéquates. Prenons un exemple pour clarifier ceci : dans le cas où le marqueur de la main est mis en mouvement pour créer une situation « ambiguë ». On le bouge alors d'une position X/Y à une seconde position X_1/Y_1 . Que se passe-t-il exactement dans l'algorithme ? Le nouvel emplacement du marqueur est catégorisé et donne lieu à une action. Mais il peut y avoir de nombreuses façons d'atteindre la position X_1/Y_1 avec les degrés de liberté du robot, et ces façons, uniquement avec les données visuelles, ne sont pas séparables.

L'apprentissage va donc en sélectionner une, généralement en moyennant tous les cas ayant été rencontrés. Le résultat est qu'il n'existe après apprentissage qu'une unique posture pour atteindre X_1/Y_1 si on ne considère que l'aspect visuel.

Ceci peut donner lieu à des mouvements non pertinents et bien peu humains, car le robot peut alors se mettre à bouger énormément ses différents axes pour que son extrémité de bras passe de X/Y à X_1/Y_1 selon que la posture en question soit proche ou non d'une posture de X/Y .

Non seulement ceci pose problème d'un point de vue pragmatique et du point de vue de la comparaison avec les mouvements humains, et donc un problème dans l'interaction si l'on postule qu'une certaine similarité dans les mouvements permet une meilleure résonance entre robot et humain ; mais il faut aussi noter que cela aboutit au fait que l'ensemble de positions possibles en X_1/Y_1 n'existe alors pas dans les catégories d'actions qui forment notre « simili-cortex moteur », car seule une de ces postures existe et est bien liée à une action menant à X_1/Y_1 ce qui rend cet espace moteur assez pauvre si l'objectif est de construire un espace moteur capable d'être réutilisé par d'autres algorithmes.

Mais que se passe-t-il en présence d'une information proprioceptive ? Dans ce cas, la couche multimodale peut catégoriser des ensembles différents pour toutes les positions existantes en X_1/Y_1 puisque, si elles ne diffèrent toujours pas par les informations visuelles, elles se distinguent par contre par l'information proprioceptive. Ces ensembles vont pouvoir être liés à des actions différentes menant en différentes postures pour X_1/Y_1 , donc le problème de la présence de ces actions sur notre simili-cortex moteur est résolu.

Qu'en est-il du phénomène émergent d'imitation et du mouvement qui en découle dans ce cadre ? Nous sommes donc toujours à notre position de départ X/Y et le marqueur ambigu est déplacé à une position X_1/Y_1 . Les catégories qui vont réagir à cette situation sont alors les catégories qui sont le plus en commun avec une position de l'extrémité du bras en X_1/Y_1 et à la position actuelle proprioceptive. Le résultat c'est que la posture finale du robot qui va être sélectionnée sera une position qui mènera l'extrémité du bras en, X_1/Y_1 mais qui sera aussi le plus proche possible de la position qu'avaient les axes en X/Y . Le résultat est donc que l'on sélectionne automatiquement et implicitement la position en X_1/Y_1 qui demandera le moins de changement depuis celle que l'on avait en X/Y .

Ce résultat implicite est lourd de conséquences :

- Tout d'abord, il devient possible, sans information visuelle supplémentaire, d'agir avec des positions très différentes du bras selon le contexte en fonction de ce qui est le plus pertinent.
- Le « simili-cortex moteur » utilise ce jeu de position et le met donc à disposition pour d'autres systèmes qui voudraient l'utiliser en activant des zones de cet ensemble de catégories.
- Cela signifie également que le robot fait nativement des mouvements cohérents et minimalistes sans avoir besoin de mécanismes de minimisation de l'énergie. Ici cette minimisation est implicite.
- Ce dernier point entraîne que les mouvements sont aussi plus « humain » et donc plus aptes à être reconnus comme tels par une personne interagissant avec le robot. Ce qui est un point essentiel pour l'interaction.

La proprioception est donc primordiale dans cette architecture pour que l'ensemble de ces phénomènes apparaissent.

On peut néanmoins se poser la question maintenant de la force ou du flou de cette proprioception. Dans cette thèse nous avons fixé cette force et ce flou de façon arbitraire pour éviter d'un côté d'avoir une situation où son impact sur les catégories serait nul et d'un autre côté une situation où cet impact serait trop fort et trop précis, ce qui aboutirait à une disparition d'ambiguïté totale.

Néanmoins au vu de l'importance du phénomène nous pensons ici qu'il y aurait matière à mener des études supplémentaires sur ces paramètres de flou et de forces. Mais également sur la nature de ce flou en faisant des parallèles avec la proprioception du muscle chez l'humain.

Cas plus réaliste d'ambiguïté perceptive Les cas que nous avons abordés pour l'émergence du pointage ou de l'imitation sont des cas très simplifiés dans lesquels on a supposé que le marqueur représentant la main du robot était retiré pour faire place à un nouveau marqueur (la main de l'humain ou un objet en mouvement). Cependant dans un contexte plus réaliste le robot devrait voir et sa main et le marqueur ambigu : on peut alors se poser la question de l'impact de ce double marqueur sur le phénomène émergent.

En réalité, les phénomènes émergents présentés ici survivent à des expériences avec deux marqueurs (que nous appellerons marqueurs 1 et 2), mais ne présentent plus la même dynamique. Si l'on reprend l'analogie des ressorts, tout se passe comme si chaque marqueur dans l'espace visuel tirait la commande dans son sens. D'un point de vue plus mathématique, la combinaison des deux stimuli visuels activent des emplacements de catégories différents qui se retrouvent normés et donc moins activés que si un seul ensemble de catégories avait été activé, le résultat est une activation des deux catégories multimodales associées au doublet « position ressentie + position marqueur 1 » et au doublet « position ressentie + position marqueur 2 », mais plus faiblement que lorsqu'elles sont seules.

Les deux étant liées à des commandes en co-contraction dans l'espace d'action on obtient donc une combinaison linéaire de ces deux commandes. Ce qui revient finalement à dire que le robot tente de

« viser » un endroit qui n'est plus le marqueur, mais un marqueur virtuel intermédiaire positionné entre les marqueurs 1 et 2. La dynamique du mouvement est alors logiquement plus lente.

Bien entendu, une fois que le bras bouge, ce marqueur visuel se déplace également et donc le bras finit par atteindre réellement le marqueur final ambigu, néanmoins il aura mis plus de temps.

Il faut également noter que cette dynamique peut énormément changer selon le système de tracking visuel existant et le système de vision implémenté. Si le robot regarde un marqueur lointain du marqueur présent sur sa main alors il n'aura plus le marqueur de sa main dans son champ visuel, la dynamique du mouvement au départ sera alors rapide, puis plus lente quand le marqueur revient dans le champ visuel.

De même, si le système de vision implémente une répartition de bâtonnets et cônes simulés proche de ce que l'on trouve chez l'humain, avec un système de cellules ganglionnaires, alors l'intensité d'un stimulus variera selon ce que le robot regarde : le marqueur sur la main ou le marqueur ambigu. Si le robot fixe le marqueur ambigu, alors celui sur sa main, en bordure de champ visuel par exemple se retrouvera moins représenté dans un espace égocentrique hypothétique (neurones répondants de façon plus floue et étant moins nombreux). Le résultat est une dynamique proche de celle avec un unique marqueur durant la majeure partie du trajet, mais plus lente à la fin quand les deux marqueurs se retrouvent dans la même zone visuelle.

On a implémenté durant cette thèse un système analogue utilisant un modèle des liens entre cellules bipolaires et ganglionnaires afin d'introduire un « flou » sur les données en bordure du champ visuel du robot. L'effet était qualitativement celui attendu puis ce système n'a plus été réanalysé. Cela mériterait sans doute une analyse quantitative plus conséquente.

En conclusion : la présence de deux marqueurs dans un système simple peut aboutir à une dynamique plus lente, mais sans changer la nature du mouvement, par contre dans un système plus complexe la dynamique peut être inchangée en fonction des conditions ou varier en fonction de la direction du regard du robot à ce moment-là.

Lien potentiel avec d'autres phénomènes présents dans le développement de l'enfant

L'effet Simon « L'effet Simon » (SIMON et WOLF, 1963), (ou « *Simon effect* » en anglais) est un effet qui montre qu'un mouvement guidé par un stimulus visuel est toujours plus rapide dans la partie du champ visuel « cohérente » que dans la partie incohérente. En clair : si l'on demande à un sujet de faire un mouvement vers la droite au déclenchement d'un stimulus présent sur la gauche alors son mouvement sera lent. À l'inverse si l'on demande à un sujet de faire le même mouvement pour un stimulus présent sur la droite alors son mouvement sera rapide. Cet effet se généralise à d'autres modalités, même quand celles-ci n'ont aucun rapport avec la tâche. Si l'on demande à une personne de faire des mouvements vers un stimulus qui peut être à gauche ou à droite, alors ces mouvements seront plus rapides si un son est entendu dans le même sens que le stimulus que dans le sens contraire.

L'algorithme que nous avons présenté pourrait apporter avec quelques changements un embryon de démonstration à l'effet Simon. En effet, si notre hypothèse sur la formation du cortex est justifiée, alors celui-ci se développerait comme dans notre algorithme en commençant par établir des liens entre les stimuli créés par l'enfant (ou le robot) et les commandes déclenchant ces stimuli. De fait, les premiers liens construits sont des liens orientés : si un stimulus est présent à gauche le robot en déduit qu'il s'agit d'un stimulus qui est forcément de son fait et corrige cette erreur en agissant à gauche.

Une fois l'ambiguïté perceptive disparut et des phénomènes d'inhibition du mouvement construit dans le cerveau de l'enfant on peut imaginer qu'une partie de ces liens persistent dans cette relation topologique : la personne devient donc capable de mouvements divers, peu importe le stimulus, mais conserverait tout de même des activations plus importantes quand celles-ci sont cohérentes.

La démonstration de cet effet nécessiterait d'implémenter ce dont nous avons déjà discuté précédemment : la perte de l'ambiguïté perceptive.

Comparaison avec d'autres modèles et apports

Les modèles d'imitation et pointage par renforcement Il existe de nombreux modèles qui reproduisent les phénomènes d'imitations et de pointages d'objets. Par exemple pour reproduire le comportement d'imitation on trouve plusieurs travaux utilisant l'apprentissage par renforcement (par exemple (KOBBER et PETERS, 2010)). Mais c'est également le cas pour le pointage d'objet (CHAO et al., 2014) ou pour atteindre des points dans l'espace (BERTHIER, 1996).

Toute cette catégorie de modèle repose sur des algorithmes d'apprentissage par renforcement et nécessite soit des récompenses finales (réussite de l'atteinte ou non de l'objet par exemple) soit des récompenses liées à la forme de la trajectoire, ou encore : sont utilisés dans des paradigmes d'apprentissage par démonstration.

Certains de ces modèles, par exemple (KOBBER et PETERS, 2010) nécessitent une démonstration puis un traitement hors ligne des données et une création artificielle de nouvelles données à partir des données d'origines pour nourrir l'apprentissage. D'autres nécessitent une récompense explicite sur le fait de toucher ou non un point.

Nous employons une approche en complète opposition de celles-ci : Notre modèle emploie une approche qui nous semble plus simple, plus en accord avec le domaine de la neurobiologie et de la psychologie de l'enfance. Il ne nécessite ni récompense, ni démonstration, ni calcul hors ligne, ni données inventées par combinaison pour nourrir l'apprentissage. Il converge dans des temps acceptables en ligne (30 min à 2 h) en permettant de construire les commandes capables d'atteindre des positions dans l'espace, d'imiter un partenaire ou de pointer vers un objet.

Modèles de type RFWR, LWPR, GMM, etc. De nombreux modèles existent qui viennent faire correspondre des espaces sensoriels à des lois motrices en profitant de stratégies d'adaptation des noyaux de catégorisation pour correspondre au mieux aux données. Par rapport à ces modèles, l'un des avantages de notre proposition est ici de pouvoir continuer à compter sur des informations « neutres » dans toutes les étapes de l'apprentissage. Le résultat est probablement moins précis que ce que l'on pourrait obtenir en le combinant avec des adaptations de noyaux, mais aussi plus souple si de nouvelles données reentraient en jeu. Pour un modèle destiné à servir de base à la construction d'autres modèles l'utilisant nous pensons qu'il est plus adapté d'éviter les adaptations de noyaux ou de ne les employer qu'avec parcimonie afin d'éviter des sur-spécialisations.

Un autre aspect est que du point de vue de la bio-inspiration notre modèle et ses apprentissages simples et strictement locaux nous paraissent plus compatibles avec ce que l'on connaît du fonctionnement de ces zones chez l'humain que certains de ces modèles qui emploient des stratégies de rétropropagation du gradient.

Similitude avec les modèles de contrôleurs dynamiques De nombreux modèles ont tenté de définir un comportement dynamique du mouvement, souvent sans apprentissage, mais en s'inspirant des caractéristiques du contrôle moteur neuronal chez l'humain (GEORGOPOULOS et al., 1986).

On trouve une grande variété de modèles dans cette discipline. Un courant représentatif s'inspire par exemple de l'outil des champs de neurones dynamiques (DNF) définis par Amari (AMARI, 1977) en les appliquant à la robotique (SCHÖNER et al., 1995). Certains simulent ainsi des phénomènes moteurs existant aussi dans le développement de l'enfant, comme la persistance des gestes (THELEN et al., 2001) dans les expériences dites de « A-non-B » dans lesquels un enfant continue d'essayer d'atteindre un endroit A même après avoir vu que l'objet qui s'y trouvait était maintenant placé à l'endroit B ¹.

1. ↑ En réalité, chez l'enfant comme chez l'animal les études récentes (TOPÁL et al., 2008 ; TOPÁL et al., 2009 ; GERGELY et al., 2016) montrent que l'erreur serait plutôt due à un excès de confiance envers l'humain. Un humain étant associé, comme la mère, à une forte confiance, déclencherait l'erreur, mais réaliser l'expérience avec une ficelle qui tire les objets à la place d'un humain n'aboutira pas à une erreur, les interprétations « dynamiques » de l'erreur « A-non-B » sont donc a priori obsolètes, comme la plupart des autres interprétations qui furent faites du phénomène précédemment, incluant l'interprétation de Piaget

Ils ont été utilisés pour modéliser les mouvements du bras en robotique et pour expliquer de nombreux comportements moteurs, (GAUSSIER et al., 1998 ; IOSSIFIDIS et SCHONER, 2004 ; ANDRY et al., 2004 ; REIMANN et al., 2011) mais aussi pour expliquer des processus pas uniquement moteurs, mais portant sur la mémoire ou l'attention visuelle (JOHNSON et al., 2009 ; ROUGIER, 2009 ; FIX et al., 2011).

On pourra noter certains modèles majeurs comme les modèles « *Vite and Flete* » de Bullock et Grossberg (BULLOCK et GROSSBERG, 1989) qui modélisent une boucle sensorimotrice pour atteindre des positions dans l'espace avec un bras. Ce modèle connut plusieurs applications et fut affiné ensuite (HERSCH et BILLARD, 2006).

Ces modèles sont très similaires au modèle que nous présentons dans nos travaux, en effet même si nos travaux n'utilisent pas de champs neuronaux dynamiques d'un point de vue strict, les propriétés de ce type de modèle restent assez proches comme l'ont montré (GAUSSIER et al., 1998) et (ANDRY et al., 2004). Par ailleurs, il est tout à fait possible de remplacer les champs de catégorisation utilisés ici par des champs neuronaux dynamiques portant un apprentissage de type Kohonen.

Un modèle dynamique en particulier nous paraît avoir de l'importance pour imaginer le fonctionnement de notre architecture, il s'agit du modèle « *Yuragi* ».

Comparaison avec le Yuragi Le modèle *Yuragi* (FUKUYORI et al., 2008) ne se base pas sur un concept de réseau de neurones, mais il a par contre l'avantage de résumer en une équation un système dynamique, ce qui permet des analogies. Le système *Yuragi* se base sur les équations de Langevin (LANGEVIN, 1908) et propose une équation du type :

$$\Delta \cdot \dot{x} = \xi \cdot f(x) + \eta$$

Où x dépeint l'état du système et f dépeint une fonction définissant des *bassins d'attractions*. η quant à lui est un terme variable représentant l'exploration et ξ un terme multiplicatif que l'on peut faire varier pour prendre en compte plus ou moins fortement les bassins d'attraction. Le principe est très souple, car il permet en définissant f de construire un ensemble de points attracteurs importants dans l'espace et en faisant varier ξ et η d'explorer cette espace et de changer d'attracteur. Le principe a notamment permis d'illustrer la commande de robot mobile (NURZAMAN et al., 2009) ou de bras robotiques (SUGAHARA, 2010).

Une fois le champ d'attracteur construit tout se passe comme si les attracteurs appliquaient des forces sur l'effecteur, le tirant vers un attracteur ou un autre en fonction de sa position par rapport à ces attracteurs. Il y a alors une similitude forte avec nos travaux dans lesquels nous apprenons des commandes qui forment des attracteurs et dans lesquels quelques attracteurs sont capables de « tirer » l'effecteur pour atteindre des positions intermédiaires, comme nous l'avons illustré sur la figure 4.26.

Nous proposons donc que notre système soit une implantation neuronale du principe du *Yuragi*.

Apport vis-à-vis des travaux précédents Les travaux employant des architectures de liens perception-actions sont aujourd'hui assez nombreux. Des architectures proches ont ainsi été théorisées (GAUSSIER et ZREHEN, 1995) puis appliquées sur des bras robotiques électriques (ANDRY et al., 2001), sur des bras hydrauliques simples (RAZAKARIVONY et al., 2010), en navigation autonome (GAUSSIER et al., 1997), dans le contexte d'imitation d'expressions faciales (BOUCENNA et al., 2010) et de la reproduction de tâche (RENGERVÉ et al., 2010) pour enfin être intégrés dans des modèles plus complexes capables de séquences d'actions (RENVERGÉ, 2013).

Notre architecture partage une boucle commune avec ces modèles, mais nous avons tenté de répondre à certains des problèmes existants dans ces précédentes approches. Par exemple, les premières architectures comme (RAZAKARIVONY et al., 2010) aboutissaient à un contrôle assez peu souple et parfois saccadé, car elles employaient un paradigme d'apprentissage de type « un neurone vers un neurone ». Une catégorie uniquement pouvait alors être activée et donnait ainsi lieu à une commande en position sans aucune capacité d'interpolation.

Notre système ne fonctionne pas du tout ainsi, l'intégralité du système est au contraire pensée dans une chaîne d'apprentissage mettant en jeu une représentation des données plutôt proche du concept de *sparse coding* ou des codages en population avec un modèle qui conserve le paradigme des « fréquences

moyennes de décharge » sur l'ensemble des couches, fournissant ainsi à chaque couche des signaux analogiques sur lesquels se baser pour l'apprentissage, ce qui aboutit à de bonnes qualités d'interpolation (voir paragraphe 4.3.4)

Ce type d'apprentissage de « un neurone vers un neurone » est resté présent ensuite dans nombre des travaux suivants, on le retrouve ainsi dans (RENGERVÉ et al., 2010) et (RENVERGÉ, 2013). Antoine Renvergé dans sa thèse récente (RENVERGÉ, 2013) note ainsi les différents problèmes auxquels cette stratégie peut aboutir, notamment une explosion de la combinatoire qui aboutit à une explosion du nombre de catégories. Notre approche résout ce problème en employant sur l'intégralité de la chaîne un codage ouvert non filtré et des rehaussements de contraste à la place des mécanismes de *Winner-Take-All*.

Dans la plupart des approches, on note également que les apprentissages se déroulent sur les positions brutes via un PID et non sur des doublets de co-contractions contrôlant directement un signal électrique de commande, ce qui donne lieu à quelques problèmes notamment le fait, comme nous l'avons déjà évoqué, qu'un apprentissage sur une donnée de position brute aboutit à un comportement erroné tant que l'apprentissage n'a pas convergé.

D'autres travaux emploient ce système de co-contraction. (RENVERGÉ, 2013) par exemple, introduit, pour d'autres raisons, la notion de système de co-contraction qui a inspiré notre modèle, mais il n'isole pas les propriétés de ce type de système et n'en profite pas : dans ses implémentations en pratique il ne commandera qu'une des composantes de la co-contraction en déduisant la seconde par normalisation. ($A^+ = 1 - A^-$). Le problème de cette simplification est qu'elle détruit tout l'intérêt de la co-contraction qui est de permettre d'apprendre A^+ et A^- séparément, cela annule également l'avantage de la co-contraction de tendre vers une position juste dès le départ, au contraire : en déduisant A^+ à partir de A^- on se ramène à un contrôle en position classique avec les mêmes défauts.

Dans notre approche au contraire nous reprenons cette idée de modèle en co-contraction, mais en isolant les propriétés et en permettant au système un contrôle libre des composantes A^+ et A^- . Ceci permet au système de faire varier les forces utilisées en profitant des phénomènes émergents de la mécanique du robot, et cela implicitement. Mais cela permet également de converger dès l'instant 0 vers une position juste tout en ayant des forces faibles, comme nous l'avons montré dans les sections sur le contrôle musculaire et précédemment dans la discussion, ce qui est idéal pour l'interaction.

Contrairement aux systèmes précédents, nous ne passons par aucun système intermédiaire de type PID pour assurer le contrôle, grâce à notre analogie entre la vitesse angulaire due à la co-contraction et la commande électrique du système, la sortie de notre réseau de neurones est ici directement branchée à la commande électrique du robot, sans filtres. Ce qui est inédit, à notre connaissance, dans l'application de ce type de modèle.

Pour finir sur ces aspects, notons que les implémentations récentes du système utilisaient parfois également des rehaussements de contraste de type « doux » (*Instar* ici) comme dans notre implémentation, mais en employant les équations de convergence finale de ces réseaux au lieu des équations dynamiques. La suppression de la dynamique dans ce processus réduisait alors considérablement les propriétés de l'*Instar* quant à sa résistance au bruit et à son traitement de signaux successifs de façon cohérente. Ceci aboutissait donc à un bruit qui se transmettait sans aucun traitement des couches visuelles aux catégories, ces dernières étant calibrées pour ne réagir qu'une à une on assistait à des « sauts » du système entre ses commandes ou à des tremblements au cours du comportement. L'implémentation présentée ici, en employant les équations d'origine de l'*Instar* tout en utilisant des catégories souples sans employer de mécanismes de sélection de type *Winner Take All* permet de supprimer complètement ces défauts et d'aboutir à un contrôle souple.

Notons que pour compenser ces défauts, certaines implémentations, notamment (RENVERGÉ, 2013) ont ajouté une implémentation de système *Yuragi* (tel que décrit précédemment) qui vient filtrer la commande finale du système pour aboutir à quelque chose de plus souple. Notre implémentation n'est pas dépendante de tels systèmes supplémentaires et devient même assez semblable aux systèmes de type *Yuragi* sans pourtant avoir besoin de l'implémenter explicitement.

En conclusion, l'architecture présentée ici est à la fois plus minimaliste (pas de sous-systèmes supplémentaires de type *Yuragi* ou PID), mais présente également de meilleures propriétés (pas de sauts

dus au bruit, convergence rapide vers la position désirée, convergence lente en force) tout en résolvant certains des problèmes pointés dans les précédents travaux (explosion combinatoire) et en ouvrant de nouvelles portes quant aux applications (apprentissage de la force).

Critique de l'architecture et perspectives

Architecture visuelle Le versant vision de notre architecture robotique est très peu développé et il s'agit sans doute d'une des faiblesses du système. Même si nous avons justifié précédemment les raisons de ce choix (pas le cœur du sujet de thèse, cohérent avec ce que l'on sait de la vision chez le jeune enfant dans un certain cadre, etc.) il reste que pour étudier des phénomènes plus complexes il devient nécessaire d'élargir le concept de cette architecture à des systèmes visuels plus complets et plus proches de la neurobiologie.

L'un des exemples typiques de ce qui peut en découler est que la prise en compte de l'intégralité du mouvement du bras, et non juste de l'effecteur terminal complexifie le problème. En fournissant plus de détail aux couches de catégorisation, l'information visuelle sur la posture du bras viendrait alors compléter l'information de la proprioception qui contient elle aussi des informations de posture. Le résultat sur le comportement dans un contexte d'ambiguïté perceptive est difficile à juger sans faire le test en pratique puisque ce sont désormais l'intégralité des mouvements du corps et des bras de l'expérimentateur qui déclencheront le phénomène. On obtiendrait très probablement un résultat moins « propre » que celui que l'on peut observer sur les vidéos du système actuel, mais sûrement plus en accord avec le phénomène réel neurobiologique.

Un autre aspect que nous avons simplifié est la projection de la vision dans un espace égocentrique. Les travaux récents en neurobiologie semblent indiquer qu'il n'existe pas d'espace « central » égocentrique. Mais qu'au contraire chaque partie du cerveau fonctionnerait dans son propre espace, souvent en lien avec son entrée principale ou sa sortie principale, et ferait les transformations adéquates des informations venant d'autres espaces pour travailler avec. Ainsi, le colliculus supérieur, qui reçoit une entrée principale de la rétine via le Corps Géniculé Latéral, et dont les sorties se transmettent aux muscles des yeux et de la tête, fonctionne principalement dans un référentiel rétinocentrique. Pourtant il va recevoir des informations centrées sur la tête via les parties traitant le signal sonore, et il va alors faire les transformations idoines durant la projection de ces parties vers ses couches d'entrées. Les travaux récents dans le domaine montrent que l'ensemble du cerveau pourrait fonctionner sur ce modèle donc : avec chaque partie dans un espace qui lui est propre et aucun espace « global » véritable.

Des travaux sont donc probablement nécessaires pour aboutir à un traitement de la position des stimuli sans faire de projection égocentrique explicite. Travail d'autant plus intéressant que les données de la neurobiologie tendent à montrer que les informations proprioceptives de la tête et du cou ne sont pas utilisées par le cerveau (des lésions sur ces récepteurs n'ont aucun impact), ce qui amène à la conclusion que pour obtenir une information facile à lier avec les commandes motrices, le système nerveux central doit probablement recourir à des mécanismes plus subtils (exploitant les données de commandes envoyées par le contrôle au niveau du colliculus par exemple) .

L'étude de ces mécanismes pourrait rendre compte de phénomènes dans le lien entre perception et action que nous ne prenons pas du tout en compte dans notre cadre.

Commande œil-cou De la même manière, la commande gérant les yeux et le cou de notre modèle est peu crédible dans un contexte neurobiologique et mériterait d'autres travaux, car cette commande, par rebond sur les entrées sensorielles, est une part importante de la boucle sensorimotrice. Un système plus solide de commande œil-cou, notamment via la modélisation du colliculus supérieur et des réflexes liant les yeux et le cou, donnerait probablement des résultats intéressants.

Couches de catégorisations Dans nos travaux, nous avons tenté d'améliorer les précédentes propositions de ce type d'architectures en travaillant sur les types de catégorisation et en privilégiant au maximum des catégorisations et filtrages capables de renvoyer les informations de façon riche et sans filtrage excessif. Pour ce faire, nous avons privilégié des modèles abandonnant les concepts de

distances, très simplifiés, pour revenir sur des traitements plus neurobiologiques de l'information grâce à la normalisation des surfaces d'entrées synaptiques.

Si l'impact sur les résultats est plutôt bon, avec une propagation naturelle de signaux faible en cas d'environnement peu connu, qui aboutit alors à une force peu importante et à une moindre résistance au bruit en début d'apprentissage, les systèmes implémentés restent néanmoins simplistes au regard de leurs véritables équivalents neurobiologiques et des propriétés de ces équivalents. Notamment si l'on prend en compte le fait que beaucoup de couches de traitement du cortex fonctionnent via des retro-inhibitions de type *basket-cell* qui viennent alors introduire des aspects dynamiques intéressants dans les traitements des signaux, qui ne sont pas du tout pris en compte ici.

On peut également discuter des impacts des méta-paramètres de notre modèle, qui sont la conséquence du fait qu'il s'agisse de modèles très simplifiés. Par exemple la largeur des gaussiennes va avoir un impact sur l'apprentissage. Des gaussiennes qui ne se recoupent pas assez vont faire tendre l'algorithme sur le cas « une catégorie → une commande ».

À l'inverse, des gaussiennes trop larges vont aboutir à des catégories qui se recoupent beaucoup et qui demandent donc énormément de temps pour l'apprentissage : un apprentissage tentant de généraliser une loi de commande générale en prenant en compte énormément d'information sera très lent par rapport à un apprentissage utilisant moins d'information.

Si on devait reprendre l'analogie du ressort, c'est un peu comme si à chaque position l'algorithme tentait de donner une valeur à chaque ressort. Chaque ressort « tirant » l'effecteur vers une posture. Le but de l'algorithme étant de trouver la bonne combinaison de valeur de ressort pour que l'ensemble des positions sur le territoire considéré soient le plus juste possible.

Néanmoins, cela n'est pas toujours possible, car certaines informations peuvent être contradictoires et l'algorithme converger alors vers une réaction lente face à des centaines de petits ressorts le tirant dans des postures contradictoires et impossibles à calibrer de façon juste.

D'un point de vue plus mathématique : il est nécessaire de conserver une certaine séparation des données lors de leur projection dans un espace plus grand (la catégorisation) pour que cette projection soit exploitable. Si les données se recoupent trop après projection la projection est inefficace.

Ici, il peut être intéressant de s'inspirer en partie de ce que font les algorithmes adaptant les largeurs des noyaux de catégorisations ou encore d'étudier plus en détail les mécanismes de la neurobiologie sur ce sujet de l'adaptation des noyaux, car le « choix » de ces largeurs de noyaux a finalement une grande influence dans notre cas et demande trop d'apports de compétence de l'expérimentateur, ce qui va à l'encontre de notre philosophie visant à créer des algorithmes autonomes et développementaux.

Un autre méta-paramètre à une grande importance dans nos algorithmes de catégorisation c'est le « N » de la normalisation des surfaces synaptiques, qui demande à l'expérimentateur d'avoir une connaissance a priori des entrées de la couche en question et de la valeur de leur norme en moyenne. On pourrait imaginer étudier plus finement les mécanismes de cette normalisation dans la neurobiologie pour qu'il ne soit pas nécessaire de fixer ce facteur à la main et que chaque lien s'adapte seul aux signaux entrant malgré des amplitudes de signaux très différentes.

Liens entre la sortie du « simili-cortex moteur » et la co-contraction Dans notre modèle, le lien final d'apprentissage est construit par un simple LMS. Ceci a quelques conséquences. Tout d'abord, on n'utilise qu'une unique valeur analogique par muscle. Comme nous l'avons montré dans la partie sur le contrôle musculaire, ceci est valide, mais dans un certain cadre seulement. Les signaux descendant depuis les motoneurones supérieurs viennent commander plusieurs unités musculaires, et alors ce signal analogique va recruter, en fonction de son intensité, des motoneurones de plus en plus connectés à des fibres musculaires et de tailles différentes. Tout en augmentant au passage l'intensité du signal délivré par chaque motoneurone déjà recruté. Ce double phénomène tend à montrer que l'approximation de signaux analogiques est valide, mais il ne faut pas oublier pour autant qu'elle n'est valide que pour un ensemble d'unités motrices, mais qu'il existe plusieurs de ces ensembles pour un même muscle. En réalité, il semble donc tout de même y avoir un certain découpage de la commande en sous-parties.

Ce découpage est plus problématique qu'on ne le pense, car il implique que plusieurs commandes

peuvent avoir le même effet, indépendamment les unes des autres, ce qui ne rend pas impossible l'apprentissage, mais le complexifie notablement et demande une exploration minutieuse de ces différents cas.

Ceci amène également la question de la modélisation de la couche multimodale finale. Plusieurs neurones de cette couche peuvent être liés à des neurones aux effets similaires. Il peut ainsi être judicieux de modéliser plus finement ces couches, notamment en considérant ces propriétés d'interconnexion dans la couche elle-même. Nous savons que de telles propriétés dans les parties de traitement visuel de catégorisation ou dans le cortex moteur chez l'humain. Le problème pourrait alors grandement se trouver simplifié lors de l'apprentissage. Une catégorie déjà explorée et liée à une commande A ayant l'effet 1 et une autre catégorie déjà explorée liée à une commande B ayant l'effet 1 également, se retrouveraient de temps à autre activées en même temps et finiraient par former un lien entre-elles. Ceci aboutirait à un groupe de catégories liées qui activeront les commandes A comme B, ce qui simplifierait ainsi considérablement, au fur et à mesure de l'apprentissage, le problème de multiples commandes donnant des effets similaires, puisque toutes les commandes aux effets similaires se retrouveraient liées.

Un autre défaut qui nous semble critique dans notre système est qu'en utilisant une sortie de type LMS on simplifie l'apprentissage en séparant clairement la sortie désirée du signal entrant, mais en sacrifiant au passage *l'interaction durant l'apprentissage*. De fait, on utilise assez peu en pratique les bonnes propriétés du système musculaire vis-à-vis du bruit : une exploration implicite dans les zones où l'apprentissage est encore en cours, mais faible là où l'apprentissage a déjà été fait. Pour profiter pleinement de tous ces aspects, il existe plusieurs solutions. On peut aller vers des modèles de types champ de neurones à commander, chacun commandant une portion de la force (voir la partie suivante à ce propos). Et on peut travailler plus finement les lois d'apprentissages de ces modèles pour être indépendant de tout « signal désiré » tout en ayant les mêmes effets. En clair, il faut un modèle capable à la fois de subir un signal d'exploration et de prendre en compte ce signal dans l'association en cours. Ce qui nous amènerait plutôt vers des modèles d'apprentissage de type Hebbien ou BCM.

Discussion sur certains aspects neurobiologiques

Positions du « cortex moteur » dans le modèle Dans les précédentes sections, on a souvent fait la comparaison entre le rôle tenu par l'une des couches de notre modèle et celui tenu par le cortex moteur humain, ceci afin d'illustrer comment l'apprentissage de liens sensorimoteurs pouvait aboutir à la construction d'un ensemble cohérent de commandes liées à un espace moteur. Toutefois, nous n'avons pas véritablement détaillé cette analogie et analysé sa pertinence.

Si l'on reprend le schéma de l'architecture présenté sur la figure 4.14 on remarque que la sortie motrice (couche D) ne tient pas vraiment lieu de « cortex moteur », car elle ne fait que représenter la commande brute. Ce qui tient lieu de « cortex moteur » est l'ensemble de catégories présent en amont, sur la couche C. Néanmoins on peut alors se poser la question de la pertinence de cette analogie puisqu'on a également fait l'hypothèse que cette couche tenait lieu dans un premier temps de couche d'association multimodale. Alors quelle hypothèse est la plus juste dans notre cas ?

En réalité, il n'est pas forcément problématique de considérer que le cortex moteur chez l'humain fonctionne comme une carte auto-organisée et donc présenterait les caractéristiques, d'un point de vue « outil », d'un catégorisateur de ces entrées. Le cortex moteur chez l'humain fonctionne en effet avec de nombreux liens proximaux et distants, qui ont en plus des propriétés différentes (inhibition/activation) similaires dans l'idée aux cartes de Kohonen classiques. Des travaux récents en neurobiologie ont démontré, par exemple, l'importance des connexions latérales, très nombreuses (chez le chat par exemple : (CAPADAY et al., 2009)). De plus, le cortex moteur présente une topologie bien connue des ordres moteurs, chaque zone contrôlant telle ou telle zone du corps humain.

On peut se poser la question de la pertinence de considérer qu'une carte auto-organisée s'auto-organise en fonction d'ordres moteurs alors que ses entrées sont strictement sensorielles. Toutefois, comme on l'a vu ce sont plutôt les sorties de notre catégorisateur qui définissent ce que représentent ses neurones d'un point de vue comportemental et ses sorties dans notre cas sont des ordres moteurs liés à l'espace atteignable du robot. De plus, une fois ces sorties construites, si on fait l'hypothèse que l'on est sur une

carte auto-organisée, alors d'autres parties du système nerveux central peuvent se mettre à les utiliser et la carte se retrouverait à s'organiser de façon pertinente autour de ses actions les plus utilisées.

Toutefois, il reste des failles dans cette analogie vis-à-vis de notre modèle et c'est l'absence, dans ce cas, de véritable couche d'intégration multimodale. Un modèle plus complet comporterait alors une couche supplémentaire : une couche d'intégration multimodale entre la couche B et C qui servirait uniquement de couche associative, et la couche du « cortex moteur », elle, se retrouverait décalée.

Ceci n'a aucune utilité sur la version actuelle du modèle seul, car l'information ne changerait pas entre les deux couches, mais ceci est toutefois utile si l'on considère un modèle plus complet puisque ça permettra à d'autres parties du modèle d'utiliser le « simili-cortex moteur », qui va se réorganiser selon ces utilisations, sans impacter la couche de catégorisation multimodale, qui resterait alors disponible sans changement majeur. Ce qui aboutirait alors à une réelle distinction d'organisation entre les deux couches tout en permettant à la couche multimodale de rester disponible pour d'autres parties du modèle sans pour autant être dépendante de l'organisation des ordres moteurs.

Dans notre hypothèse, cette proposition de modèle est donc plus juste dans un cas général. Retenons donc, en résumé, que sur notre modèle nous faisons une analogie entre le rôle tenu par la couche de catégorisation finale dans notre modèle et le rôle tenu par le cortex moteur chez l'humain, mais tout en considérant également que cette couche finale tient lieu de couche d'intégration multimodale, selon le point de vue. Un double point de vue qui n'a de sens ici que parce qu'il s'agit d'un modèle simplifié, sur un modèle plus général il faudrait donc distinguer les deux parties.

Les neurones miroirs, une analogie ? La découverte des neurones miroirs a eu énormément d'échos dans la communauté scientifique. Découverts dans les années 90 par l'équipe de Giacomo Rizzolatti ([RIZZOLATTI et al., 1999](#)), les neurones miroirs sont des neurones qui réagissent autant aux actions de l'agent considéré qu'aux mêmes actions observées chez autrui. Depuis, elles ont été mises en avant dans de très nombreuses études, les associant à des phénomènes de résonances et d'imitation motrice ou à des phénomènes moteurs connus pour leurs fonctions sociales (tels que les bâillements). Un courant important les suppose liés à la notion d'empathie chez l'humain (et animaux) par leur capacité à interpréter les actions d'autrui comme les siennes propres. ([GALLESE et GOLDMAN, 1998](#) ; [PRESTON et DE WAAL, 2002](#) ; [DECETY, 2002](#))

Il y a une analogie importante entre les neurones miroirs et le modèle présenté ici dans le sens où celui-ci n'est finalement composé que de neurones miroirs puisque le robot est incapable de distinguer ses actions de celles d'autrui, ce qui est la source de l'ambiguïté perceptive et donc de l'action dans notre cas.

On peut donc proposer que cette architecture soit un modèle possible expliquant la formation des neurones miroirs chez l'humain et qui semble cadrer avec les découvertes de la neurobiologie à ce sujet. Les neurones miroirs ont en effet été trouvés dans les parties en amont du cortex moteur, mais aussi dans les parties du lobe pariétal connues pour leur fonction d'associativité multimodale comme dans le cas de notre couche finale associatrice.

On peut tenter d'ailleurs de faire des prédictions en employant ce modèle. Si cette hypothèse est justifiée, on devrait trouver des neurones miroirs dans toute la chaîne de traitement de la vision au cortex moteur, sauf que des systèmes secondaires viennent, dès la petite enfance, prendre le relais pour inhiber les processus d'imitation immédiate ou alors celle-ci disparaît par la précision des données entrantes qui vient lever l'ambiguïté finale (voir section 8) . Dans les deux cas, la chaîne se retrouve coupée à la dernière étape du processus, ce qui permet à l'enfant de quitter les phases de « réaction » pour entrer dans celle de l'action. Si ce modèle est correct, il deviendrait alors possible de découvrir la présence de neurones miroirs jusque dans le cortex moteur primaire chez les petits bébés après le premier mois (temps de l'apparition des phénomènes d'imitation), puis qui perdraient ensuite progressivement leurs qualités de neurones miroirs au fur et à mesure des stades du développement décrit par Piaget, c'est-à-dire à partir de 18 mois / 2 ans, lorsque l'enfant commence à employer l'imitation différée au lieu de se reposer sur une personne présente servant de modèle.

4.4 | Conclusion

Dans ce chapitre, nous avons analysé plusieurs architectures de contrôle robotique afin de permettre un contrôle du robot Tino détaillé dans la partie précédente dans un contexte d'interaction avec un environnement réel ou avec un partenaire.

Nous avons tout d'abord illustré une architecture simple tentant de profiter directement des phénomènes émergents du contrôle mis en évidence dans la partie précédente, puis nous avons critiqué cet algorithme.

Pour prendre en compte les critiques formulées nous avons utilisé un modèle de contrôle musculaire par co-contraction et nous avons détaillé, puis illustré, en simulation, ses propriétés intéressantes dans notre cadre, telles qu'une intensité de force relative à la qualité de l'apprentissage dans une zone, une exploration implicitement plus forte dans les zones où l'apprentissage est faible, ou encore une capacité à stabiliser très tôt durant l'apprentissage la position d'équilibre induite par la co-contraction ceci tout en proposant une capacité à aboutir à des commandes pertinentes lorsque l'on somme des ensembles de co-contraction.

En exploitant ce système de contrôle intermédiaire, nous avons conçu une architecture qui tire parti de ces propriétés et notamment de cette capacité naturelle de somation de commandes, ceci afin de permettre un apprentissage capable de concevoir naturellement des commandes interpolées par la somation pondérée de sous-ensembles.

Nous avons ensuite illustré certaines capacités de l'algorithme, comme l'émergence des phénomènes de pointage et d'imitation dans un cadre d'ambiguïté perceptive et nous avons proposé que l'algorithme puisse apporter une explication à la formation de ces phénomènes.

Nous avons également discuté le fait que plusieurs couches de l'algorithme ont des analogies avec les neurones miroirs et que certaines convergent vers un ensemble de catégories représentant des commandes motrices pertinentes dans l'espace atteignable du robot, et donc qu'elles peuvent servir de base réutilisable pour d'autres algorithmes. Nous avons fait un parallèle entre le rôle que tient cette couche dans l'algorithme et le rôle que peut tenir le cortex moteur chez l'humain. Gardons toutefois à l'esprit que le cortex moteur chez l'humain a une structure bien plus complexe.

Dans la partie suivante, nous analyserons comment exploiter un ensemble d'actions déjà préconçues, par analogie avec celui construit dans cette couche, et via un algorithme d'apprentissage par renforcement, afin d'aboutir à des trajectoires capables de prendre en compte de façon plus pertinente l'environnement et le partenaire tout en intégrant implicitement dans leurs formes des informations capables d'aider le partenaire à déchiffrer l'intention du robot.

III

Troisième partie :

**Architecture de contrôle par
renforcement**

Ganglions de la base : propriétés et modélisations

Sommaire du chapitre :

5.1. Introduction	122
5.2. Les noyaux des ganglions de la base et leurs influences mutuelles	123
5.2.1. Voie directe, indirecte et liens topologiques : du striatum au thalamus	124
5.3. Les boucles parallèles des ganglions de la base	127
5.4. L'apprentissage dans les ganglions de la base	128
5.4.1. Substance noire compacte et aire tegmentale ventrale	128
5.4.2. Dopamine, récompense négative et récepteur de type D1/D2	129
5.5. L'apprentissage dans les ganglions de la base, modèle et similitude avec l'apprentissage par renforcement	131
5.5.1. Lien entre les ganglions de la base et les algorithmes de type TD-Lambda	131
5.5.2. Les modèles de type acteur-critique : un modèle général de l'apprentissage dans les ganglions de la base	132
5.6. Conclusion	135

5.1 | Introduction

Dans les précédents chapitres, nous avons illustré certains phénomènes qu'il est possible d'obtenir gratuitement grâce aux propriétés de la mécanique d'un robot (Partie I) et grâce aux simulations de boucles motrices de bas niveau présentes chez l'humain (Partie II). Nous avons également montré comment construire, par-dessus ces ensembles, un mécanisme cohérent capable, par apprentissage, de constituer un répertoire d'actions pertinent et de manifester des phénomènes émergents que l'on retrouve chez l'enfant et que l'on suppose primordial pour le développement de l'humain et son interaction avec l'environnement ou un partenaire, tels que l'imitation et le pointage.

Comme nous l'avons vu dans les sections discutant des autres algorithmes du genre, nous avons critiqué l'utilisation des algorithmes d'apprentissages par renforcement, notamment car il apparaissait que dans ces contextes l'usage de ces algorithmes était superflu : d'autres processus, plus simples et élémentaires, plus rapides, aboutissaient en effet gratuitement aux mêmes résultats. L'usage d'un système d'apprentissage par renforcement ne nous semblait alors pas justifié dans ces contextes.

Toutefois, il n'est pas possible d'ignorer l'apprentissage par renforcement dans d'autres contextes. Certains processus dépendant de l'effort, de la prise en compte de la douleur ou encore du plaisir sont manifestement liés à l'apprentissage par renforcement et ne peuvent donc être simulés autrement qu'en mettant en place ce type d'algorithme.

On s'intéresse notamment à deux aspects essentiels dans le cadre de la coopération Homme-machine :

- Le premier est la prise en compte de l'environnement et des obstacles dans les trajectoires et dynamiques du mouvement d'un point de vue pragmatique. Dans un cadre purement robotique, un robot en interaction doit savoir éviter les obstacles ainsi que les humains présents et donc prendre en compte l'environnement l'entourant pour ajuster ses trajectoires afin d'éviter les collisions. Mais il doit également prendre en compte les contraintes énergétiques des mouvements afin d'éviter de forcer à des endroits où cela est inutile par exemple. Autant de comportements qu'il n'est pas possible d'obtenir avec les algorithmes des chapitres précédents seuls.
- Le second aspect important est, comme nous l'avons évoqué dans la section 3.2.1 du chapitre précédent, que les êtres humains emploient les dynamiques de mouvements pour « lire » les intentions de l'autre. Un humain voulant donner un objet à un autre aura donc des trajectoires très différentes d'un humain posant un objet sur une table. On peut alors imaginer des algorithmes présentant les mêmes propriétés dans leurs dynamiques sous-jacentes. En posant l'hypothèse que ce n'est pas la dynamique « précise » qui est importante, mais plutôt les raisons sous-jacentes qui ont conduit à cette dynamique (et donc, d'une certaine manière, la forme globale des trajectoires dans un certain contexte) il devient alors nécessaire que la logique qui sous-tend les actions menant à un mouvement soit analogue entre l'être humain et un algorithme désireux de communiquer avec lui, afin que l'humain en interaction arrive à lire les « intentions » du robot simplement par l'observation de ses mouvements.

Ces deux aspects nous amènent à nous poser la question générale de la prise en compte de l'environnement et du partenaire dans un algorithme d'interaction Homme-machine. Nous posons ici l'hypothèse qu'un grand nombre de ces aspects peuvent être résolus en employant de l'apprentissage par renforcement simplement en considérant des notions de douleurs, de récompenses et d'énergie dépensée.

Dans cette partie de la thèse, nous chercherons donc à employer des stratégies d'apprentissage par renforcement via un modèle bio-inspiré des ganglions de la base (vu sous l'angle de l'apprentissage). Comme nous l'avons vu dans la petite introduction aux ganglions de la base ayant eu lieu dans la section 3.3.5 du chapitre 3 ceux-ci sont en effet communément considérés comme le centre de l'apprentissage par renforcement dans le cerveau, notamment via l'interaction entre le striatum et la Substance Noire compacte (SNc) qui délivre la dopamine dans le cerveau, souvent vue comme l'un des neuromédiateurs les plus importants de l'apprentissage. Ceci explique que nous nous intéressons ici aux ganglions de la base comme base d'inspiration pour notre modèle.

5.2 | Les noyaux des ganglions de la base et leurs influences mutuelles

Nous avons déjà abordé une vue d'ensemble du rôle des ganglions de la base, parfois appelés « *noyaux gris centraux* » dans la littérature, dans la section 3.3.5 du chapitre précédent. Mais cette vue d'ensemble est insuffisante pour en saisir réellement le fonctionnement et en établir un modèle, nous aborderons donc ici une vue globale des mécanismes existants dans les ganglions de la base au vu de la littérature.

Les ganglions de la base sont une structure positionnée au centre du cerveau et constituée d'un ensemble de noyaux, certains aux frontières mal définies, que l'on peut séparer en quelques gros ensembles selon leurs caractéristiques physiques :

- Le noyau caudé
- Le Putamen
- Le Noyau Sous Thalamique (« NST », ou « corps de Lyus »)
- Le Globus Pallidus (« GP », ou Pallidum)
- La Substance Noire (Substantia Nigra : SN)

Ces noyaux sont disposés de manière assez particulière comme on peut le voir sur le schéma 5.1. Le noyau caudé constitue ainsi le tour de cette structure particulière en forme de lentille entourée par un cerceau. Le putamen et le globus pallidus forment le centre de ce système et constituent le cœur de cette lentille. De fait, anatomiquement ils sont parfois regroupés sous le nom de « *noyau lenticulaire* ».

On peut ajouter quatre noyaux un peu en retrait par rapport à ceux présentés précédemment, mais qui conservent toutefois de forts liens avec les ganglions de la base :

- Le noyau accumbens
- Le bulbe olfactif
- L'amygdale
- L'Aire Tegmentale Ventrale (ATV)

Ceci constitue la séparation anatomique, mais il est souvent d'usage d'utiliser, plutôt, une séparation fonctionnelle de ces noyaux en fonction de leurs entrées et sortie et des types de liaisons que certains noyaux ou parties de noyaux entretiennent tout en prenant en compte leurs caractéristiques histologiques (structure des tissus, présence ou non de divers médiateurs chimiques, etc.).

On obtient alors la séparation fonctionnelle suivante :

- Le Striatum : Il va constituer la « porte d'entrée » des ganglions de la base. Il est constitué anatomiquement du noyau caudé et du putamen, qui vont constituer ce que l'on appelle le *Striatum dorsal*, et du noyau accumbens ainsi que d'une partie du bulbe olfactif (selon les auteurs), qui constitueront tous deux le *Striatum ventral*. De façon générale, le Striatum dorsal et le ventral partagent suffisamment de propriétés pour être intégrés dans une même structure fonctionnelle. Mais ils auront par contre une circuiterie un peu différente. L'une de ces propriétés, commune aux structures composant le Striatum, est qu'elles se présentent anatomiquement comme étant « striées » et constituent dans tous les cas une porte d'entrée entre l'extérieur et les autres noyaux. Le Striatum peut être séparé en deux ensembles selon le type de cellule qui le constitue. Le *Striosome* et le *Matriosome*. Ces deux ensembles sont intermêlés : le « striosome » est ainsi constitué de neurones éparpillés dans la matrice de neurones les entourant, elle-même nommée « matriosome » (ou parfois simplement « matrice »), et ces deux ensembles entretiennent des connexions radicalement différentes que nous aborderons plus tard dans cette partie.
- Le pallidum interne (Globus Pallidus interne : GPi) et le pallidum externe (Globus Pallidus externe : GPe). Ces deux noyaux ont une structure assez proche, mais sont impliqués dans des boucles différentes.
- Le Noyau Sous Thalamique (NST), qui ne subit pas de découpage supplémentaire.

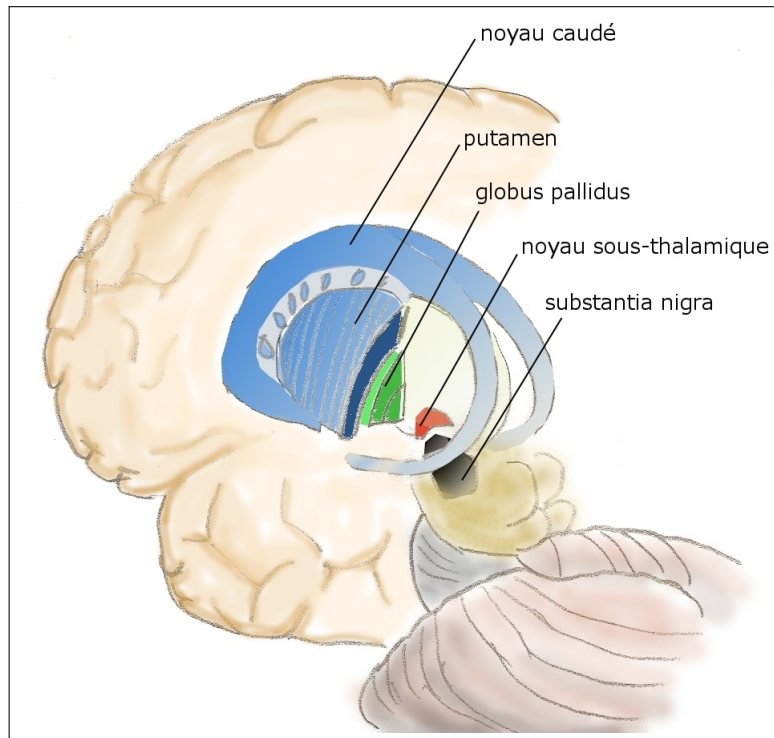


FIGURE 5.1. — Schéma représentant la disposition des ganglions de la base dans le cerveau, notons la disposition très particulière du noyau caudé et du putamen qui leur permet de recevoir des entrées de zones très diverses dans le cerveau. Image récupérée sous licence creative common CC BY-SA 3.0.

- La substance noire est divisée en deux parties radicalement différentes : la Substance Noire compacte (Substantia Nigra pars compacta : SNc [à ne pas confondre avec le Système Nerveux Central, SNC]), qui est constituée de neurones dopaminergiques, et la Substance Noire réticulée (Substantia Nigra pars reticulata : SNr). Ces deux parties de la substance noire entretiennent des liens avec l'extérieur radicalement différents. La SNr est parfois séparée en SNr et SNI (Substantia Nigra lateralis) quand on étudie la connexion un peu spécifique de certains neurones de la SNr avec le colliculus supérieur, on regroupe alors ces neurones sous le nom de SNI.

Il est courant, dans la littérature, de considérer que ces noyaux vont former deux réseaux principaux qui vont agir différemment sur le thalamus, thalamus qui joue le rôle de relai des perceptions vers le cortex ou de relai des actions du SNC vers les ordres moteurs, mais qui peut également servir de relai de communication entre différentes zones corticales. Ces deux voies d'action sont les voies *directes et indirectes* de la sélection de l'action dans les ganglions de la base.

5.2.1. Voie directe, indirecte et liens topologiques : du striatum au thalamus

Très tôt dans l'histoire de l'étude des ganglions de la base, les connexions entre certains noyaux ont été isolées et un schéma global a été imaginé pour les représenter et ainsi identifier leur rôle. Le résultat de cette représentation donne un réseau séparé en deux « voies » : la « voie directe » et la « voie indirecte », que l'on représente sur la figure 5.2.

Les deux voies sont les résultats d'un jeu d'inhibitions et d'activations qui prend en compte deux facteurs limitants : les noyaux de sortie des ganglions de la base sont exclusivement des noyaux qui inhibent le thalamus, « la porte de sortie des ganglions de la base ». Le striatum, « la porte d'entrée », n'est lui-même capable que d'inhiber les noyaux intermédiaires des ganglions de la base. Il se forme alors deux chaînes : la première, la voie directe, est une inhibition qui part du striatum (Plus précisément : du matrisome du striatum) et qui vient inhiber les deux noyaux de sorties aux propriétés similaires (SNr et GPi). Ce faisant, cette voie lève l'inhibition qui pèse sur les noyaux thalamiques ce qui a pour conséquence d'exciter, indirectement, les noyaux thalamiques.

La voie indirecte utilise deux relais pour avoir l'effet inverse. Le striatum inhibe la partie externe du pallidum (GPe) qui va inhiber les noyaux sous-thalamiques. Les noyaux sous-thalamiques se retrouvent donc, indirectement toujours, excités. Ils vont alors projeter sur les noyaux de sorties des ganglions de la base pour les exciter. Ceux-ci vont alors inhiber plus fortement les noyaux thalamiques.

En définitive, sur le thalamus, la voie directe a une action excitatrice (par rapport à un état « normal ») et la voie indirecte une action inhibitrice.

Ces deux voies restent encore vues aujourd'hui comme étant des approximations justes dans la littérature moderne, mais elles doivent être complétées. En effet, les connexions de ces noyaux entre eux ne sont pas que déterminées par leurs capacités à exciter ou inhiber. Ces connexions ont en effets des motifs topologiques qui viennent transformer l'information lors du passage d'un noyau à un autre et de nombreuses connexions récurrentes existent qui viennent compliquer l'analyse. De façon générale, il existe de très nombreux modèles dans la littérature qui étudient l'impact de la topologie dans la connexion de ces noyaux, et nous n'en ferons pas une revue complète ici. Si ces modèles sont généralement en opposition sur le détail ils ont tendance à se rejoindre sur le résultat du système global. Résumons donc simplement cette littérature par le fait que ces connexions topologiques dans les voies directes et indirectes semblent donner l'équivalent d'un système de type « *Winner-Take-All* » (sélection d'un gagnant dans une couche neuronale), mais avec de meilleures propriétés (voir l'exemple d'un de ces modèles dans les travaux de Benoit Girard (GIRARD et al., 2002; GIRARD et al., 2003)). Les ganglions de la base seraient alors capables de faire ressortir des informations au niveau du striatum, qui se retrouveraient ensuite projetées topologiquement dans les deux sous-ensembles que sont les voies directes et indirectes, le résultat final sur le thalamus à la fin étant la sélection d'une action par l'inhibition de toutes les autres plus ou moins fortement. Comme les connexions au niveau du striatum sont capables d'apprentissage, comme nous le verrons ensuite, le striatum est alors capable, selon les cas, d'ajuster les « niveaux » des différents composants de ce *Winner-Take-All* via un apprentissage différent sur la voie directe et indirecte. Ceci permet, selon les cas, de sélectionner une action précisément ou au contraire de laisser de nombreuses ouvertures et de ne sélectionner que de façon souple une action.

Schématiquement, on peut simplifier ainsi : le striatum va donner lieu à deux représentations de

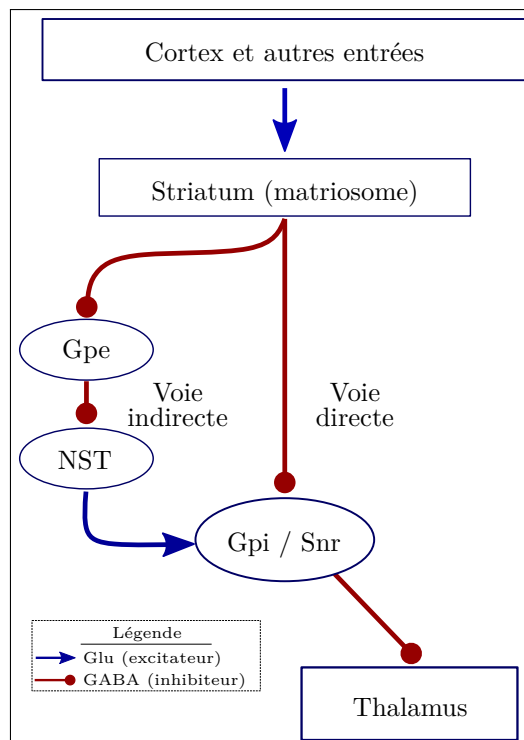


FIGURE 5.2. – *Voies directe et indirecte. La voie directe inhibe un noyau inhibiteur du thalamus. De fait, elle lève l'inhibition. La voie indirecte inhibe un noyau inhibant NST. De fait, elle l'excite indirectement. Le NST excite ensuite les noyaux de sorties et inhibe donc le thalamus. La voie directe a donc une action « excitatrice », et la voie indirecte une action inhibitrice.*

l'ensemble de données, l'une sur la voie directe et l'autre sur la voie indirecte. Ces ensembles auront des comportements que l'on considère généralement comme contraire (si l'un met en avant des données, l'autre forme des « trous » non actifs sur ces données). Ces données vont ensuite subir différentes transformations à travers les liens des deux voies (à savoir qu'il peut exister, dans les modèles plus complexes, des communications et des liens récurrents dans ces voies) pour finalement s'assembler sur la couche finale. Si le résultat est par exemple un ensemble très actif pour chaque neurone, mais très peu actif seulement pour quelques-uns, alors le résultat sur le thalamus sera un filtrage de seulement quelques neurones pertinents. Et c'est donc ainsi, très schématiquement, que les ganglions de la base sélectionnent les données.

Pour une bibliographie plus complète sur la connectivité et les modèles dans les ganglions de la base, on pourra se référer à (KACEM, 2011) et à (MOUSTAFA et al., 2014).

5.3 | Les boucles parallèles des ganglions de la base

Les ganglions de la base, comme le cervelet, forment des boucles parallèles nombreuses qui vont avoir des implications dans différentes zones du cerveau. C'est-à-dire que chaque noyau va être subdivisé en zone, et ces zones ne vont communiquer qu'avec des zones proches du noyau suivant, jusqu'à ne toucher que certaines zones des sorties. Ainsi, des zones en entrée vont correspondre à d'autres zones spécifiques sur les noyaux de sorties.

Toutes les zones du thalamus qui sont touchées en sortie des ganglions de la base ne sont donc pas impliquées dans les mêmes zones du cerveau. Certaines peuvent être des données venant du cortex et y retournant, d'autres peuvent mener à d'autres noyaux spécifiques, d'autres encore peuvent être des sorties ou mener au cervelet et au pont cérébral, etc. Certains noyaux des ganglions projettent parfois vers d'autres sorties que le thalamus, même si c'est très rare. C'est par exemple le cas du SNr, dont la partie « lateralis » (ou Substantia Nigra lateralis) projette directement sur le colliculus sans passer par un relai thalamique.

De façon générale, les ganglions de la base vont donc se retrouver au cœur de différents processus qui vont « utiliser » des sections parallèles des ganglions de la base. Un peu comme si l'on considérait de très nombreux ganglions de la base indépendants.

On est alors face au même problème que celui que nous avons évoqué dans le cadre du cervelet dans la section 3.3.8, à savoir : doit-on alors considérer les ganglions de la base comme étant dans un rôle absolument différent pour chaque boucle ? Ou doit-on plutôt les considérer, comme le préconisait Kenji Doya dans le cas du cervelet (DOYA, 2000), comme étant plutôt un outil générique effectuant une opération sur lequel viennent se connecter de nombreuses zones quand elles ont effectivement besoin de cette opération ?

La conception en tant qu'outil a de nombreuses vertus : elle évite de voir une zone uniquement dans le carcan de la boucle que l'on étudie dans son propre cadre applicatif et force à voir cette zone comme étant le fruit de la sélection génétique d'une opération à réaliser qui s'est avérée nécessaire à l'ensemble des boucles connectées. Cela permet de séparer le problème et surtout d'éviter de mal agglomérer le problème. Par exemple si l'on a un outil générique effectuant l'opération $A \rightarrow B \rightarrow f(C)$ et que l'on considère une boucle $A \rightarrow C \rightarrow \text{Outil} \rightarrow D \rightarrow A$, le risque en ne regardant que la boucle sans analyser les délimitations et le rôle de l'outil générique dans un premier temps c'est de « dérouler », en quelque sorte, la suite de transformations de la manière $A \rightarrow C \rightarrow A \rightarrow B \rightarrow f(C) \rightarrow D \rightarrow A$. Pour finalement tenter d'interpréter des « sections » de cette suite de transformations en oubliant les délimitations biologiques et les rôles de chaque partie dus à la sélection génétique, pour au contraire chercher des significations à des portions de cette suite de transformations à cheval sur plusieurs zones biologiques (par exemple, $C \rightarrow A \rightarrow B$) qui ne sont pas forcément pertinentes et qui peuvent amener à de mauvaises interprétations.

La conception en tant qu'outil évite donc cet écueil en se forçant à se poser la question de la délimitation neurobiologique considérée dans un contexte d'opération générique effectuée vis-à-vis de ses entrées et de ses sorties.

Dans la suite de ces travaux, nous considérerons donc les ganglions de la base comme un outil dont les entrées peuvent être impliquées dans différentes boucles, tout comme ses sorties, mais dont l'opération générale est commune à toutes ces boucles et identiques pour chacune.

5.4 | L'apprentissage dans les ganglions de la base

Comme nous l'avons introduit dans la section 3.3.5, les ganglions de la base sont en effet l'un des lieux les mieux identifiés concernant l'apprentissage chez l'humain, et tout particulièrement via l'émission d'un neuromédiateur spécifique : la dopamine. Notons que ce cas est assez unique dans l'étude du cerveau humain, car les émissions dopaminergiques ont été très tôt mises en évidence comme des signaux qui suivent de façon assez proche les lois théoriques d'apprentissage par renforcement conçues dans le cadre des réseaux de neurones artificiels. C'est donc l'une des rares zones du cerveau dont on a une idée assez solide du rôle et du fonctionnement global d'un point de vue apprentissage.

Mais qu'en est-il de la manière dont est fait cet apprentissage ? Même si l'on a une bonne idée du rôle des ganglions de la base et du résultat, les détails du processus restent un peu plus subtils, nous allons donner dans cette section une rapide description de certains des processus de l'apprentissage dans ces zones.

5.4.1. Substance noire compacte et aire tegmentale ventrale

La Substance Noire compacte (SNc) et l'Aire Tegmentale Ventrale (ATV) sont les deux zones dopaminergiques liées aux ganglions de la base. Ces zones reçoivent des afférences du striatum (dorsal ou ventral, selon que ce soit SNc ou ATV), principalement de la partie « striosome » du striatum, constituée de neurones particuliers au comportement assez différent de la « matrice » l'entourant, notamment d'un point de vue apprentissage, mais également en ce qui concerne leurs connexions entrantes et sortantes. Leur emplacement éparpillé dans le striatum a très tôt fait soupçonner qu'ils agiraient comme des intégrateurs des signaux circulant dans le striatum et redirigeant ces informations traitées à destination de SNc et ATV. On ne connaît par contre pas encore bien la nature des projections entre le striosome et les aires dopaminergiques et, si les neurones de la matrice entourant le striosome, GABAergiques, ont dans un premier temps fait soupçonner que ce serait principalement des liens GABAergiques, de nombreux travaux récents découvrent en réalité des liens de types très variables entre les deux structures (CRITTENDEN et al., 2016).

Mais ces deux zones dopaminergiques reçoivent également des signaux de nombreuses autres afférences notamment de la sortie des ganglions de la base via la partie « reticulata » de la substance noire (SNr), accolée à la partie compacta, qui vient fournir un signal inhibiteur. Ces neurones reçoivent aussi un signal excitateur depuis le noyau pédonculopontin (SCARNATI et al., 1986) et le noyau sous thalamique (ZHOU, 2016), qui sont soupçonnés servir dans ce cas de relai à un signal venant du striatum. (BROWN et al., 1999).

À côté de cela, l'aire tegmentale ventrale et la substance noire reçoivent des entrées provenant de différentes zones liées à la gestion de la récompense, de la peur, etc. qui sont connues pour servir d'intermédiaires entre certains états du corps (détection de dangers, satiété, etc.) et les signaux neuronaux notamment depuis l'amygdale et l'hypothalamus.

Ensuite ces deux zones vont projeter la dopamine de façon très large dans le cortex, dans divers noyaux, ainsi que sur le striatum.

Le cas du striatum est un peu particulier, car la dopamine va réagir différemment au contact des neurones du striosome et du matrisosome. Dans le matrisosome, elle va de plus réagir différemment selon que les récepteurs soient appelés « D1 » ou « D2 ». L'un des récepteurs étant spécifique à la voie directe et l'autre à la voie indirecte.

Distinction ATV/SNc Dans la littérature, il existe des débats quant à la pertinence de la distinction entre ATV et SNc. Les deux zones étant disposées côte à côte et ne possédant pas de frontière anatomique précise. De plus, leurs neurones sont du même type. La distinction est généralement faite sur le type de projection, ATV et SNc ayant des entrées légèrement différentes et projetant vers des zones également différentes. Leurs projections pour le striatum sont ainsi différentes avec une projection sur le striatum ventral (plutôt impliqué dans des boucles limbiques) pour ATV et plutôt vers le striatum dorsal (plutôt impliqué dans les boucles motrices) pour SNc. Dans notre hypothèse de « striatum outil »,

nous préfererons donc ne pas faire de distinction entre ces deux zones. Ainsi, comme dans beaucoup de modèles, nous considérerons SNc et ATV comme une entité unique.

5.4.2. Dopamine, récompense négative et récepteur de type D1/D2

La dopamine agit de façon différente sur les neurones. Elle va ainsi augmenter le cycle de l'adénylate cyclase sur les neurones équipés de récepteurs de type D1 et le diminuer sur les neurones équipés de récepteurs de types D2. Ce cycle est important, car l'adénylate cyclase est l'un des intermédiaires à l'établissement d'une LTP (Long Term Potentiation) dans un neurone, qui peut aboutir à un renforcement des liens qui y sont connectés, ce composé est donc au cœur des mécanismes d'apprentissage.

Plusieurs choses vont advenir suite à cette chaîne de réaction : les neurones vont se mettre à être plus ou moins réceptifs immédiatement à des activations extérieures. La dopamine joue alors un rôle de régulateur immédiat des voies directes et indirectes (notons que dans la littérature récente, ce rôle est remis en question). Ensuite, ceci va déclencher une LTP chez les neurones pour lesquels le cycle de l'adénylate cyclase est activé et, à l'inverse, cela va favoriser des LTD en bloquant ce cycle. La LTD arrive en effet quand il y a une certaine activité du neurone, mais trop faible pour déclencher le cycle de l'adénylate cyclase, généralement quand l'activité dans le neurone est trop faible pour déclencher une émission de forts potentiels d'actions. En bloquant artificiellement l'adénylate cyclase, les conditions d'une LTD sont réunies même dans des cas où la LTP aurait dû être prédominante. La dopamine joue alors le rôle de modificateur de l'apprentissage sur les neurones considérés. Soit en positif, soit en négatif, selon qu'ils soient équipés de récepteurs de type D1 ou D2.

Mais il est important de noter que tout ceci se fait par rapport à une ligne de base d'un niveau de dopamine tonique. Il a ainsi été montré que des pauses dans ce signal dopaminergique correspondent à ce que l'on entend par « récompenses négatives » dans l'algorithmique de l'apprentissage par renforcement (DI GIOVANNI *et al.*, 2009). Le processus exact qui conduirait à ceci n'est pas forcément clair dans la littérature : il existe en effet plusieurs processus distincts pouvant l'expliquer. Tout d'abord, il existe une habitude au niveau du nombre de récepteurs D1 et D2 sur un neurone et donc une habitude à la ligne de base du niveau de dopamine (SCHELER, 2004). Ensuite, des « trous » dans le signal conduisent à une absence d'excitation du cycle adénylate cyclase (pour les neurones disposant de récepteurs D1) ou à une absence d'inhibition de ce cycle (pour les neurones disposant de récepteurs D2). Sauf que cette excitation (ou inhibition) continue due aux niveaux toniques de dopamine a auparavant donné lieu à une habitude dans les processus chimiques du neurone. Un « trou » dans ce signal conduit donc mécaniquement à l'effet inverse : une sur-activité de cette boucle, ou une sous-activité.

Ce type de processus explique que la dopamine peut traduire un signal qui va plutôt favoriser un renforcement ou un oubli selon son niveau par rapport à sa ligne de base.

On est donc devant un double processus qui selon le niveau de dopamine va :

- Favoriser l'apprentissage pour des valeurs supérieures à sa ligne de base sur les récepteurs D1.
- Le défavoriser sur des valeurs inférieures à sa ligne de base sur les récepteurs D1.
- Défavoriser l'apprentissage pour des valeurs supérieures à sa ligne de base sur les récepteurs D2.
- Le favoriser pour des valeurs inférieures à sa ligne de base sur les récepteurs D2.

En réalité, les études récentes ont tendance à montrer que, dans certaines zones du moins, « Favoriser l'apprentissage » dans la liste ci-dessus peut s'entendre comme favoriser les renforcements tout comme les oublis (LEMON *et MANAHAN-VAUGHAN*, 2006). C'est-à-dire accélérer tous les processus d'apprentissage en cours indépendamment de leurs « signes ». Néanmoins, le mécanisme sous-tendant ces résultats n'est pas encore clair.

Notons que cela reste cohérent. Si l'objectif est de renforcer la chaîne de liens qui aboutit à une action, alors il est pertinent de renforcer à la fois les processus d'apprentissage « positif » qui avaient lieu par défaut (corrélation entre les entrées et les sorties) et à la fois les processus négatifs (désapprentissage quand non-corrélation entre les entrées et les sorties).

Néanmoins, dans la suite, nous suivrons plutôt l'hypothèse « classique » dans le cas des ganglions de la base, à savoir qu'un niveau de dopamine supérieur à la ligne de base a tendance à ne favoriser

qu'un apprentissage « positif » dans le cas D1 et « négatif » dans le cas D2. Et inversement dans le cas inférieur à la ligne de base. On peut résumer ce que l'on sait maintenant sur les ganglions de la base sur le schéma 5.3.

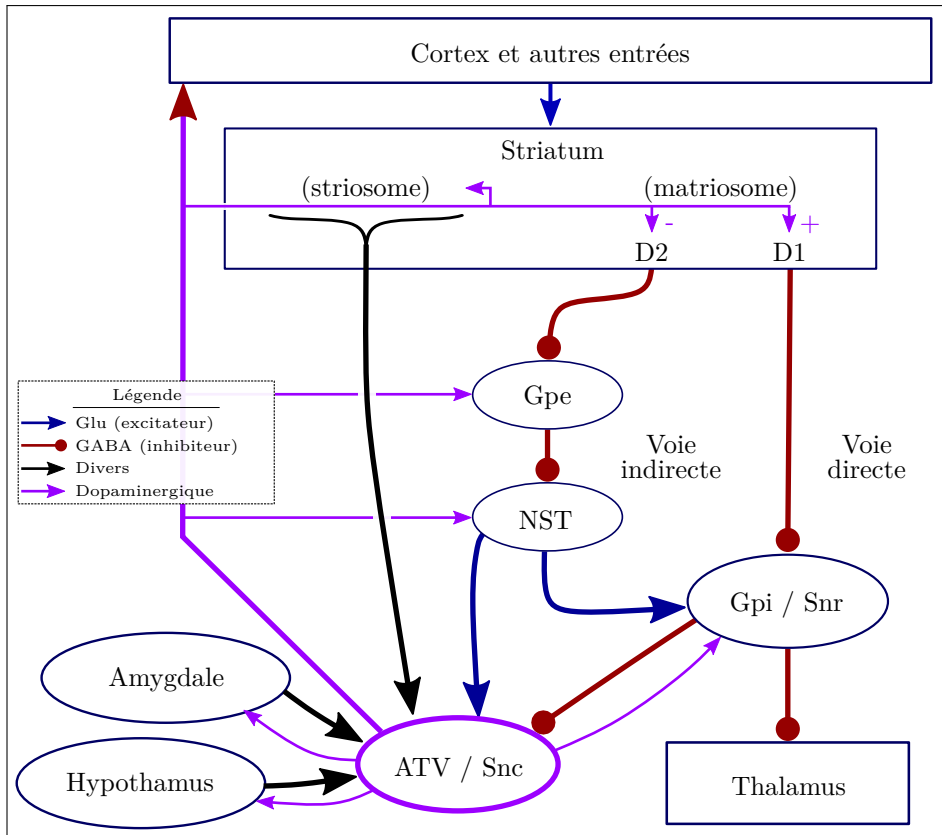


FIGURE 5.3. — Réseau d'apprentissage dans les ganglions de la base. Les noyaux SNc/ATV reçoivent des afférences depuis de nombreux noyaux et notamment depuis le striosome, ils projettent de la dopamine sur presque tous les noyaux impliqués dans les ganglions et sur de nombreuses zones du cerveau.

5.5 | L'apprentissage dans les ganglions de la base, modèle et similitude avec l'apprentissage par renforcement

Comme nous l'avons vu, les systèmes d'apprentissage dans les ganglions de la base sont assez particuliers. Néanmoins, les détails vus dans les sections précédentes ne sont pas suffisants pour avoir une idée générale du fonctionnement des ganglions de la base. En réalité, l'idée générale de fonctionnement apparaît lorsque l'on enregistre des signaux au niveau d'ATV, SNc et du striosome et permet de faire un lien avec les modèles d'apprentissage par renforcement théorique.

5.5.1. Lien entre les ganglions de la base et les algorithmes de type TD-Lambda

Rappelons les deux équations fondamentales de l'apprentissage par renforcement. Le TD-Lambda, qui est un modèle d'apprentissage simple dont il n'existe pas de démonstration mathématique d'une convergence vers une solution optimale, mais qui reste très utilisé en pratique, est défini par l'équation :

$$V(s) \leftarrow V(s) + \alpha \cdot (r + \gamma \cdot V(s') - V(s)) \quad (5.1)$$

Avec s et s' représentant respectivement l'état passé et l'état courant, $V(s)$ représentant la prédiction de récompense lorsque l'on se trouve à l'état s . α le facteur représentant l'apprentissage et γ un facteur affectant la propagation de l'erreur à travers les états.

Le Q-learning, qui est une amélioration du concept en subdivisant les états en couple « état-action » est défini par :

$$Q(s, a) \leftarrow Q(s, a) + \alpha \cdot (r + \gamma \cdot \max_{a'} (Q(s', a')) - Q(s, a)) \quad (5.2)$$

Où a représente une des actions associées à un état s . Et $Q(s, a)$ la prédiction de récompense.

Remarquons que les deux équations ont des similitudes avec les équations de convergence classiques des réseaux de neurones, car elles se présentent sous la forme $A(s) = A(s) + \alpha \cdot \text{erreur}$. Dans le cas de ces équations, on appelle ces erreurs « *erreur de prédiction de la récompense* », car elles se présentent sous la forme :

$$\begin{aligned} \text{erreur} = & (\text{récompense_actuelle} + \gamma \cdot \text{prédiction_des_récompenses_futures}) \\ & - \text{ancienne_prédiction_de_la_récompense} \end{aligned} \quad (5.3)$$

En observant les émissions de dopamine dans la substance noire et dans ATV il fut découvert assez tôt que la dopamine avait tendance à être émise de façon globale (tous les neurones, ou au moins de larges groupes, déchargeant en même temps) et qu'elle avait de très fortes similitudes avec l'erreur d'une équation de type TD-Lambda (MONTAGUE et al., 1993 ; SCHULTZ et al., 1997 ; SUTTON, 1988 ; DOYA, 1999 ; BAYER et GLIMCHER, 2005).

C'est cette relation qui a permis de cerner le rôle des ganglions de la base comme un centre de l'apprentissage par renforcement. L'erreur de prédiction est ainsi supposée être calculée au niveau de la substance noire pour servir de signal d'apprentissage cohérent sur l'ensemble du cortex et sur les parties des ganglions servant à calculer cette erreur.

Néanmoins, on peut ensuite se demander comment cela se déroule un peu plus en pratique, certaines composantes de ces équations sont par exemple des opérations implicites : quand on souhaite utiliser l'équation du TD-Lambda dans un cadre d'action, à chaque palier on doit appliquer une étape supplémentaire pour trouver le maximum des prédictions de récompenses des états avoisinants. Idem pour le cas du Q-learning : il faut pour chaque état rechercher le maximum des prédictions associées aux « états-actions » correspondant pour pouvoir agir. On a donc des équations d'apprentissage séparées du processus de sélection de l'action.

En pratique, comment peut-on interpréter les réseaux des ganglions de la base pour cadrer avec ce que l'on sait sur le lien entre dopamine et apprentissage par renforcement et que cela nous donne un modèle cohérent ?

5.5.2. Les modèles de type acteur-critique : un modèle général de l'apprentissage dans les ganglions de la base

Une interprétation célèbre du fonctionnement des ganglions de la base au regard de l'apprentissage par renforcement est le modèle « *acteur-critique* » des ganglions de la base. Dans cette interprétation (voir schéma 5.4), les ganglions de la base sont vus comme deux systèmes parallèles interdépendants :

- une partie « *acteur* » qui va agir via les voies directes et indirectes pour sélectionner des actions (dans un contexte moteur), ou, plus généralement, pour sélectionner les sorties thalamiques les plus pertinentes compte tenu de ce qui est activé sur le thalamus à ce moment-là et du contexte. Que ces relais soient moteurs ou non.
- une partie « *critique* » qui va, compte tenu des récompenses, tenter de prédire la récompense future, et compiler ensuite une erreur de prédiction une fois la récompense réelle obtenue. L'erreur de prédiction de récompense étant alors utilisée pour améliorer la prédiction la prochaine fois que l'on se retrouvera dans le même cas, mais également pour améliorer le comportement de l'acteur.

Prenons un exemple concret pour illustrer le fonctionnement du modèle. La partie acteur, en fonction d'un contexte donné par le cortex et des relais thalamiques entrants va transformer le signal via les voies directes et indirectes pour donner en sortie un banc de données à inhiber qui vont alors venir inhiber le thalamus, ne sélectionnant ainsi que certaines données pertinentes du thalamus. Dans un contexte moteur, par exemple, cela va donner lieu à une action sélectionnée, car le thalamus projette ensuite sur le cortex moteur ou relai les ordres du cortex moteur vers les effecteurs du corps.

En parallèle de cette chaîne d'événement, la partie « *critique* » utilise les données d'entrées pour tenter de prédire une récompense future.

Quand l'action est choisie et exécutée, cela donne alors lieu à une récompense réelle (ou non) et à une nouvelle prédiction de récompense. Celles-ci vont alors être comparées à la prédiction faite précédemment, et il en sera déduit une erreur de prédiction. Cette erreur de prédiction représente la différence entre ce que l'on espérait à l'état précédent et ce que l'on a réellement eu. Si cette différence est négative, alors on ajuste notre prédiction pour être moins « *déçu* » la fois suivante et on utilise ce même signal pour corriger l'acteur afin de l'empêcher de prendre cette mauvaise décision d'action la fois suivante. Et inversement si le différentiel est positif : dans ce cas on a été surpris « *en bien* » et on motive l'acteur à refaire ce genre de décision tout en affinant la prédiction pour qu'elle nous permette de prédire en avance la récompense (ce qui permettra de la propager entre les états et donc ce qui aidera à guider les actions vers elle).

Notons ici que ce double processus est un peu différent des équations du TD-Lambda et du Q-learning, car dans ces dernières il n'y a pas « *d'acteur* » indépendant, l'action est simplement déduite des résultats de la prédiction de récompense dans les états adjacents. Mais cela demande des prérequis : savoir ce que sont

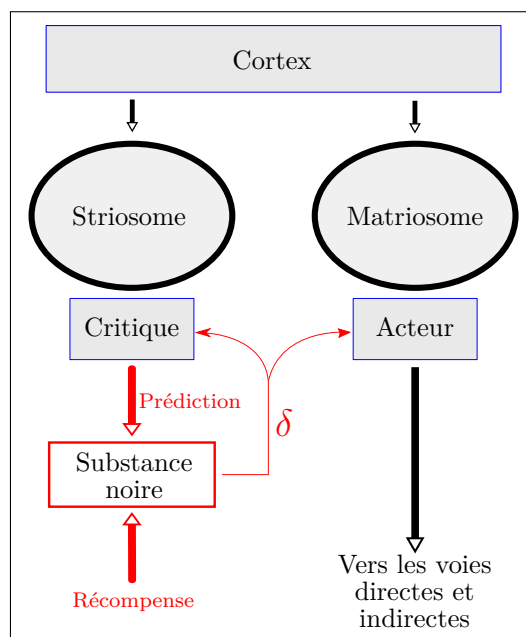


FIGURE 5.4. — Le modèle « *acteur-critique* » des ganglions de la base est une interprétation du fonctionnement des ganglions de la base composée de deux voies parallèles qui vont s'influencer l'une l'autre.

les états adjacents par exemple, puis l'action y menant, savoir comparer ces résultats, etc. Autant de prérequis simples dans une implémentation informatique discrète, mais plus complexe dans un système de réseaux de neurones. Le modèle acteur-critique permet de se passer de ces prérequis en dédoublant le système d'apprentissage.

En dehors de ce point important, le système est parfaitement analogue. L'état s dans une équation de type TD-Lambda correspond à la couche d'entrée et la prédiction $V(s)$ correspond à un élément utilisant la couche d'entrée pour déduire $V(s)$ à partir de s . Des éléments supplémentaires viennent se rajouter pour calculer l'erreur de prédiction pour obtenir finalement l'intégralité d'une équation de type TD-Lambda. Notons que ce n'est pas analogue exactement à une équation de type Q-learning, car aucun de ces éléments ne permet d'obtenir le maximum des états-actions. Toutefois, on peut avoir des entre-deux avec des états-actions en entrée du réseau, sans le maximum de l'équation, mais ceci restera donc plus proche du TD-Lambda que du Q-learning.

Comparaison avec la neurobiologie

Nous avons vu les similitudes entre le modèle acteur-critique et les équations de type TD-lambda et Q-learning, mais est-il vraiment analogue à ce que l'on sait de la neurobiologie ?

Tout d'abord sur les points les plus évidents il l'est assez bien : la partie « critique » du modèle correspond à ce que l'on connaît de la liaison striosome→SNc (ou ATV) renforcée par l'ensemble des connexions entrantes à la SNc qui lui permettent de construire le signal dopaminergique. Le signal dopaminergique lui-même présente de fortes analogies avec le signal de récompense dans le modèle et ressemble donc à une erreur de prédiction de la récompense.

La partie actrice quant à elle correspond à la liaison entre le striatum et son système de sélection de l'action (voie directe, voie indirecte).

Dans le détail, il y a tout de même quelques questionnements à se poser sur le mode de calcul aboutissant au signal dopaminergique. Si le signal dopaminergique correspond à une erreur de prédiction de la récompense alors d'où vient la récompense d'origine ? Où est calculée la prédiction de récompense et quel système est capable de soustraire la prédiction actuelle à la prédiction passée ? En résumé si l'on se ramène à l'équation du TD-Lambda (équation (5.1)) alors on sait que la dopamine correspond à :

$$Da \equiv r + \gamma \cdot V(s') - V(s) \quad (5.4)$$

Mais on ne connaît pas l'origine de chaque terme.

En réalité, il y a plusieurs éléments dans la littérature récente qui semble indiquer que les neurones du striosome encoderaient au moins $V(s')$ et peut-être $r + V(s')$ (YOSHIZAWA et al., 2018). Même s'il est probable que des parties de la récompense/punition soient directement projetées sur ATV et SNc depuis les noyaux qui forment les interfaces avec l'environnement et le corps (Hypothalamus, bulbes olfactifs, amygdale, etc.). L'hypothèse du striosome comme encodeur de la prédiction de récompense est maintenant assez classique dans les modèles des ganglions de la base (HOUK et al., 1995 ; DOYA, 2000 ; ITO et DOYA, 2015)

L'aspect le plus problématique finalement vient de la soustraction finale entre la récompense ajoutée à la prédiction actuelle et l'ancienne prédiction. Mais des recherches récentes (FUJIYAMA et al., 2011) semblent indiquer que le striosome projetterait non seulement sur la SNc, mais également sur SNr/GPi comme relai.

Cette découverte cadrerait avec la mise en évidence précédemment d'un circuit de désinhibition partant du striatum, passant par SNr en l'inhibant puis arrivant sur SNc via des liens inhibiteurs. La double inhibition aboutissant à une excitation indirecte des neurones dopaminergiques. (TEPPER et al., 1995 ; PALADINI et al., 1999 ; TEPPER et LEE, 2007 ; BRAZHNİK et al., 2008)

Il est alors possible que la SNr joue le rôle de médiateur temporel pour cette soustraction même s'il est assez étrange que l'on obtienne dans ce cas une équation inversée du type $V(s) - V(s')$ si l'on fait la supposition que les neurones de la zone du striosome représentent $V(s)$. Ce qui serait incohérent avec ce que l'on connaît de la sortie dopaminergique. Une solution serait que le chemin indirect soit ici plus rapide que le chemin direct, aussi étonnant que cela puisse paraître.

D'autres solutions sont envisageables en utilisant des interneurons, nous verrons une proposition dans ce chapitre.

Pour une comparaison générale plus fine entre les modèles de types TD et les ganglions de la base on pourra se reporter à (GLIMCHER, 2011) et pour une discussion plus complète sur les problèmes du modèle acteur-critique, notamment vis-à-vis du signe de $V(s) - V(s')$, ainsi que pour une revue générale de l'histoire de ces modèles : voir (JOEL et al., 2002).

Enfin, le dernier point important vient de l'apprentissage, en effet au moment où nous calculons le signal d'erreur de prédiction, on est déjà dans l'état d'après alors que cette erreur de prédiction doit s'appliquer, pour être juste, à la prédiction, et à l'action, de l'état précédent. Si l'on se reporte à l'équation (5.1) on observe en effet qu'au moment du calcul on doit être dans s' pour pouvoir calculer $V(s')$ alors que l'équation s'applique à modifier $V(s)$ et non $V(s')$.

Cet aspect peut se régler relativement simplement en introduisant une notion de mémoire retard dans les liens synaptiques. On sait en effet que la dopamine entraîne une cascade d'apprentissage si et seulement si certains facteurs chimiques représentant les activations sont présents dans les liens synaptiques, et on peut alors faire la supposition que ces facteurs ne représentent pas les activations actuelles, mais une intégration dans le temps plutôt représentative des activations passées. Finalement, l'apprentissage a donc bien lieu avec la dopamine calculée dans le « présent » selon les activations du « présent », mais il s'applique aux liens activés récemment dans le passé, et non à ceux qui viennent d'être activés.

Pour finir notons que la plupart des modèles acteurs critiques simplifient la relation entre la ligne de base de la dopamine et D1/D2. En effet, le système d'apprentissage inversé D1/D2 est cohérent dans le sens où la voie desservie par D1 et celle desservie par D2 représente des données inverses (l'une active, l'autre inhibe) il est donc cohérent qu'une récompense positive active plus la voie D1 et, au contraire, vienne activer moins la voie D2 pour l'empêcher d'inhiber cette situation. Néanmoins, dans un modèle simplifié computationnel on peut douter de l'intérêt d'étudier ceci dans un premier temps, en effet on peut voir cette séparation comme l'incapacité du système biologique, en général, à porter des valeurs positives et négatives par le même canal. Ce qui oblige automatiquement à former deux canaux séparés dans tous les cas, deux canaux pour l'action (indirect/direct) et deux pour l'apprentissage afin de les distinguer (D1/D2). Mais ces deux canaux ne sont pas strictement nécessaires sur un modèle computationnel simplifié dans lequel les valeurs négatives sont autorisées.

Dans le même esprit, la variation autour de la ligne de base de la dopamine et les phénomènes contraires qui adviennent alors peuvent être simplifiés en considérant que la dopamine peut être positive ou négative et que ces phénomènes d'adaptation à la ligne de base n'existent pas. Même si, là encore, on sacrifie sans doute certaines propriétés intéressantes.

5.6 | Conclusion

Dans ce chapitre nous avons discuté des raisons qui nous ont décidés à recourir à l'apprentissage par renforcement et nous avons également analysé de façon succincte le fonctionnement de l'une des structures du cerveau les plus impliquées dans l'apprentissage par renforcement : les ganglions de la base. Nous avons décrit cette structure tant du point de vue de l'apprentissage que du point de vue des mécanismes connus poussant à la considérer comme un centre de *sélection de l'action*.

Nous avons tracé un parallèle entre les systèmes d'apprentissage par renforcement classique comme le TD-Lambda ou le Q-learning, et le fonctionnement des ganglions de la base.

Nous avons ensuite analysé les raisons qui poussent à considérer un système acteur-critique comme une bonne modélisation, en première approche, des ganglions de la base. Pour ce faire, nous avons analysé quelles sont les hypothèses à poser pour que cette approximation soit juste.

Dans le chapitre suivant, nous proposerons un modèle acteur-critique et nous l'appliquerons sur des environnements simulés dans divers cadres expérimentaux afin de démontrer l'intérêt de ces systèmes dans l'interaction, tout particulièrement pour biaiser des trajectoires et décisions en fonction du contexte, ce qui pourrait permettre d'aider à *la lecture des intentions de l'autre*.

Apprentissage par renforcement via un modèle acteur-critique

Sommaire du chapitre :

6.1. Introduction	137
6.2. Modèle acteur-critique	138
6.3. Premier environnement simulé, contexte et résultats	142
6.3.1. Description de l'environnement	142
6.3.2. Validation du fonctionnement	143
6.3.3. Contexte et choix d'expérience	144
6.3.4. Premiers résultats et variante sans punition explicite	145
6.3.5. Contrôle en force : la problématique de l'apprentissage incrémental et du format des sorties	147
6.3.6. « Contexte social » : résultats préliminaires	155
6.3.7. Conclusion	158
6.4. Second environnement : simulation 2D d'un bras robotique	159
6.4.1. Description technique de la simulation du bras	159
6.4.2. Simulation du bras et résultats	165
6.5. Analyse et critiques	168
6.5.1. L'influence des méta-paramètres et leurs crédibilités neurobiologiques	168
6.5.2. Rapidité de convergence, un frein au développement sur robot ?	169
6.5.3. L'hypothèse sociale, une hypothèse simpliste qui mériterait d'être affinée	169
6.6. Conclusion	171

6.1 | Introduction

Dans le contexte d'une interaction avec un environnement, la prise en compte des frontières de cet environnement, de l'importance d'éviter les collisions, etc. est importante. Dans la partie précédente, le modèle qui fut présenté pour le contrôle d'un robot ne permettait pas de prendre en compte ces aspects.

En parallèle, dans le contexte d'une interaction sociale, la capacité à prédire l'intention de l'autre est primordiale. Les trajectoires des mouvements humains pourraient servir à aider cette prédiction, notamment en étant variable selon le contexte social, ce qui permettrait à un partenaire de tirer une conclusion quant à l'intention de l'agent par la simple observation de ses trajectoires.

Dans ce chapitre, nous souhaitons apporter des éléments de réponses aux deux problématiques via l'utilisation d'un modèle inspiré de l'étude, dans le chapitre précédent, des ganglions de la base et des modèles acteur-critiques. Nous supposerons en effet que les variations de trajectoires induites par ce type de modèle pourront rendre compte des phénomènes à la fois d'interaction avec l'environnement, et à la fois de certains phénomènes d'interaction sociale.

Comme nous l'avons vu dans la section 1.2 les robots ont de nombreux avantages quand il s'agit d'interaction et d'exploitation de caractéristiques motrices implicites toutefois ils portent malgré tout par nature un inconvénient majeur dans le cadre d'algorithmes d'apprentissages : le temps d'expérimentation nécessaire pour accumuler une expérience suffisante est très long dans le cadre d'algorithmes basés sur une exploration longue de l'environnement, tels que ces algorithmes d'apprentissage par renforcement. Comme nous l'avons évoqué : dans le cadre d'algorithme par renforcement ce temps long peut alors rendre caduque toutes tentatives d'implémentation sur robot, ou encore, peut forcer l'expérimentateur à recourir à des « astuces » pour permettre l'utilisation de ce type d'algorithme sur robot. Ces astuces sacrifiant malheureusement souvent au passage de nombreux aspects intéressants de l'apprentissage en interaction avec un robot (voir discussion de la section 1.2).

Nous prenons donc ici en compte ces éléments en considérant qu'il n'est probablement pas intéressant de développer ces algorithmes d'emblée sur robot. Nous sacrifions donc volontairement et en connaissance de cause ici les avantages du développement sur robot réel au profit d'un développement en environnement simulé. Afin de valider l'algorithme, nous le testerons alors sur des environnements simulés de complexité croissante.

6.2 | Modèle acteur-critique

En se basant sur ce que l'on a observé du fonctionnement des ganglions de la base dans le chapitre précédent et sur les modèles de type acteur-critique, on peut proposer un modèle acteur-critique dans l'idée d'en réaliser un qui soit compatible avec notre approche et nos précédents travaux. On présente le modèle général sur la figure 6.1.

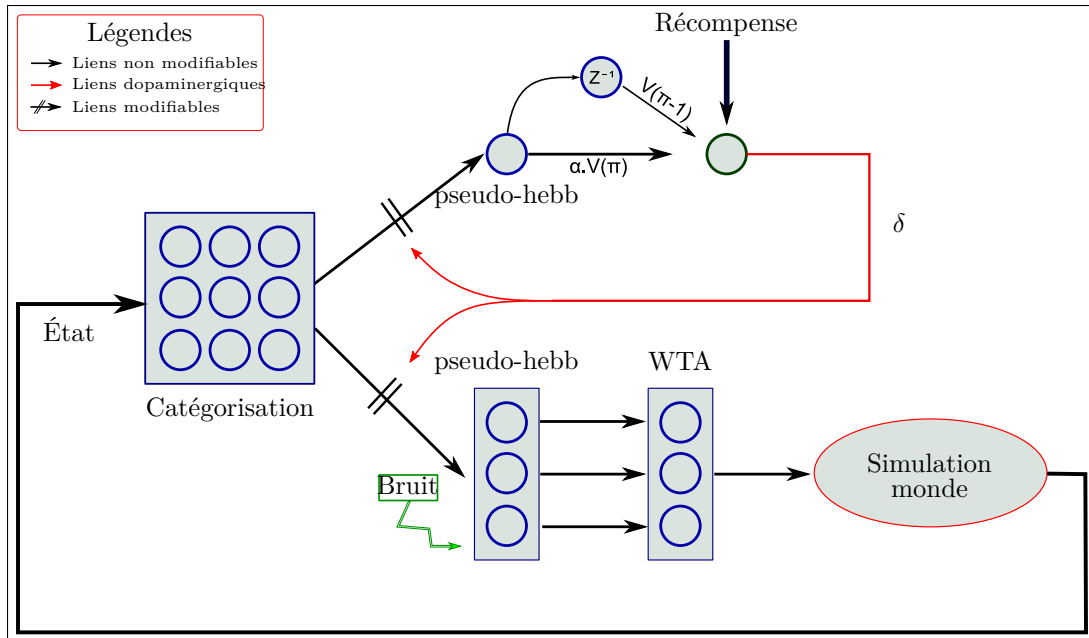


FIGURE 6.1. – Schéma de l'implémentation d'un modèle acteur critique.

Le modèle est composé selon l'hypothèse acteur-critique en deux branches séparées, la branche acteur qui prend les décisions d'action, et la branche critique qui calcule le signal de renforcement.

La première couche du système est composée d'une couche de catégorisation, de prime abord pour des raisons pragmatiques : le modèle peut ainsi fonctionner seul sans y adjoindre d'autres réseaux dans un premier temps, néanmoins ce n'est pas non plus contradictoire avec ce que nous apprend la littérature : certaines propriétés du striatum sont en effet proche de ce que l'on imagine d'un réseau ayant des propriétés de catégorisation, il a notamment été attesté la présence d'une topologie et de forts liens récurrents à l'intérieur même de la structure.

Selon les simulations, nous avons employé ici des catégorisations assez différentes. Sur la première simulation 2D que nous aborderons, le contrôle d'une bille, la simplicité des données entrantes nous a fait recourir à une organisation topologique simple de type Kohonen, qui pouvait alors converger assez rapidement. Pour les simulations suivantes nous avons employé nos réseaux similaires au K-mean que nous avons déjà décrit dans les parties précédentes, notamment dans les équations (4.1), (4.2) et (4.3).

La couche suivante est un peu particulière, car elle constitue le cœur du système. On emploie pour cette couche des équations de réactions de neurones formels classiques :

$$A_i = \sum_{\forall j} W_{ij} * E_j \quad (6.1)$$

Où E_j est le neurone j de l'entrée. Avec un calcul de la sortie du neurone en fonction de son activité A_i :

$$S_i = f(A_i) \quad (6.2)$$

Où f est une fonction d'activation non linéaire dont la sortie est définie tel que :

$$f(A_i) = \begin{cases} 1 & A_i \geq 1 \\ A_i & 0 \leq A_i \leq 1 \\ 0 & A_i \leq 0 \end{cases} \quad (6.3)$$

On utilise ensuite un apprentissage un peu particulier pour ces deux couches, car, comme nous l'avons vu sur l'équation du TD-Lambda (équation (5.1)) et la discussion précédente sur les modèles acteur-critique, il est important que les apprentissages se déroulent sur une image des activités passées, et non sur l'activité synaptique actuelle.

On suppose donc l'existence d'un élément chimique qui intègre les co-activations passées sous la forme :

$$Ele(W_{ij}) = \alpha * (E_j * f(A_i))_{t-1} + \alpha * \beta * (E_j * f(A_i))_{t-2} + \alpha * \beta * \beta * (E_j * f(A_i))_{t-3} \dots \quad (6.4)$$

L'équation est réalisée de façon récursive en pratique en multipliant par β l'ancien résultat de $Ele(W_{ij})$ pour le mettre à jour. α est un facteur multiplicatif global qui va désigner le taux de prise en compte du passé en général dans Ele . Ele est une fonction dont les valeurs sont spécifiques à chaque lien W_{ij} (on suppose l'élément en question présent au niveau du lien synaptique). $E_j * f(A_i)$ représente l'ancienne co-activation, d'un point de vue biologique on peut dire qu'elle représente une activité des récepteurs NMDA qui réagissent à la dépolarisation du neurone uniquement si celle-ci s'accompagne d'activités sur la synapse. C'est ce récepteur qui va ensuite entraîner la chaîne d'apprentissage. On fait ici la supposition que Ele représente les conséquences de cette chaîne au niveau pré-synaptique. β représente la pente de diminution à chaque itération (l'oubli du passé).

Dans ces travaux, les facteurs pour α et β n'ont pas eu de grande influence. α doit être un peu inférieur à 1, mais peut être compensé par la vitesse d'apprentissage par la suite. β a une influence un peu plus importante, car il agit sur la profondeur du passé qui sera prise en compte. Il est important ici de noter qu'en faisant cela on s'éloigne de l'équation classique d'un TD-Lambda qui se contente de propager l'apprentissage sur l'état précédent. Si on devait la réécrire, l'équation résultante serait alors :

$$\begin{aligned} V(s) &\Leftarrow V(s) + \alpha \cdot (r + \gamma \cdot V(s') - V(s)) \\ V(s-1) &\Leftarrow V(s-1) + \beta \cdot \alpha \cdot (r + \gamma \cdot V(s') - V(s)) \\ V(s-2) &\Leftarrow V(s-2) + \beta^2 \cdot \alpha \cdot (r + \gamma \cdot V(s') - V(s)) \\ V(s-3) &\Leftarrow V(s-3) + \beta^3 \cdot \alpha \cdot (r + \gamma \cdot V(s') - V(s)) \\ V(s-4) &\Leftarrow V(s-4) + \beta^4 \cdot \alpha \cdot (r + \gamma \cdot V(s') - V(s)) \end{aligned} \quad (6.5)$$

Mais en plus il faut bien comprendre qu'ici on fait une analogie entre $s-1$ et $t-1$, ce qui n'est pas totalement juste, car un état peut ne pas évoluer entre deux intervalles de temps. La solution que nous avons adoptée dans un premier temps est de considérer que les pics d'émissions dopaminergiques ne sont pris en compte que pour des changements dans l'activité des états d'entrée. Ce qui peut s'expliquer par beaucoup de mécanismes divers puisque les neurones ont tendance à réagir de façon différentielle (un changement des activités d'entrées entraîne un apprentissage plus intense), et que l'on peut également considérer les nombreuses boucles inhibitrices existantes entre le reste des ganglions de la base et SNC, qui ont peut-être pour rôle d'inhiber les pics d'activités dopaminergiques lorsqu'il y a stagnation d'un état.

Notons également qu'en pratique cela peut ne pas poser de problèmes. En effet dans nos tests qualitatifs le retrait de cette fonction n'a pas changé qualitativement les résultats, mais a plutôt demandé un apprentissage plus long. On peut en plus noter qu'il est peu probable que deux états soient strictement identiques entre deux intervalles de temps dans un contexte biologique vu que beaucoup d'entrées considérées peuvent être dépendantes du temps. On pense notamment aux entrées de l'hippocampe sur le striatum, qui sont connues pour représenter des transitions entre états, et donc des informations qui ne sont par essence que temporaires.

En résumé, on fait ici deux suppositions en tout, qui se sont vérifiées en pratique :

- En moyenne, l'équation est pertinente même si l'état d'entrée ne change que peu dans un certain intervalle de temps et même sans le processus d'adaptation de la dopamine émise selon la variation des états.
- En moyenne, même si la nouvelle forme de l'équation présentée dans l'équation (6.5) est assez éloignée du TD-Lambda d'origine, nous faisons la supposition qu'elle reste pertinente pour des facteurs β suffisamment petits. En effet, par l'absurde on se rend bien compte que si $\beta = 1.0$ alors l'équation devient clairement fautive et ne peut pas converger.

Enfin, notons une petite subtilité dans le cas de la partie critique. En effet, comme nous employons une dopamine qui peut-être indifféremment positive et négative afin d'éviter de simuler les subtilités des récepteurs D1 et D2 nous modifions alors l'équation de cette partie-là pour autoriser ce genre de sortie. Dans le cas de la partie critique, l'équation (6.3) devient alors :

$$f(A_i) = A_i \tag{6.6}$$

Sans filtrage supplémentaire particulier. On autorise ainsi la dopamine à être aussi forte ou faible que souhaité dans le négatif et positif. Ceci permet de se rapprocher des effets de ligne de base et du potentiel émetteur en groupe des champs de neurones constituant la SNC.

De même, une modification de l'équation d'apprentissage est nécessaire pour prendre en compte cet aspect. En effet, la notion de co-activation n'a plus forcément de sens sur un signal négatif, le risque étant alors, face à une dopamine négative tentant de faire désapprendre l'état actuel, d'aboutir à une multiplication de deux signaux négatifs (la dopamine entrante et la co-activation négative) ce qui ferait croire l'apprentissage, soit l'inverse de l'effet voulu. D'autant plus que la notion de co-activation négative n'a pas vraiment de sens.

Pour compenser ces défauts introduits par notre simplification on modifie l'équation d'apprentissage (6.4) pour ne pas prendre en compte l'effet de co-activation. On se base ici sur l'idée que la dopamine a toujours une activité (c'est l'activité que l'on dit « tonique » de la dopamine, ce qui est une propriété spécifique des neurones dopaminergiques) et que les neurones dopaminergiques sont donc toujours susceptibles d'apprendre face à un signal extérieur. L'autre point de vue sur la question est de considérer qu'il existe deux types d'apprentissages dopaminergiques dans la littérature : l'un dépend de l'état du neurone, et notamment des activités passées qui ont conduit, ou non, à une chaîne de renforcement de type LTP ou LTD. Et dans ce cas, la dopamine ne fait que renforcer ou non le processus. L'autre type quant à lui n'est par contre pas fonction de l'état d'activité du neurone. Dans la littérature, on parle alors d'effet « cAMP independent » (cAMP pour *Cyclic adenosine monophosphate* est une molécule impliquée dans la chaîne d'apprentissage lorsqu'un neurone subit une dépolarisation en parallèle d'une excitation).

Dans les deux cas, ceci peut donc justifier notre modification, l'équation (6.4) devient, dans le cas de la partie critique :

$$Ele(W_{ij}) = \alpha * (E_j)_{t-1} + \alpha * \beta * (E_j)_{t-2} + \alpha * \beta * \beta * (E_j)_{t-3}... \tag{6.7}$$

Où l'on suppose donc l'activité de $f(A_i)$ constante, ou du moins suffisante pour que l'apprentissage ait lieu.

Ensuite, maintenant que nous disposons de ce facteur Ele pour la partie critique et acteur, nous l'appliquons à une équation d'apprentissage classique tenant compte d'une neuromodulation dopaminergique :

$$W_{ij}(t) = W_{ij}(t-1) + \lambda * Ele(W_{ij}) * \delta \tag{6.8}$$

Où $Ele(W_{ij})$ est l'activité du facteur chimique supposé représentant ce qu'il y a à apprendre au niveau de la synapse et δ la dopamine délivrée. λ est un facteur d'apprentissage qui doit être suffisamment petit pour que la convergence soit efficace.

Les parties suivantes de l'algorithme sont les simples applications de ce que l'on a vu des modèles acteurs critiques et de la structure des ganglions de la base dans les sections précédentes : on calcule

la dopamine en rajoutant à la prédiction de récompense (multipliée par un facteur pour ne prendre en compte qu'une portion de la prédiction future, comme le facteur γ dans l'équation TD) la récompense instantanée R et en y soustrayant l'ancienne prédiction. Du côté de l'acteur, l'apprentissage se fait sur plusieurs neurones indépendants, ce qui permet d'aboutir à un champ de neurone représentant une action. On applique ensuite la simplification classique des voies directes et indirectes comme étant un type de *Winner Take All*.

Il faut être ici conscient qu'il s'agit d'une grossière simplification, et nous verrons que la structure de cette sortie est assez importante dans notre cadre, surtout quand on se pose la question d'un contrôle de forces. Nous testerons donc plusieurs genres de sorties dans les résultats que nous allons présenter.

6.3 | Premier environnement simulé, contexte et résultats

Dans cette partie nous décrivons les premiers travaux effectués avec un modèle simple d'une bille évoluant dans un environnement en deux dimensions et contrôlable par des forces quelconques. Ce modèle très simpliste a vocation ici à modéliser l'effecteur terminal d'un bras vu de côté. Nous utiliserons ce modèle pour étudier les trajectoires produites par le système acteur-critique sous diverses contraintes afin de jauger de sa capacité à converger vers des comportements proches de ceux de l'humain dans une même situation.

Nous avons privilégié dans cette thèse des environnements simulés minimalistes afin de permettre une exécution très rapide de l'environnement et donc des temps de simulation réduits.

6.3.1. Description de l'environnement

Cet environnement est un modèle physique en deux dimensions vu de côté et représenté par ses axes x et y , dans lequel une bille ponctuelle de masse m peut se mouvoir (voir figure 6.2). La bille est soumise à son poids et on y applique simplement les équations de Newton qui en découlent.

Par souci de simplicité, on borne l'environnement sur une distance d'un mètre virtuel en hauteur comme en largeur. La bille peut être guidée par des forces aux composantes projetées sur x et y que l'on pourra contrôler. On fait ici la très grosse simplification que certains problèmes posés par le contrôle des forces ponctuelles sur une bille sont similaires aux contrôles du bras. En réalité, l'hypothèse est grossière et on se servira simplement de cet environnement pour démontrer certains concepts que l'on appliquera sur une simulation plus complexe ensuite.

Dans cet environnement, on pourra appliquer un jeu de friction qui viendra ralentir la bille selon des zones que l'on pourra choisir selon une équation de frottement fluide relatif à la vitesse.

Par ailleurs, la bille subit des forces qui viennent annuler son mouvement lors de son impact sur les bordures de l'environnement. On considérera son élasticité nulle et il n'y a donc aucun rebond.

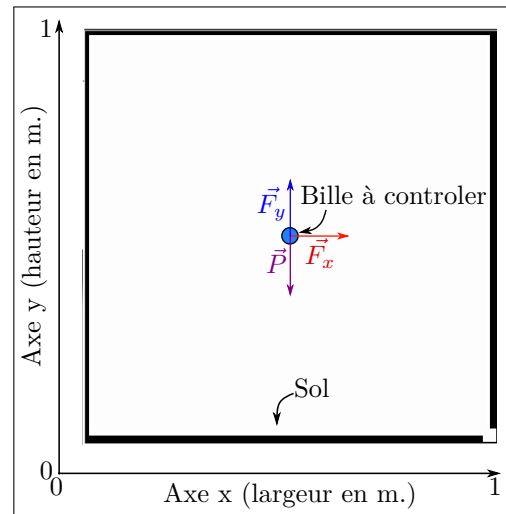


FIGURE 6.2. – Simulation d'un environnement en deux dimensions

Implémentation

On cherche à calculer la position de la bille en fonction des forces appliquées. D'après la 2^e loi de Newton :

$$\begin{aligned} \sum \overrightarrow{\text{forces}} &= m \cdot \vec{a} \\ \Leftrightarrow \vec{F} + \vec{P} &= m \cdot \vec{a} \\ \Leftrightarrow \vec{a} &= \frac{1}{m} \vec{F} + \vec{g} \end{aligned} \quad (6.9)$$

avec m la masse de la bille (non nulle), \vec{a} l'accélération instantanée de la bille, \vec{F} la force appliquée pour déplacer la bille, \vec{P} le poids subi par la bille et \vec{g} l'intensité de la pesanteur (on fixe $g = 9,81 \text{ m.s}^{-2}$).

En projetant sur les axes x et y , on obtient :

$$\begin{cases} a_x = \frac{F_x}{m} \\ a_y = \frac{F_y}{m} - g \end{cases} \quad (6.10)$$

On utilise ensuite la méthode d'Euler pour la résolution d'équations différentielles afin de calculer la vitesse :

$$\begin{aligned} & \begin{cases} v_x(t + \Delta t) = v_x(t) + a_x \cdot \Delta t \\ v_y(t + \Delta t) = v_y(t) + a_y \cdot \Delta t \end{cases} \\ & \Leftrightarrow \begin{cases} v_x(t + \Delta t) = v_x(t) + \frac{F_x}{m} \Delta t \\ v_y(t + \Delta t) = v_y(t) + \frac{F_y}{m} \Delta t - g \cdot \Delta t \end{cases} \end{aligned} \quad (6.11)$$

avec Δt le pas de temps de la simulation.

On utilise finalement la même méthode pour déterminer la position grâce à la vitesse et à la position précédente :

$$\begin{cases} x(t + \Delta t) = x(t) + v_x \cdot \Delta t \\ y(t + \Delta t) = y(t) + v_y \cdot \Delta t \end{cases} \quad (6.12)$$

Ainsi, nous pouvons calculer la position de la balle en fonction des forces qui lui sont appliquées. C'est cette position qui servira ensuite de donnée entrante à l'algorithme.

6.3.2. Validation du fonctionnement

Pour valider le fonctionnement de l'algorithme, on connecte l'algorithme proposé à l'environnement avec une sortie constituée de 9 forces possibles discrétisée (haut-bas-droite-gauche, leurs combinaisons ainsi qu'une action nulle) et en lui présentant simplement les positions X/Y de la balle comme données entrantes de l'environnement, soit le cas le plus simple possible comme contexte de test.

On positionne ensuite une récompense dans l'environnement. À cet endroit l'algorithme reçoit une récompense instantanée et l'expérience est réinitialisée. La balle est, elle, positionnée à une position de départ à l'opposé de la récompense et sa vitesse est nulle (et le restera pour toutes les expériences présentées sur cet environnement).

On teste l'algorithme dans deux situations, dans un premier test on valide la partie « critique » de l'algorithme en considérant une récompense placée en hauteur et en ne s'occupant pas particulièrement de la trajectoire. On cherche ici à vérifier si la propagation de récompense due à la structure de l'algorithme similaire à un TD-Lambda est bien fonctionnelle. On peut observer le résultat de la propagation de la récompense ensuite en présentant artificiellement des entrées à l'algorithme pour récupérer les valeurs de la partie « critique ».

Le résultat est présenté sur la figure 6.3 où l'on peut effectivement observer une propagation de récompense dans les zones les plus proches statistiquement des chemins allant du point de départ à la récompense.

Notons d'ores et déjà ici une propriété intéressante de ce type d'algorithme : l'exploration et l'exploitation y sont implicitement liées puisqu'au fur et à mesure

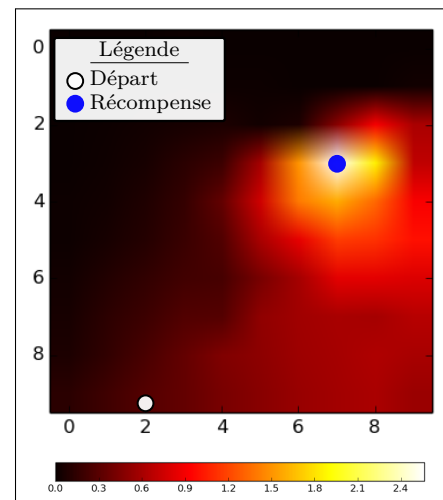


FIGURE 6.3. — Résultat de la propagation de récompense visualisé en deux dimensions

de l'apprentissage le « bruit » finit par être trop faible par rapport aux actions apprises pour pouvoir gagner dans la partie « sélection » de l'algorithme. Ceci aboutit au fait qu'implicitement l'algorithme a tendance à guider de plus en plus son exploration au fur et à mesure de la convergence vers un chemin unique optimum menant à la récompense.

Qualitativement, on peut observer que l'algorithme peut mettre longtemps à se stabiliser sur un unique chemin, mais qu'en pratique il va choisir des chemins valides, quoiqu'instables, bien plus rapidement. Un autre aspect intéressant est que l'exploration est forcément guidée par les lois de l'environnement lorsque le bruit est constant. Ici par exemple la bille est soumise à la gravité, et l'exploration de positions en hauteurs demande donc de maintenir durant plusieurs itérations une activité ayant une composante en y positive. Statistiquement, ceci arrive assez rarement et l'algorithme a donc du mal à explorer les positions en hauteur et a tendance à se limiter à un espace atteignable restreint par la gravité.

Cette propriété n'est pas tout à fait anodine dans le cas d'un bras humain et donc également dans le cas de la robotique humanoïde, car elle signifie que si les premières explorations ne sont que le fruit d'un type de « bruit » sur les décisions, sans être guidé spécifiquement, alors l'exploration aura tendance à se concentrer autour de l'espace ayant le plus bas potentiel. Dans le cas de l'humain ou du bébé : sur les parties basses de son environnement. Ceci est assez pratique si l'on considère que beaucoup d'objet et jouets avec lesquels l'enfant peut interagir dans un premier temps sont avant tout au sol et qu'au contraire il est peu intéressant d'explorer les zones au-dessus de sa tête en premier lieu. Cela a l'avantage de concentrer l'exploration dans la zone la plus utile dans un premier temps et donc d'optimiser implicitement l'apprentissage.

Nous validons ensuite l'algorithme dans un second cas en nous intéressant aux trajectoires. Nous plaçons la récompense sur le sol à l'opposé de l'environnement et nous observons si l'algorithme est bel et bien capable de converger vers une trajectoire fixe pour l'atteindre. On présente le résultat sur la figure 6.4. Cette figure, qui constitue en fait la juxtaposition de plusieurs trajectoires après convergence, nous permet de montrer que l'algorithme suit bien une trajectoire fixe sans aucune variation après convergence dans un cas très simple. On observe également qu'il a bien convergé vers le chemin optimal « évident » qui est la ligne droite dans ce cadre, et non vers d'autres types de trajectoires.

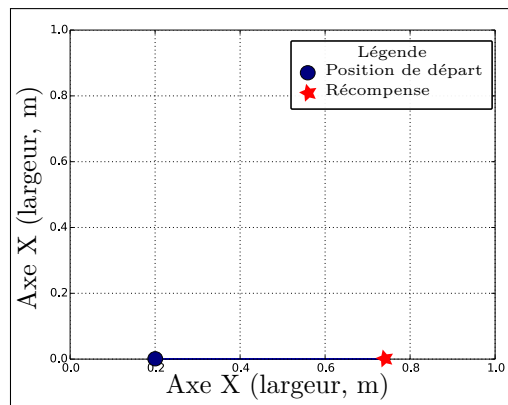


FIGURE 6.4. — *Convergence de l'algorithme vers la ligne droite : sans frottements ni punitions, le modèle va du point de départ à la récompense en ligne droite.*

6.3.3. Contexte et choix d'expérience

Comme nous l'avons abordé dans les parties précédentes, le mouvement humain a été très largement étudié à travers de nombreux types de modèles, et nous avons montré que certains aspects de ce mouvement peuvent s'obtenir via le corps, la structure des muscles, des boucles réflexes, et des algorithmes associatifs simples. Toutefois, de nombreuses propriétés du mouvement humain ne peuvent être expliquées en se limitant à ces aspects. Il est par exemple impossible d'expliquer de nombreux effets présents sur le mouvement humain dans des cadres d'interactions avec d'autres humains, tel que les expériences d'Ioanna Georgiou et collègues (GEORGIU et al., 2007) qui montre que l'intention, donner un objet à un humain ou le poser au sol, fait varier les formes de trajectoires. Comme nous l'avons abordé dans la section 3.2.1.

Il est également impossible de comprendre les interactions entre le mouvement et l'environnement. En quoi l'environnement vient-il changer la dynamique et la forme du mouvement et comment cet environnement est-il pris en compte ?

De nombreux algorithmes s'attachent à reproduire les formes exactes du mouvement en usant d'apprentissage par démonstration, parfois d'ailleurs avec le même type de modèle que celui présenté ici. Ici, nous allons essayer de démontrer dans cette première partie l'influence de l'environnement sur un mouvement dont les actions sont discrètes, dans un premier temps, en usant simplement de récompenses et punitions ponctuelles relatives à l'environnement.

Pour ce faire, nous allons étudier dans un premier temps ce qui change la forme de la trajectoire lorsque l'on souhaite prendre ou déposer un objet sur un plan.

Modification de l'environnement

Dans un premier temps, on teste une hypothèse simple : les frottements d'une surface sont une source de douleur lorsque l'on réalise un mouvement proche du sol ou d'un plan.

On veut alors vérifier qu'avec simplement cette hypothèse et la récompense, on est capable de converger vers des trajectoires humaines typiques en forme de cloche que l'humain produit lorsqu'à partir d'un plan il tente d'atteindre un objet sur le plan et disposé plus loin. L'environnement est alors modifié comme exposé sur la figure 6.5 pour inclure une zone de frottement. La punition est conçue pour être relative à une équation de frottement classique dépendante de la vitesse tel que :

$$\text{si } Y_{bille} < hauteur_{zone} \text{ alors} \\ | \quad p = \beta \cdot \dot{X}_{bille}$$

Où p est la punition appliquée à chaque instant qui vient se soustraire à la récompense instantanée R dans le modèle et où \dot{X}_{bille} représente la vitesse de la bille selon l'axe des X et Y_{bille} la position de la bille selon l'axe des Y . Le seuil est ici représentatif de la hauteur de la zone puisque l'on souhaite punir l'effecteur quand il frotte le sol. On choisit cette hauteur arbitrairement à $hauteur_{zone} = 0,01$ m, c'est-à-dire que l'on considère qu'il y a frottement dès que la bille se retrouve en dessous de 1 cm de hauteur. Ceci nous évite simplement les phénomènes calculatoires qui rendent complexe la comparaison exacte de deux points représentant un nombre à virgule dans les langages informatiques. De fait ici le « contact » est défini de manière un peu souple, mais choisi comme suffisamment faible pour que cela n'impacte pas le résultat.

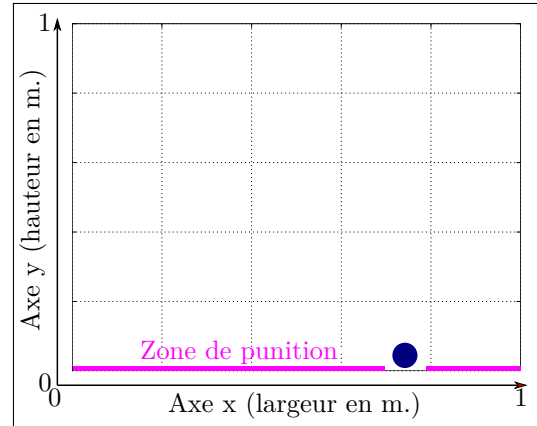


FIGURE 6.5. — Environnement modifié avec frottement en y incorporant une zone de punition lorsque la bille frotte au sol.

6.3.4. Premiers résultats et variante sans punition explicite

Après convergence de l'algorithme, on obtient effectivement le résultat attendu de courbes typiques « en cloche » tel qu'elles existent chez l'humain dans ce type de situation. On peut observer le résultat sur la figure 6.6.

On y observe bien le comportement de l'algorithme attendu, à savoir qu'il tente d'éviter dans son contrôle la zone qui pourrait lui occasionner de la douleur via le frottement. On peut observer les différentes zones du contrôle grâce aux flèches représentant les forces et leurs directions.

Nous avons donc bien un réseau capable de prendre en compte l'environnement dans son contrôle pour atteindre un objet. Notons que l'entrée des « états » est ici composée uniquement de la punition, l'algorithme arrive à se stabiliser malgré le fait que l'inertie peut rendre les conséquences des actions très différentes selon l'état « position » en cours, selon les actions passées. Probablement parce que la

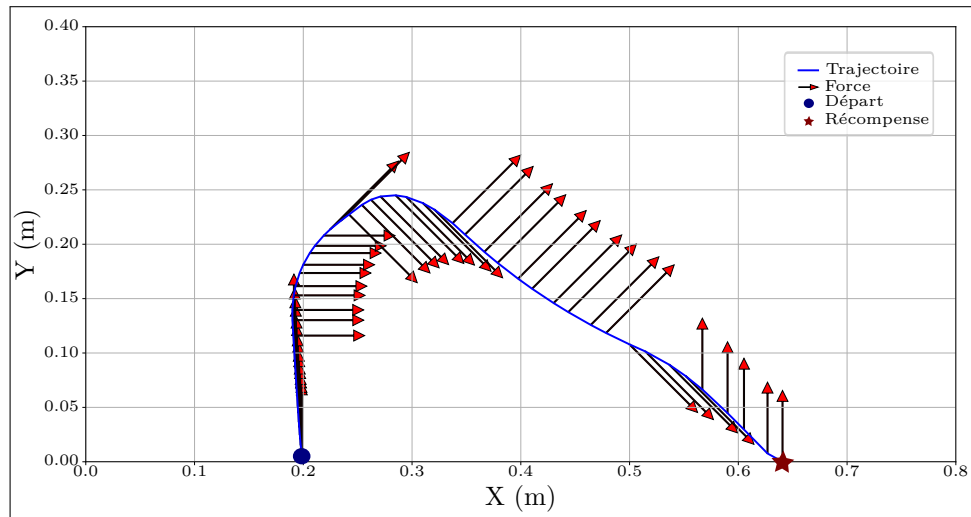


FIGURE 6.6. — *Trajectoire obtenue après convergence. On obtient une trajectoire « en cloche » typique des trajectoires humaines dans la même situation.*

convergence pas-à-pas de l'algorithme a tendance à réduire peu à peu l'ensemble des actions passées possibles pour un état donné.

Malgré le bon fonctionnement de l'algorithme, pour obtenir ce résultat on a dû supposer l'existence d'une zone de punition, si l'hypothèse semble crédible est-elle réellement nécessaire ?

Utilisation implicite de la rapidité du chemin à parcourir au lieu de punitions

On peut ainsi proposer une variante de l'expérience dans laquelle il n'y a plus de punition, mais un ralentissement de l'objet lorsqu'il est au contact en suivant une équation classique de frottement de type :

$$\begin{array}{l} \text{si } Y_{bille} < hauteur_{zone} \text{ alors} \\ | \quad Fx_{frot} = -k * \dot{X}_{bille} \end{array}$$

Où Fx_{frot} vient s'ajouter aux forces déjà appliquées sur la bille (contrôle et gravité) et k un facteur de frottement. $hauteur_{zone}$ est fixé de la même manière que précédemment.

On peut observer le résultat après convergence sur la figure 6.7

Ici on profite d'une propriété émergente intéressante de notre algorithme. Comme les états sont dépendant du temps dans notre algorithme, automatiquement, si l'on se retrouve ralenti sur une portion de trajectoire alors on considérera plus de pas de temps sur cette portion et donc plus d'états. Ce qui veut aussi dire que la récompense devra être propagée sur plus d'états. De façon imagée, si on doit emprunter deux chemins parallèles, l'un rapide et l'un lent, alors le rapide sera parcouru en deux pas de temps et l'algorithme verra une succession d'états menant à la récompense $A \rightarrow B \rightarrow R$. De fait, la récompense se propagera depuis R vers B et ensuite vers A et ne sera atténuée, dans la convergence, que deux fois. Sur le chemin lent, on parcourra plutôt un chemin du type $A \rightarrow A' \rightarrow B \rightarrow B' \rightarrow R$ avec, donc, des états intermédiaires dus à l'élongation du temps de trajet. La récompense R se retrouvera alors atténuée quatre fois au lieu de deux. Cela signifie qu'en A , l'acteur ayant le choix d'emprunter l'un de ces deux chemins, se verra récompenser plus fortement sur le chemin rapide que sur le chemin lent.

De fait, il aura tendance à emprunter le chemin rapide. On voit bien ici comment ceci modifie la propagation dans notre environnement : le chemin en cloche est plus long, certes, mais plus rapide, que la

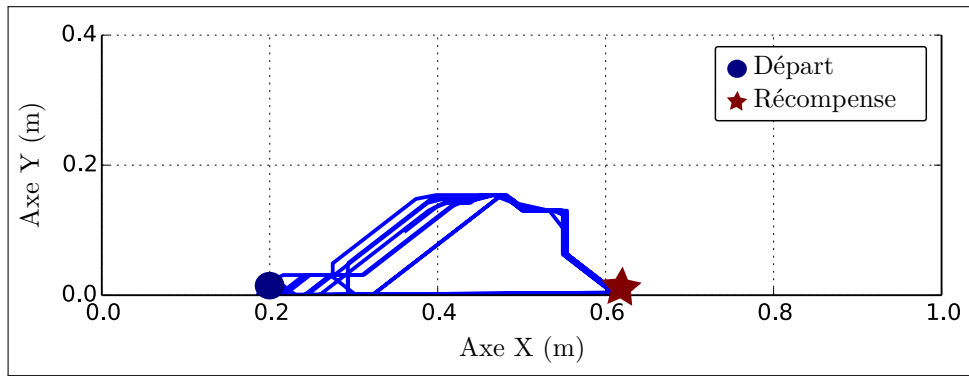


FIGURE 6.7. — *Résultat dans le cas d'un frottement sans punition. On converge vers une courbe en cloche typique bien que l'on ait retiré les punitions explicites. Ici seul le temps de parcours fait qu'implicitement le chemin le plus court est sélectionné. La courbe n'a pas été capturée dans la même série d'expériences que 6.6 et pas avec les mêmes paramètres. Le peu d'échantillonnage explique ici l'aspect saccadé.*

ligne droite qui se retrouve, elle, ralentit par la force de frottement. Ceci dépend évidemment du facteur β de la force de frottement, si celui-ci est trop faible le système préférera la ligne droite. L'implication de ce phénomène est assez importante puisque cela nous assure que cet algorithme est capable de préférer les actions menant aux résultats avec le moins de détours lorsque ce n'est pas nécessaire, ce qui est une caractéristique essentielle des actions humaines. Et cela sans qu'il soit nécessaire de rajouter de facteurs explicites à ce propos dans l'algorithme.

6.3.5. Contrôle en force : la problématique de l'apprentissage incrémental et du format des sorties

Nous avons présenté ici un algorithme dont la sortie est immédiatement un pavage d'un espace moteur en vitesse très simple représentant une seule force par direction. Néanmoins, d'autres types de sorties peuvent être imaginés, notamment pour tenter de permettre à l'algorithme de contrôler différentes forces. Mais fonctionnent-elles tout aussi bien ? Par exemple, on peut faire en sorte que l'algorithme apprenne directement des forces f_x et f_y au lieu de lui fournir un pavage en sortie, sans sélection particulière de l'action, ce qui aurait le mérite de permettre un contrôle en force fin sans avoir à segmenter le problème.

Mais on s'attend à ce que le résultat ne converge pas : en effet, comme le modèle est un modèle d'apprentissage incrémental, si l'on exécute par hasard une action f_x/f_y dans un état s et que cette action nous amène dans un état s_2 qui est considéré comme positif et donne donc lieu à une récompense : alors le modèle d'apprentissage fait que l'algorithme aura appris à associer l'état s avec une portion de cette action : $\alpha \cdot f_x/\alpha \cdot f_y$. La fois suivante, lorsqu'il se retrouvera

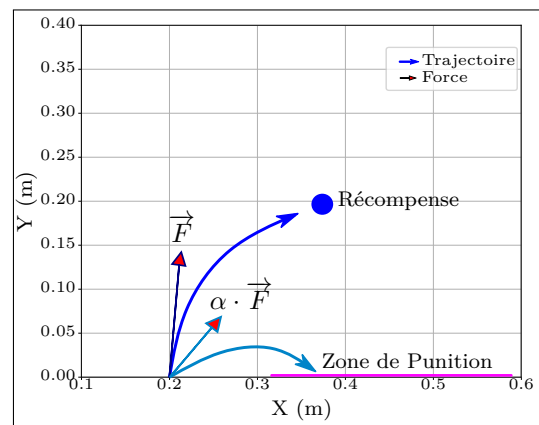


FIGURE 6.8. — *Illustration du problème de la commande directe en force sans segmentation : si en explorant on découvre une force à même de transporter l'effecteur dans un état de récompense alors la fois suivante la portion de force apprise peut être insuffisante et nous amener dans un état négatif, annulant l'apprentissage.*

dans l'état s il exécutera alors le doublet de forces $\alpha \cdot f_x / \alpha \cdot f_y$. Toutefois, ce ne sont que des portions de la force nécessaire pour atteindre l'état positif s_2 , et donc, en prenant en compte la gravité et la vitesse initiale de la bille, il est tout à fait possible que le contrôle rate l'objectif et tombe dans un état « s_3 » qui peut tout à fait, par exemple, être un lieu de punition. On illustre ce problème sur la figure 6.8.

De fait, l'algorithme désapprend ce qu'il a appris puisque $\alpha \cdot f_x / \alpha \cdot f_y$ l'a mené à une action négative. La fois d'après, quand il rencontrera par hasard une fois de plus la combinaison de force f_x / f_y associée à l'état s , cette chaîne d'événement aura encore lieu. Ce qui causera une oscillation de l'algorithme, mais sans jamais converger.

Si on tente l'expérience en pratique on obtient le résultat observable sur la figure 6.9.

Le résultat est un peu plus positif que ce qui était attendu. On s'attendait à ce que l'algorithme soit incapable de converger vers une trajectoire fixe, et c'est bien le cas, mais l'algorithme arrive tout de même à rejoindre la cible la majeure partie du temps, même si sa trajectoire est très instable d'une fois sur l'autre. Pour nous, cela montre bien à quel point ce type d'algorithme est robuste, mais cela reste trop instable pour valider la solution.

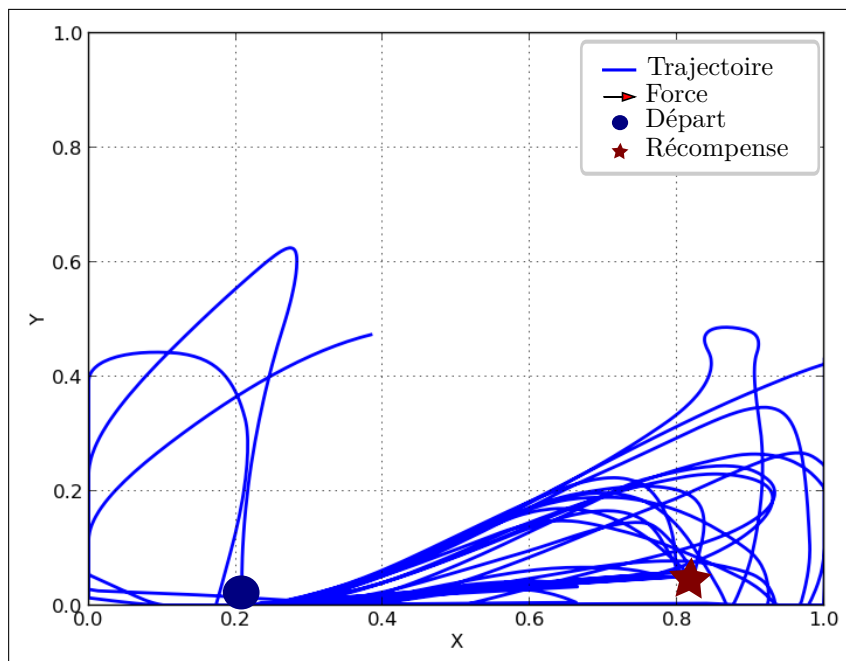


FIGURE 6.9. — *Résultat dans le cas d'un effecteur analogique non segmenté : on peut observer le défaut de ce type d'effecteur : l'algorithme ne converge pas vers un unique chemin optimal. Notons tout de même que contrairement à ce qui était attendu, l'algorithme arrive à se diriger grossièrement vers la récompense.*

Pour compenser ceci, on peut proposer un autre type d'effecteur, un effecteur dans un espace pavé, mais représentant des forces différentes dans chaque direction. On ne présente pas le résultat intermédiaire ici, mais discutons simplement du résultat qualitatif : ce type d'effecteur fonctionne parfaitement et permet effectivement de contrôler plusieurs forces, néanmoins il présente un désavantage assez important.

En effet si on considère ce type d'effecteur alors une force \vec{F} est représentée par une action totalement différente d'une force \vec{F}' très proche. De fait, si deux forces proches mènent à un même état,

l'apprentissage de l'une ne profitera pas à l'apprentissage de l'autre, ce seront deux apprentissages totalement séparés. Ainsi dans un monde analogique, il peut y avoir de nombreuses situations ambiguës où, selon la finesse du pavage des forces proposées pour l'acteur on aboutit, pour un même état, à plusieurs apprentissages en parallèle visant à atteindre le même point avec des forces légèrement différentes.

Cela fonctionne évidemment après un long temps de convergence, mais cela pose un problème de longueur d'apprentissage, qui se trouve multiplié par le nombre de forces « ambiguës » qui sont proches dans l'espace d'action. Intuitivement, cela choque également qu'une force proche d'une autre soit absolument indépendante de sa cousine vis-à-vis de l'espace d'action et qu'elle ne profite donc aucunement de l'apprentissage de sa cousine.

Une solution que l'on propose pour répondre à ce problème est alors de s'inspirer des commandes musculaires chez l'humain, dans lesquelles chaque fibre musculaire produit une petite portion de la force musculaire totale. On suppose alors qu'il en est de même ici au niveau de la commande : chaque élément ne représentant qu'une portion de force qui peut s'additionner aux autres pour former un tout.

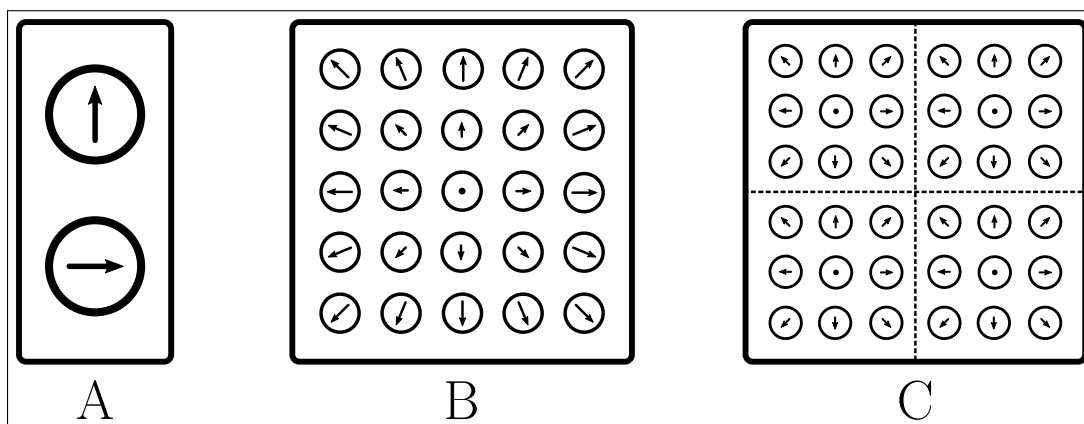


FIGURE 6.10. – On présente ici trois classes d'effecteur analogique capable de commande en force. La classe A n'est pas soumise à une sélection, c'est la simple commande des forces F_x et F_y . La classe B pave l'espace des forces possibles selon leurs amplitudes et leurs directions, la classe C fonctionne par petites parties séparées : chaque partie est capable de sélection, mais indépendamment des autres, chacune ne représente qu'une petite amplitude de force, c'est la somme des gagnants de chaque sous-partie qui va donner une force totale pour la commande.

Ce type d'effecteur (représenté sur la figure 6.10 dans la catégorie C) comprend les deux avantages de l'effecteur analogique (A) et de l'effecteur « pavé » (B) : comme l'effecteur analogique, une force \vec{F} est contenue dans la force \vec{F}' plus grande. Ainsi si l'apprentissage oscille entre deux forces proches, \vec{F} et \vec{F}' qui mènent au même état elles seront très proches sur cet effecteur : la force \vec{F} activera par exemple 3 unités dans une direction, alors que la force \vec{F}' activera les 3 mêmes unités + une nouvelle unité. Ceci permet à un apprentissage ayant lié un état s à \vec{F} de réutiliser les liens qui ont déjà convergé pour lier s à \vec{F}' , ce qui est bien plus cohérent que l'effecteur « pavé ». Par contre, contrairement à l'effecteur analogique, et comme l'effecteur pavé, la force est discrétisée et chaque unité de force peut alors être sélectionnée vis-à-vis de ses voisines.

Pour que cet effecteur fonctionne au mieux il est préférable que l'ordre d'activation de chaque unité soit toujours le même afin qu'une force \vec{F} ne soit représentable que par un unique ensemble de neurones, ce qui simplifie l'apprentissage. Ceci n'est pas un problème vis-à-vis de la cohérence avec la neurobiologie a priori, car comme nous l'avons vu : ce type de comportement existe dans l'ordre d'activation des fibres musculaires (voir section 4.2.1 dans la partie précédente), qui ne s'activent que dans un ordre précis

relatif à leurs tailles. Ce qui est donc très pratique dans ce cadre pour éviter l'apprentissage de deux motifs différents représentant une unique force. Néanmoins, cela ne joue que sur le temps d'apprentissage.

Pour implémenter cet effecteur, on crée en réalité plusieurs effecteurs indépendants en accord avec le principe illustré sur la figure 6.10 dans le cas de l'effecteur de type C. On illustre le réseau résultant sur la figure 6.11.

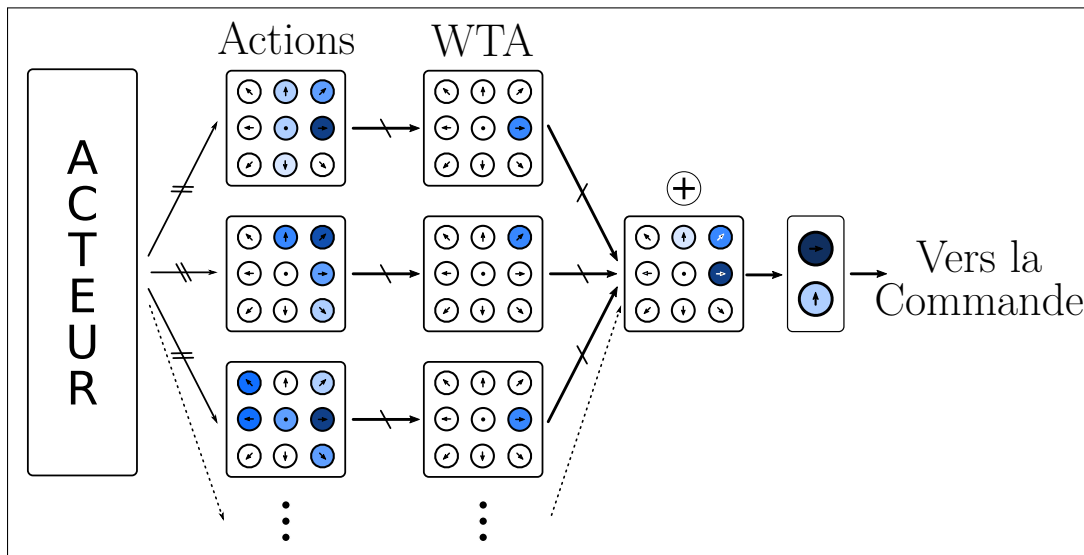


FIGURE 6.11. – Réseau acteur pour un effecteur analogique par partie : chaque acteur est indépendant et ne gère qu'une portion de la force, les gagnants sont ensuite sélectionnés dans des réseaux de types WTA indépendants, puis sommés et enfin re-projetés sur le doublet F_x/F_y afin de fournir une commande au système simulé. On ne représente que 3 sous-réseaux par simplicité, mais le réseau peut être aussi fin que désiré.

Si on implémente ce réseau et qu'on le place dans la même situation de test de réaction à l'environnement que celles des précédentes courbes alors on obtient le résultat présenté sur la figure 6.12. Sur cette figure, on observe bien la capacité du réseau à atteindre la récompense et à se stabiliser sur celle-ci. On observe bien que le réseau est capable d'utiliser les acteurs analogiques, comme on peut le voir à la taille variable des flèches représentant les vecteurs forces ainsi qu'aux angles quelconques qui sont utilisés (contrairement à la figure 6.6 qui utilisait un pavage grossier des directions, ici la combinaison de petites forces élémentaires permet un contrôle bien plus fin.

Cas de récompenses non ponctuelles

Nous avons vu que l'algorithme est capable de converger vers une trajectoire unique dans le cas de récompenses occupant une faible zone de l'espace par rapport à la largeur du mouvement (on simplifie ici en parlant de récompenses « ponctuelles », mais elles ne le sont pas vraiment). Néanmoins que se passe-t-il si la récompense n'est pas ponctuelle et couvre une large zone de l'espace ?

Dans le cas d'une commande avec un pavage des forces et des directions peu riches, comme c'est le cas pour les expériences précédant cette section, la courbe continue à converger vers une unique trajectoire (parfois plus, selon la position de la zone de récompense et sa largeur, mais en général une seule). On peut l'expliquer, car pour converger vers plusieurs trajectoires il faut être capable de fournir différentes successions de forces qui se valent d'un point de vue propagation de la récompense sur les trajectoires générées, et qui atteignent toutes la récompense. Deux conditions qui sont impossibles dans le cas d'un pavage grossier des forces.

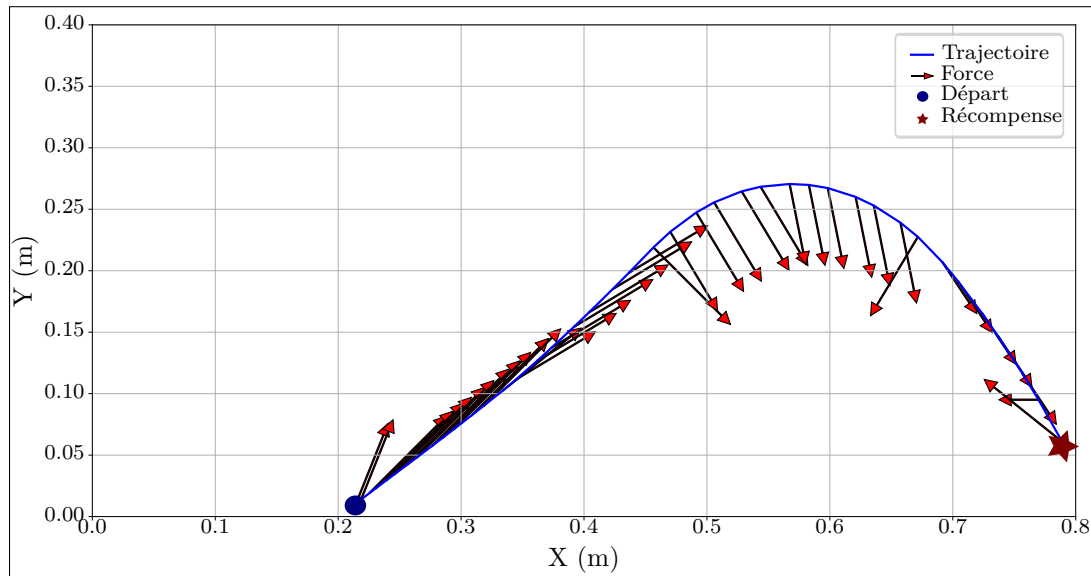


FIGURE 6.12. – *Résultat de trajectoire pour un effecteur analogique par partie, on utilise ici une variante comptant 10 sous réseaux acteurs indépendants.*

Mais que se passe-t-il avec le pavage plus fin qui permet l'expression de multiples forces telle que nous l'avons vue dans cette section ? La commande a alors tout loisir à fournir des trajectoires légèrement différentes, trop peu différenciées pour que la convergence en sélectionne une plutôt qu'une autre, et qui atteigne toute la zone de récompense.

On peut l'illustrer sur la figure 6.13. Dans cette expérience on définit une zone de récompense et on observe après convergence la trajectoire obtenue sur plusieurs expériences de suite. On observe alors bien que chaque trajectoire est souvent légèrement différente de la précédente. L'algorithme converge alors vers un ensemble d'action qui mène à la zone de récompense sans choisir précisément une unique trajectoire optimale.

Notons que c'est néanmoins très dépendant des conditions de l'expérience : un choix de taux d'apprentissage très faible combiné à une expérience très longue avec beaucoup de bruit permet de converger vers moins de trajectoires ou même une trajectoire unique. C'est dû au fait que le réseau tente implicitement de réduire le temps du trajet vers la zone de récompense (tout en évitant le sol), de facto il n'existe donc en théorie qu'une unique trajectoire optimale, même si vis-à-vis des autres trajectoires elles ne se démarquent que très peu. Si l'on souhaite atteindre cette trajectoire unique dans ce type de cas alors un long apprentissage avec un fort bruit et un très faible taux d'apprentissage est nécessaire.

Influence de la finesse de la catégorisation des données entrantes

Comme on est devant un système qui effectue des choix en fonction des données d'entrées, ceux-ci sont automatiquement influencés par la nature de ces données entrantes. Dans notre cas, les précédentes expériences ont été réalisées grâce à un pavage assez grossier des données d'entrée, qui ne permet pas forcément au système d'avoir un indicateur précis sur sa position. On peut tester en quoi un changement de la précision du pavage peut changer la trajectoire du mouvement obtenu dans notre cas, on illustre le résultat sur la figure 6.14.

Ici, on a imposé une granularité 10 fois plus importante sur la couche catégorisatrice. On peut observer que le changement dans la granularité des données change effectivement énormément la trajectoire obtenue. Cela peut être problématique dans le cadre de la création d'un algorithme bio-inspiré qui vise à mimer certaines caractéristiques des trajectoires humaines puisque la granularité des données chez l'être humain est difficile à estimer. Gardons donc à l'esprit pour la suite que l'on ne peut donc pas

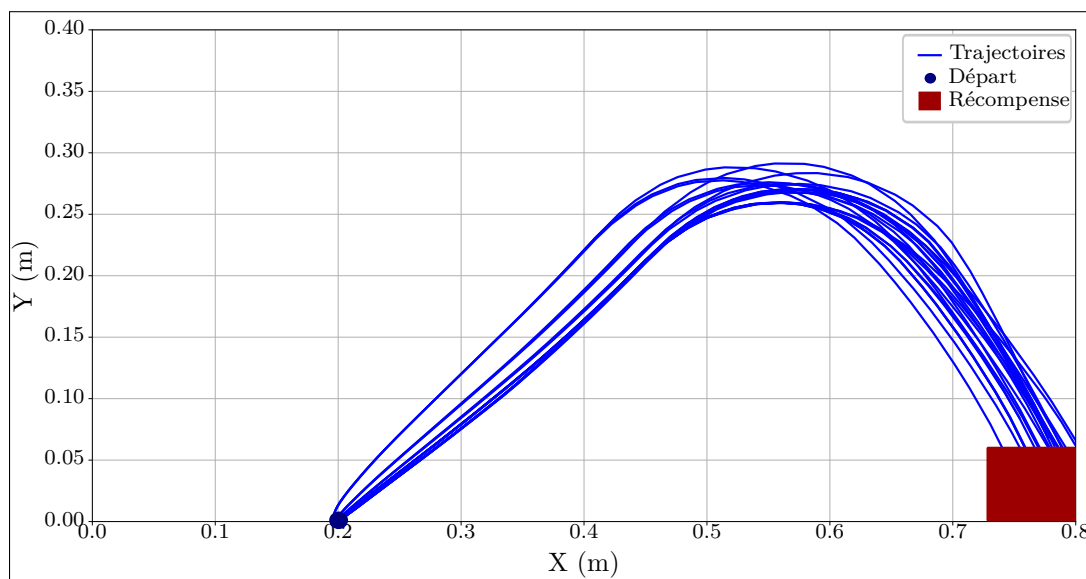


FIGURE 6.13. – *Résultat des trajectoires pour un effecteur analogique et une zone de récompense.*

réellement faire d'analogie précise entre ces courbes et les trajectoires humaines, mais tout au plus noter des tendances.

Notons également qu'ici le changement du type de catégorisation influence beaucoup l'opération effectuée par l'algorithme. Dans les cas précédents, on se contentait d'une catégorisation simple avec un unique état actif sortant, ce qui était le cas le plus simple imaginable pour un système catégorisateur. Mais dans le cas présent, on fournit au système acteur-critique l'ensemble des activations de catégories. De fait, un état est maintenant un ensemble d'activations réparties sur un champ neuronal et non plus un neurone isolé. Cela signifie aussi que les opérations d'apprentissage effectuées sur un état toucheront de fait d'autres états voisins qui partagent certaines de ces activités. On s'éloigne donc un peu plus de l'analogie avec l'équation classique du TD-Lambda. Notons que la largeur du noyau sélectionnée pour représenter les catégories devient alors impactante dans cette variante de l'algorithme.

Conséquences d'un monde « analogique » sur le comportement résultant de ce type de système

Le fait d'évoluer dans une simulation soumise aux lois de la physique, d'utiliser en entrée des données plus complexes et de contrôler l'effecteur avec un ensemble de forces introduit une certaine imprécision dans les conséquences des actions. Une action effectuée à un même emplacement va avoir des conséquences dépendantes de l'inertie et de la vitesse du solide, mais va être également soumise aux imprécisions de la catégorisation en entrée. De fait, il peut arriver qu'une même action aboutisse à atteindre la récompense dans certains contextes et pas dans d'autres.

Les conséquences de ceci sont que l'algorithme est instable par nature. Prenons un exemple : si on considère une action A dans un contexte S , menant à la récompense R qui est située sur un état adjacent à S . Dans le fonctionnement classique de l'algorithme le parcours régulier du chemin allant de S à R va affiner la prédiction et renforcer les actions y menant. Les erreurs faites par l'algorithme suite au bruit introduit pour l'exploration, en passant, par exemple, de S à S' (S' étant un état adjacent à S sans récompense) seront alors vues comme des « déceptions » et donneront lieu à un renforcement négatif qui rendra moins probable les actions partant de S et menant à S' .

L'algorithme se stabilise alors quand le poids sur l'action A menant de S à R est supérieur à l'amplitude du bruit maximum. Dans ce cas, l'algorithme prendra alors toujours le même chemin. En parallèle,

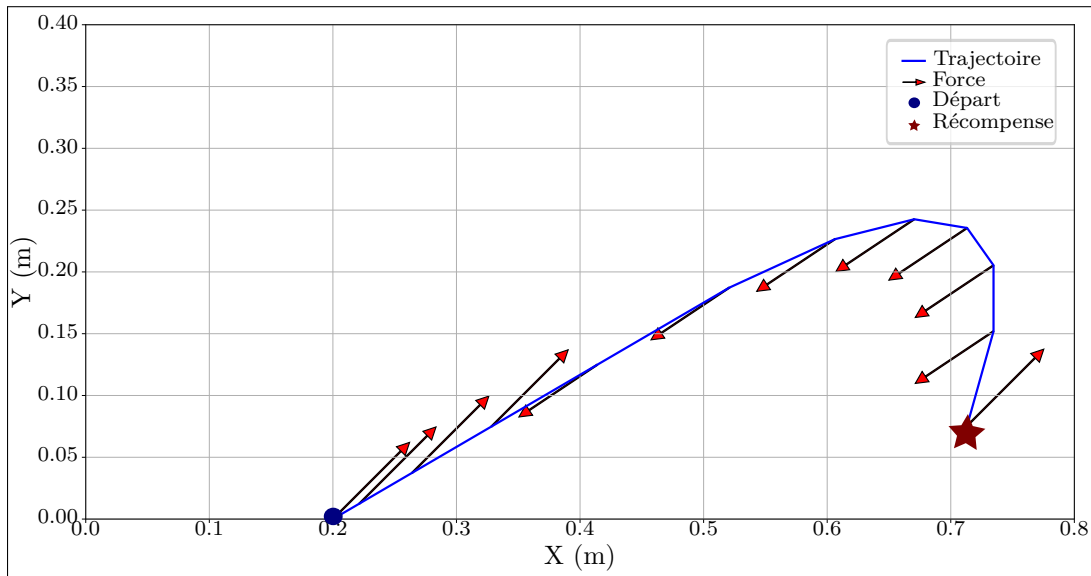


FIGURE 6.14. — *Résultat des trajectoires pour un effecteur analogique et une catégorisation plus fine.*

plus les choix de l'action A sont récurrents, moins il y aura de variabilité et plus l'estimation de la récompense future sera précise. Si l'action est renforcée plus fortement que la prédiction alors le temps que la prédiction converge l'action sera renforcée un peu au-delà du seuil du bruit, sinon l'action sera renforcée pile au niveau du dépassement du seuil du bruit.

Il est ici important de noter cet aspect : l'action ne continue plus à apprendre au-delà d'un certain seuil une fois que l'algorithme se stabilise.

Mais que se passe-t-il dans un cas où des erreurs peuvent continuer à se glisser dans la conséquence des actions ? S'il arrive une fois de temps en temps que l'action A mène à un état qui ne donne pas la récompense, alors l'algorithme désapprendra l'action A à ce moment-là. Cela aura alors tendance à faire reculer l'apprentissage. Selon les cas cela peut alors refaire basculer l'apprentissage en dessous du seuil du bruit de l'exploration, ce qui donnera lieu à un retour à une exploration potentielle aux alentours de cet état, puis à une reconvergence.

L'algorithme se place donc en réalité légèrement au-delà de la limite de l'exploration, selon les cas il pourra basculer et revenir dans un mode d'exploration. Ce qui peut être vu comme une propriété négative ou positive selon le contexte. Ceci explique donc que certaines trajectoires peuvent être instables dans le temps avec ce type de modèle déployé dans un monde « analogique » : Plus le monde est complexe, moins il a de chance d'être représenté fidèlement en entrée de l'algorithme, plus les erreurs seront possibles et on obtient alors un algorithme qui tend à se placer à la limite du choix entre exploration et exploitation : exploitant la majeure partie du temps et « décrochant » une fois de temps en temps pour revenir de façon temporaire dans un état d'exploration.

Optimisation de critères énergétiques

Dans les précédentes sections, nous avons établi que l'algorithme était capable de prendre en compte l'environnement pour converger vers des trajectoires typiques des humains dans certaines situations d'interaction avec l'environnement, notamment dans le cadre de la prise d'objet. Nous avons montré que l'on était capable d'adapter l'algorithme à la commande en force afin qu'il soit compatible avec nos précédents algorithmes pour le contrôle de robot réel. Toutefois jusqu'ici l'algorithme n'a pas vraiment d'intérêt à utiliser en détail les forces en question puisqu'aucun critère énergétique n'est établi. L'algorithme se contente donc d'éviter le sol en convergeant vers une trajectoire qui réduit implicitement la

distance parcourue, mais sans forcément tenter d'économiser ses forces. Toutefois, dans le cas humain la dépense de moindre énergie est une évidence pour éviter de produire des mouvements inutiles et dans le cas robotique il s'agit d'un critère pragmatique pour éviter une dépense énergétique trop intense lorsque ce n'est pas nécessaire.

Nous avons montré dans la partie précédente (section 8) que certains apprentissages simples sont capables de converger vers des mouvements qui sont implicitement « les plus simples », et donc les moins énergétiques, en profitant des mécanismes de séparation en distance des premières couches de l'algorithme. Néanmoins si cela fonctionne dans les cas simples ce n'est plus le cas dans le cas d'une trajectoire plus complexe.

Nous testons donc ici la faculté de l'algorithme à s'adapter à ce type de critères. Une implémentation rigoureuse demanderait probablement un modèle du muscle et des efforts musculaires complets, mais dans notre cadre nous avons voulu nous limiter à un test simple.

Nous modélisons donc cette énergie dépensée comme la somme des forces dépensées à un instant t , étant donné que l'on considère que chaque force émise est en quelque sorte une action des muscles sur l'effecteur terminal du bras, alors chaque force est une dépense énergétique. On implémente donc simplement ce critère sous la forme d'une punition instantanée qui vient se soustraire à la récompense :

$$P = \sum^n \|\vec{F}_n\| \quad (6.13)$$

Si la fatigue est trop importante par rapport à la récompense on notera que la récompense est alors trop peu intéressante pour motiver la dépense d'énergie. Comment le système humain arrive à calibrer les niveaux de « punition » face à l'effort par rapport aux récompenses est une question intéressante dont la résolution sera, selon nous, nécessaire à la création de futurs systèmes de contrôles bio-inspirés (pour des débuts de réponses sur les mécanismes de fatigue chez l'humain, on peut se reporter à la revue très complète du sujet (TAYLOR et al., 2016)). Toutefois, dans notre cas nous avons voulu faire simple et nous avons donc calibré manuellement cette punition pour qu'elle soit très inférieure au seuil de « non-action » à partir duquel le système ne fait plus rien.

Le résultat une fois implémenté peut être observé sur la figure 6.15

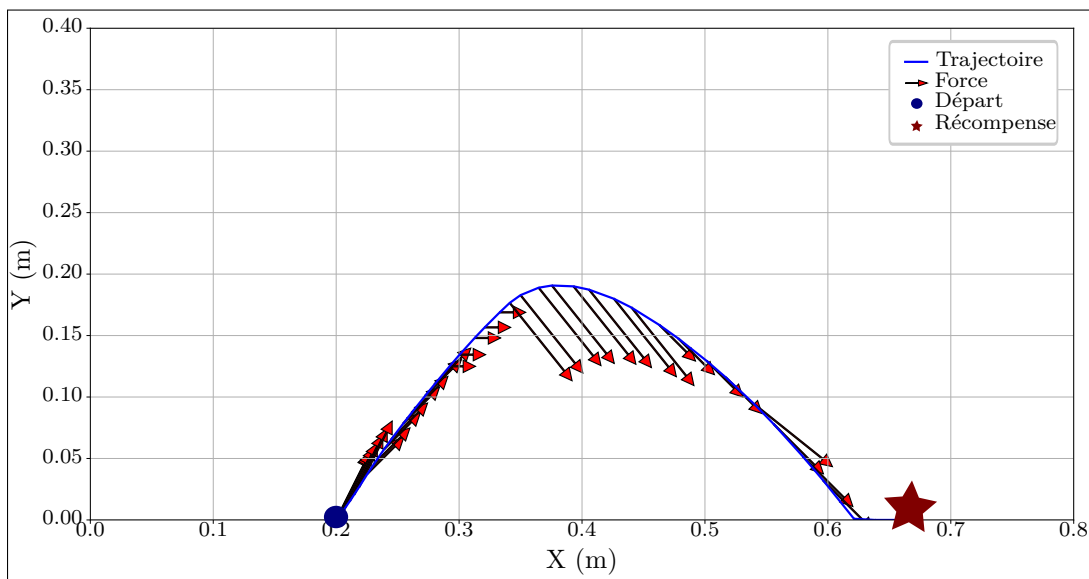


FIGURE 6.15. – *Résultat de trajectoire pour un effecteur analogique et un critère d'énergie.*

Comme on peut l'observer sur cette figure les forces employées sont bien plus modestes et de fait la courbe emprunte une trajectoire plus minimaliste pour atteindre la cible. Notons d'ailleurs que sur la

fin l'algorithme préfère faire frotter légèrement l'effecteur au sol, quitte à avoir une punition, plutôt que de dépenser de l'énergie pour l'éviter, la punition sur cette petite portion de courbe étant trop faible pour valoir la peine de choisir une trajectoire plus coûteuse en énergie.

On peut comparer les deux situations avec effecteur analogique sans et avec un critère d'énergie pour illustrer l'économie réalisée. On trace donc l'intégration au cours d'un trajet des énergies déployées sur la figure 6.15 et sur la figure 6.12.

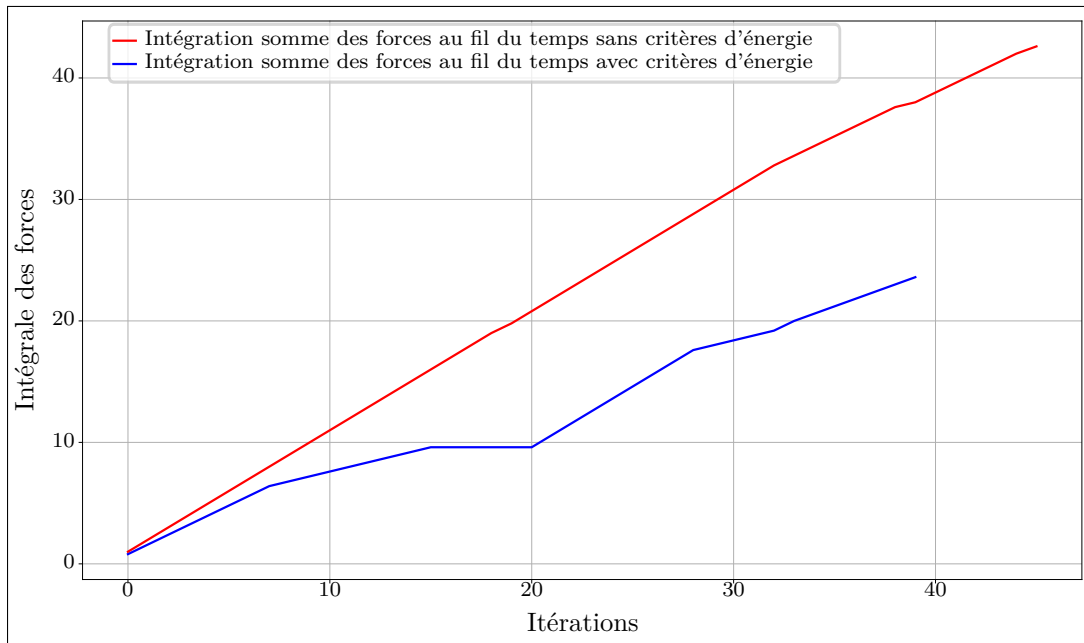


FIGURE 6.16. – Sur cette figure on montre l'intégration des forces au cours du temps pour une trajectoire avec critère d'énergie (ce qui correspond à la courbe de trajectoire 6.15) et sans (ce qui correspond à la courbe de trajectoire 6.12).

Comme on peut l'observer sur cette figure la somme des forces déployées est bien inférieure dans le cas avec critère pour atteindre la même cible, on observe également que le trajet dure sensiblement moins longtemps du fait de cette économie, ce qui n'est pas forcément un résultat évident au premier abord : cela doit en effet probablement dépendre des conditions expérimentales, on imagine bien que dans certains cas le système préférera emprunter une route plus longue en temps, mais demandant moins d'effort plutôt qu'une route rapide et demandant des efforts.

6.3.6. « Contexte social » : résultats préliminaires

Nous avons vu que l'algorithme est capable de s'adapter à l'environnement. Mais qu'en est-il de prendre en compte un contexte « social » dans ses mouvements ?

Comme nous en avons déjà discuté dans la partie 3.2.1 ainsi que dans l'introduction de cette partie, nous formulons l'hypothèse que l'être humain a une tendance à communiquer implicitement à travers la nature et la dynamique de ses actions. Un autre humain observant ses actions pourra ainsi être capable de lire son intention à travers l'observation de ses mouvements.

Ainsi en se basant sur l'image soulevée par l'expérience de pensée de Jacob et Jeannerod (JACOB et JEANNEROD, 2005) : si la dynamique de mouvement d'une personne dépend de son intention (tuer son patient ou le soigner) alors une autre personne dans la pièce peut agir en fonction de cette information implicite en observant son mouvement (regarder le médecin ou l'écartier pour sauver le patient par exemple). Jacob et Jeannerod font l'hypothèse que l'intention ne change pas vraiment les dynamiques de

trajectoires et donc que cette distinction est impossible. Mais des expériences plus tardives (GEORGIU et al., 2007; BECCHIO et al., 2008; LEWKOWICZ et al., 2013) ont bien démontré dans divers contextes que l'intention (pour une même tâche) avait bien tendance à faire varier les dynamiques observées.

L'expérience de Cristina Becchio et collègues (BECCHIO et al., 2008) est intéressante car elle se déroule dans le même contexte expérimental que notre simulation : l'humain est placé devant une table et doit déplacer un objet d'un point A à un point B. Cristina Becchio et collègue comparent dans leur expérience le cas où l'humain place l'objet dans la main d'un autre humain ou sur la table et en déduisent qu'une variation de trajectoire existe entre les deux cadres. La trajectoire observée lorsque l'on place un objet dans la main d'un autre humain est ainsi plus « prudente » et notamment plus haute que lorsqu'on place l'objet sur une table.

Peut-on apporter un début d'interprétation de cette expérience avec notre simulation ?

Pour ce faire, on propose deux interprétations du phénomène dans un contexte social :

- Dans la première interprétation, on retourne la problématique en considérant les conséquences des actions dans les deux cas du point de vue de l'humain. Dans le cas où l'on donne un objet à un humain, l'échec est en effet plus coûteux que si on le pose sur une table. Le poser sur une table implique qu'en cas d'erreur la seule chose à faire est de le pousser de côté pour ajuster la position de l'objet. L'ajustement final de la trajectoire est ainsi assez peu coûteux dans le processus et un manque de précision n'a pas de grandes conséquences. À l'inverse, rater la main de l'humain à qui l'on veut donner l'objet implique une confusion des deux systèmes qui s'étaient fixé un point de rendez-vous, ce qui implique potentiellement un mouvement des deux acteurs pour rattraper l'erreur, une perte de temps, etc. On peut également le voir comme quelque chose de plus haut niveau : rater la tâche en présence d'un autre humain à qui on doit donner l'objet impliquerait un genre de « punition sociale » due au jugement de l'autre humain sur nos actions.
- La seconde interprétation est assez proche, mais ne recourt pas à des concepts de punitions sociales : dans cette interprétation, on met plutôt en avant le fait qu'un être humain implique un contexte différent : un être humain est en effet un point d'arrivée qui peut être soumis à une erreur. L'humain étant plus imprévisible qu'un objet inerte il est alors plus prudent, pour réussir la tâche, de prendre son temps pour ne pas rater une éventuelle variation de la position de la main du partenaire. Une trajectoire trop rapide par rapport à l'inertie du système de contrôle et à ses capacités ferait ainsi qu'il serait impossible de corriger un mouvement pour ajuster suffisamment sa position en cas d'erreur du partenaire (qui déplacerait sa main à la dernière minute par exemple).

Les deux interprétations donnent lieu à deux protocoles différents, le premier protocole va simplement modéliser la punition autour de la récompense comme plus importante que dans le contexte non social (« punition sociale » + confusion possible). Il s'agit d'une simplification importante, mais qui nous semble pertinente.

La seconde interprétation nécessite de mettre en place un lieu de récompense soumis à une erreur de position (se déplaçant aléatoirement au cours du temps autour d'un point fixe par exemple). Si ce second protocole est plus « propre », car il ne nécessite que de prendre en compte la différence de contexte sans faire d'interprétation telle que les « récompenses sociales », il est néanmoins plus complexe à mettre en place, car il nécessite d'avoir un algorithme capable :

- De « voir » la cible, ce qui n'est pas notre cas ici.
- De converger pour prendre en compte la position de cette cible dans son action, ce qui peut être bien plus long comme apprentissage. (Car on demande alors à l'algorithme un travail supplémentaire, qui est celui réalisé par l'algorithme de la partie 2 de cette thèse. La solution élégante à long terme serait sans doute de combiner les deux algorithmes pour profiter de l'un et de l'autre).
- D'avoir un temps de traitement de l'information (vision vers action) pertinent par rapport à ce que l'on connaît de l'être humain pour que l'algorithme apprenne à se montrer prudent face à des phénomènes trop rapides pour sa capture d'information.

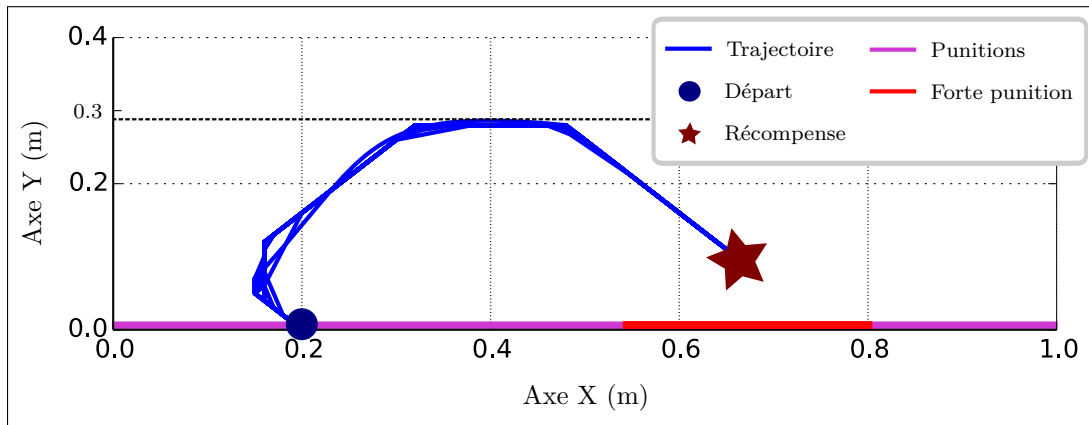


FIGURE 6.17. – Sur cette figure, on trace la trajectoire après convergence dans le cas d'un environnement modifié pour simuler simplement un contexte « social » de passage d'objet à un autre être humain, dont les conséquences seraient importantes en cas d'échec. La zone violette est la zone de punition classique pour simuler les effets des frottements et la zone rouge est une zone de plus forte punition avec une particularité : pour simuler le fait qu'en cas d'échec la confusion fait perdre du temps aux deux humains pour se remettre d'accord et se repasser l'objet, l'expérience est tout simplement recommencée (la bille se téléporte à sa position de départ) si la bille touche la zone rouge.

Malheureusement dans cette simulation simple nous ne disposons pas de ces trois caractéristiques essentielles à la modélisation du second protocole. Nous pouvons toutefois nous rabattre sur le premier protocole qui reste parfaitement modélisable.

Pour ce faire, nous proposons en pratique un protocole simple dans lequel une erreur à proximité de la récompense conduit à une punition plus importante et à un redémarrage de l'expérience depuis le début (comme lorsque la bille touche l'emplacement de la récompense), contrairement à l'erreur sur la table qui ne conduit qu'à une punition normale et permet de continuer les mouvements. On observe le protocole et le résultat après convergence sur la figure 6.17.

Globalement, la courbe est la même, toutefois si l'on compare les hauteurs des trajectoires avec celles disposant du protocole classique (on est ici dans un cas avec un effecteur classique non analogique pour cette expérience, on compare donc avec les trajectoires obtenues avec un effecteur classique). On obtient une augmentation d'une hauteur d'environ 0,24 m en moyenne à 0,29 m. À première vue, on pourrait dire que l'augmentation de hauteur est suffisamment faible pour ne pas être significative, toutefois si on compare avec les résultats obtenus sur les humains par Cristina Becchio et collègues (BECCHIO et al., 2008) on note que les augmentations d'amplitudes relatives sont très similaires.

Si on trace les amplitudes relatives de notre simulation et des travaux de Cristina Becchio et collègues en prenant pour références à chaque fois le cas « non social » (les valeurs absolues sont peu intéressantes ici, notre modèle étant trop rustique pour qu'une concordance se dégage) alors on peut observer une similitude entre les augmentations relatives de notre simulation et celle des expériences sur humains. (voir figure 6.18)

Néanmoins, il faut noter que dans ce modèle l'importance des punitions sur les deux zones a évidemment un impact sur la différence de hauteur. Même si dans nos tests la hauteur maximale dépasse rarement celle-ci.

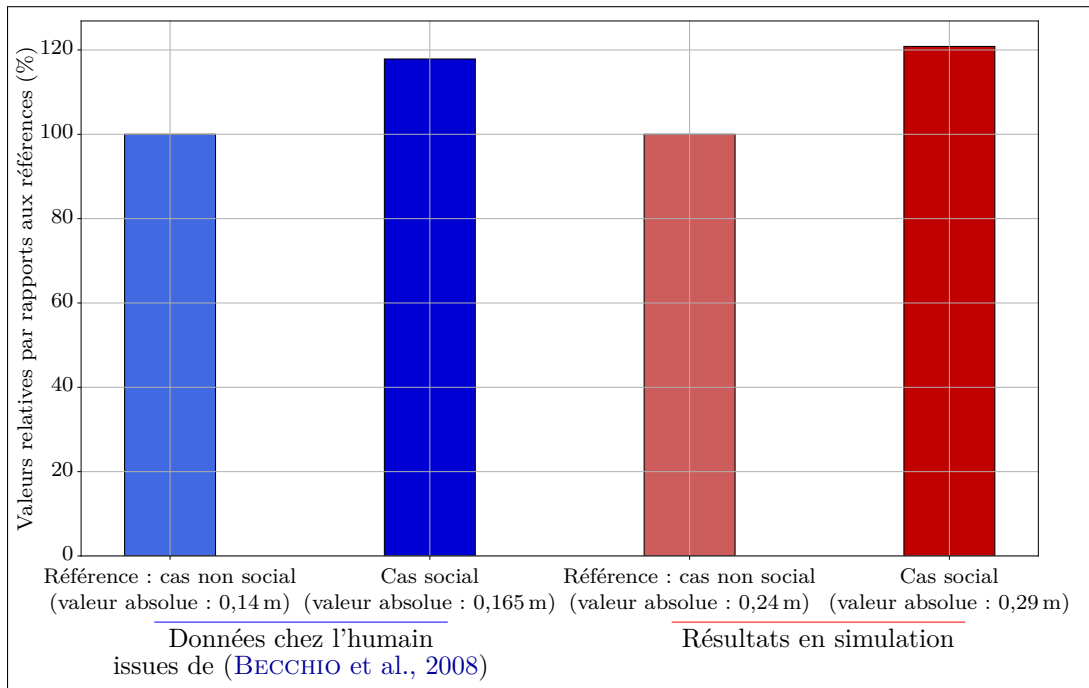


FIGURE 6.18. – Sur cette figure, on compare les hauteurs relatives des trajectoires dans deux cas : notre simulation (en rouge) et les données de (BECCHIO et al., 2008), issues des expériences sur humains dans un contexte similaire. On peut observer que l'augmentation relative est assez similaire.

6.3.7. Conclusion

Dans cette section, nous avons abordé une démonstration simple de la réaction de l'algorithme à une problématique de déplacement d'objet dans le cadre d'une simulation très simplifiée. Nous avons montré que globalement l'algorithme est robuste à différentes situations, même si ses trajectoires précises sont assez sensibles aux paramètres. Nous avons montré que l'algorithme est capable de s'adapter à des cas de récompenses non ponctuelles, à une minimisation de son énergie dépensée, et qu'il donne des résultats cohérents avec les mouvements humains dans ces situations.

Enfin, nous avons proposé une simulation simple d'un « cas social » proche des expériences de Becchio et collègues et montré que malgré la simplicité de la simulation et des hypothèses sur les aspects sociaux on obtenait le même ordre de grandeur de variations entre les cas non-sociaux et sociaux.

Dans la section suivante, nous proposerons un environnement plus complexe pour valider cet algorithme sur un système que l'on conservera simple, mais tout de même plus proche des commandes d'un robot réel.

6.4 | Second environnement : simulation 2D d'un bras robotique

Pour mettre en place un modèle plus complexe, on choisit de simuler un bras en deux dimensions vu de côté et soumis aux lois de la physique du mouvement. Cette partie de la thèse a été réalisée en collaboration avec Zinedine Galouz, dans le cadre de son stage de Master 2 encadré dans le cadre de cette thèse (GALOUZ, 2018).

Nous nous sommes inspirés dans cette section des travaux de Rod Cross sur les modèles de bras dans le cadre de l'étude du mouvement des lancers de balles au tennis (CROSS, 2011).

6.4.1. Description technique de la simulation du bras

Le bras est simulé à l'aide d'un modèle de pendule double sur lequel on peut appliquer un couple à chaque segment à tout moment. Dans cette partie, on décrit la logique régissant cette simulation et les équations mises en place pour son fonctionnement.

Pendule double

Le pendule double est un système à comportement chaotique, c'est-à-dire que l'on ne peut pas observer de périodicité dans son comportement et qu'il dépend au contraire fortement des conditions initiales (LEVIEN et TAN, 1993).

Ce pendule est constitué de deux segments : 1 (l'arrière-bras) et 2 (l'avant-bras). Tous ont un certain angle par rapport à la verticale (respectivement θ et ϕ), une longueur (L_1 et L_2), une masse (m_1 et m_2), un centre de gravité (G_1 et G_2) des coordonnées (x_1, y_1) et (x_2, y_2) et une longueur entre l'origine du segment et le centre de gravité (h_1 et h_2) (voir figure 6.19). Le point correspondant à l'épaule est noté P_0 (on note ses coordonnées (x_E, y_E)), le point correspondant au coude est noté P_1 et le point correspondant à la main est noté P_2 .

On cherche à caractériser ce système par ses équations afin de pouvoir le simuler et agir dessus par l'application de couples sur les deux segments. On va d'abord commencer par exprimer la position des points G_1 et G_2 en fonction de la position du point P_0 , de la longueur des segments et de l'angle des articulations en fonction du temps.

$$\begin{cases} x_1(t) = x_E + h_1 \sin(\theta(t)) \\ y_1(t) = y_E - h_1 \cos(\theta(t)) \\ x_2(t) = x_E + L_1 \sin(\theta(t)) + h_2 \sin(\phi(t)) \\ y_2(t) = y_E - L_1 \cos(\theta(t)) - h_2 \cos(\phi(t)) \end{cases} \quad (6.14)$$

On dérive ensuite ces expressions par rapport au temps

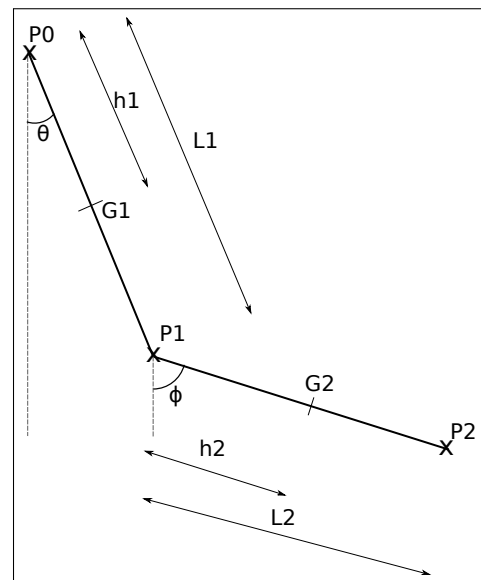


FIGURE 6.19. — Schéma du pendule double et des valeurs utilisées dans les équations du bras.

afin d'obtenir la vitesse des points G_1 et G_2 .

$$\begin{cases} v_{x_1}(t) = \dot{x}_1(t) = v_{x_E} + h_1\dot{\theta}(t)\cos(\theta(t)) \\ v_{y_1}(t) = \dot{y}_1(t) = v_{y_E} + h_1\dot{\theta}(t)\sin(\theta(t)) \\ v_{x_2}(t) = \dot{x}_2(t) = v_{x_E} + L_1\dot{\theta}(t)\cos(\theta(t)) + h_2\dot{\phi}(t)\cos(\phi(t)) \\ v_{y_2}(t) = \dot{y}_2(t) = v_{y_E} + L_1\dot{\theta}(t)\sin(\theta(t)) + h_2\dot{\phi}(t)\sin(\phi(t)) \end{cases} \quad (6.15)$$

Avec v_E la vitesse du point P_0 , $\dot{\theta}$ la vitesse angulaire de l'épaule et $\dot{\phi}$ la vitesse angulaire du coude. On cherche ensuite à obtenir l'accélération des points G_1 et G_2 en dérivant les expressions de la vitesse précédemment trouvées par rapport au temps.

$$\begin{cases} a_{x_1}(t) = \ddot{x}_1(t) = a_{x_E} + h_1\ddot{\theta}(t)\cos(\theta(t)) - h_1\dot{\theta}^2\sin(\theta(t)) \\ a_{y_1}(t) = \ddot{y}_1(t) = a_{y_E} + h_1\ddot{\theta}(t)\sin(\theta(t)) + h_1\dot{\theta}^2\cos(\theta(t)) \\ a_{x_2}(t) = \ddot{x}_2(t) = a_{x_E} + L_1\ddot{\theta}\cos(\theta(t)) - L_1\dot{\theta}^2\sin(\theta(t)) \\ \quad + h_2\ddot{\phi}\cos(\phi(t)) - h_2\dot{\phi}^2\sin(\phi(t)) \\ a_{y_2}(t) = \ddot{y}_2(t) = a_{y_E} + L_1\ddot{\theta}\sin(\theta(t)) + L_1\dot{\theta}^2\cos(\theta(t)) \\ \quad + h_2\ddot{\phi}\sin(\phi(t)) + h_2\dot{\phi}^2\cos(\phi(t)) \end{cases} \quad (6.16)$$

Avec a_E l'accélération du point P_0 , $\ddot{\theta}$ l'accélération angulaire de l'épaule et $\ddot{\phi}$ l'accélération angulaire du coude. On effectue maintenant le bilan des forces appliquées sur les deux segments du bras au niveau du centre de gravité (voir figure 6.20). Le segment 1 subit son poids P_1 , la force de réaction F_1 exercée par l'origine du bras sur le segment et l'opposé de la force de réaction F_2 que le segment 1 exerce sur le segment 2 par le biais du coude (3^e loi de Newton). Le segment 2, quant à lui, subit son propre poids P_2 et la force de réaction F_2 exercée par le segment 1 sur le segment 2 via le coude.

D'après la 2^e loi de Newton :

$$\Sigma \overrightarrow{F_{ext}} = m \cdot \overrightarrow{a} \quad (6.17)$$

En séparant pour les deux segments du bras, on obtient :

$$\begin{aligned} P_1 + F_1 - F_2 &= m_1 \cdot a_1 \\ P_2 + F_2 &= m_2 \cdot a_2 \end{aligned} \quad (6.18)$$

Avec $P_1 = m_1 \cdot g$ et $P_2 = m_2 \cdot g$.

On cherche ensuite à exprimer F_1 et F_2 en fonction du poids et de l'accélération. On traite séparément les composantes x et y .

$$\begin{cases} F_{x_1} - F_{x_2} = m_1 a_{x_1} \\ F_{y_1} - F_{y_2} = m_1 g + m_1 a_{y_1} \\ F_{x_2} = m_2 a_{x_2} \\ F_{y_2} = m_2 g + m_2 a_{y_2} \end{cases} \quad (6.19)$$

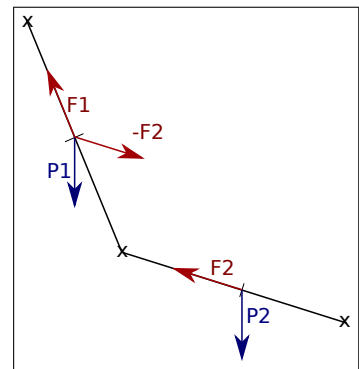


FIGURE 6.20. — Bilan des forces appliquées sur les segments du bras au niveau de leur centre de gravité.

En remplaçant l'accélération par son expression, on obtient :

$$\left\{ \begin{array}{l} F_{x_1} - F_{x_2} = m_1(a_{x_E} + h_1\ddot{\theta} \cos(\theta(t)) - h_1\dot{\theta}^2 \sin(\theta(t))) \\ F_{y_1} - F_{y_2} = m_1g + m_1(a_{y_E} + h_1\ddot{\theta} \sin(\theta(t)) + h_1\dot{\theta}^2 \cos(\theta(t))) \\ F_{x_2} = m_2(a_{x_E} + L_1\ddot{\theta} \cos(\theta(t)) - L_1\dot{\theta}^2 \sin(\theta(t)) \\ \quad + h_2\ddot{\phi} \cos(\phi(t)) - h_2\dot{\phi}^2 \sin(\phi(t))) \\ F_{y_2} = m_2g + m_2(a_{y_E} + L_1\ddot{\theta} \sin(\theta(t)) + L_1\dot{\theta}^2 \cos(\theta(t)) \\ \quad + h_2\ddot{\phi} \sin(\phi(t)) + h_2\dot{\phi}^2 \cos(\phi(t))) \end{array} \right. \quad (6.20)$$

Par la suite, pour alléger les formules, nous ne faisons pas apparaître « (t) » pour les termes dépendant du temps.

Traitons dans un premier temps le segment 2. La 2^e loi de Newton appliquée en dynamique de rotation nous indique que la somme des moments des forces est égale au produit du moment d'inertie et de l'accélération angulaire du système (le segment 2 en l'occurrence). On introduit C_2 qui est le couple que l'on va vouloir faire appliquer au système sur le segment 2 (la commande de l'avant-bras). Le moment du poids est nul, car la distance entre le pivot et le point d'application de la force est nulle (elle s'applique sur le centre de gravité qui est aussi le pivot du système).

$$\begin{aligned} C_2 + \vec{M}_{\vec{F}_2/G_2} &= I_{cm,2} \cdot \ddot{\phi} \\ \Rightarrow C_2 + \vec{F}_2 \wedge \overrightarrow{G_2P_1} &= I_{cm,2} \cdot \ddot{\phi} \end{aligned} \quad (6.21)$$

On résout le produit vectoriel en projetant \vec{F}_2 sur les axes x et y , on ne s'intéresse qu'à la norme du produit.

$$C_2 + F_{x_2} h_2 \cos \phi + F_{y_2} h_2 \sin \phi = I_{cm,2} \ddot{\phi} \quad (6.22)$$

En remplaçant F_{x_2} et F_{y_2} par leurs expressions précédemment trouvées :

$$\begin{aligned} C_2 + m_2 h_2 L_1 \dot{\theta}^2 \sin(\phi - \theta) + m_2 h_2 [a_{x_E} \cos(\phi) + (g + a_{y_E}) \sin(\phi)] \\ + m_2 h_2 L_1 \ddot{\theta} \cos(\phi - \theta) = (I_{cm,2} - m_2 h_2^2) \ddot{\phi} \end{aligned} \quad (6.23)$$

Pour simplifier, on pose :

- $I_2 = I_{cm,2} - m_2 h_2^2$
- $B = m_2 h_2 L_1$
- $Q = C_2 + B \dot{\theta}^2 \sin \beta + R [a_{x_E} \cos \phi + (g + a_{y_E}) \sin \phi]$
- $R = m_2 h_2$
- $\beta = \phi - \theta$

On obtient donc :

$$Q + B \ddot{\theta} \cos(\beta) = I_2 \ddot{\phi} \quad (6.24)$$

On effectue ensuite le même processus pour le premier segment. Le couple appliqué sur le segment 2 se répercute de façon opposée au segment 1.

$$C_1 - C_2 + \vec{F}_1 \wedge \overrightarrow{G_1P_0} + \overrightarrow{-F_2} \wedge \overrightarrow{G_1P_1} = I_{cm,1} \ddot{\theta} \quad (6.25)$$

Avec $I_{cm,1}$ étant le moment d'inertie du segment 1. En termes de normes, on obtient :

$$\begin{aligned} C_1 - C_2 + F_{x_1} h_1 \cos(\theta) + F_{y_1} h_1 \sin(\theta) \\ + F_{x_2} (L_1 - h_1) \cos(\theta) + F_{y_2} (L_1 - h_1) \sin(\theta) = I_{cm,1} \ddot{\theta} \end{aligned} \quad (6.26)$$

En remplaçant F_{x_1} , F_{y_1} , F_{x_2} et F_{y_2} par leurs expressions respectives :

$$C_1 - C_2 - m_2 h_2 L_1 \dot{\phi}^2 \sin(\phi - \theta) + (m_1 h_1 + m_2 L_1)[a_{x_E} \cos(\theta) + (g + a_{y_E}) \sin(\theta)] + m_2 h_2 L_1 \cos(\phi - \theta) \ddot{\phi} = (I_{cm,1} - m_1 h_1^2 - m_2 L_1^2) \ddot{\theta} \quad (6.27)$$

On simplifie l'équation avec :

$$\begin{aligned} - A &= I_{cm,1} - m_1 h_1^2 - m_2 L_1^2 \\ - P &= C_1 - C_2 - B \dot{\phi}^2 \sin \beta + S[a_{x_E} \cos \theta + (g + a_{y_E}) \sin \theta] \\ - S &= m_1 h_1 + m_2 L_1 \end{aligned}$$

$$A \ddot{\theta} = P + B \cos(\beta) \ddot{\phi} \quad (6.28)$$

Finalement, en combinant l'équation (6.24) avec l'équation (6.28), on obtient les expressions d'accélération angulaires suivantes :

$$\ddot{\theta} = \frac{PI_2 + BQ \cos(\beta)}{AI_2 - B^2 \cos^2(\beta)} \quad (6.29)$$

$$\ddot{\phi} = \frac{AQ + BP \cos(\beta)}{AI_2 - B^2 \cos^2(\beta)}$$

Pour obtenir les vitesses angulaires des deux segments ainsi que leurs angles, il suffit d'intégrer numériquement ces expressions en fonction du temps. Pour cela, on peut utiliser la méthode d'Euler, mais nous préférons ici utiliser la méthode de Runge-Kutta pour plus de précision (RUNGE, 1895; KUTTA, 1901).

Pour le premier segment, on note $f(\theta, \phi, \dot{\theta}, \dot{\phi})$ l'expression de $\ddot{\theta}$ et h le pas de temps de la simulation. On calcule ensuite :

$$\begin{aligned} - k_1 &= f(\theta, \phi, \dot{\theta}, \dot{\phi}) \\ - k_2 &= f(\theta + \frac{h}{2} \dot{\theta}, \phi, \dot{\theta} + \frac{h}{2} k_1, \dot{\phi}) \\ - k_3 &= f(\theta + \frac{h}{2} \dot{\theta} + \frac{h^2}{4} k_1, \phi, \dot{\theta} + \frac{h}{2} k_2, \dot{\phi}) \\ - k_4 &= f(\theta + h \dot{\theta} + \frac{h^2}{2} k_2, \phi, \dot{\theta} + h k_3, \dot{\phi}) \end{aligned}$$

À partir de cela, nous pouvons calculer la variation d'angle $\Delta\theta$ depuis l'itération précédente et la variation de vitesse angulaire $\Delta\dot{\theta}$ par :

$$\Delta\theta = h \dot{\theta} + \frac{h^2}{6} (k_1 + k_2 + k_3) \quad (6.30)$$

et :

$$\Delta\dot{\theta} = \frac{h}{6} (k_1 + 2k_2 + 2k_3 + k_4) \quad (6.31)$$

De même, pour le second segment, on note $f(\theta, \phi, \dot{\theta}, \dot{\phi})$ l'expression de $\ddot{\phi}$.

$$\begin{aligned} - k_1 &= f(\theta, \phi, \dot{\theta}, \dot{\phi}) \\ - k_2 &= f(\theta, \phi + \frac{h}{2} \dot{\phi}, \dot{\theta}, \dot{\phi} + \frac{h}{2} k_1) \\ - k_3 &= f(\theta, \phi + \frac{h}{2} \dot{\phi} + \frac{h^2}{4} k_1, \dot{\theta}, \dot{\phi} + \frac{h}{2} k_2) \\ - k_4 &= f(\theta, \phi + h \dot{\phi} + \frac{h^2}{2} k_2, \dot{\theta}, \dot{\phi} + h k_3) \end{aligned}$$

Puis on peut calculer :

$$\Delta\phi = h \dot{\phi} + \frac{h^2}{6} (k_1 + k_2 + k_3) \quad (6.32)$$

et :

$$\Delta\dot{\phi} = \frac{h}{6} (k_1 + 2k_2 + 2k_3 + k_4) \quad (6.33)$$

Au final, grâce à $\Delta\theta$ et $\Delta\phi$, on peut calculer la nouvelle position du bras grâce à l'équation (6.14).

Réaction du support et force d'impact

Le but de notre expérience étant de faire prendre un objet sur une table au bras, il est important de modéliser la réaction du support qu'applique la table sur le bras afin que celui-ci ne puisse pas passer au travers de la table, mais plutôt qu'il glisse dessus (voir figure 6.21). On modélise alors la réaction du support et la force d'impact.

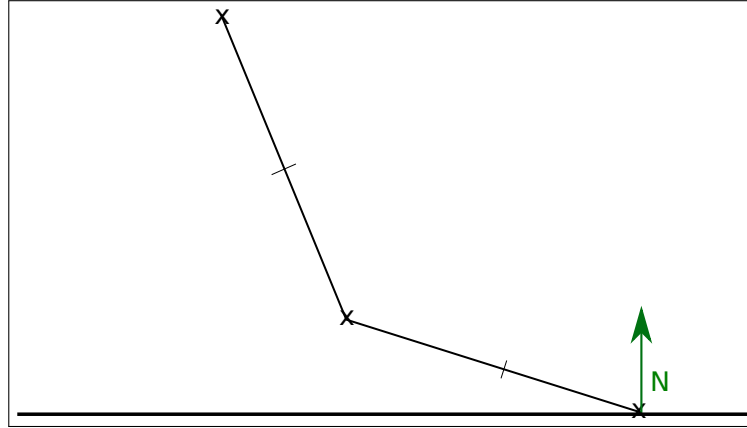


FIGURE 6.21. — La force de réaction du support appliquée par la table sur le bras, normale à la surface de la table.

Afin de simplifier les calculs, on considère que la masse d'un segment est répartie en deux boules situées aux extrémités du segment (voir figure 6.22).

La force de réaction du support s'applique sur l'extrémité du segment 2 (le point P_2) et doit donc compenser :

- Le poids du point de contact P_2 : F_{P_2}
- La force induite par le couple C_1 : $F_{C_{1y}}$
- La force induite par le couple C_2 : $F_{C_{2y}}$
- Le poids appliqué par le segment 1 sur le segment 2 : $F_{P_{1/2y}}$

Pour le poids de P_2 , nous avons simplement :

$$P_2 = \frac{m_2 g}{2} \quad (6.34)$$

Ensuite, pour calculer F_{C_1} , il faut traiter le couple comme deux forces de directions opposées appliquées de part et d'autre du segment (voir figure 6.23a). On a donc :

$$\begin{aligned} C_1 &= \overrightarrow{G_1 P_0} \wedge \overrightarrow{F_{C_1}} + \overrightarrow{G_1 P_1} \wedge \overrightarrow{F_{C_1}} \\ &= h_1 \cdot F_{C_1} + (L_1 - h_1) \cdot F_{C_1} \\ &= L_1 \cdot F_{C_1} \\ \Leftrightarrow F_{C_1} &= \frac{C_1}{L_1} \end{aligned} \quad (6.35)$$

En projetant sur l'axe y :

$$F_{C_{1y}} = \frac{C_1}{L_1} \cdot \sin(\theta) \quad (6.36)$$

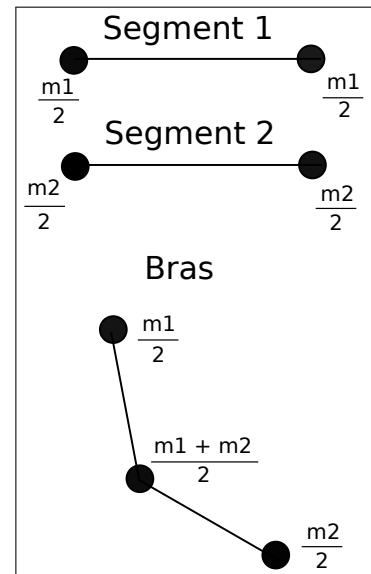


FIGURE 6.22. — La répartition des masses sur les deux segments de bras. Le point P_1 (correspondant au coude) a donc comme masse équivalente $\frac{m_1+m_2}{2}$.

De la même façon, on calcule :

$$\begin{aligned}
 C_2 &= \overrightarrow{G_2 P_1} \wedge \overrightarrow{F_{C_2}} + \overrightarrow{G_2 P_2} \wedge \overrightarrow{F_{C_2}} \\
 &= h_2 \cdot F_{C_2} + (L_2 - h_2) \cdot F_{C_2} \\
 &= L_2 \cdot F_{C_2} \\
 \Leftrightarrow F_{C_2} &= \frac{C_2}{L_2}
 \end{aligned} \tag{6.37}$$

En projetant sur l'axe y :

$$F_{C_{2y}} = \frac{C_2}{L_2} \cdot \sin(\phi) \tag{6.38}$$

Afin de calculer la force appliquée sur le segment 2 par le segment 1, on décompose celle-ci en deux vecteurs orthogonaux \vec{A} et \vec{B} (voir figure 6.23b). On a donc $\overrightarrow{F_{P1/2}} = \vec{A} + \vec{B}$. Le vecteur \vec{A} est annulé par la force $\overrightarrow{F_1}$ de réaction du solide appliquée sur le segment 1 par l'origine du bras. On a donc :

$$F_{P1/2} = \frac{m_1 + m_2}{2} \cdot g \cdot \sin(\theta) \tag{6.39}$$

En projetant sur y , on obtient :

$$F_{P1/2y} = \frac{m_1 + m_2}{2} \cdot g \cdot \sin^2(\theta) \tag{6.40}$$

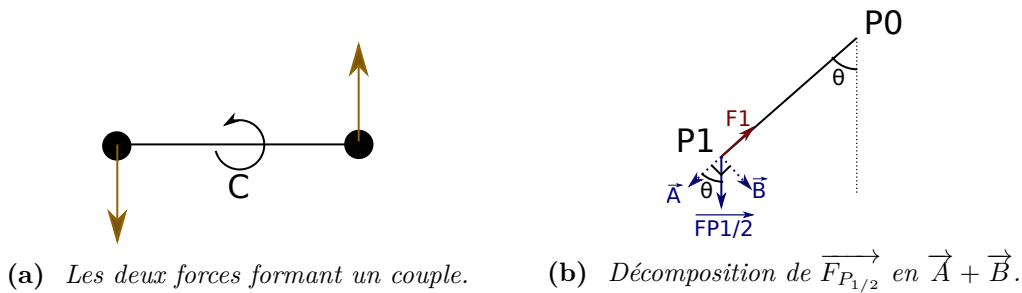


FIGURE 6.23.

Au final, on obtient l'expression de la force de réaction du support suivante :

$$N = \frac{m_2 g}{2} + \frac{C_1}{L_1} \cdot \sin(\theta) + \frac{C_2}{L_2} \cdot \sin(\phi) + \frac{m_1 + m_2}{2} \cdot g \cdot \sin^2(\theta) \tag{6.41}$$

Pour compléter cette réaction du support, il est nécessaire d'introduire une force d'impact \vec{I} qui va annuler la vitesse à laquelle arrive le bras avant de toucher la table (sans cela, le bras passerait au travers de la table en conservant ensuite une vitesse constante). Soit a_{yP_2} l'accélération du point P_2 , on a :

$$I = -\frac{m_2}{2} \cdot a_{yP_2} \tag{6.42}$$

Ces deux forces interviennent dans le calcul de F_{y_2} et se répercutent donc sur le calcul de F_{y_2} :

$$F_{y_2} = P_2 + m_2 \cdot a_{y_2} - N - I \tag{6.43}$$

Et :

$$\begin{aligned}
 F_{y_1} - F_{y_2} &= P_1 + m_1 \cdot a_{y_1} \\
 \Leftrightarrow F_{y_1} &= P_1 + P_2 + m_1 \cdot a_{y_1} + m_2 \cdot a_{y_2} - N - I
 \end{aligned} \tag{6.44}$$

Ces forces interviennent également dans la somme des moments des forces appliquées sur le segment 2 :

$$C_2 + \vec{F}_2 \wedge \vec{G_2P_1} + \vec{N} \wedge \vec{G_2P_2} + \vec{I} \wedge \vec{G_2P_2} = I_{cm,2}\ddot{\phi} \quad (6.45)$$

En projetant sur les axes x et y :

$$C_2 + F_{x_2}h_2 \cos(\phi) + F_{y_2}h_2 \sin(\phi) + N(L_2 - h_2) \sin(\phi) + I(L_2 - h_2) \quad (6.46)$$

Au final, on obtient les expressions de Q et de P suivantes :

$$\begin{aligned} Q &= C_2 + B\dot{\theta}^2 \sin \beta + R[a_{x_E} \cos \phi + (g + a_{y_E}) \sin \phi] - Nh_2 \sin(\phi) - Ih_2 \sin(\phi) \\ &\quad - N(L_2 - h_2) \sin(\phi) - I(L_2 - h_2) \sin(\phi) \\ &= C_2 + B\dot{\theta}^2 \sin \beta + R[a_{x_E} \cos \phi + (g + a_{y_E}) \sin \phi] - NL_2 \sin(\phi) - IL_2 \sin(\phi) \\ P &= C_1 - C_2 - B\dot{\phi}^2 \sin \beta + S[a_{x_E} \cos \theta + (g + a_{y_E}) \sin \theta] - Nh_2 \sin(\phi) - Ih_2 \sin(\phi) \end{aligned} \quad (6.47)$$

Butées articulaires

Afin que le système simule un bras de manière réaliste, il est important de lui fixer des butées articulaires. Les valeurs de ces butées ont été déterminées de manière empirique, en expérimentant sur le bras d'un expérimentateur.

L'angle de l'épaule peut aller de -90° à 180° par rapport à la verticale et l'angle du coude peut aller de 0° à 165° par rapport à l'angle de l'épaule (voir figure 6.24) .

6.4.2. Simulation du bras et résultats

Maintenant que l'on a une simulation de mouvement du bras, on peut lui appliquer le modèle d'apprentissage. Pour ce faire, nous avons réalisé quelques modifications dans la structure du modèle.

Tout d'abord, comme on commande deux couples on sépare la commande en deux pour plus de simplicité au lieu de paver toutes les combinaisons d'actions possibles. Dans un premier temps on teste ce modèle simple appliqué à la simulation dans le cas classique d'une récompense posée sur (ou proche de) la surface et d'un frottement qui est source de récompenses négatives lorsque l'effecteur frotte le sol (ici la « main »).

On obtient le résultat présenté sur la figure 6.25 qui montre bien que le comportement obtenu précédemment dans l'interaction avec l'environnement via la simulation précédente se généralise bien sur cette nouvelle simulation.

Par la suite, nous combinons les différentes variantes que nous avons expérimentées précédemment. Ainsi on intègre la finesse de catégorisation vue dans la section 6.3.5 tout en l'améliorant afin d'y intégrer une catégorisation, non seulement de la position seule (ici du couple θ / ϕ), mais également une catégorisation de la position et de la vitesse ($\dot{\theta} / \dot{\phi}$), ceci permettant de ne plus compter uniquement sur le phénomène de convergence pas à pas (décrit section 6.3.4) pour réussir à converger vers une trajectoire stable malgré les variations d'inerties. On intègre ensuite également un contrôle en force tel que vu dans la section 6.3.5 en se limitant ici à trois forces différentes par contrôleur par souci de simplicité (ce qui donne 6 contrôleurs indépendants en tout).

Notons un point intéressant ici : la présence de plus de finesse dans le contrôle, contrairement à ce que l'on pourrait penser au premier abord, a donné lieu à une réduction importante du temps de convergence du système, probablement parce que la capacité à atteindre la récompense par des chemins plus diversifiés permet d'éviter d'attendre de découvrir un chemin trop spécifique pour converger. On intègre également une légère punition relative à l'énergie dépensée pour le contrôle comme vu dans la

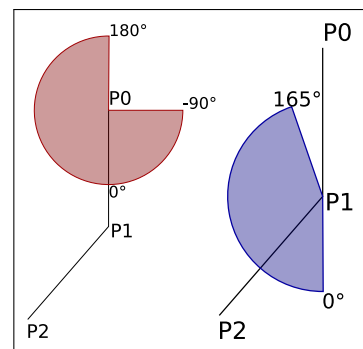


FIGURE 6.24. — Amplitude de mouvement de l'épaule (de -90° à 180°) et amplitude de mouvement du coude (de 0° à 165° par rapport à l'angle de l'épaule).

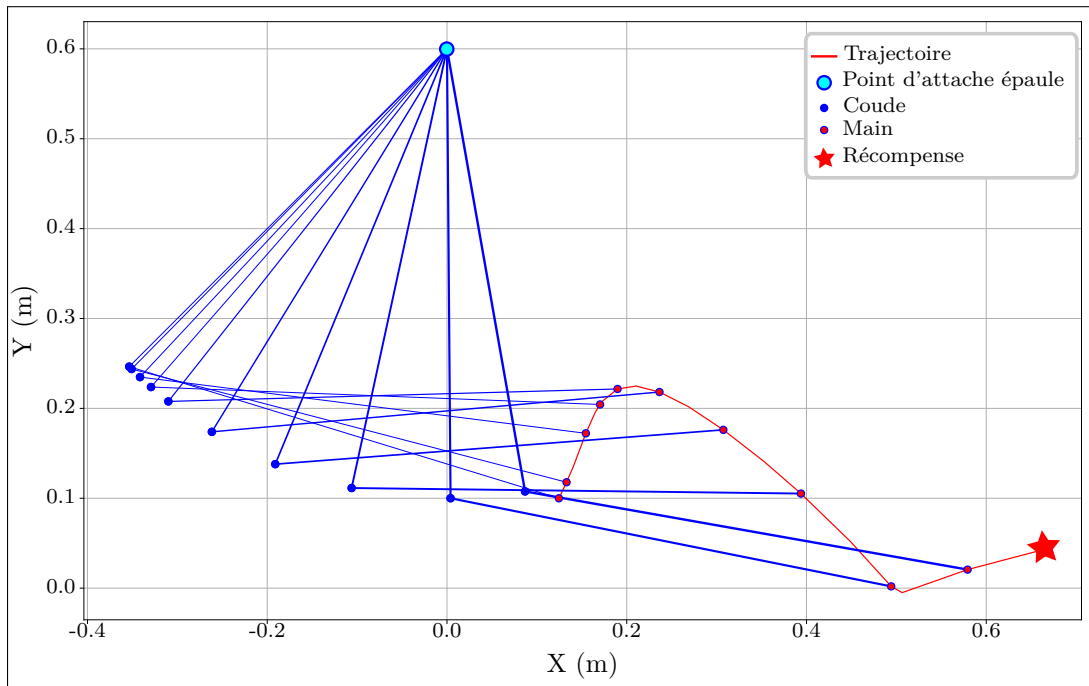


FIGURE 6.25. — Sur cette figure, on peut observer la progression du bras simulé après convergence. On observe bien la trajectoire typique en cloche afin d’atteindre la cible.

section 6.3.5 afin d’éviter de sélectionner n’importe quelle trajectoire dès lors qu’elle atteint la récompense.

Par souci de simplicité, on n’intègre par contre pas le ralentissement de l’effecteur via les frottements en lieu et place de la récompense négative vu que nous avons montré que les deux sont équivalents dans le résultat final.

Une fois ces aspects intégrés on peut tester ce modèle intégré sur la simulation du bras dans le contexte habituel d’une cible à atteindre sur le sol et d’une punition en cas de frottement. On obtient alors la courbe présentée sur la figure 6.26.

On peut observer que l’acteur exploite bien ici les deux effecteurs et leurs conséquences. En effet dans un premier temps la rotation de l’avant-bras permet d’éviter de frotter contre le sol. Puis la rotation du bras pour se rapprocher de la récompense induit un abaissement de l’effecteur. Mais celui-ci n’est pas compensé immédiatement pour ne pas induire de perte d’énergie inutile et l’avant-bras ne compense alors que la gravité pour ne se laisser tomber qu’à la dernière minute. La composition des deux mouvements donne cette trajectoire un peu particulière.

On applique ensuite le même cas « social » que celui que nous avons posé dans nos hypothèses formulées plus en détail dans la section 6.3.6 : une zone restreinte de forte punition autour de la récompense, servant à matérialiser l’échec en cas de passage d’objet à un partenaire. On observe le résultat sur la figure 6.27.

On peut y observer la trajectoire bien plus haute et prudente de la « main » simulée dans ce contexte. Notons d’ailleurs ici que la trajectoire est bien plus haute proportionnellement au résultat précédent que les résultats chez l’humain. Ce qui montre la faiblesse de notre hypothèse simplifiée de contexte social : il rend la différence de trajectoires très dépendante du contexte et du détail de l’hypothèse.

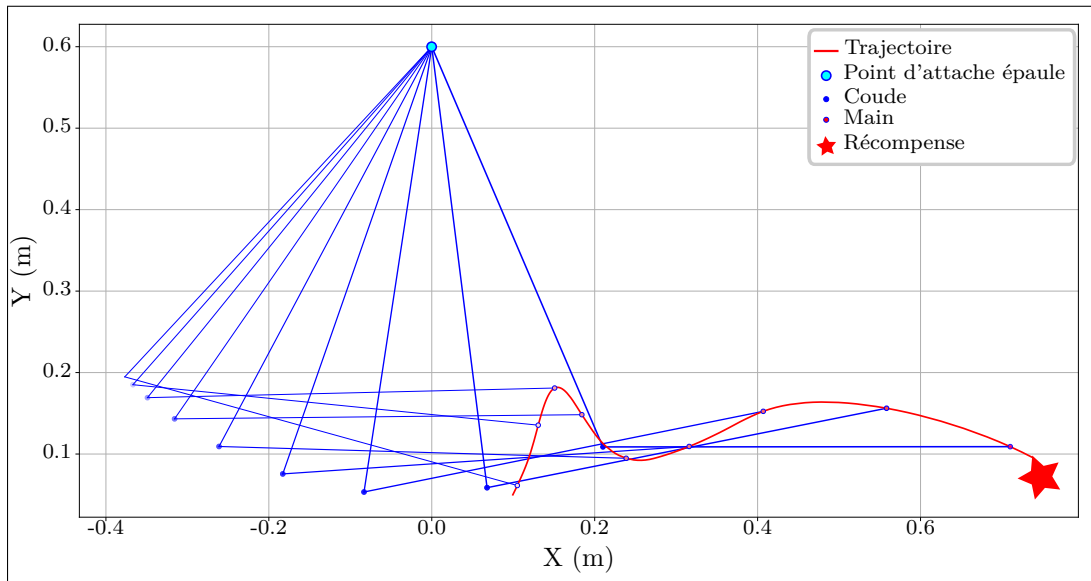


FIGURE 6.26. – Sur cette figure, on peut observer la progression du bras simulé après convergence avec le modèle intégrant toutes les variantes présentées précédemment.

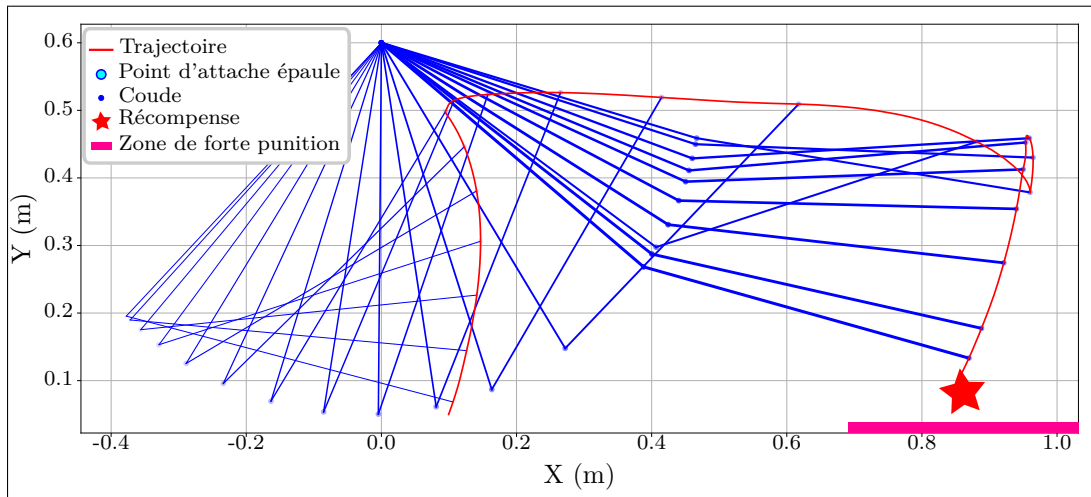


FIGURE 6.27. – Sur cette figure on peut observer la progression du bras simulé après convergence avec le modèle intégrant toutes les variantes présentées précédemment, mais cette fois-ci dans un contexte de passage d'objet à un partenaire, simplifié d'après nos hypothèses présentées section 6.3.6.

6.5 | Analyse et critiques

Dans ce chapitre, nous avons présenté nos travaux visant à créer et tester un algorithme basé sur un modèle acteur-critique inspiré du fonctionnement des ganglions de la base chez l'humain. L'objectif ayant été de montrer en quoi ce genre d'algorithme peut aider nos systèmes à interagir avec l'environnement de façon naturelle en intégrant dans leurs trajectoires des éléments également présents chez l'être humain lors d'interactions avec l'environnement.

Nous avons fait quelques hypothèses visant à modéliser simplement un contexte « social » et avons comparé l'ordre de grandeur des résultats avec les données existantes chez l'humain. Nous concluons de nos tests qu'il est tout à fait possible d'utiliser ce type d'algorithme pour transcrire une intention « sociale » dans le mouvement d'un robot. Ce qui pourrait amener à une meilleure interaction entre l'Homme et le robot.

Malgré ces aspects, il reste que l'algorithme présenté est très critiquable sur bien des points. Nous en discutons dans cette section.

6.5.1. L'influence des méta-paramètres et leurs crédibilités neurobiologiques

Le modèle que nous avons présenté ici est dépendant de ses paramètres de fonctionnement, plus que le modèle que nous avons présenté dans la partie précédente. De nombreux paramètres sont ici présents et peuvent jouer sur la forme des trajectoires. Tout d'abord, il y a le format des sorties et des entrées, s'ils jouent sur la trajectoire, comme nous l'avons montré, ils n'impactent pas forcément le fonctionnement du modèle en lui-même de façon critique, mais ensuite beaucoup d'autres paramètres peuvent impacter plus sévèrement le fonctionnement du modèle, nous les détaillons ici.

Paramètre de propagation de la récompense

Les paramètres de propagation de la récompense, et notamment le paramètre de prise en compte de l'estimation future dans le calcul de la récompense, ont un impact fort sur le fonctionnement de l'algorithme. Ce paramètre est assez critique, car il va agir sur le nombre d'états successifs qui peuvent transmettre une récompense. Si le paramètre est trop petit, la prédiction de récompense se verra diminuer très vite en s'éloignant de la récompense, et à l'inverse s'il est très proche de son maximum elle se propagera très loin.

Ce changement de comportement est assez critique, car cela signifie qu'un mauvais choix interdira de prendre en compte une récompense « lointaine », d'autant plus que les états sont ici des successions temporelles.

Ceci pose quelques problèmes de cohérence avec ce que l'on sait des comportements humains, puisque dans notre modèle, des événements liés entre eux, mais éloignés dans le temps (marcher d'un point A à un point B éloigné par exemple) ne pourront jamais se propager de récompenses de manière efficace.

On peut résoudre partiellement le problème en agissant sur ce paramètre en fonction de la différence entre les états, mais il reste néanmoins qu'il s'agit d'un mécanisme assez peu souple que l'on s'imagine mal exister seul chez l'être humain au vu de ses limitations. D'un point de vue neurobiologique, comme nous en avons discuté précédemment, l'emplacement neurobiologique de chaque membre de l'équation du calcul du signal dopaminergique est encore sujet à beaucoup de débats et il est donc difficile d'utiliser l'état de l'art de la neurobiologie pour trancher. Nous nous contenterons donc de signaler que cet aspect est probablement le point le plus faible du modèle, qui mériterait un travail conséquent de bibliographie et d'études pour converger vers une variante plus solide et bio-inspirée.

Signal de fatigue

Nous introduisons dans notre modèle la prise en compte d'un critère énergétique via une récompense négative censée transcrire un signal de fatigue musculaire. En réalité, cette prise en compte d'une fatigue

en tant que récompense négative est quelque chose d'assez sommaire. Les mécanismes de fatigue chez l'humain sont assez complexes et impactent toute la chaîne de commande depuis le muscle jusqu'aux motoneurons inférieurs pour enfin atteindre le système nerveux central (TAYLOR et al., 2016). La façon dont cette fatigue vient agir sur les ganglions de la base est encore aujourd'hui un sujet de débat qui n'est pas tranché. Néanmoins, on peut douter du fait que cela se traduise par une simple récompense négative, les effets de bords étant en effet trop nombreux (système « flemmard » qui préfère ne faire aucune action si récompense trop basse ou lointaine) pour être cohérent si existant tel quel chez l'humain. Ici encore il faudrait étudier plus en profondeur ce sujet spécifique en vue d'améliorer le système existant et d'implémenter un système de prise en compte de l'énergie sur robot qui soit pertinent.

6.5.2. Rapidité de convergence, un frein au développement sur robot ?

Nous avons assez peu abordé quantitativement la question de la rapidité de convergence dans nos cas applicatifs. Mais il suffit de dire que pour converger il faut bien souvent plusieurs milliers d'expériences complètes (une trajectoire réussie d'un point A à un point B) pour se rendre compte que l'implémentation du modèle actuel sur robot reste problématique, ce qui était attendu et l'une des raisons pour lesquelles nous avons décidé de nous orienter vers des modèles de simulations simples.

Néanmoins, il existe sans doute plusieurs pistes qui pourraient amener à améliorer légèrement la situation. Tout d'abord, la rapidité pour converger est très différente du temps mis pour obtenir des trajectoires instables, mais valides et orientées vers un but. Les orientations de l'effecteur dans une direction pertinente intervenant bien en amont de la convergence. Ensuite, certains des paramètres que nous avons utilisés pour accélérer la convergence (la propagation sur plusieurs états passés par exemple) ont une influence positive sur le temps de convergence. D'autres paramètres, comme la mise en place d'effecteurs analogiques, ont aussi eu comme conséquence d'améliorer le temps de convergence.

Néanmoins, il reste des interrogations ici encore vis-à-vis de la crédibilité biologique d'un tel système dans le cas général. L'humain est en effet capable de changer de comportement, du moins temporairement, assez rapidement après plusieurs expériences positives ou négatives inattendues. Bien plus rapidement que les ordres de grandeur pour optimiser son comportement par la suite. On peut alors se demander dans quelle mesure la mise en place de système d'apprentissage concurrent, plus rapide, est nécessaire pour représenter ce type de comportement. Ou alors on peut supposer que le fonctionnement des ganglions de la base est capable d'effectuer ce changement comportemental brusque grâce à des mécanismes de modulation implicite de l'apprentissage en fonction de la nouveauté de la situation.

Une autre façon de le voir est de penser que les changements comportementaux selon le contexte ne sont pas des liens aussi simples que dans le modèle présenté ici. Si l'on imagine, par exemple, des modèles plus complexes basés sur des réseaux chaotiques, le passage d'un état à un autre peut se faire avec un apprentissage très léger capable d'induire un déséquilibre basculant le système d'un attracteur à un autre. Ce type de modèle pourrait être une explication aux basculements de comportements rapides sans qu'il ne soit alors nécessaire de disposer d'autres systèmes d'apprentissages.

6.5.3. L'hypothèse sociale, une hypothèse simpliste qui mériterait d'être affinée

Dans nos simulations, nous avons posé une hypothèse pour aboutir à des résultats dans un contexte social. Notre hypothèse se basait sur le fait qu'un échec dans des conditions de passage d'objet à un humain impliquait plus de facteurs négatifs que de rater le fait de poser l'objet sur une surface. Cette hypothèse, même si elle donne les résultats attendus, est également assez sommaire. En réalité, il aurait été plus juste de simuler le second cas proposé dans la section 6.3.6, c'est-à-dire une simulation prenant en compte une boucle complète entre la cible perçue et l'action afin de pouvoir réagir à une cible capable de mouvement. Nous n'avons pas mis en place cette simulation dans la thèse pour diverses raisons. Tout d'abord d'un point de vue pragmatique la prise en compte de la position de la cible est une convergence

autrement plus complexe que ce que nous avons testé jusqu'ici. Si elle ne pose théoriquement pas de problèmes particuliers, le temps de convergence, lui, serait probablement démultiplié. Les simulations étant déjà longues à converger ceci nous a retenus d'affiner nos travaux dans cette direction. Ensuite, cela nous a paru inutile sur le long terme. Sur le long terme, l'objectif est d'utiliser cette simulation et de la mêler au modèle proposé dans la partie 2. Nous avons vu que le modèle proposé dans la partie 2 est capable de construire un pavage d'actions pertinentes et devient capable de suivre une cible après un temps d'apprentissage assez court. Ce premier modèle simplifierait donc énormément le problème à résoudre pour le modèle présenté ici dans un contexte de cible mouvante et c'est donc plus probablement dans cette direction qu'il valait mieux se diriger. Néanmoins, le modèle de la partie 2 nécessite encore plusieurs travaux afin de le rendre compatible avec le modèle présenté ici, il faudrait notamment, dans un premier temps, apporter des éléments de réponses aux points qui ont été soulevés dans la section 4.3.4 de la partie 2 avant d'envisager un modèle commun.

6.6 | Conclusion

Dans cette partie, nous avons voulu apporter des éléments de réponses à la question de l'adaptabilité de la commande motrice aux conditions environnementales et à la présence de partenaires dans le but de compléter les défauts du modèle proposé dans la partie 2 vis-à-vis de l'interaction Homme-robot dans un environnement réel.

Pour ce faire, nous avons développé une architecture acteur-critique basée sur un modèle des ganglions de la base chez l'humain et nous l'avons appliqué sur deux simulations simples. Nous avons ensuite formulé des hypothèses sur les cas d'interactions sociales en nous basant sur les résultats de la littérature étudiant le mouvement humain et nous les avons appliqués à nos simulations.

Nous avons ainsi montré que dans ces hypothèses le mouvement était capable de refléter la présence d'un partenaire dans ses caractéristiques, notamment la volonté d'interagir ou non avec un partenaire.

Il reste probablement encore beaucoup de voies d'amélioration au modèle présenté ici, il est notamment possible d'étudier plus en détail les aspects neurobiologiques pour résoudre certains des défauts discutés précédemment, et il doit également être possible d'apporter des pistes à une implémentation sur robot qui ne sacrifierait pas les bonnes propriétés du modèle. De même, il doit être possible d'améliorer considérablement les hypothèses « sociales » développées ici pour obtenir un résultat dépendant moins des hypothèses de travail.

Pour finir, vu qu'il s'agissait d'un premier jet de modélisation sur ces aspects, nous n'avons pas comparé, en dehors des travaux de Cristina Becchio et collègues, les aspects quantitatifs des formes de trajectoires dans ce type de cadre chez l'humain principalement parce que les algorithmes présentés ici sont encore trop rustiques pour que la comparaison soit pertinente. Néanmoins, si ces algorithmes sont développés autour de modèles plus solides dans le futur, il faudrait très probablement aboutir à des comparaisons quantitatives.

IV

Quatrième partie :

**Synchronisation non-intentionnelle et
réseaux oscillants**

Synchronisation non-intentionnelle, impact chez l'humain et usage en robotique

Sommaire du chapitre :

7.1. Introduction	174
7.2. La synchronisation non-intentionnelle chez l'humain	175
7.2.1. Modèle de synchronisation pour la coordination	176
7.2.2. Diverses démonstrations de la validité du modèle HKB	177
7.2.3. L'importance du type de couplage	178
7.3. Synchronisation et robotique	180
7.3.1. Introduction : des applications et des questions très diversifiées	180
7.3.2. Travaux de caractérisation de la synchronisation non-intentionnelle chez l'humain	181
7.3.3. Les réseaux de neurones oscillants, un modèle pour la synchronisation en robotique	185
7.3.4. Discussion sur la synchronisation en robotique	188
7.4. Conclusion. L'étude des réseaux oscillants, un intérêt pour la robotique interactive? . . .	194

7.1 | Introduction

Si on considère le système nerveux humain comme un ensemble de filtres, de boucles et de réflexes alors on peut le voir comme une somme de processus dynamiques interconnectés qui réagissent face aux stimuli du monde extérieur. On peut ainsi imaginer que de nombreuses propriétés du comportement humain, notamment dans l'interaction avec un partenaire ou l'environnement, découlent des propriétés dynamiques du système modélisé.

Dans les précédentes parties nous avons, par exemple, introduit deux modèles principaux d'apprentissage qui sont destinés à coopérer, l'un sur robot et l'autre en simulation. Ces modèles se contentent toutefois de faire correspondre une entrée donnée à une sortie spécifique. Dans le premier cas présenté, il s'agissait d'un apprentissage associatif classique qui nous permettait, dans l'interaction avec le monde, de construire un pavage des actions possibles pour le robot et de démontrer l'apparition, via ce processus, de phénomènes présents chez l'humain tels que le pointage ou l'imitation. Dans le second cas, nous nous étions arrêtés sur la question de l'usage de ces primitives motrices, une fois construites, via un réseau d'apprentissage par renforcement dans un environnement où peuvent être présents des humains. Toutefois, aucun des deux cas considérés ne possédait réellement de dynamique « propre » capable de se coupler avec l'environnement, aspect que l'on sait pourtant importants pour l'interaction de l'humain avec son environnement ou un partenaire. De fait, dans le cas de ces deux précédents modèles, l'étude de ce qu'aurait pu apporter à l'interaction la présence de dynamiques intrinsèques était difficilement possible.

Pour répondre à la problématique des propriétés dynamiques d'un système de contrôle et de l'apport potentiel que peut avoir ce type de propriété dans l'interaction, nous abordons dans cette partie un processus dynamique simple qui, une fois mis en situation d'interaction, peut aboutir à un phénomène d'interaction avec l'environnement ou avec un partenaire : la synchronisation non-intentionnelle. Ce phénomène présente l'avantage de pouvoir être présent dans chaque partie du cerveau. C'est une « brique » essentielle dont il convient d'étudier l'importance et la possible intégration avec les autres modèles proposés dans cette thèse et avec les modèles d'apprentissage existant en IA en général.

Dans un premier temps dans ce chapitre nous présenterons le phénomène en question ainsi que les travaux déjà réalisés à ce sujet. Nous exposerons les travaux que nous avons menés pour préciser et caractériser ce phénomène dans l'interaction avec l'humain. Nous aborderons ensuite les limites des hypothèses couramment formulées dans l'usage et l'étude de ce phénomène en robotique, ainsi que les limites techniques des outils classiquement employés pour le modéliser. Enfin, nous proposerons plusieurs modèles basés sur la littérature de neurobiologie récente afin de réduire les limites techniques de ces outils et ainsi permettre leur intégration dans des modèles de contrôle moteur.

7.2 | La synchronisation non-intentionnelle chez l'humain

Les interactions sociales chez l'humain peuvent donner lieu à des phénomènes de coordination, que ce soit deux personnes discutant ensemble et échangeant ainsi via des successions de phrases et pauses, qui peuvent être vues comme autant de signaux implicites permettant de comprendre l'autre et de continuer la discussion, ou que ce soit en portant un objet à deux ou en marchant côte à côte à la même vitesse. Les interactions sociales de ce type sont généralement regroupées sous le nom de *coordinations interpersonnelles* (HARI et KUJALA, 2009).

Parmi celles-ci, deux catégories principales sont généralement isolées dans la littérature, la similarité comportementale (de l'anglais « *behavioural matching* ») et la synchronisation interactionnelle (BERNIERI et ROSENTHAL, 1991). La similarité comportementale fait principalement référence aux coordinations interpersonnelles dites « orientées vers un but » tandis que la synchronisation interactionnelle fait plutôt référence aux propriétés dynamiques (rythmiques, cycliques) des coordinations entre humains.

Dans cette seconde sous-catégorie que sont les synchronisations interactionnelles, de nombreux travaux ont postulé ou démontré l'omniprésence et l'importance des phénomènes de synchronisation chez l'humain (MCNEILL, 1997; OULLIER et al., 2008), chez d'autres espèces (CLAYTON et al., 2005) et dans la nature en général (GLASS, 2001). Certains de ces travaux se sont attachés à la synchronisation dite « intentionnelle », c'est-à-dire volontaire de la part des deux partenaires, généralement orientées vers un but (on pourrait citer la danse (LARGE, 2000), ou encore le port d'un objet lourd avec un partenaire (MARSH et al., 2009)). D'autres travaux se sont attachés aux aspects « non-intentionnels » de ces synchronisations entre humains, parmi lesquels :

- La coordination entre deux personnes qui marchent et qui synchronisent le mouvement de leurs jambes non-intentionnellement (ZIVOTOFSKY et HAUSDORFF, 2007; ULZEN et al., 2008).
- La synchronisation des postures durant une discussion entre deux personnes (SHOCKLEY et al., 2003; SCHMIDT et RICHARDSON, 2008).
- La synchronisation des applaudissements durant un spectacle. (NÉDA et al., 2000).
- La synchronisation de deux personnes discutant entre elles pour passer naturellement d'un interlocuteur à l'autre (*turn-taking*) (WILSON et WILSON, 2005). Celle-ci est présente dès le plus jeune âge (MILLER et OLLS, 1980).
- La synchronisation des rythmes cardiaques et respiratoires entre deux partenaires en interaction (WATANABE et OKUBO, 1997).

Notons que cette synchronisation non-intentionnelle apparaît même si l'on demande spécifiquement au sujet de ne pas se synchroniser (ISSARTEL et al., 2007), ce qui tend à démontrer qu'il s'agit d'une propriété intrinsèque du système à un niveau élémentaire. Notons également que la limite entre « intentionnelle » et « non-intentionnelle » peut être assez floue dans certains des cas présentés ci-dessus. Pour finir, le phénomène de synchronisation comportementale en général est loin d'être limité à l'humain, on le retrouve un peu partout dans la nature, du clignotement des lucioles (MIROLLO et STROGATZ, 1990) au chant des grillons (WALKER, 1969).

Les deux types de synchronisation, intentionnelle et non-intentionnelle, se distinguent chez l'humain, non seulement par la volonté qui y est associée, mais aussi par le « coût attentionnel » qui en résulte. Ainsi il a été démontré que les synchronisations non-intentionnelles pouvaient s'accompagner d'un coût attentionnel réduit par rapport aux variantes intentionnelles (TEMPRADO et LAURENT, 2004). Ce phénomène est intéressant à deux points de vue :

- Le premier avantage potentiel des synchronisations non-intentionnelles est leur aspect peu coûteux pour l'attention qui permettrait alors leurs usages dans des applications d'interaction ou d'interfaçage Homme-machine. En effet, ces interactions ne portent pas préjudice à la tâche en cours, qui peut bénéficier de la pleine attention de la personne. De fait, s'il était possible de les exploiter elles pourraient alors devenir un levier utile pour aider, orienter ou améliorer la tâche de

l'humain en interaction avec la machine (SCHMIDT et RICHARDSON, 2008). Même si la manière de l'employer pour ces cadres reste hypothétique.

- Le second avantage est qu'au-delà de leur coût attentionnel faible certaines hypothèses laissent supposer que la synchronisation non-intentionnelle pourrait avoir des avantages directs pour la vie au jour le jour de l'être humain et ses interactions. La synchronisation des jambes de deux personnes qui marchent ensemble par exemple, dans le cas de jambes de tailles approximativement identiques, pourrait aider ces personnes à implicitement évoluer à vitesse identique. Non seulement le coût attentionnel faible de ce processus permettrait alors à l'humain de faire une autre activité (tenir une discussion par exemple), mais en plus il lui économise une action qui aurait pu lui demander des ressources attentionnelles supplémentaires potentiellement importantes (ajuster sa vitesse).

Ce second point est particulièrement intéressant dans notre approche, car il signifie que de nombreux phénomènes d'interaction avec l'environnement ou avec le partenaire pourraient être offerts « gratuitement » par ce phénomène de synchronisation non-intentionnelle. Isoler la source de ce phénomène et réussir à l'intégrer à des architectures de contrôle robotique en IA signifierait alors, si ces hypothèses sont valides, d'être capable de répliquer cet apport gratuit dans de nombreux domaines d'interaction.

7.2.1. Modèle de synchronisation pour la coordination

Dans le cadre de son étude de la coordination entre les deux mains d'un humain durant une tâche conjointe, J. A. Scott Kelso observe que si l'on demande à une personne de faire un mouvement des doigts de façon répétitive en antiphase, alors en augmentant peu à peu la fréquence du mouvement le sujet va se mettre à resynchroniser ses doigts sur la même phase (KELSO, 1995). Kelso propose alors que ces propriétés présentent des similitudes avec celles de deux oscillateurs couplés, un concept permettant l'expression de la synchronisation et que l'on retrouve sous de nombreuses formes analogues en science dans divers cadres dans lesquels un transfert d'information existe entre plusieurs processus dynamiques (PIKOVSKY et al., 2003).

Plus tard, ce modèle sera étendu au-delà de la coordination d'un humain avec lui-même pour se généraliser aux coordinations sociales entre humains, en prenant pour hypothèse que le principe est similaire (SCHMIDT et al., 1990).

Le modèle oscillant proposé par Kelso et collègues est connu sous le nom de modèle « HKB » (Haken, Kelso et Bunz) et met en jeu deux oscillateurs mécaniques (ayant une masse, une longueur, etc.) s'échangeant de l'énergie via un facteur appelé le couplage. Un phénomène qu'on peut observer simplement en posant deux oscillateurs suffisamment proches l'un de l'autre sur une table : les vibrations transmises de l'un à l'autre via leurs socles respectifs à travers la surface de la table suffiront à transmettre une énergie qui induira alors une perturbation mutuelle des deux oscillateurs (voir figure 7.1).

À terme, cette perturbation aboutira à une synchronisation si les fréquences propres des deux systèmes sont suffisamment proches. On parle « *d'accrochage de phase* » quand on rentre dans le phénomène de synchronisation et de « *décrochage de phase* » quand on en sort. L'importance du phénomène varie selon la force et la phase de la perturbation (SCHÖNER, 1990) et le phénomène est ensuite définitif (le système ne revient pas aux phases d'origine si la perturbation s'arrête (KAY et al., 1991)) : on parle de « mémoire », car le système conserve ainsi une trace des événements qui l'ont perturbé dans le passé. Une fois le système stabilisé il est alors caractérisé par sa différence de phase, qui se maintient constante (on parle donc de synchronisation) (PIKOVSKY et al., 2001). Les conditions initiales des deux systèmes et leurs propriétés définiront alors la différence de phase finale.

Plus le couplage est faible plus les oscillateurs sont indépendants. Au contraire, plus le couplage est fort plus ils sont dépendants l'un de l'autre. La particularité du phénomène est que la synchronisation peut avoir lieu pour des oscillations d'amplitudes très différentes entre les deux oscillateurs lorsque le couplage est faible. Cet aspect est intéressant, car il signifie qu'un oscillateur présentant une certaine forme d'oscillation peut se synchroniser sur un phénomène extérieur tout en conservant sa propre forme, et cela même si le phénomène extérieur est de forme très différente.

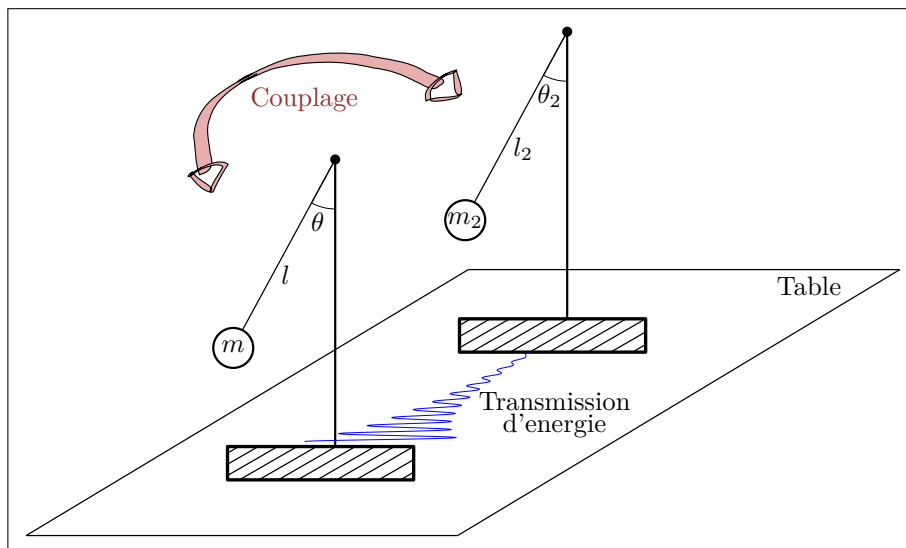


FIGURE 7.1. — *Modèle des effets de synchronisation dans les coordinations interpersonnelles. Deux oscillateurs s'échangeant de l'énergie via un médium (ici les vibrations de la table) vont s'influencer mutuellement et sous certaines conditions, notamment lorsque leurs fréquences propres sont proches, ils vont se synchroniser.)*

7.2.2. Diverses démonstrations de la validité du modèle HKB

Si le modèle d'un humain se comportant comme un ou des ensembles d'oscillateurs est justifié alors on doit pouvoir le valider chez l'Homme dans des expériences simples. De nombreux travaux récents se sont ainsi attachés à démontrer la capacité de simples phénomènes dynamiques non humains à venir perturber le mouvement humain.

Ainsi Bingham (BINGHAM, 2004) s'est attaché à démontrer en théorie que la vision pourrait être une source essentielle de la perturbation des mouvements humains. Dit autrement, cela signifie que la table dans le modèle des deux oscillateurs interagissant entre eux via un médium, peut être remplacée par la vision. La vision humaine permettant de transmettre les informations de l'« oscillateur » extérieur jusqu'à l'« oscillateur » de l'humain, expliquant ainsi la possibilité d'un effet d'entraînement (voir figure 7.2).

Il doit donc être possible d'impacter les mouvements humains par l'excitation rythmique du champ visuel. Et c'est bien ce qu'ont pu observer de très nombreux auteurs en utilisant divers protocoles tels qu'un point mouvant dans le champ visuel de façon rythmique (PEPER et BEEK, 1998; SCHMIDT et al., 2007; RUGY et al., 2008; HAJNAL et al., 2012). De façon générale, les mouvements humains se synchronisent involontairement avec le stimulus visuel, quelle que soit la tâche qu'ils exécutent et on observe bien les phénomènes d'accrochage et décrochage en phase selon que les différences de fréquences propres soient importantes ou non. Ceci démontre deux choses :

- Qu'un humain possède bien une certaine fréquence propre associée à ses mouvements, du moins en moyenne et en considérant type de mouvement par type de mouvement.
- Et que le modèle d'oscillateurs ayant des influences mutuelles est, en première approche, un modèle pertinent de l'interaction entre l'environnement et l'humain ou entre deux partenaires.

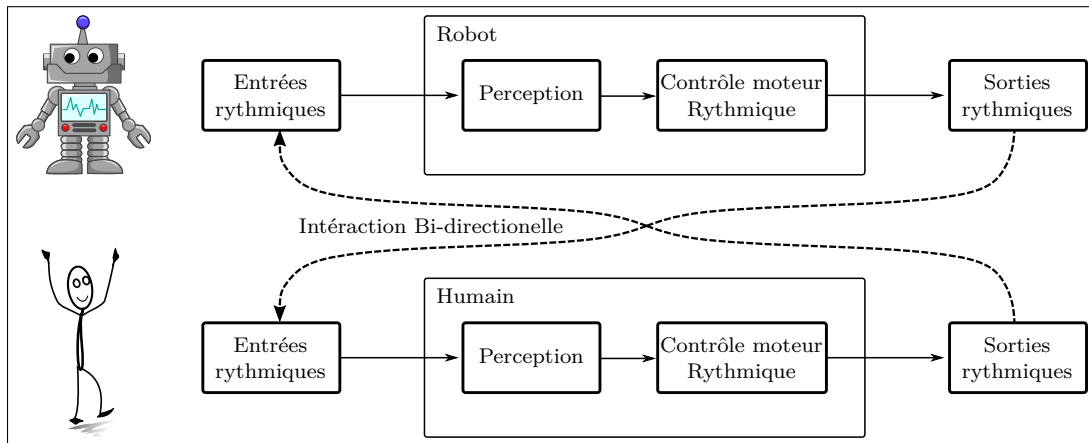


FIGURE 7.2. – Dans un cas d'interaction bidirectionnelle (ici entre un humain et un robot, mais le principe est le même entre deux humains), la vision, et les perceptions en général, deviennent le médium de transmission d'informations entre les deux systèmes « oscillants ».

7.2.3. L'importance du type de couplage

Pour étudier les perturbations dynamiques du mouvement humain, on peut mettre en place une approche simple : on se contente classiquement d'une influence unidirectionnelle. Un point mouvant dans le champ visuel humain vient perturber le mouvement humain durant la réalisation d'une certaine tâche. Mais ce point possède, lui, une fréquence fixe.

Néanmoins, ce protocole a des limites, car dans la réalité beaucoup de phénomènes dynamiques présents dans l'environnement sont le fruit de mouvement induit par des êtres vivants, surtout dans le cadre de l'évolution d'un système dans un environnement où les humains sont présents. Ceci implique que dans la plupart des cas le générateur de ces perturbations est également perturbé en retour par notre propre mouvement. De fait, il est plus pertinent de modéliser le protocole selon les hypothèses du modèle HKB initial : une influence mutuelle, plutôt que comme un système où l'influence ne serait qu'unidirectionnelle (PIKOVSKY et al., 2001 ; SCHMIDT et RICHARDSON, 2008).

C'est dans ce cadre que la robotique trouve naturellement sa place dans l'étude de l'influence des perturbations dynamiques sur le mouvement humain. Non seulement ces études peuvent être utiles à l'amélioration des modèles de contrôle robotiques dans l'espoir de pouvoir améliorer l'interaction du robot avec son environnement et l'humain, mais le robot devient également une pièce essentielle des études du point de vue du chercheur travaillant sur le mouvement humain. En effet dans un cadre d'influence bidirectionnelle il est nécessaire de disposer d'un système capable d'action dans le champ visuel humain et également capable de réagir dynamiquement à des stimuli venant de l'humain. Un robot est un système répondant parfaitement à ces contraintes.

Impact de l'aspect humanoïde dans la synchronisation non-intentionnelle entre humains

Malgré le fait que certains travaux ont reproduit des perturbations motrices avec de simples points mouvants, l'impact de la forme de l'agent artificiel ou du robot sur la synchronisation a longtemps été débattu (MARIN et al., 2009). En effet, certaines théories de la résonance motrice considèrent que pour avoir un impact sur nos propres mouvements seuls les agents « humanoïdes » sont pertinents. C'est une généralisation du point de vue particulier que nous avons présenté dans la partie II de cette thèse : le lien entre la perception et l'action passe par des neurones miroirs qui se construisent sur un phénomène d'ambiguïté perceptive (selon notre théorie). Dans ce cas, les neurones répondant à des stimuli proches

de ceux de notre propre corps seront ceux qui se retrouveront les plus liés à l'espace d'action. De fait, le mouvement perçu d'un système présentant des ressemblances avec l'humain a théoriquement plus de chance de voir sa dynamique impacter l'espace d'action. Dit autrement : son couplage sera plus fort.

À l'extrême de ce point de vue, James M. Kilner et collègues ont détecté un effet dans leurs expériences uniquement si le bras était humain et non robotique (KILNER et al., 2003). Néanmoins, l'expérience fut reproduite par la suite par d'autres équipes en utilisant un robot humanoïde complet au lieu d'un bras désincarné et cette fois le même effet fut détecté que l'agent soit humain ou non (OZTOP et al., 2005). Ludovic Marin et collègues ont proposé que le phénomène puisse être dû à un gradient de « non-humain » à « humain » avec tous les intermédiaires entre ces deux extrémités (MARIN et al., 2009). Le bras robotique désincarné de l'expérience de James M. Kilner serait alors trop « non-humain » pour que l'effet soit visible.

Toutefois, gardons à l'esprit que d'autres expériences ont reproduit les résultats de synchronisation ou de perturbation du mouvement moteur en général avec de simples points ou objets mouvants dans le champ visuel. On peut probablement en tirer une conclusion plus tempérée sur la base de celle de Ludovic Marin et collègues : les stimuli « désincarnés » de simples points et objets mouvants causeraient bien une perturbation motrice, mais avec un couplage très faible, ce qui fait que leurs effets ne seraient observables que sous certaines conditions. Ceci rend l'observation du résultat très dépendante du contexte et peut expliquer que Kilner et collègues n'observent aucun effet dans leur cadre. À l'inverse, les stimuli produits par des agents humains ou humanoïdes auraient un couplage plus fort, ce qui permet d'en observer le résultat dans de plus larges conditions, notamment pour des fréquences propres de mouvement plus éloignées de celles de l'humain. Il y aurait donc bien un gradient, mais partant d'un impact non nul pour un objet artificiel.

L'expérience de Qiming Shen et collègues semble confirmer ce point de vue. Ils ont en effet comparé divers cadres dans lesquels un humain se synchronise sur différents « partenaires » ou objet (pendule, humains, robots, etc.) et en ont conclu que les synchronisations sont plus fortes pour des agents de forme humanoïde que dans les autres cas. (SHEN et al., 2010).

7.3 | Synchronisation et robotique

7.3.1. Introduction : des applications et des questions très diversifiées

Trois points ressortent de notre analyse : l'importance pour l'expérimentateur visant à isoler les propriétés du système humain d'utiliser un robot ; l'importance pour le roboticien d'isoler les propriétés du système humain et de les implémenter sur robot afin d'améliorer les interactions Hommes-robots ; la nécessité d'avoir un agent « humanoïde » pour maximiser le couplage.

L'ensemble de ces points a donné lieu à des objectifs très différents et à de nombreuses applications en robotiques dont on propose une séparation en quelques branches :

- Les expériences en robotiques visant à utiliser la synchronisation en général pour améliorer le résultat dans un contexte donné.
- Les expériences en robotique servant de « support » à des expériences de psychologie nécessitant un robot pour étudier la synchronisation chez l'humain.
- Les expériences en robotique cherchant à utiliser la synchronisation non-intentionnelle en partant de l'hypothèse que si elle s'avère utile à l'humain alors elle pourra s'avérer utile au robot en interaction.

Le premier groupe propose en général un usage de la synchronisation à but pragmatique sans particulièrement s'étendre sur le fait qu'elle soit intentionnelle ou non, qu'elle respecte les critères tels que les bornes d'accrochage et décrochage ou non, et généralement en étudiant plutôt les propriétés qui découlent de son usage dans des tâches nécessitant une synchronisation avec un phénomène extérieur.

Christopher Crick et collègues, par exemple, ont utilisé un robot batteur pour jouer de la batterie dans un groupe afin d'étudier comment les capteurs sensoriels du robot pouvaient permettre d'établir une synchronisation entre lui et les autres joueurs malgré un environnement bruyé. (CRICK et al., 2006). Les travaux de Tomofumi Kasuga et Minoru Hashimoto quant à eux (KASUGA et HASHIMOTO, 2005) ont cherché à utiliser des oscillateurs neuronaux pour qu'un robot puisse serrer la main à un être humain de façon harmonieuse. Artem Melnyk et collègues, sur la même problématique, ont appliqués des stratégies d'apprentissage à des réseaux formés de plusieurs oscillateurs (Central Pattern Generator) (MELNYK et al., 2013 ; MELNYK et al., 2014 ; MELNYK, 2014).

Le second groupe quant à lui, cherche à utiliser des implémentations de la synchronisation en robotique afin de servir d'expérience support à l'étude de la synchronisation chez l'humain. Généralement, il prend place dans des problématiques liées au milieu de la psychologie et on observe souvent que dans ce groupe la catégorisation du phénomène observé (intentionnelle, non-intentionnelle, borne d'accrochage, etc.) est plus précise. L'idée étant pour le roboticien qu'un apport aux études sur la synchronisation humaine pourra servir, par rebond, à sa propre discipline en permettant d'identifier des phénomènes utilisables sur robot.

Dans ce groupe on peut notamment citer Marek P. Michalowski et collègues qui ont ainsi utilisé un robot danseur pour étudier les effets de synchronisation entre humains et robot (MICHALOWSKI et al., 2007). À leurs suites, les travaux de Khursheed Hasnain et ses collègues (HASNAIN et al., 2013) ont été basés sur l'utilisation du robot Nao dans un cadre d'interaction avec l'humain afin de reproduire les phénomènes de synchronisation entre deux partenaires lors de mouvement du bras. Ceux-ci ont confirmé qu'un humain est bien incapable de s'empêcher de se synchroniser lorsque les fréquences sont suffisamment proches comme l'avait également démontré Johan Issartel et collègues (ISSARTEL et al., 2007). Dans cette catégorie d'expérience, on décrira une contribution réalisée dans le cadre de cette thèse, à l'issue d'un partenariat avec le laboratoire Euromov.

La troisième catégorie quant à elle tente d'utiliser la synchronisation, non-intentionnelle spécifiquement, afin d'illustrer les usages possibles à ce phénomène en se basant sur les hypothèses postulant ses apports chez l'humain. Les travaux de Ken Prepin et d'Arnaud Revel (PREPIN et REVEL, 2007) puis d'Arnaud Revel et Pierre Andry (REVEL et ANDRY, 2009), d'Arnaud Blanchard et Jacqueline Nadel (BLANCHARD et NADEL, 2009) ou encore de Ken Prepin et Philippe Gaussier (PREPIN et GAUSSIER,

2010) ont permis de démontrer la formation d'un lien de synchronisation entre un partenaire humain et un partenaire robotique dans divers cadres et d'essayer d'y trouver des usages robotiques. (tour de parole notamment)

Plus récemment, les travaux de Caroline Grand ont donné lieu à des applications de la synchronisation du mouvement des jambes en simulant les jambes par des bâtonnets bougeant rythmiquement sur un robot mobile en interaction avec un humain durant une marche (HASNAIN et al., 2014; LESUEUR-GRAND, 2017) ou en utilisant la synchronisation comme un signal de renforcement (GRAND et al., 2014).

Les travaux récents d'Eva Ansermin auxquels nous avons collaboré (ANSERMIN et al., 2016), puis de Mélanie Jouaiti et Patrick Henaff (JOUAITI et HENAFF, 2018) ont étendu le concept du processus de synchronisation à des mouvements plus complexes, en utilisant notamment des bancs d'oscillateurs afin de montrer comment la synchronisation non-intentionnelle peut être utilisée pour apprendre des trajectoires en robotiques. Ce dernier ensemble de travaux rapproche le domaine d'étude de la synchronisation de celui des travaux d'apprentissage de trajectoires à partir de bancs d'oscillateurs, et du domaine de l'étude des Central Pattern Generator (CPG) (DOYA et YOSHIZAWA, 1989; KALVERAM, 1998; IJSPEERT et al., 2002; SEO et al., 2010). Ces derniers utilisent eux aussi des méthodes d'apprentissage, mais en ne citant que rarement la synchronisation non-intentionnelle tout en l'employant en pratique de façon implicite dans leurs architectures afin d'aboutir à un résultat stable.

En pratique, ces groupes sont assez perméables, notamment quant aux méthodes qu'ils utilisent, et ils partagent de nombreux objectifs communs. Certains se positionnent explicitement sur deux points de vue distincts pour de mêmes expériences.

Ceci aboutit souvent à l'emploi d'outils communs ou à la mise en place de protocoles similaires dans des contextes scientifiques un peu différents. Ainsi, les travaux de Christopher Crick et collègues (CRICK et al., 2006) ou d'Artem Melnyk et collègues ont de fortes similitudes avec les travaux employant la synchronisation non-intentionnelle pour l'apprentissage d'Eva Ansermin (ANSERMIN et al., 2016) ou de Mélanie Jouaiti (JOUAITI et HENAFF, 2018), qui ont tous de fortes similitudes d'un point de vue modèle et outil avec les travaux illustrant l'apprentissage dans les CPG, quoique les objectifs soient présentés différemment.

Retenons donc que l'usage de la synchronisation en robotique a abouti à plusieurs directions de recherche différentes, mais qui peuvent parfois employer les mêmes outils et les mêmes approches.

Conclusion

Le modèle de l'humain comme étant un système dynamique interagissant avec l'extérieur a donné lieu à de nombreux travaux qui ont réussi à démontrer la validité et l'omniprésence des effets découlant du modèle. Notamment la présence des effets de synchronisation sur les mouvements humains.

Ces effets sont soupçonnés d'avoir des conséquences importantes sur les interactions entre humains tout en résolvant des problèmes d'interaction avec l'environnement de façon implicite, économisant ainsi des ressources mentales. Ils sont donc susceptibles d'être utilisés dans nos modèles de contrôle robotique afin d'en répliquer les bonnes propriétés.

En parallèle, la nécessité de mettre en place des protocoles utilisant des agents artificiels capables de perceptions et d'actions a mené à l'usage de la robotique dans les expériences d'études de ces effets chez l'humain. Ce qui a mené naturellement à des coopérations entre ces différents domaines de recherche.

Dans la section suivante, nous détaillerons une expérience menée dans le cadre d'une collaboration avec le laboratoire Euromov et qui a visé à la caractérisation plus précise des effets de synchronisations non-intentionnelles chez l'humain.

7.3.2. Travaux de caractérisation de la synchronisation non-intentionnelle chez l'humain

Comme nous en avons discuté, la robotique s'est positionnée de plusieurs manières dans l'étude des effets de synchronisation. L'un de ces positionnements est d'aider à la caractérisation du phénomène

de synchronisation chez l'humain dans le cadre d'interaction entre Hommes et robot, ceci afin d'aider à l'identification de propriétés de la synchronisation qui pourraient s'avérer utiles par rebond aux modèles que nous développons en robotique interactive. Dans ce contexte, nous présentons ici une expérience menée en partenariat avec le laboratoire Euromov qui vise à préciser certains des paramètres des phénomènes existants quand un humain se synchronise avec un agent artificiel.

Les travaux précédents travaillant dans le cadre de la synchronisation non-intentionnelle n'avaient en effet pas caractérisé finement les limites de celle-ci dans un cadre d'interaction Homme-robot.

Cette expérience a été réalisée en partenariat avec le laboratoire Euromov de l'université de Montpellier, spécialisé dans l'étude du mouvement humain. Cette partie de la thèse a été réalisée avec Eva Ansermin (Laboratoire ETIS) et Emmanuel Madieu (Laboratoire Euromov) dans le cadre de leurs stages de master 2 respectif (ANSERMIN, 2015 ; MADIEU, 2015a ; MADIEU, 2015b).

Méthode et outils

Dans cette expérience, nous cherchons à caractériser le détail des phénomènes de synchronisation dans une interaction Homme-robot. Pour ce faire, un protocole d'interaction a été établi dans lequel le robot sera un oscillateur forcé et où l'on cherche à entraîner l'humain sur cet oscillateur. Le robot Nao est disposé dans une pièce, assis sur une chaise, et le sujet est placé sur une table, les jambes pendantes, face au robot. Le dispositif peut être visualisé en photo sur la figure 7.3.

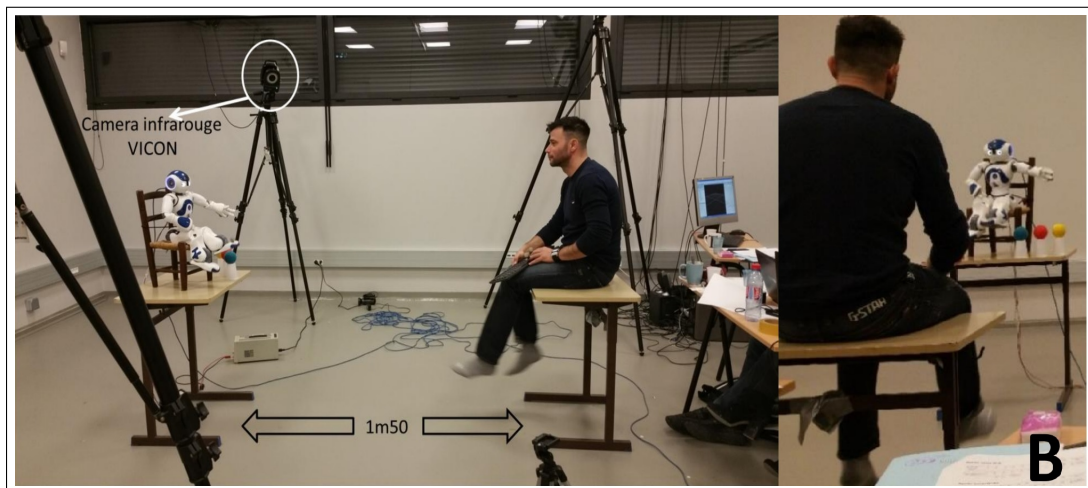


FIGURE 7.3. – Photographie du protocole d'interaction. On peut observer ici la tâche principale dite « de couverture » ou de réaction de la part du participant. Dès que Nao présentait avec son bras une des trois couleurs placées devant lui, les sujets devaient appuyer le plus vite possible sur le bouton de la même couleur sur le clavier en leur possession. C'était le mouvement des jambes du sujet qui était en réalité mesuré dans notre expérience grâce aux caméras de capture 3D.

Le but est d'arriver à déclencher un phénomène de synchronisation non-intentionnelle. Pour ce faire, on place le sujet dans une situation où il est aisé d'effectuer un mouvement rythmique non volontaire, comme ici, le fait de ne pas avoir les jambes qui ne touche pas le sol.

Pour s'assurer que le mouvement est non-intentionnel, les sujets sont naïfs quant au but réel de l'expérience et on leur demande de réaliser une tâche de couverture qui n'est pas la tâche réellement mesurée. Ici cette tâche consistait à appuyer le plus vite possible sur une touche (jaune, rouge ou vert) correspondant aux couleurs d'une cible désignée par le robot Nao avec son bras (voir partie B de la figure 7.3 ou figure 7.4).

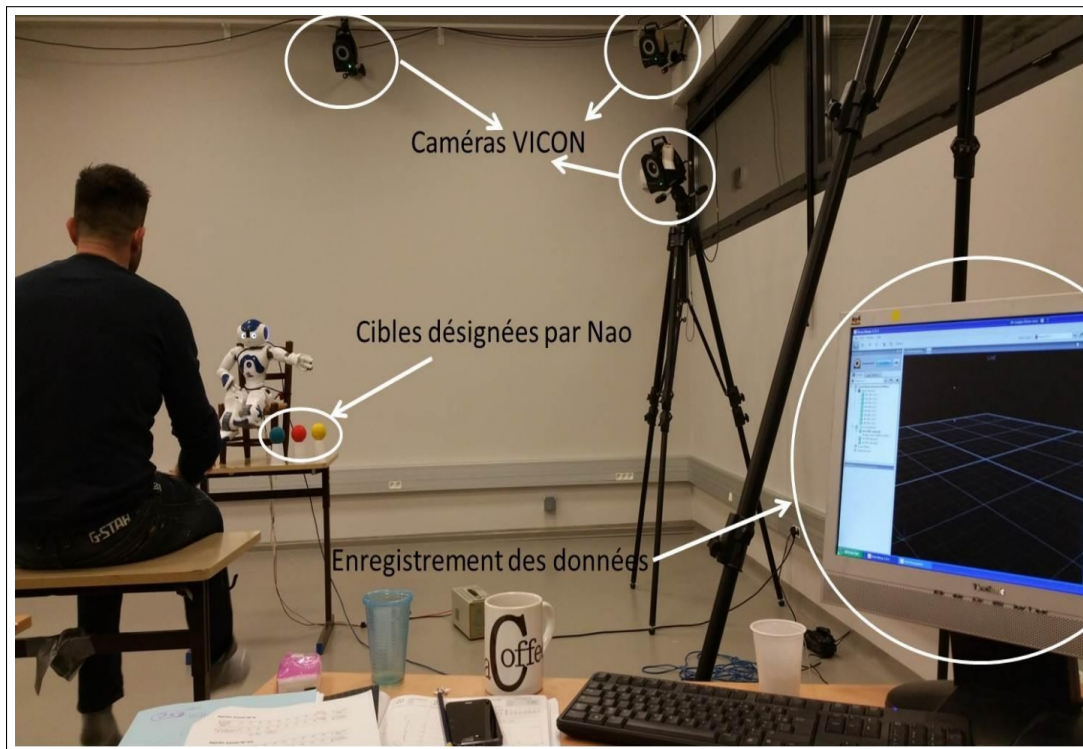


FIGURE 7.4. – Photographie du protocole d'interaction depuis le terminal de l'expérimentateur. On peut voir les boules pointées par le robot qui constituent une partie de la tâche de couverture.

Pour induire le mouvement naturellement on se place ici dans le cadre où le couplage est théoriquement le plus fort : le robot est en effet humanoïde et ce sont les jambes du robot qui sont contrôlées en oscillation dans notre cas (en antiphase l'une de l'autre afin de paraître le plus naturel possible). Comme la tâche est construite pour que les jambes du sujet oscillent naturellement, on s'attend à déclencher le maximum de résonance motrice entre le sujet et le robot et ainsi avoir un facteur de couplage suffisamment fort pour que le phénomène soit mesurable.

Pour éviter que les sujets ne se doutent de quelque chose en observant les variations de rythme des jambes du robot on leur précise qu'en réalité les jambes mouvantes du robot sont là pour les déconcentrer et qu'on cherche justement à mesurer l'impact de ces perturbations sur leur capacité à presser la bonne touche du clavier.

Durant la tâche on mesurera, grâce à un ensemble de caméras de capture en trois dimensions de type Vicon (les mêmes que celles décrites dans le dispositif de capture de mouvement de la partie I) le rythme des jambes du sujet ainsi que celui des jambes de Nao. Deux variables indépendantes ont ainsi été mesurées : l'ordre de présentation des fréquences de Nao (2 modalités : ascendante et descendante) et les paliers de fréquence d'oscillation des jambes de Nao en fonction de la fréquence préférentielle du sujet (9 modalités : de -20% à +20% par pas de 5%, voir figure 7.5).

Les variables dépendantes ont été l'index de synchronisation (mesure de la stabilité de la coordination non-intentionnelle) et la pente de la phase relative (mesure de l'entraînement/la synchronisation entre les jambes de l'humain et du robot).

Pour pouvoir effectuer ces décalages relatifs depuis la fréquence préférentielle des sujets on effectue dans un premier temps une mesure de cette fréquence préférentielle en demandant aux sujets de bouger les jambes comme ils le veulent afin de pouvoir maintenir ce rythme durant longtemps sans se fatiguer. L'idée étant d'insinuer qu'il s'agit d'une tâche destinée à les déconcentrer de la tâche « principale ».

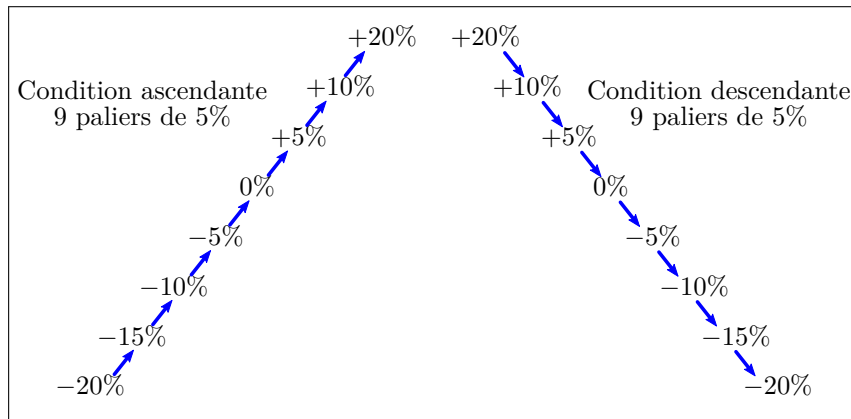


FIGURE 7.5. – Paliers de fréquences mesurées

Ceci nous permet de mesurer la fréquence préférentielle d'oscillation des jambes de chaque sujet avant l'expérience. Ceci afin d'ajuster les paliers en conséquence.

Pour finir, notons que dans cette expérience nous avons fait participer 21 sujets : 8 hommes (M : 27,6 ans, écart type : $\pm 6,9$) et 13 femmes (M : 32,1 ans, écart type : $\pm 6,5$). Les sujets étaient naïfs quant au but réel de l'étude. Leur participation était volontaire et non rémunérée.

Résultats

Qualitativement, les résultats ont bien montré une synchronisation autour de la fréquence préférentielle des sujets et une désynchronisation passé un certain seuil (voir figure 7.6 pour un exemple de l'accrochage-décrochage en phase sur un sujet). On peut observer les résultats du taux de synchronisation moyen sur la figure 7.7. Néanmoins, les résultats varient dans le détail si les paliers sont présentés dans un ordre ascendant ou descendant.

Une Anova sur l'index de synchronisation a révélé un effet significatif de la différence de fréquence ($F(8,88) = 14,33, p < 0,05, h2p = 0,57$). L'analyse post hoc de Newmal-Keuls a également démontré que les paliers de fréquence à 0% et 5% sont significativement différents de tous les autres paliers. Ce résultat indique que la synchronisation entre le robot et l'humain est significativement plus élevée lorsque la différence de fréquence entre les deux agents est faible.

Une Anova sur la pente de la dérivée de la phase relative a révélé un effet significatif du palier ($F(8,88) = 100,96, p < 0,05, h2p = 0,90$) et une interaction significative Ordre x Palier ($F(8,88) = 3,24, p < 0,05, h2p = 0,32$). Plus précisément, le post hoc de Newmal-Keuls a montré que dans la condition ascendante (Figure 7.8) les paliers de 5% et 10% de la fréquence préférentielle induisent un accrochage de fréquence significativement plus important que dans les autres conditions. Tandis que dans la condition descendante ce sont les paliers 0% et 5% qui induisent un accrochage de fréquence significativement plus important que dans les autres conditions.

Ces résultats révèlent qu'un effet d'entraînement se produit bien lorsque la fréquence d'oscillation de Nao est proche de la fréquence préférentielle de l'humain (SCHMIDT et RICHARDSON, 2008). Cependant, le seuil critique d'entraînement a été trouvé à +5% de la fréquence préférentielle du sujet et non à 10% comme c'est le cas dans la littérature chez les humains. Plus précisément, un seuil critique est retrouvé à 5% en modalité ascendante et à 0% en condition descendante.

L'ensemble de ces résultats montrent qu'il y a bien un effet d'entraînement non intentionnel en vision périphérique et que celui-ci reproduit en partie (en les minorant) les caractéristiques observées. Ces résultats, outre les mémoires de Master 2 des deux étudiants en lien avec ces travaux, ont donné lieu à une communication de l'équipe d'Euromov en partenariat avec nos travaux (MADIEU et al., 2017). Un article est encore en cours d'écriture avec l'équipe d'Euromov sur le sujet.

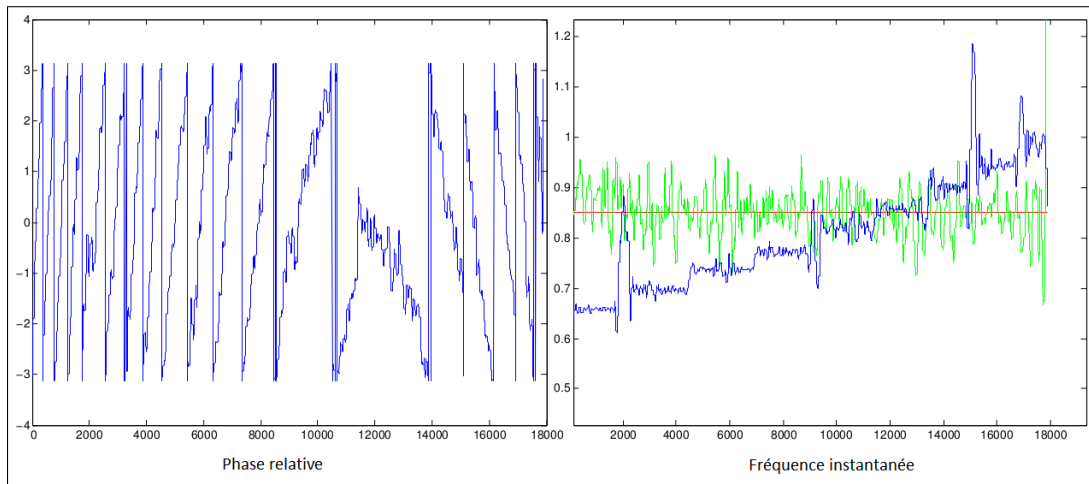


FIGURE 7.6. — *Résultat de la phase relative et de la fréquence instantanée. Durant la succession de paliers, la fréquence instantanée augmente progressivement (courbe de droite). On peut alors tracer la phase du sujet relativement à celle du robot pour identifier les phénomènes d'accrochage et de décrochage. Ici, on remarque bien la synchronisation quand on se rapproche de la zone où les fréquences du sujet et du robot sont proches.*

7.3.3. Les réseaux de neurones oscillants, un modèle pour la synchronisation en robotique

Dans l'expérience précédente nous avons précisé le phénomène du côté de l'humain, mais sans en proposer de modèle ni en l'implémentant sur le robot. Principalement car le protocole auquel on a abouti pour cette série d'expériences ne nécessitait pas dans un premier temps que l'interaction soit bidirectionnelle.

Néanmoins comme nous en avons discuté dans la section 7.3 même si les objectifs de recherche et les contextes sont assez diversifiés dans le cadre de l'étude de la synchronisation en robotique, il existe des similitudes quant aux outils et modèles employés. Notamment : la plupart de ces travaux font l'usage de modèles oscillants comme substrat des phénomènes de synchronisation.

Ils se basent notamment sur une analogie entre le modèle HKB, qui postule que l'Humain peut être vu comme un oscillateur en interaction avec l'extérieur, et les réseaux de neurones oscillants qui sont proches mathématiquement du modèle HKB et très présent dans le cerveau, ce qui en fait des candidats idéals pour modéliser la synchronisation chez l'humain.

Afin de donner un exemple du fonctionnement de ces modèles et des raisons qui ont amené à les employer de façon assez générale dans de nombreux travaux, nous détaillerons dans cette section, en tant qu'exemple, un modèle d'oscillateur neuronal qui fut employé dans d'autres travaux proches des questionnements de cette thèse, que ce soit des travaux antérieurs à cette thèse dans la même équipe (HASNAIN et al., 2013 ; GRAND et al., 2014 ; LESUEUR-GRAND, 2017) ou des travaux dans la thèse d'Eva Ansermin (ANSERMIN et al., 2016).

Le détail de ce système nous permettra de discuter de ses propriétés et des hypothèses qui ont emmené à l'emploi de ce type de systèmes dans les sections suivantes.

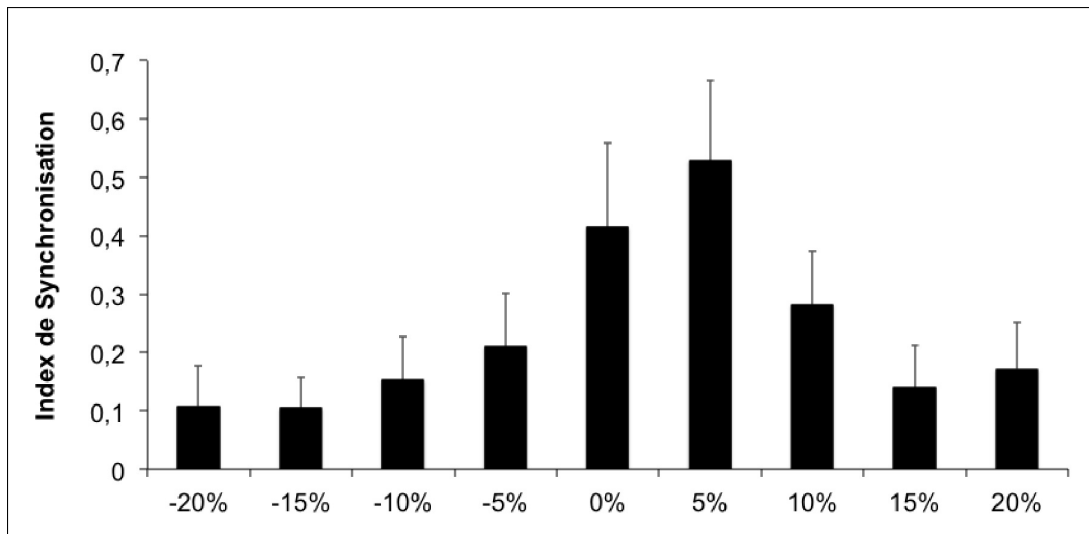


FIGURE 7.7. – Résultat du taux de synchronisation selon les fréquences. On mesure le taux de synchronisation selon les paliers de fréquences relativement à la fréquence préférentielle.

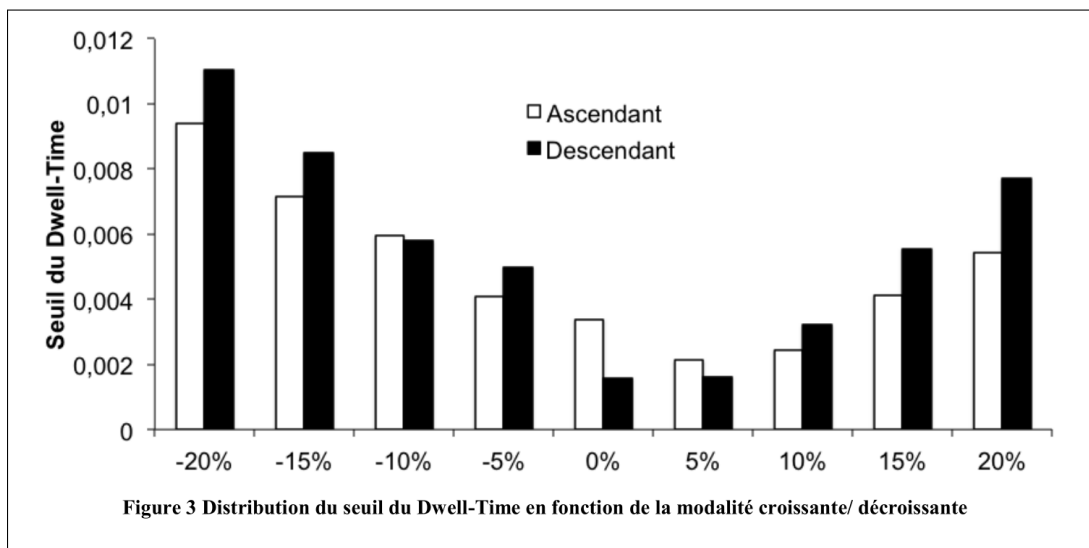


FIGURE 7.8. – Résultats pour les conditions ascendantes et descendantes.

Un exemple de modèle de réseau neuronal oscillant

Ce modèle de contrôle robotique simple a souvent été appliqué au robot Nao dans les travaux précédents de Khurshid Hasnain puis de Caroline Grand (HASNAIN, 2014 ; LESUEUR-GRAND, 2017) et développé à l'origine par Ken Prépin et Arnaud Revel (PREPIN et REVEL, 2007). Il modélise l'effecteur final du robot comme un réseau de neurones oscillants qui vient recevoir une entrée légère contenant des informations de vitesse depuis l'espace visuel du robot.

Le réseau est ainsi une vision plus bio-inspirée, mais également plus simple équationnellement, du modèle HKB utilisé généralement dans les travaux de coordination Homme-machine. Il est également plus simple que le modèle oscillant bio-inspiré de Rowat et Silverston, classiquement utilisé dans ce type de contexte (ROWAT et SELVERSTON, 1997).

L'oscillateur est composé de deux neurones formels ($N1$ et $N2$) simples qui vont tenter de s'inhiber l'un l'autre et que l'on décrit sur le schéma 7.9 et en équation par :

$$\begin{aligned} N1(n+1) &= N1(n) - \beta \cdot N2(n) + \alpha_1 \\ N2(n+1) &= N1(n) - \beta \cdot N2(n) + \alpha_2 \end{aligned} \quad (7.1)$$

Contrairement à d'autres modèles, notons que la simplicité se paye ici par la nécessité de recourir à une excitation externe constante matérialisée par α_1 et α_2 . Une fois mis en place ce modèle va osciller à une fréquence propre qui est fonction du choix de α et β ainsi que du temps séparant deux itérations.

Le réseau est ensuite déployé sur le robot Nao. Pour ce faire, certains filtrages sont nécessaires au niveau des entrées et sorties. Plus précisément, on emploie pour entrée visuelle un filtrage de la vision du robot afin d'en extraire un flot optique (HORN et SCHUNCK, 1981 ; LUCAS, KANADE et al., 1981). L'intégration spatiale de ce flot optique construira ensuite le signal perturbateur du système oscillant.

L'équation de l'oscillateur se modifiant donc en :

$$\begin{aligned} N1(n+1) &= N1(n) - \beta \cdot N2(n) + \alpha_1 + \epsilon \cdot \sum_{i=0}^{i=\max} Fl(i) \\ N2(n+1) &= N1(n) - \beta \cdot N2(n) + \alpha_2 \end{aligned} \quad (7.2)$$

Ensuite, le signal provenant de l'oscillateur est réinjecté sur l'axe à contrôler du robot Nao en l'additionnant à une valeur centrale et en multipliant le signal par une valeur ad hoc qui vont permettre à l'ensemble d'osciller harmonieusement, sans heurts, dans l'intervalle de travail de l'axe considéré ; ceci afin de faciliter l'interaction avec l'humain dans l'expérience. Le contrôle est ici purement positionnel et géré par une équation de type PID classique.

Le réseau complet peut être visualisé sur la figure 7.10.

Le modèle est ensuite déployé pour contrôler le robot et réagir face à l'extérieur dans un motif d'interaction avec l'extérieur proche de ce qui fut également proposé par Phillips (PHILLIPS-SILVER et al., 2010) et que l'on détaille sur la figure 7.2. Le robot est ainsi positionné dans l'espace visuel de l'humain et regarde l'humain, chacun pouvant donc s'influencer l'un l'autre.

Ce réseau très simple a été utilisé avec succès pour modéliser les phénomènes de synchronisation dans divers travaux précédents, il fut utilisé pour étudier la synchronisation Homme-robot dans divers cas. Notamment comme signal utilisé pour renforcer l'apprentissage dans certains cas (LESUEUR-GRAND, 2017), ou comme base pour l'étude des phénomènes d'oscillation en partenariat avec un humain (HASNAIN, 2014 ; LESUEUR-GRAND, 2017). Il fut ensuite utilisé pour former des réseaux d'oscillateurs capables de servir de base de décomposition temporelle à l'apprentissage de mouvement (ANSERMIN et al., 2016).

De façon générale, on peut donc dire que ce modèle est un modèle efficace, en première approche, d'introduction des phénomènes de synchronisation chez un robot.

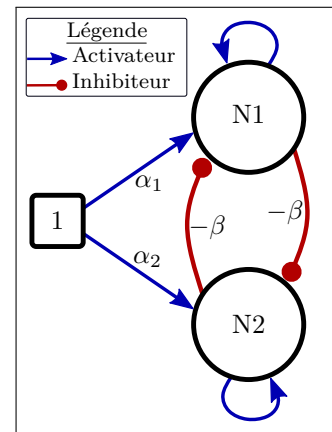


FIGURE 7.9. – Modèle oscillant simple.

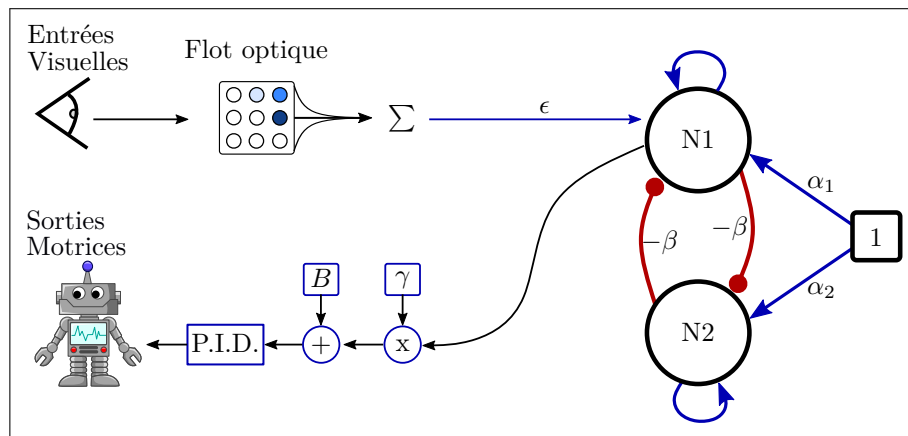


FIGURE 7.10. – Réseau de contrôle robotique pour la synchronisation immédiate

Néanmoins, les hypothèses qui ont conduit à l'utilisation de ce type de modèles, et des réseaux de neurones oscillants en général, pour représenter les effets de synchronisation, sont-elles solides ?

7.3.4. Discussion sur la synchronisation en robotique

Les effets de synchronisation entre humains sont omniprésents, et parmi ceux-ci les effets de synchronisation non-intentionnels ont soulevé de forts intérêts de recherche en psychologie par les apports potentiels qu'ils pourraient avoir vis-à-vis de l'interaction. Ce qui a amené les roboticiens à s'y intéresser.

Il est aujourd'hui de plus en plus établi que ces effets sont importants pour l'interaction humaine, par exemple, dans des protocoles visant à supprimer ces effets de synchronisation, le résultat peut alors être très négatif pour l'interaction entre humains, notamment entre la mère et l'enfant (NADEL et al., 2005a ; NADEL et al., 2005b).

Pourtant, on ne sait pas encore précisément quelle est l'utilité de la synchronisation, et particulièrement de la synchronisation non-intentionnelle chez l'humain. Si l'impact des perturbations de l'environnement en général, et de l'effet de synchronisation, en particulier, sur le mouvement humain est aujourd'hui de plus en plus identifié et caractérisé, son utilité pour l'humain reste donc encore sujette à débat.

Est-ce que ce sont des effets parasites principalement présents pour des raisons de construction du système nerveux humain, mais n'ayant aucune utilité spécifique ? Ou au contraire peut-on identifier une utilité précise à certains de ces effets ?

Certaines hypothèses semblent ainsi assez solides pour justifier de l'utilité de ces effets, notons par exemple notamment l'hypothèse de la vitesse de marche synchronisée, implicitement et sans effort supplémentaire, lors de la synchronisation du rythme de la marche chez l'humain.

Néanmoins pour le moment ces hypothèses n'ont pas donné lieu à des démonstrations. Si on reprend l'exemple de la marche, il reste ainsi à démontrer que la synchronisation est bel et bien le facteur principal de régulation de la vitesse de marche et en quoi le gain potentiel (éviter de gérer précisément son asservissement en vitesse dans une marche de groupe) est si important que cela. L'effet pourrait être un effet de bord, dont l'utilité serait alors discutable.

Le problème semble le même pour les synchronisations appliquées aux mouvements. Certaines hypothèses supposent que le phénomène pourrait ainsi induire des réponses « réflexes » d'amplitudes et de composantes fréquentielles adéquates face à un mouvement extérieur et donc aider à l'interaction de l'humain avec l'environnement. Néanmoins, il convient de noter que ce type de réaction peut tout à fait se passer de composantes dynamiques internes ou d'effets de synchronisation. Le modèle détaillé dans la partie II est ainsi capable de générer des mouvements dont la dynamique est en accord avec celle du

partenaire sans qu'il ne soit nécessaire d'implémenter de systèmes ayant une dynamique interne propre ou des réseaux oscillants et sans mettre en jeu d'effets de synchronisation.

D'autres hypothèses nous paraissent également solides pour justifier de l'intérêt de la synchronisation pour l'humain : par exemple, si on se synchronise implicitement sur l'extérieur alors le phénomène extérieur devient parfaitement prédictible, et cela de façon « gratuite ». En effet, comme on a accès à nos propres sensations et actions, il devient aisé d'utiliser celles-ci pour prédire un phénomène extérieur qui a exactement la même dynamique.

Cette facilitation de la prédiction pourrait ainsi servir à *inhiber l'espace connu*. Prenons un exemple : si on marche côte à côte avec quelqu'un, la perturbation induite par les jambes de la personne constamment mouvantes dans le champ visuel pourrait entraîner une perturbation de l'attention. Pourtant on est capable d'inhiber les perceptions de notre propre corps durant une action, et par exemple, d'éviter de prêter attention à nos propres jambes durant la marche. Si nos jambes sont parfaitement synchronisées avec celles du partenaire, il devient alors très aisé de faire d'une pierre deux coups et d'inhiber également celles du partenaire. De façon schématique, le partenaire en synchronie devient une partie de nous-même, puisqu'il agit de la même façon, et de fait on peut inhiber ou se concentrer efficacement sur les parties de cet autre « soi » pour réaliser une tâche.

Cependant, cette hypothèse, comme les autres, n'est pas appuyée par des expériences réelles. Sans doute parce que la validation de ce type d'hypothèses nécessiterait un ensemble robotique extrêmement complet, doté d'une capacité d'action, d'un processus attentionnel neurobiologiquement plausible, de processus de synchronisation implicite avec l'extérieur et d'un système neurobiologiquement plausible complet d'inhibition du connu.

De façon générale, au fur et à mesure de la réalisation de cette thèse nous nous sommes rendu compte que la plupart des hypothèses sur l'utilité de la synchronisation chez l'humain, particulièrement dans le cadre de la synchronisation non-intentionnelle, et qu'elles soient crédibles ou non, n'ont pas donné lieu à des démonstrations. Ceci nous a amenés à nous interroger sur le socle sur lequel les expériences de synchronisation en robotique sont fondées tant d'un point de vue du principe que des outils et analogies utilisés.

Les outils introduits pour la synchronisation en robotique sont-ils superflus ?

Lorsque le concept de la synchronisation non-intentionnelle et de la synchronisation de phénomènes moteurs avec un signal extérieur en général a commencé à s'étendre au champ de la robotique et de l'intelligence artificielle, il est devenu commun de formuler implicitement l'hypothèse que les réseaux de neurones oscillants seraient un substrat adéquat à l'expression des phénomènes de synchronisation.

Non seulement ces réseaux de neurones spécifiques sont de plus en plus étudiés, notamment dans le domaine de l'étude des Central Pattern Generator (CPG), mais en plus leurs propriétés permettent une analogie immédiate avec le modèle HKB traditionnellement utilisé pour modéliser les propriétés de la synchronisation et des coordinations interpersonnelles. Comme nous en avons discuté dans la section 7.3.3. De plus, certains travaux ont démontré un lien observable entre synchronisation comportementale et synchronisation de zone du cerveau avec un partenaire (PFEUTY et al., 2003) ce qui a tracé un lien entre synchronisation motrice et réseaux présent dans le cerveau.

De fait, les réseaux de neurones oscillants sont devenus assez tôt les candidats idéaux à la modélisation des phénomènes de synchronisation moteurs en robotique.

Ainsi, plusieurs auteurs travaillant sur les phénomènes de synchronisation moteurs ont proposé un lien explicite entre les phénomènes de synchronisation et les CPG (MIYAKE, 2009 ; JOUAITI et HENAFF, 2018). D'autres auteurs présentent les CPG dans leurs travaux, sans établir de lien explicite, mais en utilisant toutefois des oscillateurs neuronaux afin de modéliser des phénomènes de synchronisation dans le contrôle moteur robotique (DEGALLIER et al., 2006 ; HASNAIN, 2014 ; LESUEUR-GRAND, 2017 ; ANSERMIN et al., 2016).

Les auteurs mêlant l'étude des phénomènes de synchronisation et la robotique ont ainsi naturellement convergé vers l'utilisation de modèles neuronaux oscillants.

Mais l'analogie que ces modèles présentent naturellement avec le système HKB n'a-t-elle pas orienté implicitement les travaux menés dans ces domaines pour modéliser les phénomènes de synchronisation

moteurs chez l'humain, sans que cela ne soit réellement justifié ?

C'est la question que nous nous sommes posée à la lecture de la littérature attachée au concept d'intelligence du corps et plus particulièrement lors de la revue du domaine d'étude des marcheurs passifs. On peut noter par exemple que des travaux récents ont démontré que la marche ne nécessite pas autre chose qu'un corps bien construit additionné de quelques réflexes élémentaires (GEYER et HERR, 2010). De même, les marcheurs modernes sont construits sur ce type d'hypothèses (MCGEER et al., 1990; RAIBERT et al., 2008). Néanmoins, c'est plus particulièrement la lecture de l'étude de Zhenyu Gan et collègue qui a été la plus marquante pour nous. Cette étude (GAN et al., 2016) utilise les hypothèses émises dans le domaine des marcheurs passifs pour modéliser le squelette des chevaux à sang chaud et des chevaux islandais. Ils arrivent alors à expliquer les phénomènes, non seulement de marche, mais également le passage au trot et au galop, simplement en considérant les chevaux comme des marcheurs passifs d'une certaine constitution. Les différents rythmes étant alors des bassins d'attraction à différentes vitesses du cheval passif.

Cette étude démontre que ce qui auparavant était un rôle attribué strictement au système nerveux (notamment à des bancs d'oscillateurs neuronaux), tel que le passage d'un rythme de marche à l'autre, peut en fait n'être qu'une propriété du corps sans aucune intervention d'un système dynamique interne. En tout cas dans le cas des chevaux.

Ceci nous a amenés à nous demander si les réseaux neuronaux oscillants pouvaient de même être inutiles à la synchronisation, ou du moins, ne pas être la source principale du phénomène, de la même manière qu'ils ne le sont pas dans le cas de la génération de différents rythmes de course/marche chez le cheval. La synchronisation dans les modèles de type réseaux oscillants est en effet une propriété dynamique d'un système qui vise avant tout à générer un signal rythmique. Si ce signal rythmique peut être le fruit du corps en interaction avec l'environnement, peut-il en être de même avec la synchronisation ?

Cette hypothèse peut alors amener à une réinterprétation des résultats expérimentaux du domaine : on pourrait ainsi supposer que la synchronisation entre zones du cerveau de deux partenaires dont les comportements se synchronisent (PFEUTY et al., 2003) ne soit pas la cause de la synchronisation comportementale, mais la conséquence de processus de synchronisation mettant en jeu les corps des partenaires plutôt que des processus neuronaux ayant une dynamique interne. Vu autrement : le modèle HKB qui modélise l'humain comme un oscillateur peut être interprété comme une modélisation, en réalité, de son corps et de son système nerveux combinés, en interaction avec l'environnement. En partant de l'analogie du modèle HKB avec les réseaux de neurones oscillants présent dans le cerveau, n'a-t-on alors pas négligé que le corps pourrait en réalité être la source principale de la synchronisation ?

Démonstration sur le robot humanoïde Tino

Dans cette section, nous nous baserons sur les hypothèses de l'intelligence du corps dont nous avons discuté précédemment. L'idée est de déterminer s'il est possible d'obtenir les mêmes résultats de synchronisation que ceux qui furent observés dans les travaux utilisant des réseaux de neurones oscillants, notamment dans les travaux portant sur la synchronisation non-intentionnelle. C'est-à-dire une synchronisation d'un phénomène extérieur avec le corps quand les deux sont proches en fréquence ainsi que des phénomènes d'accrochage et décrochage de phase.

Nous n'emploierons pas ici de réseaux de neurones oscillants, mais resterons sur un système rudimentaire de réflexe en interaction avec le corps robotique.

Le modèle que nous allons présenter dans cette section est implémenté sur le robot Tino qui fut présenté en détail dans la partie I et qui servira ici de plateforme de démonstration. Comme pour les expériences avec des réseaux de neurones oscillants nous nous positionnerons dans le même type d'hypothèses : on suppose un signal représentant le système extérieur et qui vient « fuiter » jusqu'aux effecteurs du robot, mais avec une intensité faible; par simplicité on restreint l'expérience à un unique degré de liberté et on utilise un système de contrôle classique de type PID.

Pour injecter de l'énergie dans le corps en interaction avec l'environnement avec d'aboutir à un mouvement on utilise ici un réflexe qui vient réinjecter de l'énergie dans le muscle lorsque le bras atteint sa fin de course. Nous faisons une analogie avec les réflexes de protection des muscles et tendons qui viennent soustraire de l'énergie au muscle en cours de contraction, et en ajouter au muscle antagoniste

pour stopper puis inverser le mouvement quand le bras arrive en butée articulaire (réflexe d'extension et flexion). Ici on pousse le concept à l'extrême pour que le système soit autonome en injectant suffisamment d'énergie pour qu'il atteigne la butée opposée.

On simplifie le système pour la démonstration avec un réflexe parfaitement binaire que l'on adapte au contrôle en position. Le réflexe forme alors une équation d'hystérésis classique :

$$\begin{aligned} \text{si } P_s >= \text{LimiteButeeSup} \text{ alors} \\ \quad | \quad S_m = \text{Pos}(\text{min}) \\ \text{si } P_s <= \text{LimiteButeeInf} \text{ alors} \\ \quad | \quad S_m = \text{Pos}(\text{max}) \end{aligned} \quad (7.3)$$

Où P_s est la position en radian, S_m l'ordre donné au bras du robot, $\text{Pos}(\text{min})$ et $\text{Pos}(\text{max})$ sont arbitraires et au-delà des butées. Notons que le fait que le système soit contrôlé en position ou en commande brute n'a aucune influence sur le résultat qualitatif décrit ci-après. L'objectif est ici de construire un système de réflexe simple suffisant pour donner un comportement analogue à certains phénomènes rythmiques comme la marche et qui n'est dépendant que du corps (ici de la capacité du membre à atteindre en un certain temps sa butée) et d'un réflexe purement musculaire comme ce qu'on peut trouver dans les études autour des marcheurs passifs ou dynamiques citées précédemment.

La nature d'hystérésis du réflexe est justifiable, car certains senseurs chez l'humain dans les bras et jambes répondent mieux quand ceux-ci sont contractés ou en position au contraire totalement tendue. Par exemple, les organes tendineux de Golgi, qui sont sensibles aux positions extremum des membres sont parmi les éléments clés permettant le basculement entre les réflexes d'extension et de flexion (CONWAY et al., 1987).

Une fois le réflexe implémenté on transmet au système un signal qui va servir de base sur laquelle se synchroniser et qui vient s'ajouter directement sur l'effecteur avec un couplage faible représentant la « fuite » d'information de la perception vers l'action. (voir figure 7.11)

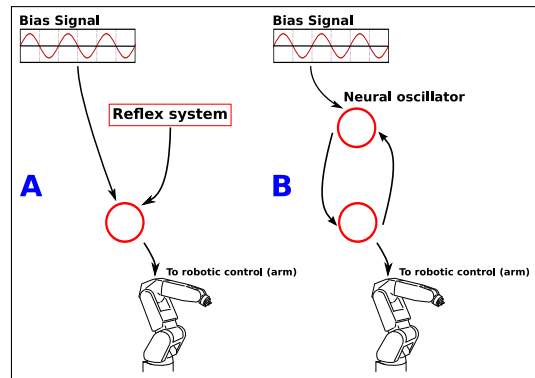


FIGURE 7.11. – Comparaison entre le système classique de synchronie par réseaux oscillants et le système employant des réflexes en lien avec le corps robotique

Résultats Les résultats ont permis de mettre en évidence les mêmes phénomènes d'accrochage et décrochage en phase que l'on observe habituellement dans les modèles employant des réseaux de neurones oscillants ou avec les modèles d'origine de type HKB.

On montre sur la figure 7.12 le détail du phénomène selon que le système soit connecté au signal extérieur « fuyant » ou non. On observe bien l'accrochage de phase uniquement dans le cas où le signal extérieur fuit vers l'effecteur du robot.

On montre ensuite sur la figure 7.13 l'évolution de la vitesse de phase (vitesse du décalage de phase entre le mouvement du bras et le signal perturbateur « fuite ») en fonction de la fréquence relative du signal perturbateur (1,0 = fréquence du signal identique à fréquence propre du système bras+réflexe). On montre sur cette figure que le phénomène de synchronisation apparaît pour des fréquences proches du mouvement des bras puis disparaît comme ce qui peut être observé chez l'humain. Le seuil d'accrochage décrochage correspond ici approximativement à $\pm 10\%$ ce qui est exactement dans la gamme de valeur qu'on trouve classiquement dans la littérature pour l'humain.

On en conclut qu'il est donc possible, avec un système de réflexes rudimentaires inspiré des réflexes d'extension et de flexion, en interaction avec un corps robotique, de retrouver les mêmes effets de synchronisations sur un signal extérieur que ceux que l'on obtient dans les modèles employant des CPG ou dans le modèle HKB, ou encore dans l'étude de la synchronisation non-intentionnelle chez l'humain.

Analyse Au premier abord, ces résultats peuvent paraître peu intéressants et peu surprenants. L'extension d'un système de réflexe pour aboutir à un système ayant une dynamique propre peut en effet

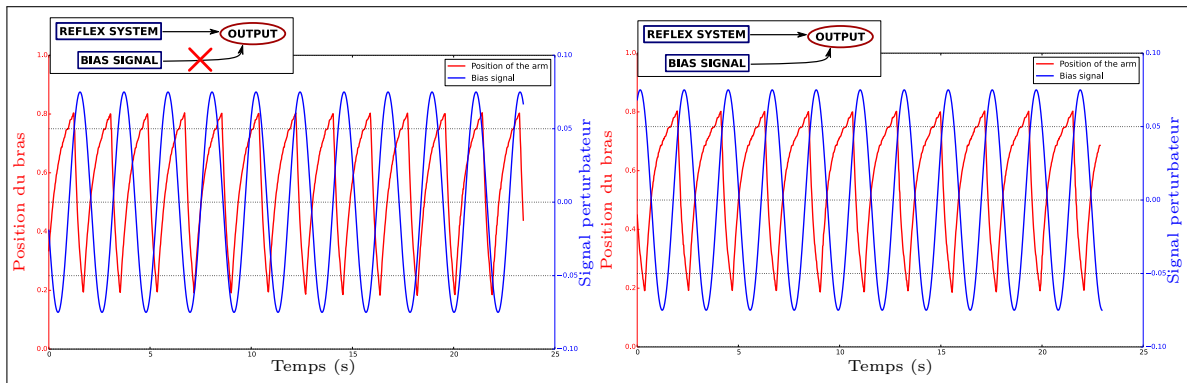


FIGURE 7.12. – Lorsque le système de signal perturbateur est déconnecté de l'effecteur (à gauche), il n'y a aucune synchronisation. Quand il est connecté (à droite), une synchronisation apparaît. Le facteur de couplage est ici faible ($\leq 5\%$)

paraître exagérée. L'implémentation du système via un système de contrôle en position peut également sembler peu concluante pour faire un parallèle avec la biologie. Et pour finir, le système étant oscillant on s'attend à ce que sa mise en équation suggère un parallèle avec le modèle HKB ce qui fait que le résultat n'est pas surprenant.

Pourtant si on s'y attarde, certains de ces points ne sont pas vraiment problématiques. Le système de réflexe par exemple est certes exagéré, mais semble cadrer avec ce que la littérature autour des marcheurs passifs, dynamiques et de l'étude de la marche suggère. Notamment, que la marche ou la course peuvent être le fruit de réflexes d'extension et de flexion mettant en jeu des connexions passant par peu de relais neuronaux. De même, les études sur les organes tendineux de Golgi montrent leur implication dans le basculement entre le réflexe de flexion et d'extension.

Par ailleurs, même si ces réflexes ne forment pas un système à la dynamique parfaitement autonome sur d'autres membres que les jambes et en toute situation, ils ont tout de même un impact sur la dynamique et la rythmique durant les actions et de fait pourront être un facteur rythmique important dans un mouvement du bras volontaire. Les effets de synchronisation que nous démontrons ici pourraient donc avoir lieu même dans un cas où ces réflexes ne sont pas autonomes. Il reste néanmoins à en faire une démonstration.

Sur les aspects de contrôle en position, comme nous l'avons vu dans la partie traitant du contrôle musculaire et du contrôle en co-contraction, la simplification d'un contrôle musculaire proche d'un contrôle asservi en position n'est pas forcément si exagérée, le système en co-contraction aboutissant naturellement à une position donnée pour un couple de forces données, et chaque muscle disposant d'un asservissement interne sur son élongation. De plus, nous avons testé le réseau ci-dessus sur un contrôle en force classique (impulsion directement sur les commandes du robot avec suffisamment d'élan pour arriver en butée) et les résultats qualitatifs furent identiques.

Enfin, même si ces résultats sont peu surprenants et attendus, leur signification sur le fond reste selon nous importante. Dans les travaux classiques autour des réseaux neuronaux oscillants et de la synchronisation l'élément qui se synchronise est complètement déconnecté du corps contrôlé et la question de la fréquence propre associée à un mouvement demeure non résolue. Pourquoi un membre contrôlé par des réseaux oscillants aurait-il une certaine fréquence propre et pas une autre ? Pourquoi un ensemble de réseaux oscillants à fréquences différentes ne se synchroniserait-il pas dans tous les cas sans fréquence propre particulière, sans effets d'accrochage et de décrochage en phase notable vu que dans tous les cas une partie au moins des oscillateurs rentre en résonance ?

Face à ces questions, les expérimentateurs se contentent en général de concevoir un réseau ad hoc oscillant à une fréquence « convenable » pour le cas d'application ciblé.

Le système présenté ici, même s'il semble analogue, porte une signification très différente et apporte un début d'éclairage à ces questions. En effet, sa fréquence propre est implicitement liée aux capacités

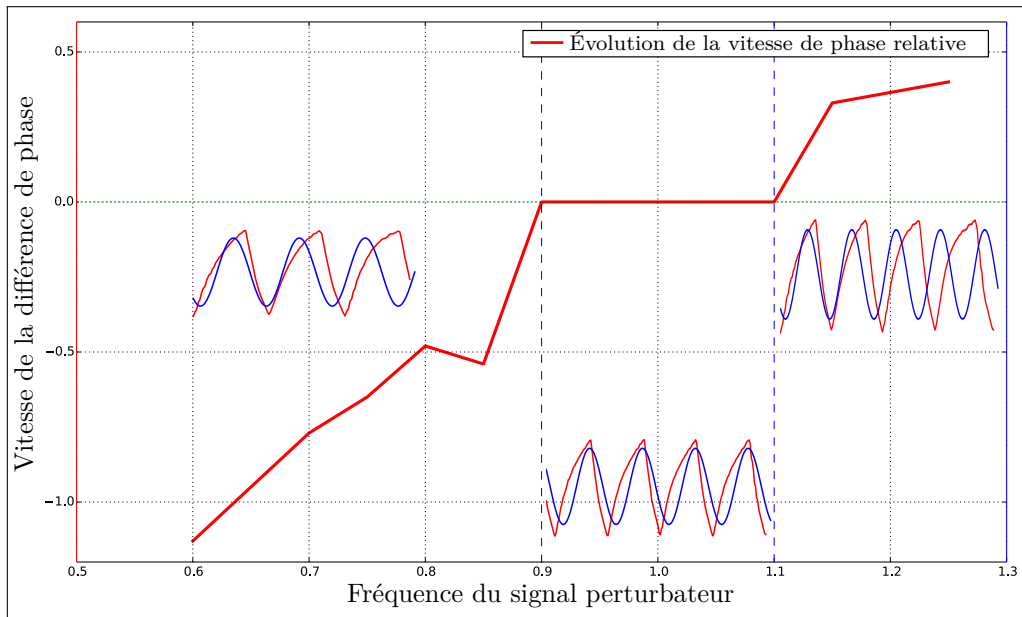


FIGURE 7.13. — *Évolution de la vitesse de déplacement de la différence de phase entre le signal perturbateur et le mouvement induit par les réflexes du bras du robot Tino, ici représentée en fonction de la fréquence relative du signal perturbateur par rapport à la fréquence propre du bras de Tino. On observe bien la plage d'accrochage en phase autour d'un point où la fréquence du signal perturbateur est égale à celle du bras du robot Tino*

du robot et à la forme de son corps ainsi qu'à la nature du réflexe. La forme du corps va ici jouer sur le temps d'atteinte des butées et donc sur la longueur d'onde du rythme propre du mouvement. La nature du réflexe de flexion/extension va directement jouer sur la quantité d'énergie réinjectée et donc sur l'importance du facteur rythmique dû au réflexe dans le mouvement.

L'ensemble permet de tracer un lien explicite entre la fréquence propre d'un mouvement, les bornes d'accrochage et de décrochage de phase et la géométrie du membre concernée. De même, l'ensemble pointe le fait que l'étude des réflexes musculaires peut être une direction de recherche à privilégier, face aux réseaux de neurones oscillants, pour l'étude de la synchronisation non-intentionnelle entre humains. On peut d'ailleurs se hasarder à quelques prédictions. Si l'approche proposée ici est justifiée, on devrait détecter des fréquences propres différentes sur des personnes ayant des morphologies différentes. Dans le cas des balancements des jambes, la taille du sujet pourrait par exemple être corrélée avec la fréquence propre du balancement. Un point qu'il faudrait sans doute vérifier dans les expérimentations à venir autour de la synchronisation.

7.4 | Conclusion. L'étude des réseaux oscillants, un intérêt pour la robotique interactive ?

Comme nous l'avons vu, la synchronisation motrice avec un stimulus extérieur peut s'expliquer sans recourir à des réseaux de neurones ayant une dynamique interne de type réseaux oscillants. Cependant, cela ne veut pas forcément signifier que les réseaux oscillants sont inintéressants dans notre cadre. La conclusion précédente signifie simplement que le lien qui fut tracé implicitement entre réseaux de neurones oscillants et effets observables de synchronisation non-intentionnelle est basée sur des hypothèses qui nous apparaissent ténues et que le phénomène peut s'expliquer de façon plus satisfaisante et plus simple par une autre approche reposant sur les propriétés du corps. Notons néanmoins que les oscillateurs neuronaux pourraient tout à fait se combiner avec notre hypothèse sur la synchronisation motrice en formant des réseaux qui viendraient s'adapter aux systèmes réflexes et au corps.

Toutefois, si l'on s'intéresse spécifiquement aux réseaux oscillants en dehors du cadre de ces effets de synchronisation non-intentionnelle alors il apparaît que leur importance dans les coordinations interpersonnelles ne peut pas être négligée et qu'ils présentent un intérêt pour l'interaction même en écartant les effets de synchronisation non intentionnelle.

En effet, si l'on reprend l'hypothèse de la partie II selon laquelle on peut construire par des liens entre perceptions et actions un pavage moteur qui peut ensuite être placé en situation de résonance motrice alors l'intérêt des systèmes oscillants devient important pour compléter avec une dynamique interne les phénomènes de résonance moteurs.

On ne peut pas nier, par exemple, l'impact des oscillateurs sur le système moteur bas niveau. Même s'ils peuvent ne pas être liés, directement ou seuls, aux phénomènes de synchronisations moteurs, ils restent néanmoins très probablement responsables de la forme de certains pans de trajectoire, de l'exécution de trajectoires complexes dans certains cadres et de routines automatiques ; autant de phénomènes qui peuvent avoir un impact fort sur le contrôle robotique notamment dans un contexte d'interaction et de lecture de l'intention de l'autre.

En effet, nous sommes parti du postulat dans cette thèse qu'avoir un corps et des dynamiques proches de celle de l'humain (voir la partie I) et des trajectoires motrices répondant aux mêmes contraintes (voir la partie III) permettrait d'amplifier les effets de résonance entre l'humain et le robot ce qui pourrait alors aider à décharger l'humain de l'effort de prédire des trajectoires et des comportements très différents du sien, aidant ainsi à l'interaction.

En suivant cette hypothèse : les réseaux de neurones artificiels oscillants pourraient alors aider à rapprocher nos architectures de contrôle robotique du fonctionnement des zones concernées du système nerveux humain, et ainsi permettre de rapprocher le comportement observé d'un robot de celui d'un humain. Ce qui pourrait aider, en définitive, à l'interaction Homme-robot.

Dans la section suivante, nous aborderons donc la question de l'intégration de systèmes oscillants à notre cadre robotique en respectant les principes énoncés dans l'introduction de la thèse : à savoir des architectures que nous concevons dans le but d'être souples et développementales tout en étant capables d'interagir avec l'extérieur dans un environnement réel pouvant être occupé par des partenaires effectuant d'autres actions.

Apprentissage dans les réseaux de neurones oscillants appliqués à la robotique

Sommaire du chapitre :

8.1. Introduction	196
8.2. L'application d'oscillateurs neuronaux dans un cadre robotique : emplacement et contraintes	197
8.2.1. Problématique de l'emplacement	197
8.2.2. Problématique du lien entre unités inhibitrices et activatrices	197
8.2.3. Le problème pragmatique de la création de multiples couples oscillants avec les modèles classiques	198
8.3. Apprentissage et adaptation dans les réseaux oscillants	199
8.4. Différentes propositions de modèles	200
8.4.1. Perspectives pour le modèle, quelles applications?	212
8.5. Conclusion	216

8.1 | Introduction

L'étude des réseaux de neurones oscillants constitue un large domaine scientifique très varié qui va de la locomotion chez divers animaux (la lamproie par exemple (GRILLNER et al., 1991)) et chez l'homme (MARDER et BUCHER, 2001), au vol des insectes (STEVENSON et KUTSCH, 1987), en passant par l'étude des phénomènes gastriques (ROWAT et SELVERSTON, 1993). Leur implication dans les processus de maintien de la respiration ou des battements de cœurs sont également étudiés (VAN DER POL et VAN DER MARK, 1928).

Beaucoup de travaux se sont focalisés sur l'étude des structures oscillantes dans le système moteur bas niveau (KIEHN et BUTT, 2003) ou dans le cervelet qui en est proche. D'autres se focalisent sur les effets oscillants qui peuvent émerger du fait de la position et de l'éloignement de certaines structures dans le cerveau, notamment dans les liens du thalamus avec le reste du cerveau (LE MASSON et al., 2002). D'autres branches encore étudient de façon plus locale l'effet que peuvent avoir les réseaux oscillants dans les systèmes de traitement sensoriel, notamment visuel (ECKHORN et al., 1988; GROSSBERG et SOMERS, 1991; ECKHORN, 1994) ou encore dans les ganglions de la base (BRITAIN et BROWN, 2014). Beaucoup de travaux du domaine traitent également des liens entre différentes zones du cortex vis-à-vis des oscillations. Pour finir, notons qu'il existe de forts liens entre l'étude des réseaux oscillants et l'étude des réseaux de neurones chaotiques (WANG, 1996).

En robotique et dans le domaine de l'intelligence artificielle en général cela a donné lieu à l'emploi de réseaux oscillants dans de nombreuses applications, notamment liées à la locomotion (IJSPEERT et al., 2007; SEO et SLOTINE, 2007; LIU et al., 2009; BARRON-ZAMBRANO et al., 2010; YU et al., 2013; NASSOUR et al., 2014) (pour une revue du domaine, voir (IJSPEERT, 2008)) ou au traitement visuel et sensoriel en général (VELICHKO et al., 2019).

Les types d'oscillateurs employés sont assez divers; nous pouvons citer les plus communs. Dans une première catégorie on peut noter l'existence d'oscillateurs mono-neuronaux : certains se basent par exemple sur des oscillateurs de Van der Pol (CARTWBIGHT, 1960) ou de Stein ou encore sur des modèles dits « *Leaky integrate and fire* » (LIF). Des modèles plus complexes existent également, modélisant des oscillations inter-compartiments au sein d'un neurone. Ce ne sont pas ces types d'oscillateurs qui seront développés dans cette thèse puisque nous nous intéressons plutôt aux phénomènes d'oscillation émergents de l'influence d'une structure sur une autre.

Dans une deuxième catégorie, il existe des modèles mettant en jeu plusieurs neurones, au minimum deux, l'un activateur et l'autre inhibiteur, mais pas seulement : Le modèle présenté précédemment (section 7.3.3) développé par Ken Prépin et Arnaud Revel (PREPIN et REVEL, 2007) est par exemple basé sur un couple de deux neurones tentant de s'inhiber l'un l'autre. Parmi les modèles historiques utilisés notons le modèle d'oscillateur bi-neuronaux de Deliang Wang (WANG, 1995; WANG et TERMAN, 1995) qui est une amélioration des modèles précédents autour des travaux de Wang (WANG et al., 1990; MALSBERG et BUHMANN, 1992). Notons également le modèle de Rowat et Silverston (ROWAT et SELVERSTON, 1997).

Ces oscillateurs bi-neuronaux nous paraissent être les plus intéressants à intégrer dans nos modèles du fait de la très large présence des couples exciteur-inhibiteur dans le cerveau, notamment dans le système moteur bas niveau. Mais quelles sont les contraintes pour l'usage de ces types d'oscillateurs bi-neuronaux dans notre contexte de travail? Nous aborderons cette question dans ce chapitre et y proposerons des solutions via l'étude de l'apprentissage dans les réseaux oscillants.

8.2 | L'application d'oscillateurs neuronaux dans un cadre robotique : emplacement et contraintes

8.2.1. Problématique de l'emplacement

L'emplacement des réseaux oscillants joue beaucoup sur la problématique associée. Des réseaux oscillants utilisés pour des traitements sensoriels préalables à un apprentissage sont une disposition courante dans les modèles simulés et en robotique. L'avantage de cette disposition. (partie A du schéma 8.1) est qu'elle ne pose pas le problème de l'apprentissage en amont du réseau, puisque celui-ci est employé à des fins de filtrage, les liens allant des entrées sensorielles au réseau sont généralement considérés comme fixes. C'est évidemment une vue de l'esprit. Dans la réalité l'apprentissage existe probablement à tous les emplacements du cerveau, donc également en amont des parties de traitement sensoriel. Néanmoins du point de vue du roboticien la conception d'un réseau présentant de multiples couples oscillants à des fins de filtrage ne pose pas de problèmes particuliers vis-à-vis de l'apprentissage en amont.

Il n'en est pas de même lorsque l'on souhaite étudier un positionnement de type « CPG » dans lequel un réseau de neurones oscillants sert de tampon entre la commande et les sorties motrices. Dans ce cas, non seulement le lien entre ce réseau et la sortie motrice pose question, mais en plus on peut se demander quelles sont les propriétés de l'apprentissage en amont du réseau pour former les liens entre la commande et le réseau oscillant (partie B du schéma 8.1). S'il existe des études évaluant l'apprentissage en amont d'un réseau oscillant c'est toutefois plus souvent pour analyser le lien entre plusieurs réseaux oscillants connectés (NISHII, 1997 ; NISHII, 1999) plutôt que pour analyser une boucle complète mêlant perception et action en robotique.

Finalement, le positionnement en mode « CPG », qui semble le plus complexe à mettre en œuvre, nous paraît plus riche quant à ses implications sur le mouvement.

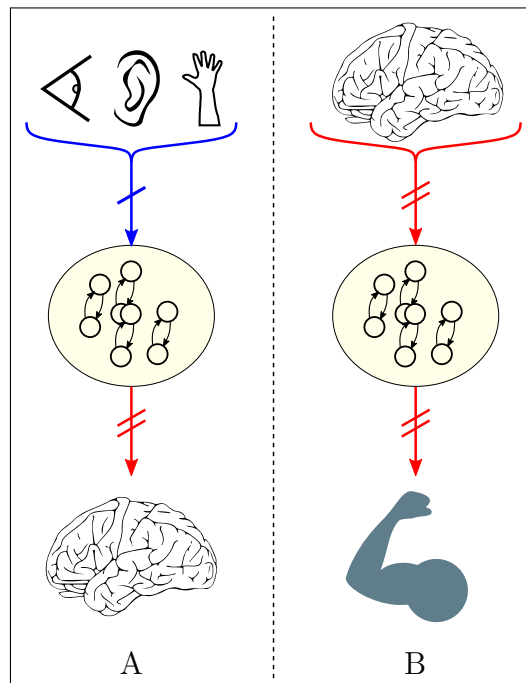


FIGURE 8.1. – *Problématique de l'emplacement pour le positionnement des CPG dans les modèles d'apprentissage. A : réseaux oscillants comme filtre sensoriel. B : positionné façon « CPG » en fin de chaîne. Lien rouge = lien variable, bleu = fixé.*

8.2.2. Problématique du lien entre unités inhibitrices et activatrices

Outre le problème du lien en amont dans un système oscillant, un second problème apparaît lorsque l'on considère des interneurons inhibiteurs servant à réguler un neurone comme cela peut être le cas sur la sortie motrice des motoneurons inférieurs (cellules de Renshaw) ou dans l'étude de l'influence des « *Basket cells* » sur les cartes topologiques du cortex visuel.

Dans un cadre développemental dans lequel le robot apprend en expérimentant dans un monde réel ces interneurons inhibiteurs vont se retrouver à inhiber des signaux en constante évolution. Il apparaît alors comme une nécessité que cette inhibition se développe de façon homogène face à l'apprentissage dans le reste du système, ceci afin de maintenir les propriétés du couple oscillant ou filtrant.

Sans cela il est complexe de définir un interneurone capable de maintenir son rôle dans un cadre

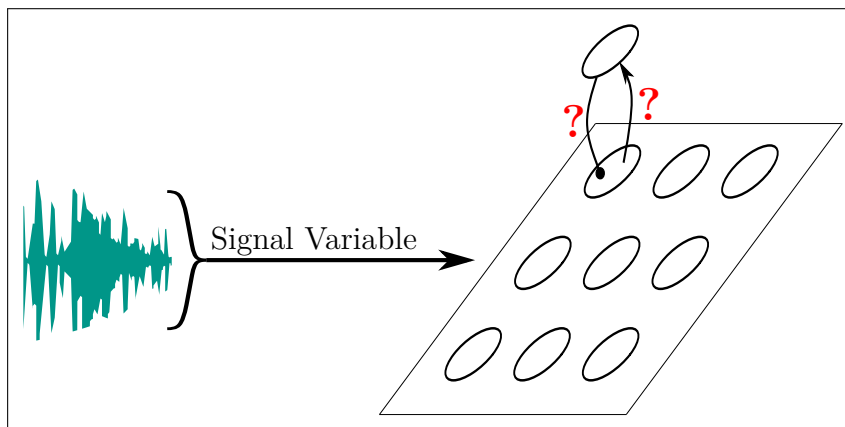


FIGURE 8.2. – Problématique de la connexion des interneurones face à un signal variable

de robotique développementale en interaction dans un environnement complexe et donc soumis à des signaux variables dont les propriétés sont en constante évolution.

8.2.3. Le problème pragmatique de la création de multiples couples oscillants avec les modèles classiques

Comme nous l'avons vu, la définition des liens entre l'activateur et l'inhibiteur se heurte à la problématique de l'évolution des signaux dans le temps pour un système développemental. Mais de façon plus générale, cela pose un problème pratique au modélisateur et au roboticien : comment constituer des ensembles oscillants qui vont conserver leur propriété d'action ou de filtrage dans un certain environnement ? Non seulement cela peut être problématique si l'environnement change ou si le système évolue par apprentissage au fur et à mesure, ce qui oblige le roboticien à ne concevoir qu'un unique ensemble de réseaux oscillants pour un ensemble d'applications. Mais cela pose également problème pour la constitution d'un large ensemble de réseaux oscillants aux propriétés différentes.

En effet, les modèles classiques de réseaux oscillants, tout comme le modèle présenté précédemment dans la section 7.3.3, ne réagissent que pour des gammes de signaux d'entrée données. Si le signal en entrée est trop faible ou trop fort, ces types de réseaux entrent généralement dans un état de sur-inhibition ou de sur-activation. Cette saturation les rendant généralement inutiles dans l'application considérée.

Les modélisateurs et roboticiens ont généralement une approche du problème au cas par cas. Chaque expérience se déroulant dans un ensemble de conditions précises, il devient alors possible de paramétrer des oscillateurs, des systèmes d'inhibition, ou encore de construire de larges bancs d'oscillateurs « à la main », de façon ad hoc, en réglant manuellement les amplitudes d'inhibition et d'activation pour que le modèle fonctionne dans les paramètres fixés.

Dans nos travaux, cette approche du problème ne nous a pas paru satisfaisante au regard des principes de travail sur lesquels nous avons basé cette thèse, notamment l'importance pour nous de concevoir des architectures capables de s'imbriquer l'une dans l'autre, pour fonctionner dans un paradigme développemental et pour être déployées ensuite sur des robots aux architectures mécaniques potentiellement différentes qui évolueront dans un environnement réel.

Cet aspect du problème nous a amenés à nous poser la question de l'apprentissage dans les réseaux oscillants.

8.3 | Apprentissage et adaptation dans les réseaux oscillants

La plupart des travaux existants autour du domaine des réseaux oscillants et qui se sont posé la question de l'apprentissage peuvent être séparés en quelques catégories.

- L'importance de l'inhibition sur l'apprentissage en amont du neurone inhibé. (D'ANGELO, 2008)
- L'apprentissage entre les entrées et l'ensemble oscillants (DOYA et YOSHIKAWA, 1991)
- L'apprentissage depuis les sorties du réseau oscillant. Cela constitue la plus grande partie du domaine et représente de très nombreux travaux que nous ne listerons pas ici.
- L'adaptation des paramètres internes aux neurones (RIGHETTI et al., 2006) ou des paramètres géants les boucles rétroactives (KUROKAWA et al., 1997), les deux aboutissant à l'adaptation des paramètres de chaque neurone indépendamment.
- L'apprentissage entre plusieurs réseaux oscillants (on parle d'oscillateurs couplés) ((NISHII, 1997) (NISHII, 1999) etc.)

On note qu'il existe également de nombreux travaux permettant d'apporter des éléments de réponse à l'apprentissage en amont d'un groupe oscillant et qui peuvent alors être exploités dans notre cadre. De même, il existe quelques travaux sur l'apprentissage dans chaque neurone indépendant présent dans le couple, ceci afin d'ajuster la fréquence du doublet. Nous remarquons par contre qu'il n'existe pas de travaux sur l'adaptation des liens excitateurs et inhibiteurs à l'intérieur d'un doublet de neurones oscillant.

Les travaux sur l'adaptation des paramètres internes au neurone considéré nous paraissent, quant à eux, intéressants, mais peu pertinents dans notre cas, car ils emploient des méthodes d'adaptation mathématiques de paramètres qui semblent peu en lien avec la neurobiologie. Par ailleurs en général ils visent un objectif précis (copier la fréquence d'entrée pour l'apprendre) qui n'est pas notre objectif qui est d'aboutir à un réseau stable en toute circonstance via l'apprentissage.

On peut alors se demander pourquoi si peu de travaux (aucun, à notre connaissance) n'existent sur l'apprentissage dans les liens inhibiteurs et activateurs d'un doublet de neurones. La raison qui semble prévaloir ici est une raison venant des connaissances historiques en neurobiologie. En effet jusqu'à récemment le consensus en neurobiologie voulait que les liens inhibiteurs n'apprennent tout simplement pas et soient donc en quelque sorte fixés génétiquement.

Ce n'est que très récemment qu'un ensemble de travaux en neurobiologie a réussi à démontrer que non seulement les liens inhibiteurs se formaient par l'apprentissage et n'étaient pas uniquement « pré-codé » par la génétique (par exemple (ALVAREZ et al., 2013) pour les cellules de Renshaw), mais qu'en plus ils se développaient de façon semblable et en suivant des règles similaires aux liens activateurs (voir (LIEN et al., 2006) et pour une revue des travaux récents : (FLORES et MÉNDEZ, 2014)).

C'est en se basant sur ces travaux récents de neurobiologie que nous avons voulu revoir les concepts de modèles oscillants et de duo excitateur-interneurone inhibiteur en général afin d'aboutir à des modèles souples, capables de s'adapter pour rester fonctionnels malgré des changements dans les dynamiques des signaux d'entrée.

8.4 | Différentes propositions de modèles

Dans cette section, nous avons abordé le problème de façon « pas à pas » en proposant des modèles de complexité croissante, en commençant par un modèle peu crédible neurobiologiquement et en avançant itérativement vers des modèles plus solides tout en faisant en sorte qu'ils restent le plus simples possible.

Modèle de base

Le modèle sélectionné pour notre base de travail est un modèle de couple inhibiteur-excitateur basé sur des neurones de type CTRNN, ce modèle est basé sur les travaux de (WANG, 1995).

La spécificité de ce modèle en comparaison des modèles plus classique basée sur les neurones de type CTRNN comme (MALSBERG et BUHMANN, 1992) et (WANG et al., 1990) est l'absence de liens récurrents inhibiteurs. Wang note en effet dans (WANG, 1995) que ce lien inhibiteur classiquement employé dans d'autres modèles ne semble pas très utile dans ses tests, le modèle étant parfaitement stable sans celui-ci.

L'équation d'origine de Wang, lié à la figure 8.3 est :

$$\tau * \dot{U} = -\alpha * U + C * f(U) - A * f(V) + S(E) \quad (8.1)$$

Pour le neurone excitateur. Et :

$$\tau_2 * \dot{V} = -\beta * V + B * f(U) \quad (8.2)$$

Pour le neurone inhibiteur.

Pour le neurone excitateur et inhibiteur respectivement : τ et τ_2 sont des paramètres d'intégration (pas forcements identiques). α et β sont les paramètres de relaxation. U et V sont les états énergétiques internes des neurones. A, B et C sont les poids pour, respectivement, le lien inhibiteur, le lien activateur, et le lien récurrent sur le neurone excitateur. E est l'entrée d'un signal extérieur sur le neurone excitateur.

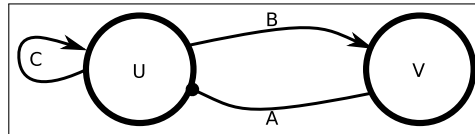


FIGURE 8.3. – *Modèle oscillant inspiré des travaux de (WANG et al., 1990)*

Pour simplifier la conception et l'étude des systèmes d'apprentissage entre ces deux neurones, nous simplifions encore une seconde fois le travail de Wang après sa première simplification. Ainsi, nous considérons qu'en présence d'une entrée constante, l'équation 8.1 peut s'écrire :

$$\tau * \dot{U} = -\alpha * U + C * f(U) + B - A * f(V) \quad (8.3)$$

Où

$$C * f(U) + B \quad (8.4)$$

Peut s'écrire :

$$E * (P1 * f(U) + P2) \quad (8.5)$$

Comme E est constant. On écrit, en réalité, une équation de modulation de la capacité de l'entrée à être prise en compte en fonction de la sortie du neurone.

Dans ce cas particulier, ceci est très proche des mécanismes de facilitation existants dans les neurones. Où $P2$ est le facteur minimum d'énergie entrant dans le système quand celui-ci n'est pas excité et où $P1$ est un facteur qui facilite l'entrée de l'énergie en fonction de l'excitation du système.

Cette simplification supprime la nécessité d’avoir un lien récurrent d’auto-excitation dans le modèle en introduisant un phénomène classique et bien connu dans les mécanismes neuronaux. Cette simplification est justifiée uniquement dans le cas d’une entrée constante, toutefois, comme l’importance de ce lien récurrent est d’induire une compétition à l’intérieur de l’oscillateur en déclenchant un effet domino on peut considérer ici que la simplification fonctionne également pour des entrées non-constantes et qu’elle peut donc se généraliser.

Les tests qualitatifs ont montré que l’oscillateur fonctionnait parfaitement malgré des entrées variables dans ce cadre. Nous conservons donc cette équation initiale comme base de travail.

Notons que le système de facilitation peut également être utilisé sur le neurone inhibiteur comme il apparait que dans les précédentes versions du modèle de Wang un lien récurrent d’activation existe également sur ce neurone. Toutefois par simplicité nous n’intégrons pas ce mécanisme de facilitation sur le neurone inhibiteur. Les tests qualitatifs ont toutefois montré que le fonctionnement ne présentait pas de différence notable.

Après cette dernière simplification les liens existants dans le modèle sont réduits au lien d’inhibition venant du neurone inhibiteur et au lien d’activation provenant du neurone activateur. Ainsi que des liens d’entrée au neurone excitateur. On aboutit donc au modèle présenté sur la figure 8.4

Les équations pour chaque neurone deviennent alors :

$$\tau * \dot{U} = -\alpha * U + S(E) * (f(U) * P1 + P2) - A * f(V) \quad (8.6)$$

Pour le neurone excitateur et :

$$\tau_2 * \dot{V} = -\beta * V + B * f(U) \quad (8.7)$$

Pour le neurone inhibiteur. Où f est une fonction sigmoïde définie par :

$$f(U) = (2 * ((1.0 / (1.0 + \exp(-4 * (U - \theta)))) - 0.5)) \quad (8.8)$$

Où θ est un seuil et où les autres paramètres sont là pour centrer la sigmoïde afin de s’assurer que la dérivée est égale à 1 à l’origine.

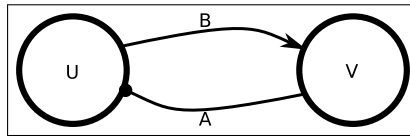


FIGURE 8.4. – *Modèle oscillant inspiré de Wang et simplifié*

Principes généraux d’apprentissage utilisés dans ces travaux

Dans cette section, nous tentons de suivre les découvertes récentes faites par (ALVAREZ et al., 2013) dans les systèmes des boucles inhibitrices entre les motoneurones et les cellules de Renshaw qui montrent que les synapses inhibitrices et excitatrices de ces boucles se développent par apprentissage et ne sont pas fixées génétiquement (voir aussi (CASTILLO, 2012) pour une approche plus générale autour de l’apprentissage des synapses excitatrices et inhibitrices dans les interneurones). Nous avons ainsi testé plusieurs règles simples d’apprentissage dans la suite de ces travaux. Pour ce faire, nous posons quelques hypothèses :

Premièrement, nous supposons que dans le cerveau les connexions synaptiques sont maintenues par des règles de stabilisation qui essaient d’empêcher un neurone de se retrouver déconnecté de toute source de signal et donc d’être « mort » ou du moins, inutile. Deuxièmement, nous supposons que les liens sont ajustés de manière à éviter une trop longue surcharge des neurones qui les rendrait peu intéressants pour représenter une information.

Gardant cela à l'esprit nous avons sélectionné deux règles d'apprentissages principales pour nos modèles : la règle d'apprentissage de type Hebb bilinéaire, qui cherche à maximiser les activités corrélées et la classique, mais solide, règle de type BCM (Elie Bienenstock, Leon Cooper, et Paul Munro) (BIENENSTOCK et al., 1981) qui a été largement utilisée et étudiée comme une bonne approximation des processus de modification synaptiques dans le cerveau. Ces deux règles, dans des conditions dynamiques, assurent qu'un réseau reste stable grâce à leurs propriétés d'ajustement.

Deuxièmement, nous basons nos travaux sur les recherches récentes sur les liens GABAergiques qui ont démontré que la modification des synapses GABAergiques se faisait via un processus très proche de celle des synapses Glutaminergiques (FLORES et MÉNDEZ, 2014).

L'idée générale dans les modèles que nous allons présenter sera que si les synapses GABAergiques présentent des caractéristiques plastiques similaires à leurs cousines glutaminergiques alors il devient possible de créer un système basé sur deux neurones liés par des liens excitateurs et inhibiteurs, régit par des règles similaires, et qui vont converger vers un système dynamique qui se maintient dans une « zone utile » à travers l'adaptation de ses liens synaptiques au vu du fait que nous savons exactement ce que nous voulons éviter dans le système (saturation, situation statique, neurone ne délivrant plus aucun signal, etc.). Nous détaillons l'application de ce principe en pratique dans les sections suivantes.

Modèle 1 : Adaptation du poids synaptique inhibiteur seul

De façon schématique, si on imagine un neurone inhibiteur qui reproduit avec un délai un signal d'un neurone en amont qu'il tente en retour d'inhiber, et si l'on considère le poids synaptique inhibiteur très faible à l'origine, alors, si l'on applique une règle classique d'apprentissage de type Hebb bilinéaire et que le couple se retrouve excité par un signal constant sur le neurone excitateur alors les poids de la synapse inhibitrice vont se mettre à augmenter à cause de la corrélation entre l'énergie présente dans la cellule inhibitrice et la forte énergie présente dans la cellule excitatrice. (voir partie A du schéma 8.5).

Pour prévenir les cas de sur-inhibition, la partie négative de la règle d'apprentissage bilinéaire est ici importante, car elle nous assure que le poids décroît en cas de sur-inhibition. En effet en cas de sur-inhibition le neurone activateur se retrouve totalement inhibé tandis que l'inhibiteur conserve une trace des activations passées, du moins pendant un temps. Il y a donc décorrélacion et réduction du poids de la synapse inhibitrice (voir partie B du schéma 8.5).

La loi de modification bilinéaire (partielle) des poids W_{ij} est :

$$\begin{array}{l} \text{si } U > \varepsilon \text{ alors} \\ | \\ \Delta W_{ij} = \delta * f(V) * U \end{array} \quad (8.9)$$

$$\begin{array}{l} \text{sinon} \\ | \\ \Delta W_{ij} = -\delta * \beta * f(V) * W_{ij} \end{array} \quad (8.10)$$

Où ε est un seuil désignant la valeur de « non-activité ». Dans notre cas nous considérerons l'activité comme trop faible pour être représentative pour $0.1 * Max_output$ où Max_output est défini comme la valeur maximum prise par $f(U)$ (ici : 1,0). Cela nous donne une valeur de 0,1 pour ε . δ est ici le pas d'apprentissage et β est le facteur d'oubli appliqué quand le potentiel d'excitation est trop faible.

L'oubli est ici proportionnel à la quantité d'énergie qui tente d'inhiber en « trop » le neurone excitateur (soit $f(V) * W_{ij}$). Ceci évite la diminution du lien quand celui-ci est déjà trop faible ou quand aucune autre activité d'inhibition n'est présente.

On peut noter que cette règle est similaire, mais qu'elle présente toutefois une différence avec la règle BCM par la persistance du processus d'oubli quand U est égal à 0. Ceci permet de s'assurer qu'une sur-inhibition, même si aucune activité n'est présente sur le neurone excitateur, mène bien à une diminution du lien d'inhibition alors que la règle BCM d'origine est neutre quand aucune activité n'est présente sur le neurone postsynaptique. Nous nous sommes inspirés pour ce type de comportement de la découverte récente d'une influence présynaptique des récepteurs NMDAR sur les liens GABAergiques (LIEN et al., 2006) (MAPELLI et al., 2016). Cette influence peut ainsi générer une LTD même en l'absence d'une activation post-synaptique neuronale (voir partie B du schéma 8.5).

Toutefois, notons que ces mêmes études montrent également que la présence de glutamate dans l'environnement entourant les liens reste nécessaire pour déclencher les récepteurs présynaptiques NMDA qui sont responsables de cette LTD afin que l'effet apparaisse. En d'autres termes, cela signifie que la réduction en cas de sur-inhibition a bel et bien lieu, mais uniquement si le neurone sur-inhibé reçoit tout de même des entrées d'excitation. Un neurone qui serait trop inhibé, mais un peu excité par une autre voie verra ainsi ses liens inhibiteurs décroître pour laisser l'excitation s'exprimer. À contrario, si le neurone ne reçoit absolument aucun signal, une sur-inhibition sera considérée comme n'ayant aucun impact et ne sera pas pénalisée.

Il s'agit d'un type d'interaction dite triadique. Cette interaction porte une signification assez riche d'un point de vue mathématique. L'équation que tente de résoudre un neurone inhibiteur tentant d'inhiber un neurone activateur, dans notre théorie, est finalement de trouver le bon poids synaptique qui aboutit à inhiber ni trop ni pas assez. En empêchant le lien de décroître en cas de sur-inhibition dans le vide, on signale au neurone qu'il est inutile de pénaliser le poids dans ce cas, car la sur-inhibition n'a alors aucun impact sur le résultat. Cela permet donc de simplifier le problème posé à la synapse pour sa convergence, le problème de convergence étant ainsi moins contraint.

Cette règle supplémentaire permet alors d'éviter une pénalisation de l'inhibition lorsqu'elle n'est pas un problème pour la transmission de l'information.

On implémente donc ceci pour changer le comportement présenté sur la partie B du schéma 8.5. Le nouveau comportement est alors illustré par les parties B' et B'' du schéma 8.6.

D'un point de vue équationnel nous changeons ici l'équation 8.10 qui devient donc l'équation 8.11 en introduisant un terme qui tient compte du glutamate extérieur comme une variable de la somme des excitations appliquées au neurone considéré : $\sum_{a=1}^n (E_{aj})$.

On suppose ici, en se basant sur les études précédemment citées, que le glutamate fuit dans l'environnement extérieur environnant, ce qui déclenche les récepteurs NMDA présynaptiques sur les liens inhibiteurs. Ce glutamate environnemental est alors représentatif de la somme des excitations récentes appliquées sur le neurone. Notre hypothèse étant qu'il est plutôt représentatif d'une moyenne temporelle de toutes les synapses glutaminergiques avoisinantes et non pas d'une interaction triadique synapse par synapse.

$$\Delta W_{ij} = -\delta * \beta * f(V) * W_{ij} * \sum_{a=1}^n (E_{aj}) \quad (8.11)$$

Cette modification assure donc que la sur-inhibition n'est pénalisée que si les informations extérieures entrantes sont existantes, mais elle n'est pas pénalisée dans les autres cas.

Notons que dans cette version le poids de la synapse activatrice est fixe. Après convergence, la caractéristique de l'oscillation n'est donc plus que dépendante du poids de la synapse activatrice et des caractéristiques des neurones, leur temps d'intégration notamment, que l'on peut voir comme correspondant à leur taille.

Si le signal d'excitation n'est pas constant, mais périodique alors la synapse inhibitrice converge vers un poids qui dépend de la corrélation entre la fréquence du signal et la fréquence naturelle du couple neuronal.

Bien sûr, ici ce « modèle 0 » ne tient pas compte de la synapse excitatrice, qui est alors considérée comme fixe. Nous montrerons dans le modèle suivant comment tenir compte de l'adaptation de cette

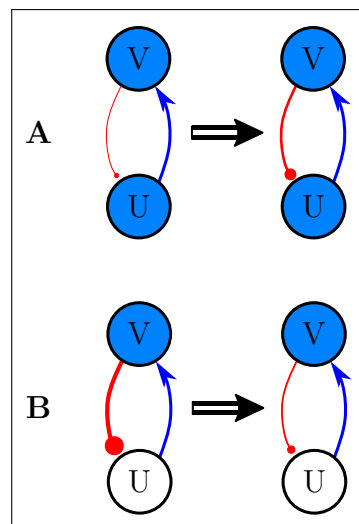


FIGURE 8.5. – Apprentissage bilinéaire dans le modèle 1, partie 1

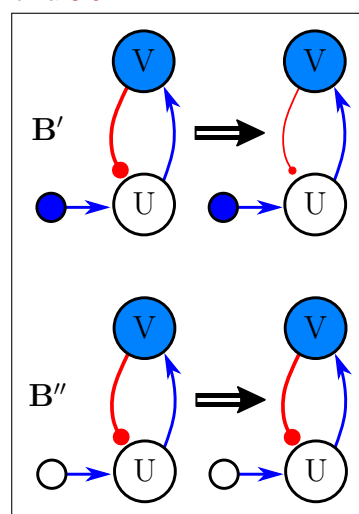


FIGURE 8.6. – Apprentissage bilinéaire dans le modèle 1, partie 2

synapse afin de réduire davantage les paramètres ad hoc à fixer lors de l'implémentation de ce type de réseau dans des architectures complexes.

Résultats Si l'on soumet le réseau à un signal constant, il va converger vers un poids de l'inhibition suffisant pour aboutir à un système oscillant. La raison en est simple. Comme on le voit sur les schémas 8.5 et 8.6, le poids du lien inhibiteur varie pour éviter que l'énergie transmise par l'activateur et arrivant dans l'inhibiteur ne cause une trop grande inhibition.

Ainsi, si on considère que le neurone inhibiteur contient une image retardée de l'activité dans le neurone excitateur, du fait des paramètres d'intégration des neurones CTRNN et du fait que le neurone excitateur est le premier à recevoir le signal excitateur, alors le poids synaptique du neurone inhibiteur tente de converger de telle sorte que l'inhibition ne cause pas de dépassement trop important en sur-inhibition tout en faisant en sorte qu'en moyenne l'inhibition soit efficace. On illustre ce basculement entre l'apprentissage et le désapprentissage sur la figure 8.7.

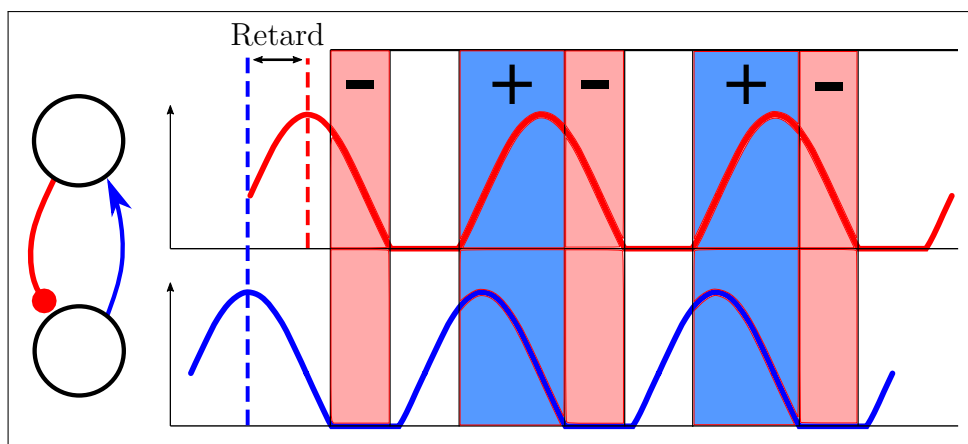


FIGURE 8.7. – Explication du processus d'équilibre durant l'apprentissage.

Sur cette figure on peut observer le signal existant sur l'activateur, qui prend une forme rythmique au fur et à mesure de l'apprentissage. Lorsque l'on converge, le processus atteint un équilibre. En effet, si l'on visualise les zones où l'apprentissage a lieu et si on les colore selon que le poids synaptique augmente ou diminue, alors on peut voir que dans le cas illustré il y a une succession de diminution et augmentation du poids synaptique.

Il faut noter que le facteur pour le désapprentissage n'est pas forcément le même que pour l'apprentissage, ce qui peut donner lieu à des formes différentes. S'ils sont égaux, le système convergera vers un temps équivalent des zones bleues et rouges sur le schéma ce qui donnera un système avec autant de sur-inhibitions que d'inhibition efficace sur un signal existant. À l'inverse, si le facteur pour le désapprentissage est très grand devant celui pour l'apprentissage, alors les zones rouges seront très réduites et l'on obtiendra un signal dont la zone nulle sera très faible et qui ressemblera à une sinusoïde.

On peut tracer pour quelques facteurs d'apprentissage l'évolution de l'apprentissage sur ce système sur la figure 8.8. On observe alors que le facteur d'apprentissage n'a pas de lien avec la valeur vers laquelle le lien inhibiteur converge, qui, dans des conditions identiques (valeur de l'entrée, paramètres des neurones, etc.), converge vers un poids synaptique unique. Par contre, le facteur d'apprentissage fait évidemment varier le temps de convergence. On peut observer également sur cette courbe que la convergence a bien lieu et qu'à la fin le poids synaptique est parfaitement stable.

Notons d'ores et déjà quelques bonnes propriétés du système. Il est capable de s'adapter à de très larges conditions d'entrée. En effet, si l'entrée est très importante par exemple, alors il faudra énormément d'inhibition pour venir la compenser. Un réseau standard serait rentré dans une situation bloquante puisque le neurone activateur et le neurone inhibiteur ne peuvent pas avoir des sorties infinies et sont bornés à 1 (dans notre cadre). Un mauvais réglage initial du poids de l'inhibition (par exemple : 1)

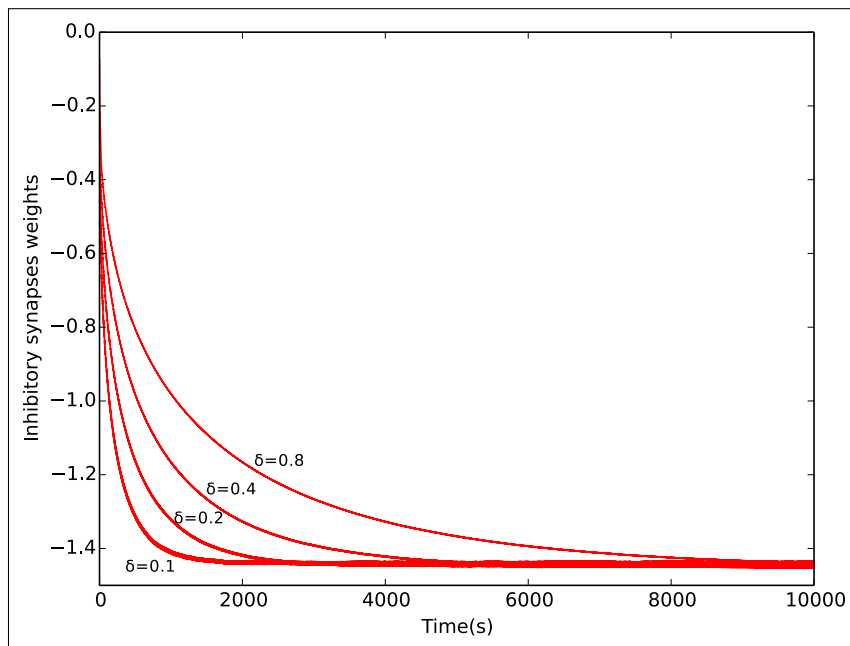


FIGURE 8.8. — Convergence du modèle avec lien inhibiteur seul en fonction de plusieurs facteurs d'apprentissage.

serait incapable de venir inhiber une entrée trop importante (10 pour continuer sur l'exemple) ce qui aboutirait à un système saturé et à un interneurone inutile. Dans notre cas, le problème n'apparaît pas puisque le poids synaptique inhibiteur va augmenter jusqu'à être capable d'inhiber le système.

Enfin si le système est soumis à une entrée dynamique et que le facteur d'apprentissage est suffisamment faible alors il aura tendance à converger pour tenter d'être dans un régime utile en moyenne sur l'intégralité des signaux qu'il reçoit. Notons que si ces signaux varient sur le très long terme, le système suivra alors pour se maintenir en moyenne dans sa plage d'utilité.

Critiques Le système reste encore dépendant de plusieurs paramètres et peut être mis en échec dans certains cas. Par exemple, il est nécessaire que le neurone inhibiteur emmagasine de l'énergie suffisamment lentement pour que l'effet de retard soit suffisant. Dans le cas contraire, le système se comporte comme un filtre et non comme une unité ayant une dynamique propre, car l'inhibition est trop rapide pour enclencher le phénomène oscillant. Ce n'est pas forcément un problème en soi, mais une adaptation du lien excitateur peut aider à faire varier la vitesse d'accumulation d'énergie dans le neurone inhibiteur en augmentant la quantité d'énergie en entrée.

Le second point critique est que le système est ici dépendant du lien excitateur qui doit être suffisant pour transmettre une énergie, mais pas trop important non plus sinon le neurone inhibiteur emmagasine de l'énergie trop rapidement.

Dans les modèles suivants, on se pose la question de l'apprentissage dans ce lien excitateur qui mène vers l'interneurone inhibiteur.

Modèle 2 et 2b : Adaptation du poids synaptique du lien excitateur et combinaison

On se pose la question de l'adaptation du poids synaptique du lien activateur. Dans un premier temps au lieu de chercher un modèle neurobiologiquement plausible d'emblée on cherche à imaginer une règle qui amènerait notre réseau dans des conditions optimales. Néanmoins, notons que la littérature

nous confirme bien que les liens qui viennent exciter les interneurones inhibiteurs peuvent subir un apprentissage (CASTILLO, 2012).

Si on analyse le problème en observant les équations de fonctionnement des deux neurones on se rend compte que le poids synaptique excitateur va impacter la *largeur d'inhibition* en saturant le neurone inhibiteur. En effet, plus le poids excitateur est fort, plus celui-ci va exciter le neurone inhibiteur. Néanmoins au-delà d'un certain seuil le neurone inhibiteur reste saturé au maximum et un poids synaptique plus grand ne fera alors qu'augmenter le temps nécessaire pour la décroissance de l'énergie accumulée dans le neurone inhibiteur. De fait, notons qu'indirectement le rôle de ce poids synaptique est lié à la largeur de l'inhibition.

Ensuite, on peut poser les conditions critiques : si le neurone inhibiteur est trop excité et accumule trop d'énergie alors il va rester longtemps activé même quand le neurone activateur est éteint depuis un certain temps. Poussée à l'extrême, cette condition est négative puisqu'elle empêche le signal de circuler pendant un temps qui peut être très long, le temps que le neurone inhibiteur retombe à un niveau énergétique acceptable.

On définit donc une règle pour ce poids synaptique : si le neurone présynaptique est désactivé et que le neurone postsynaptique est activé, il faut alors diminuer le poids synaptique en question afin de réduire peu à peu l'effet de saturation.

À l'inverse, si les deux sont activés en même temps on augmente le poids synaptique. Notons qu'on obtient alors un principe contraire à celui énoncé pour le poids synaptique inhibiteur dans la section précédente puisque cette fois-ci c'est l'absence de signal sur le neurone présynaptique, quand le neurone post-synaptique est encore activé, qui va déclencher un oubli.

La loi de modification des poids W_{ij} pour le lien excitateur est :

$$\text{si } U > \varepsilon \text{ alors} \quad \Delta W_{ij} = \delta * f(U) * V \quad (8.12)$$

$$\text{sinon} \quad \Delta W_{ij} = -\delta * \beta * V * W_{ij} \quad (8.13)$$

Où ε est le même seuil désignant la valeur de « non-activité » que dans la section précédente. δ est ici le pas d'apprentissage et β est le facteur d'oubli appliqué quand le potentiel d'excitation est trop faible.

Résultats avec le lien excitateur seul On ne présente ici que le résultat qualitatif. Tout comme pour la règle précédemment établie sur le lien inhibiteur, la règle établie sur le lien excitateur est stable et arrive à converger pour éviter une sur-activation, c'est-à-dire une activation trop importante qui conserve le neurone inhibiteur dans un état énergétique bien que le neurone excitateur soit complètement inhibé. Le principe pour la convergence est identique à celui présenté pour le lien inhibiteur seul sur le schéma 8.7. Ainsi de même que pour le lien inhibiteur variable précédemment la forme vers laquelle converge le réseau est principalement dépendante du facteur d'oubli β . Si celui-ci est très grand alors il y aura peu d'activation du neurone inhibiteur en l'absence d'activation du neurone excitateur ce qui convergera dans l'idéal vers deux formes sinusoïdes ne tombant que de façon très temporaire dans une zone nulle. À l'inverse, un facteur β faible aboutira à un comportement avec de longues plages d'inhibition.

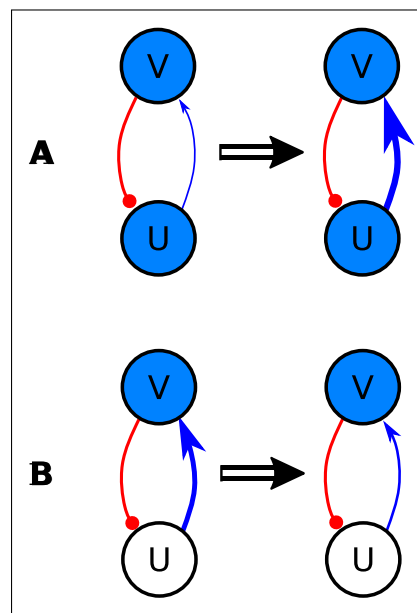


FIGURE 8.9. — Apprentissage dans le modèle 2 : mécanismes d'apprentissage sur le lien activateur.

Combinaison des deux systèmes d'apprentissage Lors de la combinaison des deux systèmes d'apprentissages le comportement du système est assez complexe en fonction des conditions et notamment en fonction des deux facteurs, β mais le système converge néanmoins toujours vers un comportement oscillant, quelles que soient les conditions que nous avons pu tester, du moins pour un neurone inhibiteur plus « lent » que le neurone activateur, sinon le système fonctionne, mais se comporte plutôt comme un filtre. En effet, contrairement au système avec un lien inhibiteur seul, il n'est pas ici possible de converger vers un cas où le neurone inhibiteur par exemple n'est pas ou trop peu activé, vu que la synapse excitatrice qui y mène s'adapte pour s'assurer que le neurone reste actif.

Les deux seuls paramètres qui viennent changer de façon notable le macro-comportement du système une fois que celui-ci a atteint la convergence sont les paramètres internes aux deux neurones. Si le neurone inhibiteur réagit avec un facteur d'accumulation d'énergie faible dans l'équation (8.6) alors le comportement sera oscillant. Au contraire si le neurone activateur réagit lentement et que l'inhibiteur réagit très rapidement, le système tend à fonctionner comme un filtre régulateur. Il s'agit néanmoins d'un résultat attendu.

Le fait que le système converge vers des comportements différents en fonction des paramètres d'excitabilité des neurones est un aspect que l'on retrouve dans la littérature. Les variations de tailles dans les cellules de Golgi du cervelet sont par exemple suspectées de servir à induire des « filtres » de nature différente, ce qui est le cas également dans notre système.

Un autre point important est que le système, sous cette forme, résout d'ores et déjà la problématique que nous nous étions posée à savoir la création de multiples bancs de filtre de façon simple ou du paramétrage des poids de l'oscillateur en fonction d'un domaine de travail. Ici, si les poids sont mal paramétrés au départ cela n'a aucune influence sur la valeur atteinte après convergence. Les deux neurones peuvent même être virtuellement déconnectés au départ (poids nul) sans que cela ne change la convergence ensuite. Nous avons donc bien un système stable capable de converger vers son maximum d'utilité et ne dépendant que de paramètres restreints.

Critique des aspects bio-inspirés La disposition du système telle qu'elle est présentée dans cette version peut être critiquée. Ici, les équations d'apprentissage pour les deux versions (excitateur et inhibiteur) ont été définies de façon ad hoc pour converger vers un régime de fonctionnement. Néanmoins, si l'équation de l'inhibiteur est basée sur un fonctionnement connu dans la neurobiologie, il n'en est pas de même de l'équation du neurone activateur qui a été purement construit de façon ad hoc.

De plus, les travaux en neurobiologie cités précédemment ont relevé que le comportement en apprentissage des deux types de liens était très similaire. Pourtant nos deux types de liens sont ici soumis à des équations d'apprentissage très différentes.

Dans le modèle suivant, nous proposons un moyen de réunir les deux équations en une seule et nous étudions l'impact de ce changement sur le comportement du modèle.

Modèle 3 : modèle unifié autour d'une règle hebbienne bilinéaire

Si on observe les deux équations des liens excitateurs et inhibiteurs (équation (8.13) et équation (8.10)) on se rend compte que chacune décrit une partie d'une règle hebbienne bilinéaire. Les règles hebbiennes bilinéaires pouvant être définies par :

$$\Delta W_{ij} = \delta \cdot ((\epsilon \cdot f(A) \cdot f(B)) - (\beta \cdot f(A)) - (\gamma \cdot f(B)) - \sigma) \quad (8.14)$$

Où A est le neurone présynaptique du lien considéré et B le neurone postsynaptique du lien considéré. δ est le facteur d'apprentissage global, ϵ le facteur de renforcement en cas de co-activation et β et γ les facteurs d'oubli relatif aux deux autres cas (quand le neurone post-synaptique est actif seul et quand le neurone présynaptique est actif seul). σ est un facteur d'oubli permanent qu'on considérera nul dans notre cas.

La règle ici a donc tendance à faire augmenter le poids synaptique en cas de co-activation, mais à le faire diminuer dans tous les autres cas. Notre équation (8.10) correspond à la partie $\epsilon \cdot f(A) \cdot f(B)$ et $\gamma \cdot f(B)$ (apprentissage quand co-activation, oubli quand le neurone post-synaptique est activé seul)

de l'équation bilinéaire tandis que notre équation 8.13 correspond plutôt à la partie $\epsilon * f(A) * f(B)$ et $\beta \cdot f(A)$ (apprentissage quand co-activation, oubli quand le neurone présynaptique est activé seul).

A priori, l'implémentation de cette règle dans notre cas pose quelques problèmes. En effet, si on rajoute à l'équation (8.10) (côté neurone inhibiteur donc) une composante qui diminue le poids synaptique inhibiteur quand le neurone présynaptique est désactivé alors si trop peu d'énergie au départ est transmise au neurone inhibiteur, le lien d'inhibition se retrouvera réduit. Ce n'est pas forcément un problème illogique pour cette partie de l'équation puisque cela revient à dire que si le neurone inhibiteur est désactivé il évite de trop inhiber les prochaines fois afin d'éviter de revenir dans cet état désactivé, ce qui reste, indirectement, pertinent dans la dynamique de la boucle. Par contre, ceci est plus problématique dans le cas du neurone exciteur. En effet, si l'on intègre une règle qui stipule qu'en cas de désactivation du neurone inhibiteur, le lien de l'exciteur vers l'interneurone inhibiteur se retrouve à diminuer, alors cela signifie que si les conditions initiales sont, par exemple, un lien exciteur dont le poids est nul ou très faible, alors le neurone inhibiteur aura au départ une énergie nulle et le poids synaptique n'augmentera donc jamais.

Il existe donc un seuil de conditions initiales à respecter pour s'assurer que le neurone inhibiteur au départ ait une activité minimum.

Dans un premier temps, afin d'éviter cette problématique, nous avons voulu tester le système en fixant un poids minimal de départ sur le lien synaptique activateur.

Résultats Comme dans la version précédente on obtient un réseau oscillant stable si l'on respecte la contrainte d'un poids minimal de départ sur le lien exciteur. Néanmoins, le réseau n'est pas tout à fait stable dans toutes les conditions comme le montre le diagramme des poids synaptiques après convergence par rapport à l'énergie moyenne du signal en entrée du réseau oscillant sur la figure 8.10.

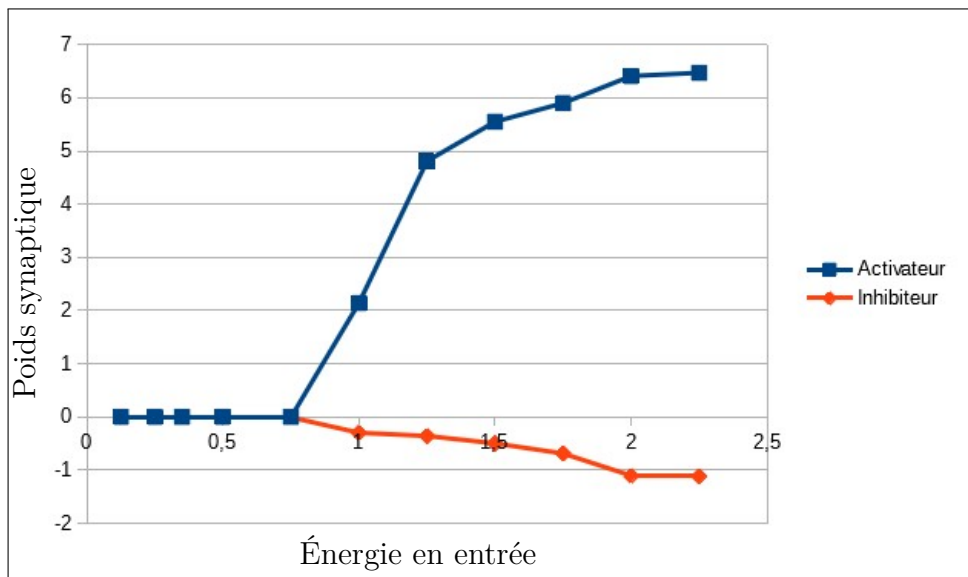


FIGURE 8.10. — *Poids synaptiques après convergence sur le modèle bilinéaire en fonction de l'énergie moyenne du signal injecté en entrée.*

Sur cette figure, nous pouvons voir que les poids s'adaptent en conséquence face à l'énergie d'entrée. Plus l'énergie d'entrée est forte en moyenne, plus les poids sont importants pour permettre une inhibition plus forte, ce qui est bien la logique que nous cherchions dans le système. Toutefois, on peut noter qu'il n'y a aucun démarrage de l'apprentissage pour des énergies trop faibles. En effet, même avec un lien inhibiteur à une valeur suffisante au départ, il existe un seuil à partir duquel l'énergie injectée n'est pas suffisante pour déclencher une activation du neurone inhibiteur. En deçà de ce seuil les deux liens

ne peuvent donc que désapprendre et converger vers 0. Il s'agit ici du principal défaut de ce système : l'existence d'un ensemble de conditions menant à la désactivation définitive du réseau.

On peut ensuite étudier la réaction de ce système vis-à-vis de la fréquence moyenne d'un signal imposée. En effet malgré l'apprentissage on s'attend à ce que le système ait une certaine préférence pour un domaine fréquentiel particulier, en lien avec les caractéristiques des deux neurones considérés.

Si on trace la convergence moyenne des poids synaptiques en fonction de la fréquence moyenne du signal considéré on obtient la courbe présentée sur la figure 8.11

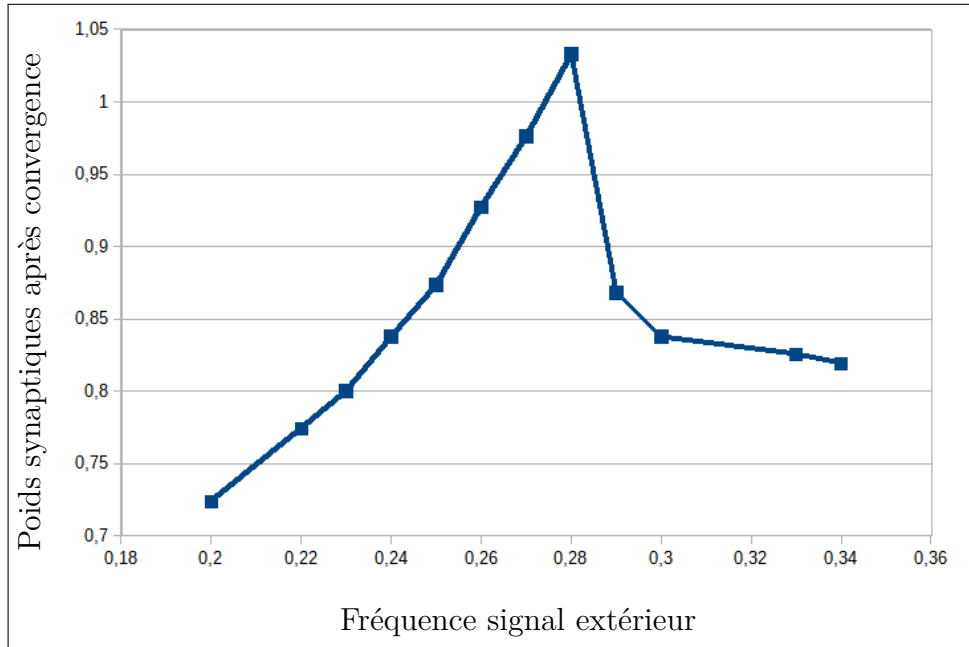


FIGURE 8.11. – Évolution des poids synaptiques en fonctions de la fréquence imposée.

Sur cette figure, on peut observer qu'il y a bien un pic représentatif à une certaine fréquence, pour laquelle la transformation effectuée par le système est maximale grâce à des poids synaptiques élevés. On pourrait appeler cette fréquence la fréquence caractéristique de l'apprentissage dans le système de neurone oscillant.

Analyse La combinaison des deux règles donne un résultat fonctionnel, mais au prix de contraintes plus importantes que pour le modèle précédent. Ici, il est ainsi nécessaire de partir d'un poids synaptique non nul pour le lien inhibiteur. L'énergie d'entrée est également contrainte selon le même principe. Il y a donc un risque de rendre le réseau inutile en cas d'énergie trop faible pendant quelque temps. Ce qui est au contraire aux principes initiaux que nous nous étions fixés.

Pour pallier ces problématiques, nous avons développé un nouveau modèle basé sur une autre loi d'apprentissage qui a le mérite de supprimer une part importante de ces défauts.

Modèle 4 : Introduction d'une loi d'apprentissage de type BCM

La loi d'apprentissage de type BCM est une loi qui fut proposé par Bienenstock, Cooper et Munro en 1981 (BIENENSTOCK et al., 1982)), pour expliquer à l'origine les mécanismes de l'apprentissage dans le cortex visuel. Depuis, la règle a été validée dans de nombreux autres cadres comme une bonne approximation des processus neurobiologiques de l'apprentissage. Il a ainsi été démontré que cette règle est aux modèles neuronaux à moyenne de décharge ce que la règle STDP, est aux modèles à spike

(IZHIKEVICH et DESAI, 2003). la règle STDP étant validée dans de très nombreux cadres cela a permis de valider en rebond la règle BCM dans ces cadres. Il a par exemple été proposé que les neurones du cervelet, qui ont un apprentissage inverse un peu particulier, pourraient suivre une règle BCM inversée (COESMANS et al., 2004).

La règle BCM représente un apprentissage en fonction de la valeur émise par le neurone postsynaptique. Si la valeur émise par ce neurone post-synaptique est en dessous d'un seuil alors il y a désapprentissage ; si cette valeur est au contraire au-dessus de ce seuil alors il y a apprentissage.

La courbe de l'apprentissage/désapprentissage est également conçue de telle sorte que si le neurone post-synaptique est inactif alors il n'y a plus aucun apprentissage. De fait, cette règle est très similaire à l'équation (8.10) que nous avons développée pour le neurone inhibiteur. On peut écrire l'équation BCM ainsi :

$$\Delta W_{ij} = \delta \cdot \phi(y_B(t), \theta(t)) \cdot y_A(t) \quad (8.15)$$

Où δ est un facteur d'apprentissage global. $y_B(t)$ correspond à la sortie du neurone post-synaptique et $y_A(t)$ à la sortie du neurone présynaptique. ϕ est la fonction qui dépend de y_B et qui permet d'apprendre ou de désapprendre en fonction de sa valeur.

La particularité de la règle BCM est que ϕ est une fonction qui peut varier en fonction d'un second facteur, ici $\theta(t)$. Ce facteur représente la moyenne des activations récentes de $y_B(t)$. La forme de la fonction ϕ est assez simple à comprendre en la traçant plutôt qu'en explicitant les paramètres détaillés de la fonction. On peut observer ce tracé sur la figure 8.12. La variation de la forme de la fonction en fonction du paramètre $\theta(t)$ est alors bien visible : si θ diminue le seuil de passage entre le désapprentissage et l'apprentissage diminue, et réciproquement.

Ce système a été introduit pour apporter une solution au problème des équations hebbiennes ou bilinéaires classiques. Ici si $y_B(t)$ est en moyenne très faible, voire égale à 0, alors le point de transition sera très faible ou égal à 0, ce qui fait que le modèle ne pourra qu'apprendre.

Inversement, un $y_B(t)$ trop fort en moyenne aura tendance à mener à un désapprentissage, car le seuil sera très élevé. On a donc un système qui tend à être naturellement stable et à privilégier une représentation des données sur le neurone.

On modifie ici quelque peu le système en changeant la définition de $\theta(t)$ au lieu de poser l'équivalence $\theta = \overline{y_B}$ on écrit θ comme une fonction linéaire de $\overline{y_B}$ sous la forme :

$$\theta = C * \overline{y_B} + D \quad (8.16)$$

On calcul alors C et D pour que θ soit égal à y_B quand y_B est égal à une valeur d'activation correcte du neurone, tout en laissant un espace de travail. Dans notre cas nous avons défini l'ensemble pour que θ soit égal à y_B quand $y_B = 0.7$ (voir figure 8.13). L'objectif ici est de guider légèrement le système pour qu'en moyenne le neurone représente une donnée. Ce n'est pas très éloigné du concept d'origine de l'équation BCM qui a été conçu pour maximiser la représentation de l'information par le neurone. Ce n'est également pas incohérent avec ce que l'on sait de la neurobiologie à ce propos.

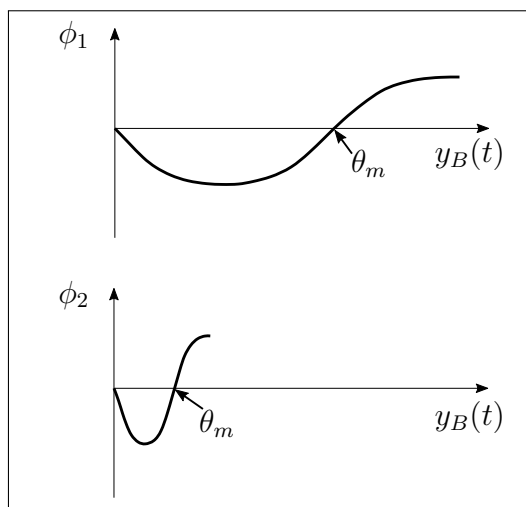


FIGURE 8.12. — Exemples de formes de la fonction ϕ dans l'équation BCM.

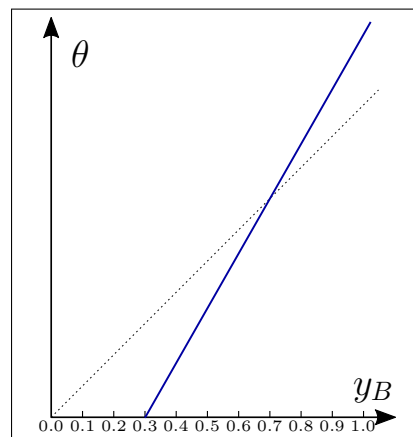


FIGURE 8.13. — Exemple de fonction θ modifiée, ici la fonction est inférieure à $\overline{y_B}$ pour $\overline{y_B} < 0.7$ et supérieure sinon.

Une fois ceci mis en place on se rend compte que la règle BCM résout beaucoup des problèmes de la partie précédente en restant équivalente sur le fond. Par exemple, une fois appliquée au lien excitateur, cette règle assure que le lien converge vers une valeur qui rend le neurone inhibiteur actif en moyenne. De même, les problèmes d'énergie minimum en entrée ou de poids synaptique minimum n'existent plus. Une sur-activation mènera également à une décroissance implicite du poids, car elle mènera à une augmentation de la moyenne θ qui aboutira à un dépassement du point équivalent à $\overline{y_B}$. De fait, si on se réfère au schéma 8.12, θ deviendra alors grand par rapport à $\overline{y_B}$ ce qui aboutira à une plus grande zone de désapprentissage. La conséquence est que le système tente ainsi indirectement d'éviter une sur-activation.

De même, la règle BCM étant nativement proche de l'équation (8.10) le lien inhibiteur converge nativement vers une valeur capable d'inhiber le neurone excitateur, à ceci près que la règle BCM ne pénalise pas la sur-inhibition quand le neurone excitateur est totalement inhibé et n'implémente pas nativement la dépendance à la moyenne des signaux d'entrée que nous avons implémentée dans l'équation (8.11).

Y a-t-il ici une incohérence d'un point de vue neurobiologique ? En réalité, ce n'est pas vraiment le cas. Un article récent de Jonathan Mapelli et collègue a étudié un mécanisme neuronal très proche qui pourrait expliquer cette incohérence (MAPELLI et al., 2016). Dans cet article, il est mis en évidence qu'il existe deux types de récepteurs distincts sur les liens GABAergiques, contrairement, a priori, aux liens Glutaminergiques. Le récepteur NMDAR postsynaptique est le récepteur classique qui a un impact sur l'apprentissage des liens glutaminergiques comme GABAergiques. C'est ce récepteur qui est concerné par l'apprentissage de type BCM avec un renforcement en cas de corrélation et un désapprentissage en cas de décorrélation.

Toutefois, il existe un second type de récepteurs NMDA, présynaptiques ici, qui va déclencher un oubli sur les liens GABAergiques. Une émission GABAergique trop forte déclenchera ainsi un oubli via le récepteur présynaptique, qui ne sera pas compensé par un apprentissage puisque le récepteur NMDA post-synaptique se retrouvera inactif du fait de l'inhibition du neurone auquel il est rattaché. Comme nous en avons déjà discuté, la présence de glutamate reste ici nécessaire.

Pour prendre en compte ces aspects, on reprend donc la règle formulée dans l'équation (8.11) pour le critère d'oubli, que l'on vient ajouter à une règle BCM classique et qui forme ainsi le nouveau fonctionnement de notre lien inhibiteur.

Résultats Ce nouveau réseau offre les mêmes avantages que le réseau précédent, mais sans les défauts dus à la quantité d'énergie minimum nécessaire. Le réseau se stabilise dans toutes les conditions que nous avons pu tester, quelles que soient les valeurs initiales des poids synaptiques ou les entrées et leur forme (variable, rythmique, constante). Contrairement au réseau précédent, il n'existe pas de cas qui viendrait rendre inutile définitivement le système. On peut étudier l'évolution du comportement du couple en fonction de plusieurs paramètres. Par exemple l'évolution de la fréquence du couple en fonction de l'énergie moyenne entrante sur la figure 8.14.

Cette évolution nous montre que dans ce modèle il n'y a pas de fréquence privilégiée, mais plutôt une zone de trou fréquentiel dans laquelle le modèle réagit lentement. Notons qu'il est possible que le pic de fréquence soit présent pour une énergie plus basse que celles tracées ici, mais nous ne l'avons pas testé dans cette série de caractérisation.

On peut étudier en parallèle l'évolution des poids synaptiques selon les mêmes conditions, que l'on présente sur la figure 8.15.

Le comportement est assez différent des modèles précédents, puisque, même si les poids ont la même forme au premier abord, il ne faut pas s'y tromper : l'un étant inhibiteur et l'autre activateur, quand l'un augmente, l'autre voit sa valeur absolue diminuer.

On peut mettre en évidence ce fait en présentant le résultat de la multiplication des deux valeurs absolues sur la figure 8.16.

Ceci montre bien que le système s'organise de façon à laisser passer de plus en plus d'énergie pour l'inhibition en fonction de l'énergie entrante sur le neurone excitateur. Même si les poids synaptiques

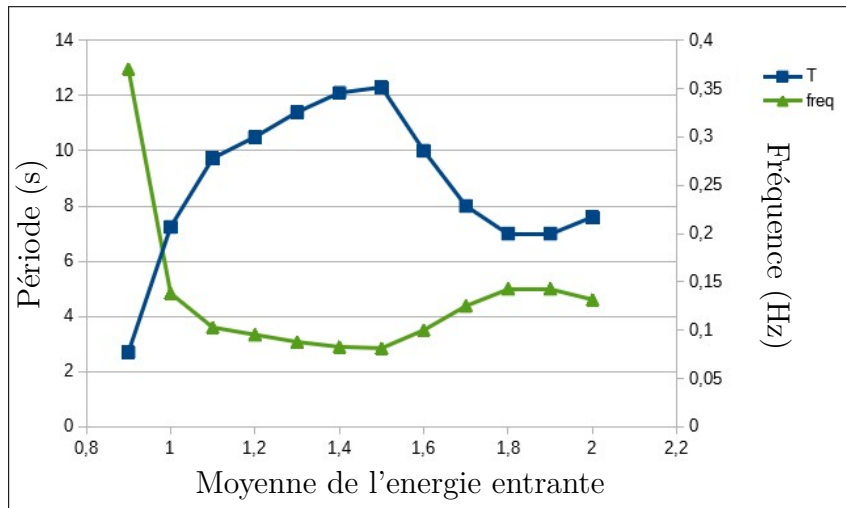


FIGURE 8.14. – *Évolution de la fréquence et de la période du système en fonction de l'énergie moyenne entrante.*

inhibiteurs et excitateurs du duo peuvent varier, si l'un diminue, l'autre augmente pour compenser cette variation et maintenir l'inhibition, ce qui était le but recherché à la conception de ce système.

On a voulu tester ensuite les conséquences d'une variation de certains paramètres internes aux neurones sur la dynamique du doublet. L'un des paramètres qui nous a paru le plus pertinent ici est le seuil d'activation du neurone inhibiteur. Dans les équations précédentes, le seuil n'apparaît pas, car il est égal à 0, mais en réalité la valeur énergétique du neurone V est modifiée par un seuil qui vient s'y soustraire avant le passage dans les fonctions d'activation. Ce seuil définit donc de façon simple la capacité d'excitation du neurone inhibiteur. Plus il est bas, plus le neurone inhibiteur est facilement excitable. On trace la variation de la fréquence du duo en fonction de ce seuil d'activation sur la figure 8.17.

On peut observer qu'il n'existe pas de situation où le réseau devient non-actif, même pour des seuils assez hauts dans lesquels le neurone inhibiteur est très difficile à exciter. Le système arrive ainsi à apprendre en faisant varier les liens pour compenser les seuils et permettre la transmission d'énergie. Cette compensation entraîne évidemment des conséquences sur la dynamique du système, mais celui-ci reste fonctionnel malgré tout.

On peut observer cette compensation en traçant la même variable composée que celle que nous avons étudiée sur la figure 8.16 : le produit des valeurs absolues des poids synaptiques en fonction de l'évolution du seuil est visible sur la figure 8.18.

L'évolution du produit sert ici d'indicateur quant à la capacité du système à transmettre une énergie dans sa boucle excitateur-inhibiteur. On observe bien que ce produit augmente linéairement avec l'augmentation du seuil. Ce qui est le résultat attendu puisque le seuil introduisant une barrière énergétique, le système doit la compenser pour rester actif.

8.4.1. Perspectives pour le modèle, quelles applications ?

Dans ce chapitre, nous avons ainsi abordé la problématique de l'intégration de réseaux oscillants dans des architectures de contrôle robotique bio-inspirées, développementales, destinées à être implémentées sur robot et à fonctionner en environnement réel. Les modèles existants et les applications qui en sont faites dans la littérature ne nous ont pas paru pertinents, car ils dépendaient trop souvent de paramètres fixés par l'expérimentateur pour être capables de fonctionner pour une situation particulière, ce qui limitait considérablement l'usage que l'on pouvait en faire dans notre cadre. Par ailleurs, les conclusions de ces études ne nous paraissaient pas non plus pertinentes du fait d'hypothèses trop peu solides : si

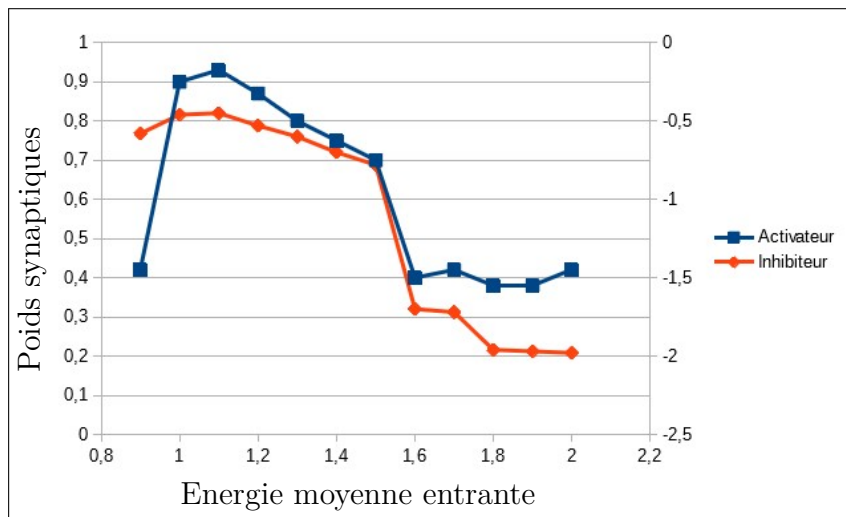


FIGURE 8.15. — Évolution des poids synaptiques en fonction de l'énergie entrante.

ces réseaux venaient à s'adapter à la réalité, alors leurs comportements dynamiques pourraient être très différents de ceux étudiés classiquement.

Pour cet ensemble de raisons, nous avons tenu à développer un réseau capable de modéliser les interactions d'un couple composé d'un neurone excitateur et d'un interneurone inhibiteur du point de vue de l'apprentissage des liens synaptiques les reliant entre eux. Ce type de couple est en effet très présent dans le système nerveux humain, notamment dans les ensembles moteurs ou dans les systèmes de traitement sensoriel.

Par l'étude des découvertes en neurobiologie faites dans le domaine de l'apprentissage dans les liens inhibiteurs et de l'apprentissage dans les couples excitateur-interneurone inhibiteurs, nous avons pu développer une succession de modèles dont la dernière itération est non seulement apte à résoudre le problème posé, mais est également plus stable et moins contrainte que les modèles précédents tout en étant plus bio-inspirée vis-à-vis des règles d'apprentissage implémentées.

Néanmoins, nous n'avons ensuite pas exploité ce modèle dans une architecture complète robotique. Nous proposons ici en perspective trois applications potentielles simples du modèle.

La première est une application aux réseaux de traitement de type Kohonen pour la vision. Si on modélise les cellules de type *Basket cells* dans les cartes topologiques visuelles comme cela est classiquement fait dans les études autour de la modélisation du cortex visuel, il doit alors être possible d'adapter ce modèle (l'interneurone inhibiteur étant le « basket-cell » qui reçoit des activations d'un ou plusieurs neurones en amont et qui en retour va en inhiber plusieurs). Ce modèle pourrait en effet apporter un élément de réponse à la largeur d'inhibition de ce type de cellule, puisqu'on s'attend à ce que les liens d'inhibition se construisent ici naturellement en fonction de la corrélation de la cellule inhibitrice avec les données des cellules voisines. Ceci pourrait donner un modèle de carte topologique auto-organisatrice dépendant de très peu de méta-paramètres et présentant de très bonnes propriétés quant à la précision des catégories selon le contexte. De plus, le modèle pourra sans doute séparer implicitement des motifs rythmiques.

La seconde proposition est d'appliquer directement le modèle au couple motoneurone-cellule de Renshaw. Comme nous l'avons déjà abordé, les cellules de Renshaw sont des cellules spécifiques inhibitrices associées à chaque motoneurone commandant le muscle. L'application de ce modèle à ce couple dans un système robotique complet usant d'un modèle de muscle pour la commande aboutirait sans doute à des comportements inédits (quoique difficile à prédire a priori). Bien entendu, pour capturer un phénomène intéressant il serait ici nécessaire de complexifier quelque peu le modèle présenté dans la partie II. Il

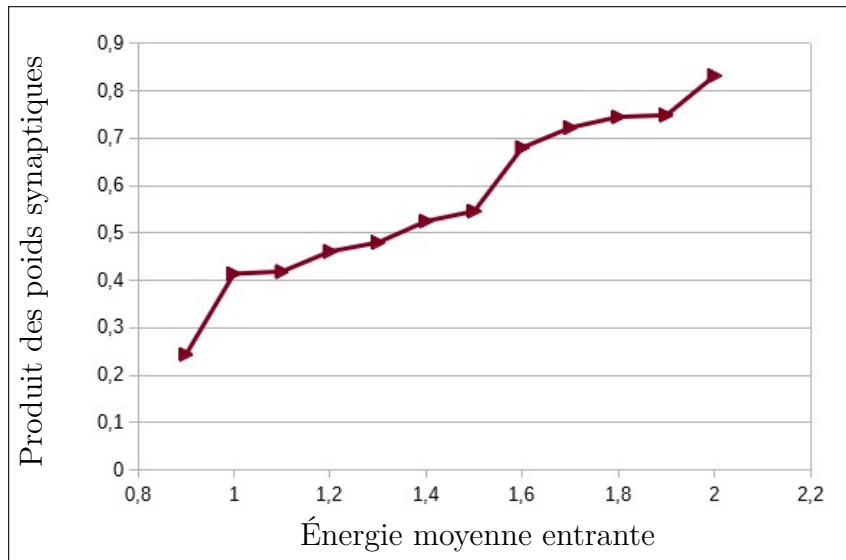


FIGURE 8.16. — *Évolution du produit des poids synaptiques du duo de neurone en fonction de l'énergie moyenne entrante.*

ne serait alors plus possible de se contenter d'une approximation analogique pour la commande sur le muscle, puisque chaque cellule de Renshaw est indépendante. Il faudrait donc modéliser plusieurs unités motrices indépendantes dont les forces s'additionnent dans le même esprit que ce que nous avons fait en simulation dans la partie III.

Une dernière proposition est d'appliquer ce modèle aux différents systèmes de séparation dynamiques des données. Ceux-ci sont basés sur des couples analogues, comme ce que l'on peut trouver dans le cervelet avec l'hypothèse des couples cellules granulaires-cellules de Golgi (voir bibliographie sur le cervelet dans la section 3.3.8 de la partie 2) comme étant des couples oscillants, ou du moins des couples de filtres, destinés à filtrer le signal d'entrée de façon extensive pour permettre un apprentissage fin ensuite (trajectoire, sensation, etc.). Ce type d'usage des bancs oscillants est abondamment présent dans la littérature et pourrait profiter directement des travaux présentés ici. Notons toutefois que c'est peut-être le résultat qui s'avèrera le moins intéressant, car même avec des paramètres « ad hoc », les résultats obtenus sur de très larges bancs d'oscillateur apprenant sont, d'après ce que nous avons parcouru de la littérature sur le sujet, très précis. Il y a donc ici peu de chance que le résultat qualitatif change en mieux, un couplet rendu inutile par de mauvais paramètres n'étant pas un problème particulier quand on dispose de milliers de couplets en parallèle avec des paramètres différents sur lesquels mener un apprentissage. Toutefois d'un point de vue pragmatique cela pourra servir à simplifier la tâche au modélisateur et à l'expérimentateur.

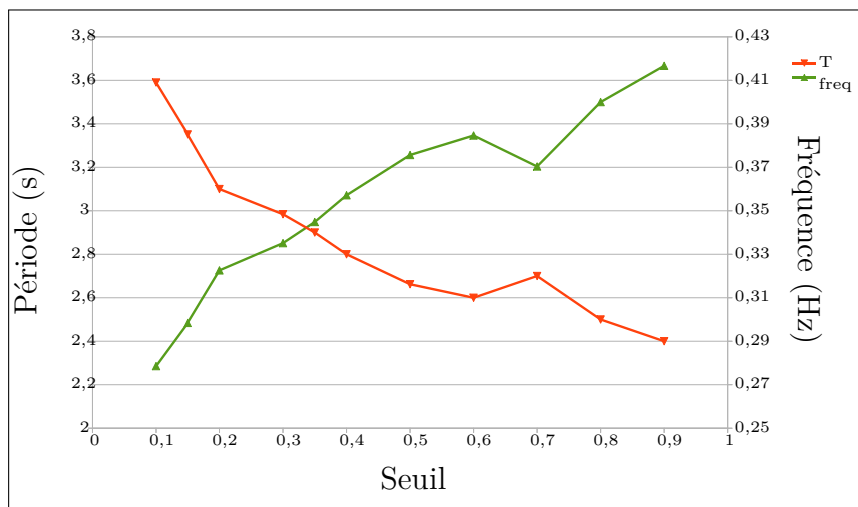


FIGURE 8.17. – Évolution de la fréquence du doublet oscillant en fonction du seuil d'activation.

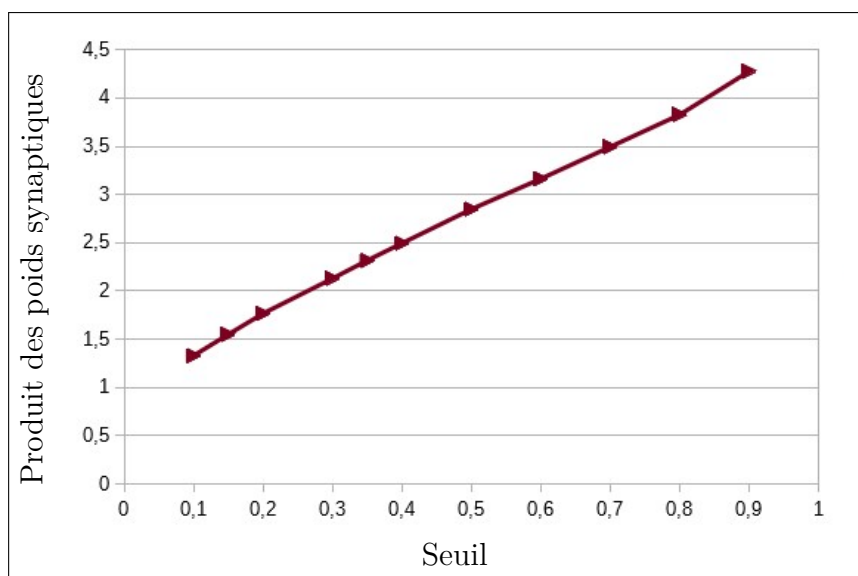


FIGURE 8.18. – Évolution du produit des poids synaptiques en fonction du seuil sur le neurone inhibiteur.

8.5 | Conclusion

Dans le chapitre précédent, nous avons abordé les avantages de la synchronisation non-intentionnelle et avons introduit nos travaux menés afin de préciser les limites de cette synchronisation chez l'humain. Nous avons par la suite critiqué les hypothèses qui ont mené à l'emploi de la synchronisation non intentionnelle en robotique et nous avons proposé un modèle implémenté sur robot afin de tester notre raisonnement.

Nous avons noté en conclusion que les liens faits implicitement et explicitement dans les travaux mêlant robotique et synchronisation non-intentionnelle, entre la synchronisation et les réseaux de neurones oscillants, sont, à notre avis, trop ténus face à d'autres hypothèses plus simples qui rendent compte des phénomènes observés.

Ceci nous a ensuite menés à séparer la problématique de la synchronisation non-intentionnelle et des réseaux oscillants. Néanmoins, les réseaux oscillants conservant des propriétés et un intérêt fort dans un cadre de robotique interactive, notamment pour se rapprocher des modes de génération des trajectoires humaines, mais également parce que leur omniprésence dans le cerveau humain appelle à comprendre leur utilité. Nous avons alors tenu à les étudier dans l'objectif de les intégrer à une architecture de contrôle robotique développementale. Ceci nous a menés à étudier les limites de ce type de système dans notre cadre et à proposer plusieurs modèles neurobiologiquement plausibles exploitant un état de l'art récent sur les apprentissages dans les liens GABAergiques. Nous avons ensuite caractérisé le modèle proposé, qui s'est avéré répondre à la problématique posée. Pour finir, nous avons proposé plusieurs applications potentielles en perspective qu'il serait intéressant de développer pour valider l'utilité et l'impact du modèle en pratique dans une architecture de contrôle robotique développementale intégrée.

Conclusion

Dans cette thèse nous nous sommes posé la question de l'architecture de contrôle nécessaire à un robot pour lui permettre d'être déployé dans un environnement réel dans lequel le robot serait capable d'exécuter des tâches en parallèle d'un humain ou en interaction avec des humains.

Pour explorer cette problématique, nous avons étudié plusieurs modèles et principes issus de travaux de robotique basés sur divers points soulevés par la psychologie, la mécanique ou la neurobiologie. Ces points, qui ont réussi à prouver leur côté essentiel pour l'interaction de l'humain avec l'environnement ou un partenaire peuvent être résumés par :

- La nécessité d'un corps bien conçu capable de faire profiter « gratuitement » de phénomènes utiles à l'interaction.
- La lecture des intentions de l'autre au travers de l'impact qu'ont les intentions sur les trajectoires humaines.
- La nécessité d'entretenir une résonance motrice (dynamique : synchronisation ou non : imitation) pour aboutir à des interactions intuitives peu coûteuses en ressource mentale.

Chacun de ces aspects avait donné lieu par le passé à de nombreux types de travaux notamment autour de l'intelligence du corps, de la prédiction de trajectoire, de l'imitation et de la synchronisation non-intentionnelle.

Dans cette thèse, nous avons cherché à nous inspirer de certains des travaux associés à ces points dans l'objectif de concevoir une architecture de contrôle bio-inspirée et développementale pour contrôler un robot capable d'interagir.

Le robot Tino qui fut la plateforme expérimentale sur laquelle nous avons travaillé pour aboutir à cet objectif est un prototype de robot hydraulique humanoïde unique en France que nous avons présenté pour la première fois dans cette thèse. De fait, nous avons présenté certains travaux de caractérisation, tant techniques que vis-à-vis de l'acceptabilité du robot, qui nous ont permis d'identifier les propriétés du robot, ses capacités d'usage dans un contexte d'interaction avec l'humain, ainsi que certains phénomènes émergents exploitables pour son contrôle.

Nous avons cherché à mettre en place sur le robot Tino une architecture combinant certains des points précédents en commençant par un contrôle sensorimoteur développemental combiné à un modèle de contrôle musculaire en co-contraction. Nous avons montré que ce type d'architecture est capable de fournir une base de travail solide. En effet, elle est non seulement capable de construire via l'exploration un « simili cortex moteur », c'est à dire un pavage des actions pertinentes pour l'interaction avec l'environnement, mais en plus cette architecture peut rendre compte de deux phénomènes importants pour les interactions interpersonnelles humaines et pour le développement de l'humain en interaction avec ses semblables : l'imitation et le pointage.

Par la suite, nous avons cherché à explorer la question de l'adaptabilité des trajectoires en fonction du contexte ou de l'intention. La lecture de l'intention de l'autre, et particulièrement du robot, étant l'un des points les plus problématiques dans l'interaction, et notamment de l'interaction Homme-machine, c'est tout naturellement que nous avons tenu à explorer les moyens d'améliorer cet aspect. Pour ce faire, nous nous sommes basés sur un résultat du domaine de la psychologie qui concluait que les trajectoires humaines sont influencées par l'intention durant l'action. De fait, il devient possible de lire l'intention par la simple observation de ces trajectoires.

L'insertion de ce phénomène en robotique nécessitait d'identifier ses mécanismes de fonctionnement chez l'humain. Les expériences de psychologie en question avaient démontré que le fait de donner un objet à un humain ou de le poser sur une table au même endroit changeait la forme et la dynamique des

trajectoires. Nous avons alors proposé, en nous inspirant de ce protocole, que « l'intention » puisse être simplement vue comme un contexte et les conséquences associées à ce contexte : ainsi, dans l'analogie du patient de Jacob et Jeannerod, le médecin cherchant à tuer son patient risque des conséquences différentes par rapport au médecin cherchant à le soigner.

Sous cette hypothèse, nous avons supposé qu'une architecture de contrôle par renforcement devait être capable de prendre en compte le contexte du point de vue du robot et donc d'aboutir à des résultats différents en fonction de l'intention. Nous avons alors démontré sur plusieurs variantes d'une architecture bio-inspirée basée sur le modèle des ganglions de la base chez l'humain que nous pouvions obtenir, sous certaines hypothèses, les mêmes résultats qualitatifs que ceux des équipes de psychologies ayant travaillé sur le sujet. Nous avons ensuite répliqué l'expérience sur un environnement simulé plus complexe basé sur un modèle de bras en deux dimensions.

Toutefois, nous n'avons pas résolu les problèmes de temps de convergence qui empêchent ce type d'architecture d'être intégrée directement dans l'architecture testée sur le robot Tino.

Par la suite, nous avons étudié un autre phénomène important du point de vue de l'interaction entre humains : la synchronisation non-intentionnelle. Nous avons présenté les spécificités de ce phénomène et nos travaux de caractérisation de ce phénomène chez l'humain en partenariat avec le laboratoire Euromov.

Jusqu'ici, en robotique, ce phénomène avait été modélisé par l'implication de réseaux oscillants. Pourtant, des travaux récents dans le domaine de l'intelligence du corps laissent entendre que de nombreuses propriétés autrefois attribuées aux réseaux de neurones oscillants pouvaient en réalité être le fruit de l'interaction du corps avec l'environnement et d'une simple impulsion. En nous basant sur ces travaux, nous avons proposé un modèle appliqué au robot Tino et inspiré des réflexes chez l'humain qui nous a permis de mettre en évidence que le même type d'effets était tout à fait observable sans l'intervention de neurones oscillants et uniquement dans le cadre d'un corps robotique mu par des réflexes basiques. Nous avons montré que les caractéristiques du phénomène obtenu étaient de plus proches, vis-à-vis des bornes d'accrochage et de décrochage en phase, de ce que l'on obtient chez l'humain. Ceci nous a amenés à remettre en cause le lien implicite tracé dans de nombreux travaux entre réseaux de neurones oscillants et synchronisation non-intentionnelle.

Sur la base de ces résultats, nous avons cherché à analyser les réseaux de neurones oscillants seuls, décorrélés de la problématique de la synchronisation motrice. Si les réseaux de neurones oscillants nous paraissaient susceptibles d'apporter des aspects intéressants à l'interaction entre le robot et son environnement ou entre le robot et l'homme c'était alors devenu plutôt du point de vue de leur capacité à servir de substrat pour la génération de trajectoires ou pour le filtrage des trajectoires et sensations, plutôt que comme substrat à la synchronisation. En effet, les réseaux oscillants restent très présents dans le cerveau humain et leur impact sur le mouvement semble faire consensus.

Toujours dans l'optique d'essayer d'obtenir des trajectoires et comportements suffisamment « humains » pour que dans l'interaction l'Homme n'ait pas besoin de beaucoup de ressources mentales afin de prédire les comportements de son partenaire robotique nous avons alors proposé d'étudier l'implémentation de réseaux de neurones oscillants dans nos architectures de contrôle.

De fait, nous avons cherché à isoler, dans les modèles existants de réseaux oscillants, les points critiques qui empêchent leurs implémentations simples dans une architecture évoluant dans un environnement complexe et en interaction avec des partenaires. Nous avons mis en avant que les réseaux oscillants traditionnels peuvent être mis en échec dans un contexte de réception de signaux perturbés et évolutifs, comme cela peut être le cas si le corps change, si l'environnement ou le partenaire varie, ou si les autres parties du modèle dans lequel ils sont intégrés évoluent eux aussi.

Nous en avons conclu que, pour utiliser ces réseaux dans notre contexte, en respectant les principes énoncés au début de cette thèse, il devenait nécessaire d'étudier l'apprentissage au sein de ces réseaux afin de les rendre aptes à être utilisés dans ce type de cadre changeant.

Nous nous sommes alors basés sur un ensemble de travaux récents en neurobiologie qui ont démontré la présence d'apprentissage dans les liens GABAergiques et tout particulièrement dans les couples de neurones excitateurs-interneurones inhibiteurs. En utilisant certains des points énoncés dans ces travaux nous avons alors proposé plusieurs modèles d'apprentissage successifs pour finir par aboutir à un réseau

stable, dont les règles sont inspirées de la neurobiologie, et qui se maintient dans une zone « utile » de fonctionnement malgré des signaux de dynamiques et d'énergie moyenne variables en entrée.

Malgré les travaux présentés ici, nous n'avons pas réussi dans cette thèse à lever l'ensemble des obstacles à l'obtention d'un modèle de contrôle unifié pour l'interaction, qui emploierait chacune des sous-parties de modèles présentés ici. Il reste ainsi plusieurs points critiques qui devraient être résolus dans des travaux futurs pour permettre cette convergence.

Dans le modèle initial d'apprentissage sensorimoteur par exemple, la simplification effectuée pour identifier les sorties de chaque muscle à une valeur analogique est par exemple incompatible avec ce que nous avons montré sur l'intérêt d'un pavage d'actions « unité musculaire par unité musculaire » dans la partie sur l'algorithme acteur-critique. Il faudrait donc adapter l'algorithme d'apprentissage sensorimoteur ainsi que la simulation du muscle pour prendre en compte cet aspect. Cela nécessiterait d'avoir une description plus fine des unités motrices chez l'humain et notamment des causes entourant le phénomène de recrutement des unités motrices par ordre d'importance en force (principe de la taille de Henneman).

Pour l'algorithme acteur-critique, la principale problématique vient du temps de convergence. Il existe probablement plusieurs méthodes pour réduire ce temps de convergence. Dans nos tests qualitatifs, il semble possible pour certains jeux de paramètres d'atteindre des temps de convergence très courts. Néanmoins, pour valider que l'on ne perd pas énormément de précision dans la convergence en employant ces jeux de paramètres, il convient sans doute de comparer méthodiquement les trajectoires obtenues après convergence sur toute une gamme de paramètres influant sur la vitesse de convergence.

Un autre aspect nécessaire est d'adapter l'algorithme au modèle musculaire en co-contraction, mais normalement cet aspect ne devrait pas poser de problèmes supplémentaires et devrait même simplifier la problématique à résoudre par l'algorithme. Enfin, la partie « acteur » de l'algorithme est encore ici très embryonnaire et mériterait sans doute d'intégrer des travaux plus précis sur le sujet.

Les aspects de synchronisation au travers du corps peuvent sans doute, eux, être intégrés assez facilement au modèle existant, mais cela nécessiterait de se pencher plus finement sur les réflexes présents chez l'humain pour en offrir une démonstration rigoureuse et permettre une intégration efficace au modèle d'apprentissage sensorimoteur.

Pour finir l'algorithme de réseaux oscillants est sans doute celui qui est le plus simple à implémenter. Nous avons déjà décrit dans la conclusion du chapitre 8 plusieurs propositions d'usages pour cet outil, la plus simple d'entre elles étant de l'implémenter comme un modèle des liens entre les motoneurones et leurs cellules inhibitrices associées de Renshaw. Le résultat qui en découlera par contre, dans l'hypothèse d'une implémentation sur l'architecture d'apprentissage sensorimotrice, est encore très hypothétique.

Par ailleurs, il reste nécessaire, pour aller dans cette direction, d'adapter l'algorithme d'apprentissage sensorimoteur à de multiples unités motrices, car la notion d'interneurone inhibiteur associé aux motoneurones n'a guère de sens dans le cadre de la simplification utilisée sur cet algorithme. L'étude des réflexes pourrait ici aussi être importante, car les cellules de Renshaw servent de relai à plusieurs réflexes visant à inhiber les motoneurones. L'effet de ces liens réflexes dans un réseau motoneurone-cellule de Renshaw dont les liens sont adaptables est là aussi très difficile à prédire.

Ces points qui restent à résoudre sont, d'après nous, parmi les plus importants pour permettre une convergence vers un modèle de contrôle robotique capable d'interaction dans un environnement complexe, avec l'humain, et en exploitant les phénomènes qui furent présentés ici et que l'on sait primordiaux chez l'humain.

Annexes

Protocole d'expérimentation au laboratoire partenaire Ureca de Lille 3

Protocoles expérimentaux TINO

Introduction:

Lorsqu'ils doivent se mouvoir ou attraper un objet sur une table, les individus vont produire des séquences d'action qui a priori peuvent être très diverses. Cependant, si l'on observe de plus près la manière avec laquelle le mouvement est réalisé, nous pouvons constater qu'il est régi par un certain nombre de lois (Fitts, Loi puissance $\frac{2}{3}$). Ces lois motrices vont ainsi gouverner l'ensemble des actions d'un humain et définir ce que nous pouvons nommer le mouvement biologique.

Le développement actuel de la robotique nous amène à penser que d'ici peu, nous pourrions interagir avec des robots ayant l'apparence humaine (tel que Neo). Dans ce contexte, nous nous interrogeons sur les facteurs pouvant influencer la qualité de l'interaction avec un agent non-humain. Plus particulièrement, nous nous centrerons sur les paramètres cinématiques pouvant favoriser l'interaction. Pour ce faire, nous mettrons en place deux études distinctes.

L'objectif de la première étude sera de fournir une évaluation des capacités d'action des différents robots en comparaison à l'humain. En particulier, nous souhaitons étudier si la cinématique de mouvement du robot hydraulique TINO est plus proche de celle d'un humain par rapport à ce que peut faire un robot électrique classique.

Par la suite, nous testerons si notre capacité à interagir avec un système dépend-t-elle de la cinématique de mouvement de ce dernier ? En d'autres termes, nous testerons l'intuitivité de l'interaction avec différents agents ayant des cinématiques de mouvement distinctes.

Etude 1: → Evaluation des capacités d'actions des différents agents

Méthode

Participants

En plus des enregistrements réalisés sur le robot TINO et le robot électrique, Quatre participants humains prendront part à l'expérience. Ils devront remplir un formulaire de consentement éclairé en adéquation avec les normes éthiques décrites par la Déclaration d'Helsinki.

Matériel

Pour mener à bien cette expérience, un ordinateur de bureau (Windows 7) muni de deux écrans sera utilisé pour la collecte. Relié à cet ordinateur, un système de capture du mouvement QUALISYS avec 4 caméras infrarouges (1 sur trépied et 3 fixées au mur à l'aide de support en bois, vissés dans le mur) sera mis en place dans le but d'enregistrer les mouvements d'agent et de l'acteur à une fréquence d'échantillonnage de 200 Hz (figure 1).

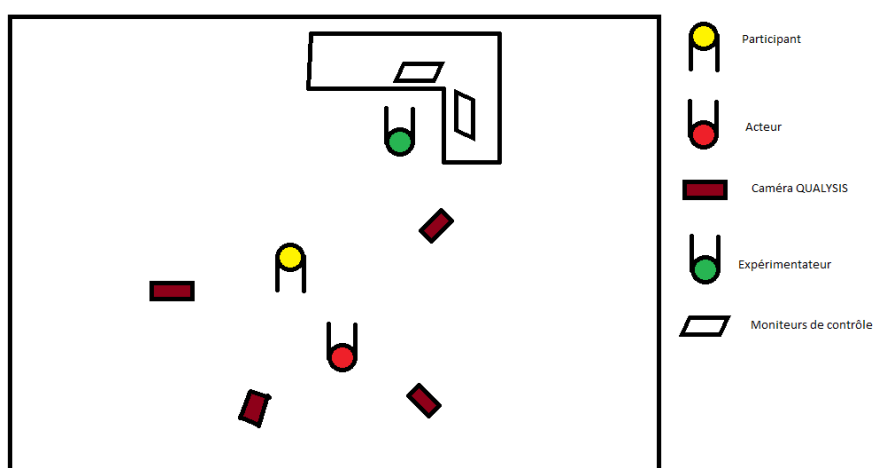


Figure 1. Set-up expérimental

Comme le présente la figure 1, un acteur sera assis en face de l'agent (participant). L'acteur, humain préalablement entraîné, servira de modèle à l'agent dans la condition d'imitation décrite plus loin. L'agent (ou participant) sera assis sur une chaise (une quinzaine de centimètre sur la gauche), face au mur et dos à l'expérimentateur, tandis que l'acteur sera également sur une chaise. Trois types d'agent seront testés : un humain, un robot électrique et le robot hydraulique (TINO). Treize marqueurs infrarouges passifs seront placés sur l'agent et 6 sur le bras gauche de l'acteur (Figure 2).

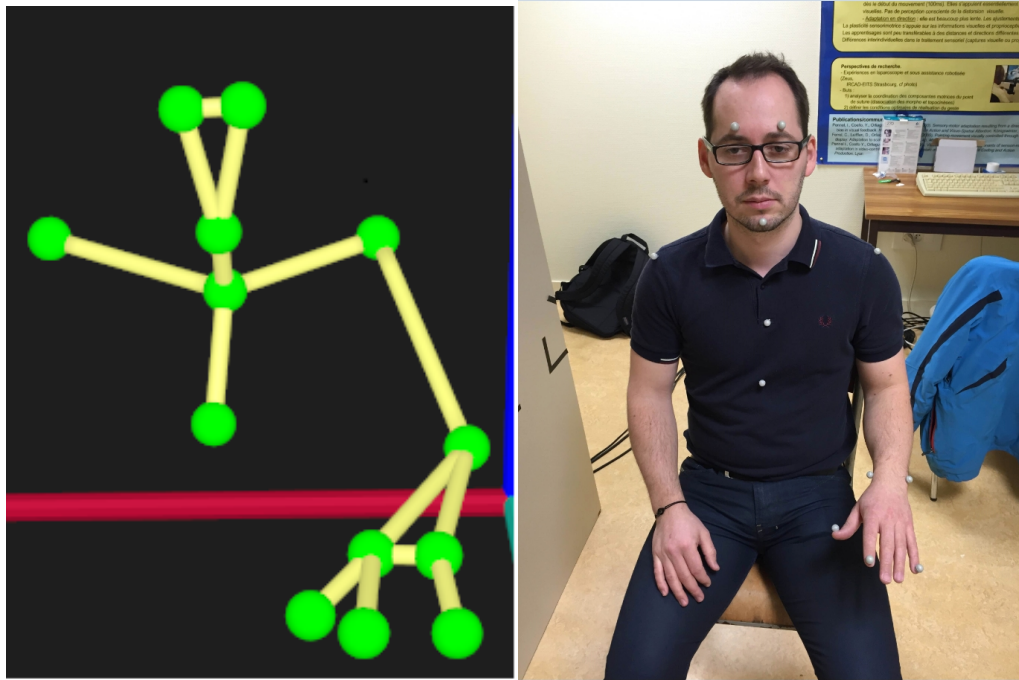


Figure 2: Modèle et position des 13 marqueurs positionnés sur l'agent: Visage (3), Epaule (2), Coude (1), Sternum (1), Plexus (1), Poignet interne et externe (2), et dernière phalange du pouce, index et auriculaire.

Procédure

Durant cette expérience, l'agent devra réaliser deux tâches différentes. En effet, ils devront effectuer des mouvements seul (S), en imitation de l'acteur (I).

Dans la condition Seul (S), le participant devra réaliser des mouvements horizontaux ou verticaux. Ces mouvements seront réalisés à 3 types de tempo. Un tempo rapide, caractérisé par un intervalle inter bip de 300ms, un tempo lent avec un intervalle inter bip de 900ms et un tempo spontané pour lequel le participant réalise le mouvement sans contrainte temporelle (environ 500-600ms). Pour les deux tempi imposés, nous utiliserons un métronome auditif fonctionnant sous Matlab caractérisé par deux sons de fréquence différente (un pour l'aller à une fréquence de 880Hz, le second pour le retour à une fréquence de 440Hz). Nous calculerons ensuite la cadence spontanée moyenne de chaque participant pour chacun des types de mouvement (H/V).

Dans la condition imitation (I), l'agent devra également réaliser des mouvements horizontaux ou verticaux suivant les 3 mêmes tempi (L/R/S). Cependant, dans cette condition, c'est l'acteur assis en face de lui qui imposera le tempo. En effet, un casque audio lui permettra d'écouter les différents tempi imposés: Lent (intervalle = 900ms), Rapide (300ms) et Spontanée (correspondant à la cadence spontanée moyenne du participant préalablement calculé dans la condition seule). Ainsi à chaque essai le participant devra imiter les mouvements produits par l'acteur.

Pour chacune des conditions, trois essais devront être effectués et, pour éviter la présence d'un effet d'ordre, les conditions seront contrebalancées pour chaque agent. A la suite de cette expérience, un debriefing sera proposé au participant afin de lui expliquer le but du projet et qu'il puisse poser des questions. Pour cette expérience, le temps de passation sera compris entre 45 minutes et 1h.

Description de l'essai

Avant chaque essai, il sera demandé à l'agent de maintenir leur bras devant eux, paume vers le bas (mvt du bras autour de l'épaule avec coude bloqué). Au signal sonore, il devra commencer son mouvement en fonction de la condition souhaitée.

Chaque essai durera 30 secondes.

Pour les conditions verticales, le participant devra faire bouger son bras gauche sur un plan vertical aux tempi demandés et de façon similaire sur un plan horizontal pour les autres conditions. Afin d'uniformiser les mouvements des participants, une limite devra être fixée en hauteur et en largeur. Cette limite sera, en hauteur, l'épaule de l'acteur et en largeur, la cuisse de l'acteur. Les participants ne recevront aucun retours sur leur performance durant l'expérience.

Analyse des données

Grâce au logiciel QUALISYS Track Manager, nous interpolerons les données manquantes puis nous exporterons les données du mouvement vers MATLAB. A l'aide de ce logiciel, les coordonnées de positions de chaque articulation seront filtrées, puis nous calculerons les vitesses, accélérations (figure 3).

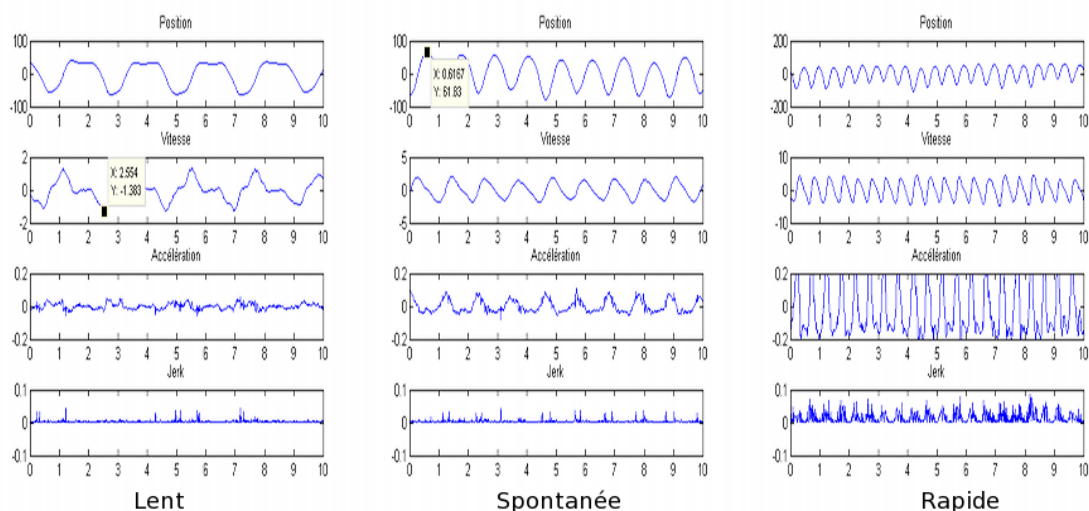


Figure 3: Variation de position, vitesse, accélération et jerk du barycentre des marqueurs de la main de l'agent humain en fonction du temps demandé (Lent, Spontanée et Rapide)

Pour la suite des analyses, nous nous centrerons particulièrement sur les coordonnées de la main pour laquelle nous calculerons le barycentre des 3 capteurs: pouce, index et auriculaire. A partir des ces données, nous calculerons l'amplitude de mouvement (correspondant à la différence entre les extremum de position) mais également plusieurs indices de fluidité du mouvement.

Le premier est le Jerk, correspond à la dérivée de l'accélération. Cette variable rend compte des variations d'accélération au cours du temps.

Le second paramètre est le time-to-peak de vitesse qui correspond au temps entre à d'initiation du mouvement et à l'instant où la vitesse du mouvement cesse de croître.

Le troisième paramètre time-to-speed0 correspondra au temps passé par chaque agent à vitesse nulle sur l'ensemble de l'essai (- de $5\text{mm}\cdot\text{s}^{-1}$).

L'ensemble de ces variables seront quantifiées pour les différents agents et pour l'ensemble des conditions expérimentales.

Résultats attendus

Si l'hypothèse de départ, à savoir que le robot hydraulique (TINO) possède une cinématique de mouvement plus proche de celle de l'homme, nous pourrions nous attendre à ce que les indices de fluidité soient semblables à ceux enregistrés pour le mouvement biologique. Par exemple, nous pensons:

- Amplitude de mouvements différentes pour les deux robots
- $\text{Time-to-peak Speed}_{\text{TINO}} - \text{Time-to-peak Speed}_{\text{Humain}} < \text{Time-to-peak Speed}_{\text{Elec}} - \text{Time-to-peak Speed}_{\text{Humain}}$
- Présence d'un tempo spontanée pour humain et TINO
 - Diminution du jerk et temps où la vitesse est nulle.

Etude 2: → Notre capacité à interagir avec un système dépend-t-elle de la cinématique de mouvement de ce dernier ?

Méthode

Participants

26 participants seront recrutés et répartis aléatoirement dans trois groupes expérimentaux (Interaction H-H; Interaction H-TINO; Interaction H-Robot). Après lecture de la lettre d'information, ils devront remplir un formulaire de consentement éclairé en adéquation avec les normes éthiques décrites par la Déclaration d'Helsinki.

Matériel

En fonction du groupe dans lequel sera réparti le participant, il devra interagir avec un agent humain ou non-humain. L'ensemble du matériel utilisé est identique à celui présenté dans l'étude 1. Ainsi un agent et un acteur équipés de capteurs passifs (13 pour l'agent et 6 pour l'acteur) sont assis et se font face. Pour rappel, l'acteur est la personne qui guide l'interaction alors que l'agent est celui qui réagit au mouvement de l'acteur. Dans cette étude, trois types d'acteur seront utilisés: Humain, TINO ou robot électrique.

Procédure

La procédure est identique à celle utilisée dans l'étude 1. En effet, l'agent devra réaliser deux tâches différentes: réalisation de mouvements seul (S), et en imitation de l'acteur (I).

Dans la condition seul, le participant devra réaliser des mouvements horizontaux ou verticaux. Ces mouvements seront réalisés à 3 types de tempo (rapide = 300ms, lent = 900ms et un tempo spontané pour lequel le participant réalise le mouvement sans contrainte temporelle (environ 500-600ms). Pour les deux tempi imposés, nous utiliserons un métronome auditif fonctionnant sous Matlab caractérisé par deux sons de fréquence différente (un pour l'aller à une fréquence de 880Hz, le second pour le retour à une fréquence de 440Hz). Nous calculerons ensuite la cadence spontanée moyenne de chaque participant pour chacun des types de mouvement (H/V).

Dans la condition imitation, l'agent devra également réaliser des mouvements horizontaux ou verticaux suivant les 3 mêmes tempi (L/R/S). Cependant, dans cette condition, c'est l'acteur assis en face de lui qui imposera le tempo. En effet, un casque audio lui permettra d'écouter les différents tempi imposés: Lent (intervalle = 900ms), Rapide (300ms) et Spontanée (correspondant à la cadence spontanée moyenne du participant préalablement calculé dans la condition seule). (-->A voir avec Cergy comment nous pouvons imposer le tempo au robot). Ainsi, à chaque essai le participant devra imiter les mouvements produits par l'acteur.

Enfin, une condition de coopération où l'acteur et l'agent devront coopérer pour se faire passer une canette de soda d'une position A à une position B sera réalisé

Pour chacune des conditions, trois essais devront être effectués et, pour éviter la présence d'un effet d'ordre, les conditions seront contrebalancées pour chaque agent. A la suite de cette expérience, un questionnaire permettant d'évaluer l'agréabilité de l'interaction sera proposé au participant avant que le but de l'expérience soit présenté

au sein d'un debriefing. Pour cette expérience, le temps de passation sera compris entre 45 minutes et 1h.

Analyse des données

L'analyse de données sera comparable à ce qui a été présenté dans l'étude 1. Grâce au logiciel QUALISYS Track Manager, nous interpolerons les données manquantes avant de les exporter vers MATLAB. Ce logiciel nous permettra de calculer les indices de fluidité du mouvement (amplitude, time-to-peak vitesse, time-to-speed0, jerk) en fonction des différentes conditions expérimentales. De plus, nous analyserons également la cinématique de mouvement de l'acteur et le décalage de phase pouvant exister entre la performance de l'acteur et celle de l'agent.

Résultats attendus

Tout d'abord, concernant les questionnaires, nous nous attendons à ce que les participants jugent comme agréable la situation d'interaction avec un acteur ayant un cinématique de mouvement proche de l'humain. Par ailleurs, concernant les enregistrements cinématiques, nous pensons que lors de cette même situation, la cinématique de mouvement des individus devrait être plus fluide (réduction jerk...). Enfin, nous pensons que le décalage de phase entre le mouvement de l'acteur et celui de l'agent, devrait être minimal.

Bibliographie

- ALBUS, James S. (1971). « A theory of cerebellar function ». In : *Mathematical Biosciences* 10.1-2, p. 25-61 (cf. p. 61).
- ALFAYAD, Samer (2009). *Robot humanoïde HYDROÏD : actionnement, structure cinématique et stratégie de contrôle*. Editions universitaires europeennes EUE. URL : <https://books.google.fr/books?id=VibWMwEACAAJ> (cf. p. 9, 24).
- (2011). *Robot Humanoïde Hydroïd*. Editions universitaires europeennes EUE. ISBN : 9783841783455. URL : <https://books.google.fr/books?id=XRIUtQAACAAJ> (cf. p. 24).
- ALFAYAD, Samer, Mohamad ELASSWAD, Ahmad IBRAHIM, Fathi Ben OUEZDOU, Arnaud BLANCHARD, Nils BEAUSSÉ et Philippe GAUSSIER (2016). « HYDROiD humanoid robot head with perception and emotion capabilities : Modeling, Design and Experimental Results ». In : *Frontiers in Robotics and AI* 3.15. ISSN : 2296-9144. DOI : 10.3389/frobt.2016.00015. URL : http://www.frontiersin.org/humanoid_robotics/10.3389/frobt.2016.00015/abstract (cf. p. 28, 30, 90).
- ALFAYAD, Samer, Fethi B. OUEZDOU, Faycal NAMOUN et Gordon GHENG (2011a). « High performance integrated electro-hydraulic actuator for robotics–Part I : Principle, prototype design and first experiments ». In : *Sensors and Actuators A : Physical* 169.1, p. 115-123 (cf. p. 18).
- (2011b). « High performance Integrated Electro-Hydraulic Actuator for robotics. Part II : Theoretical modelling, simulation, control & comparison with real measurements ». In : *Sensors and Actuators A : Physical* 169.1, p. 124-132 (cf. p. 18).
- ALVAREZ, Francisco J., Ana BENITO-GONZALEZ et Valerie C. SIEMBAB (2013). « Principles of interneuron development learned from Renshaw cells and the motoneuron recurrent inhibitory circuit ». In : *Annals of the New York Academy of Sciences* 1279.1, p. 22-31 (cf. p. 199, 201).
- AMARI, Shun-ichi (1977). « Dynamics of pattern formation in lateral-inhibition type neural fields ». In : *Biological cybernetics* 27.2, p. 77-87 (cf. p. 112).
- ANDERSEN, Birgitte Bo, Lise KORBO et Bente PAKKENBERG (1992). « A quantitative study of the human cerebellum with unbiased stereological techniques ». In : *Journal of Comparative Neurology* 326.4, p. 549-560 (cf. p. 60).
- ANDERSON, Michael L. (2010). « Neural reuse : A fundamental organizational principle of the brain ». In : *Behavioral and brain sciences* 33.4, p. 245-266 (cf. p. 59).

- ANDRY, Pierre, Philippe GAUSSIER, Sorin MOGA, Jean-Paul BANQUET et Jacqueline NADEL (2001). « Learning and communication via imitation : An autonomous robot perspective ». In : *Systems, Man and Cybernetics, Part A : Systems and Humans*, *IEEE Transactions on* 31.5, p. 431-442 (cf. p. 88, 113).
- ANDRY, Pierre, Philippe GAUSSIER et Jacqueline NADEL (2002). « Development of the First Sensory-Motor Stages : A Contribution to Imitation. » In : *From Animals to Animals 7 : Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior*. T. 7. MIT Press, p. 230 (cf. p. 100).
- (2005). « Autonomous learning and reproduction of complex sequences : a multimodal architecture for bootstrapping imitation games ». In : (cf. p. 3).
- ANDRY, Pierre, Philippe GAUSSIER, Jacqueline NADEL et Beat HIRSBRUNNER (2004). « Learning invariant sensorimotor behaviors : A developmental approach to imitation mechanisms ». In : *Adaptive behavior* 12.2, p. 117-140 (cf. p. 100, 113).
- ANSERMIN, Eva (2015). « Rapport de Master 2 ». Mém. de mast. Ensea / UCP (cf. p. 182).
- ANSERMIN, Eva, Ghiles MOSTAFAOUI, Nils BEAUSSÉ et Philippe GAUSSIER (2016). « Learning to synchronously imitate gestures using entrainment effect ». In : *International conference on simulation of adaptive behavior*. Springer, p. 219-231 (cf. p. 181, 185, 187, 189).
- ANSUINI, Caterina, Marco SANTELLO, Stefano MASSACCESI et Umberto CASTIELLO (2006). « Effects of end-goal on hand shaping ». In : *Journal of neurophysiology* 95.4, p. 2456-2465 (cf. p. 51).
- AOSAKI, Toshihiko, Minoru KIMURA et A. M. GRAYBIEL (1995). « Temporal and spatial characteristics of tonically active neurons of the primate's striatum ». In : *Journal of neurophysiology* 73.3, p. 1234-1252 (cf. p. 59).
- ARBIB, Michael A. (2003). *The handbook of brain theory and neural networks*. MIT press (cf. p. 90).
- ATKESON, Christopher G., Andrew W. MOORE et Stefan SCHAAAL (1997). « Locally weighted learning ». In : *Lazy learning*. Springer, p. 11-73 (cf. p. 53).
- ATKINSON, Janette, Oliver BRADDICK et Fleur BRADDICK (1974). « Acuity and contrast sensitivity of infant vision. » In : *Nature* (cf. p. 88).
- BACH-Y-RITA, Paul, Carter C. COLLINS, Frank A. SAUNDERS, Benjamin WHITE et Lawrence SCADDEN (1969). « Vision substitution by tactile image projection ». In : *Nature* (cf. p. 88).
- BAINBRIDGE, Lisanne (1983). « Ironies of automation ». In : *Automatica* 19.6, p. 775-779 (cf. p. 49).

- BANKS, Martin S. et Philip SALAPATEK (1981). « Infant pattern vision : A new approach based on the contrast sensitivity function ». In : *Journal of Experimental Child Psychology* 31.1, p. 1-45 (cf. p. 88).
- BARRON-ZAMBRANO, Jose Hugo, Cesar TORRES-HUITZIL et Bernard GIRAU (2010). « Hardware implementation of a CPG-based locomotion control for quadruped robots ». In : *International Conference on Artificial Neural Networks*. Springer, p. 276-285 (cf. p. 196).
- BARTNECK, Christoph, Dana KULIĆ, Elizabeth CROFT et Susana ZOGHBI (2009). « Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots ». In : *International journal of social robotics* 1.1, p. 71-81 (cf. p. 43).
- BAYER, Hannah M. et Paul W. GLIMCHER (2005). « Midbrain dopamine neurons encode a quantitative reward prediction error signal ». In : *Neuron* 47.1, p. 129-141 (cf. p. 131).
- BECCHIO, Cristina, Luisa SARTORI, Maria BULGHERONI et Umberto CASTIELLO (2008). « The case of Dr. Jekyll and Mr. Hyde : a kinematic study on social intention ». In : *Consciousness and cognition* 17.3, p. 557-564 (cf. p. 2, 51, 156-158).
- BEDNAR, James A. et K. I. CHRISTOPHER WILLIAMS (2016). « From Brain to Cognition via Computational Neuroscience ». In : MIT Press. Chap. Neural Maps : Their Function and Development (cf. p. 93).
- BEER, Randall D. (1995). « A dynamical systems perspective on agent-environment interaction ». In : *Artificial intelligence* 72.1, p. 173-215 (cf. p. 88).
- BEISER, David G., Sherwin E. HUA et James C. HOUK (1997). « Network models of the basal ganglia ». In : *Current opinion in neurobiology* 7.2, p. 185-190 (cf. p. 58).
- BELLMAN, Richard E. (1957). English. Sous la dir. de Princeton University PRESS. ISBN : 978-0691079516 (cf. p. 49, 53).
- BERNIERI, Frank J. et Robert ROSENTHAL (1991). « Interpersonal coordination : Behavior matching and interactional synchrony. » In : (cf. p. 1, 175).
- BERRET, Bastien, Christian DARLOT, Frédéric JEAN, Thierry POZZO, Charalambos PAPAXANTHIS et Jean-Paul GAUTHIER (2008). « The inactivation principle : mathematical solutions minimizing the absolute work and biological implications for the planning of arm movements ». In : *PLoS computational biology* 4.10, e1000194 (cf. p. 50).
- BERTHIER, Neil E. (1996). « Learning to reach : a mathematical model. » In : *Developmental psychology* 32.5, p. 811 (cf. p. 54, 112).
- BIENENSTOCK, Elie L., Leon N. COOPER et Paul W. MUNRO (1981). *Theory for the development of neuron selectivity : orientation specificity and binocular interaction in visual cortex*. Rapp. tech. DTIC Document (cf. p. 202).

- BIENENSTOCK, Elie L., Leon N. COOPER et Paul W. MUNRO (1982). « Theory for the development of neuron selectivity : orientation specificity and binocular interaction in visual cortex ». In : *Journal of Neuroscience* 2.1, p. 32-48 (cf. p. 209).
- BINGHAM, Geoffrey P. (2004). « A perceptually driven dynamical model of bimanual rhythmic movement (and phase perception) ». In : *Ecological Psychology* 16.1, p. 45-53 (cf. p. 177).
- BLANCHARD, Arnaud (2014). *Présentation humoristique du robot Tino 1*. URL : <https://youtu.be/C5CE84qt59k> (cf. p. 27).
- BLANCHARD, Arnaud et Jacqueline NADEL (2009). « Designing a turn-taking mechanism as a balance between familiarity and novelty ». In : *Proceedings of the Ninth International Conference on Epigenetic Robotics (EPIROB2009)*, p. 199-201 (cf. p. 3, 180).
- BOTTINI, Gabriella, H-O KARNATH, Giuseppe VALLAR, Roberto STERZI, Christopher D. FRITH, Richard S. J. FRACKOWIAK et Eraldo PAULESU (2001). « Cerebral representations for egocentric space ». In : *Brain* 124.6, p. 1182-1196 (cf. p. 89).
- BOUCENNA, Sofiane (2011). « De la reconnaissance des expressions faciales à une perception visuelle partagée : une architecture sensori-motrice pour amorcer un référencement social d'objets, de lieux ou de comportements ». Thèse de doct. (cf. p. 27).
- BOUCENNA, Sofiane, Philippe GAUSSIER, Pierre ANDRY et Laurence HAFEMEISTER (2010). « Imitation as a communication tool for online facial expression learning and recognition ». In : *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, p. 5323-5328 (cf. p. 88, 100, 113).
- BRAITENBERG, Valentino (1986). *Vehicles : Experiments in synthetic psychology*. MIT press (cf. p. 88).
- BRAITENBERG, Valentino et Roger P. ATWOOD (1958). « Morphological observations on the cerebellar cortex ». In : *Journal of Comparative Neurology* 109.1, p. 1-33 (cf. p. 60).
- BRAUD, Raphael, Ghiles MOSTAFAOUI, Ali KARAOUZENE et Philippe GAUSSIER (2014). « Simulating the Emergence of Early Physical and Social Interactions : A Developmental Route through Low Level Visuomotor Learning ». In : *From Animals to Animats 13*. Springer, p. 154-165 (cf. p. 91).
- BRAZHNIK, Elena, Fulva SHAH et James M. TEPPER (2008). « GABAergic afferents activate both GABAA and GABAB receptors in mouse substantia nigra dopaminergic neurons in vivo ». In : *Journal of Neuroscience* 28.41, p. 10386-10398 (cf. p. 133).
- BRITTAIN, John-Stuart et Peter BROWN (2014). « Oscillations and the basal ganglia : motor control and beyond ». In : *Neuroimage* 85, p. 637-647 (cf. p. 196).

- BROWN, Angela M. et Misao YAMAMOTO (1986). « Visual acuity in newborn and pre-term infants measured with grating acuity cards ». In : *American journal of ophthalmology* 102.2, p. 245-253 (cf. p. 88).
- BROWN, Joshua, Daniel BULLOCK et Stephen GROSSBERG (1999). « How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues ». In : *Journal of Neuroscience* 19.23, p. 10502-10511 (cf. p. 128).
- BULLER, A. J., John C. ECCLES et Rosamond M. ECCLES (1960). « Interactions between motoneurons and muscles in respect of the characteristic speeds of their responses ». In : *The Journal of physiology* 150.2, p. 417-439 (cf. p. 75).
- BULLOCK, Daniel et Stephen GROSSBERG (1989). « VITE and FLETE : Neural modules for trajectory formation and postural control ». In : *Advances in Psychology*. T. 62. Elsevier, p. 253-297 (cf. p. 113).
- BURNS, Yvonne, Michael O'CALLAGHAN, Brian MCDONELL et Yvonne ROGERS (2004). « Movement and motor development in ELBW infants at 1 year is related to cognitive and motor abilities at 4 years ». In : *Early human development* 80.1, p. 19-29 (cf. p. 64).
- BUSER, Pierre et Michel IMBERT (1982). *Neurophysiologie fonctionnelle : Psychophysiological sensorielle. II* (cf. p. 77).
- BUTTERWORTH, George (1999). *Neonatal imitation : Existence, mechanisms and motives*. Cambridge University Press (cf. p. 2, 100).
- CALINON, Sylvain, Florent GUENTER et Aude BILLARD (2007). « On learning, representing, and generalizing a task in a humanoid robot ». In : *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 37.2, p. 286-298 (cf. p. 53).
- CAPADAY, Charles, Christian ETHIER, Laurent BRIZZI, Attila SIK, Carl van VREESWIJK et Denis GINGRAS (2009). « On the nature of the intrinsic connectivity of the cat motor cortex : evidence for a recurrent neural network topology ». In : *Journal of neurophysiology* 102.4, p. 2131-2141 (cf. p. 117).
- CARPENTER, Gail A. et Stephen GROSSBERG (1987). « ART 2 : Self-organization of stable category recognition codes for analog input patterns ». In : *Applied optics* 26.23, p. 4919-4930 (cf. p. 91).
- CARTWBIGHT, M. L. (1960). « Balthazar van der Pol ». In : *Journal of the London Mathematical Society* 1.3, p. 367-376 (cf. p. 196).
- CASEY, B. J., Nim TOTTENHAM, Conor LISTON et Sarah DURSTON (2005). « Imaging the developing brain : what have we learned about cognitive development? » In : *Trends in cognitive sciences* 9.3, p. 104-110 (cf. p. 64).

- CASTILLO, Pablo E. (2012). « Presynaptic LTP and LTD of excitatory and inhibitory synapses ». In : *Cold Spring Harbor perspectives in biology* 4.2, a005728 (cf. p. 201, 206).
- CHANQUOY, Lucile et Isabelle NEGRO (2004). *Psychologie du développement*. Hachette Éducation (cf. p. 11).
- CHAO, Fei, Zhengshuai WANG, Changjing SHANG, Qinggang MENG, Min JIANG, Changle ZHOU et Qiang SHEN (2014). « A developmental approach to robotic pointing via human–robot interaction ». In : *Information sciences* 283, p. 288-303 (cf. p. 54, 100, 112).
- CHARTRAND, Tanya L. et John A. BARGH (1999). « The chameleon effect : the perception–behavior link and social interaction. » In : *Journal of personality and social psychology* 76.6, p. 893 (cf. p. 2).
- CHEVALIER, G. et J. M. DENIAU (1990). « Disinhibition as a basic process in the expression of striatal functions ». In : *Trends in neurosciences* 13.7, p. 277-280 (cf. p. 59).
- CHIEL, Hillel J. et Randall D. BEER (1997). « The brain has a body : adaptive behavior emerges from interactions of nervous system, body and environment ». In : *Trends in neurosciences* 20.12, p. 553-557 (cf. p. 88).
- CLAYTON, Martin, Rebecca SAGER et Udo WILL (2005). « In time with the music : the concept of entrainment and its significance for ethnomusicology. » In : *European meetings in ethnomusicology*. T. 11. Romanian Society for Ethnomusicology, p. 1-82 (cf. p. 175).
- COESMANS, Michiel, John T. WEBER, Chris I. DE ZEEUW et Christian HANSEL (2004). « Bidirectional parallel fiber plasticity in the cerebellum under climbing fiber control ». In : *Neuron* 44.4, p. 691-700 (cf. p. 210).
- COLLINS, Steve, Andy RUINA, Russ TEDRAKE et Martijn WISSE (2005). « Efficient bipedal robots based on passive-dynamic walkers ». In : *Science* 307.5712, p. 1082-1085 (cf. p. 2, 12).
- CONWAY, B. A., H. HULTBORN et O. KIEHN (1987). « Proprioceptive input resets central locomotor rhythm in the spinal cat ». In : *Experimental Brain Research* 68.3, p. 643-656 (cf. p. 191).
- CONWIT, R. A., D. STASHUK, B. TRACY, M. MCHUGH, W. F. BROWN et E. J. METTER (1999). « The relationship of motor unit size, firing rate and force ». In : *Clinical neurophysiology* 110.7, p. 1270-1275 (cf. p. 78).
- COSTILL, D. L., W. J. FINK et M. L. POLLOCK (1976). « Muscle fiber composition and enzyme activities of elite distance runners. » In : *Medicine and science in sports* 8.2, p. 96-100 (cf. p. 75).

- CRICK, Christopher, Matthew MUNZ et Brian SCASSELLATI (2006). « Synchronization in social tasks : Robotic drumming ». In : *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, p. 97-102 (cf. p. 180, 181).
- CRITTENDEN, Jill R., Paul W. TILLBERG, Michael H. RIAD, Yasuyuki SHIMA, Charles R. GERFEN, Jeffrey CURRY, David E. HOUSMAN, Sacha B. NELSON, Edward S. BOYDEN et Ann M. GRAYBIEL (2016). « Striosome–dendron bouquets highlight a unique striatonigral circuit targeting dopamine-containing neurons ». In : *Proceedings of the National Academy of Sciences* 113.40, p. 11318-11323 (cf. p. 128).
- CROSS, Rod (2011). « A double pendulum model of tennis strokes ». In : *American Journal of Physics* 79.5, p. 470-476 (cf. p. 159).
- D'ANGELO, Egidio (2008). « The critical role of Golgi cells in regulating spatio-temporal integration and plasticity at the cerebellum input stage ». In : *Frontiers in neuroscience* 2, p. 8 (cf. p. 199).
- DECETY, Jean (2002). « Naturaliser l'empathie ». In : *L'encéphale* 28.1, p. 9-20 (cf. p. 100, 118).
- DEGALLIER, Sarah, Cristina P. SANTOS, Ludovic RIGHETTI et Auke IJSPEERT (2006). « Movement generation using dynamical systems : a humanoid robot performing a drumming task ». In : *2006 6th IEEE-RAS International Conference on Humanoid Robots*. IEEE, p. 512-517 (cf. p. 189).
- DHAWALE, Ashesh K., Maurice A. SMITH et Bence P. ÖLVE CZKY (2017). « The role of variability in motor learning ». In : *Annual review of neuroscience* 40, p. 479-498 (cf. p. 86).
- DI GIOVANNI, Giuseppe, Vincenzo DI MATTEO et Ennio ESPOSITO (2009). *Birth, life and death of dopaminergic neurons in the substantia nigra*. T. 73. Springer (cf. p. 129).
- DOLLÉ, Laurent, Denis SHEYNIKHOVICH, Benoît GIRARD, Ricardo CHAVARRIAGA et Agnès GUILLOT (2010). « Path planning versus cue responding : a bio-inspired model of switching between navigation strategies ». In : *Biological cybernetics* 103.4, p. 299-317 (cf. p. 59).
- DOUTRELOUX, Jean-Paul (1992). « Contribution à l'étude du muscle : évolution de l'architecture du muscle penne au cours du développement post-embryonnaire : exemple des muscles crural et droit antérieur du lapin (*Oryctolagus cuniculus*) ». Thèse de doct. Bordeaux 2 (cf. p. 78).
- DOYA, Kenji (1999). « What are the computations of the cerebellum, the basal ganglia and the cerebral cortex ? » In : *Neural networks* 12.7-8, p. 961-974 (cf. p. 59, 131).
- (2000). « Complementary roles of basal ganglia and cerebellum in learning and motor control ». In : *Current opinion in neurobiology* 10.6, p. 732-739 (cf. p. 59, 60, 127, 133).

- DOYA, Kenji et Shuji YOSHIZAWA (1989). « Adaptive neural oscillator using continuous-time back-propagation learning ». In : *Neural Networks* 2.5, p. 375-385 (cf. p. 181).
- (1991). « Adaptive synchronization of neural and physical oscillators ». In : *NIPS*, p. 109-116 (cf. p. 199).
- DUDCHENKO, Paul A. (2015). « Spatial cognition : A Tabula rasa for the sense of direction ». In : *Current Biology* 25.4, R143-R144 (cf. p. 90).
- ECKHORN, Reinhard (1994). « Oscillatory and non-oscillatory synchronizations in the visual cortex and their possible roles in associations of visual features ». In : *Progress in brain research*. T. 102. Elsevier, p. 405-426 (cf. p. 196).
- ECKHORN, Reinhard, Roman BAUER, Wolfgang JORDAN, Michael BROSCHE, Wolfgang KRUSE, Matthias MUNK et H. J. REITBOECK (1988). « Coherent oscillations : A mechanism of feature linking in the visual cortex? ». In : *Biological cybernetics* 60.2, p. 121-130 (cf. p. 196).
- FERREY, Anne E., Tyler J. BURLEIGH et Mark J. FENSKE (2015). « Stimulus-category competition, inhibition, and affective devaluation : a novel account of the uncanny valley ». In : *Frontiers in psychology* 6, p. 249 (cf. p. 16).
- FIELD, Tiffany M., Robert WOODSON, Reena GREENBERG et Debra COHEN (1982). « Discrimination and imitation of facial expression by neonates ». In : *Science* 218.4568, p. 179-181 (cf. p. 100).
- FIX, Jérémy, Nicolas ROUGIER et Frédéric ALEXANDRE (2011). « A dynamic neural field approach to the covert and overt deployment of spatial attention ». In : *Cognitive Computation* 3.1, p. 279-293 (cf. p. 113).
- FLASH, Tamar et Neville HOGAN (1985). « The coordination of arm movements : an experimentally confirmed mathematical model ». In : *Journal of neuroscience* 5.7, p. 1688-1703 (cf. p. 49).
- FLORES, Carmen E. et Pablo MÉNDEZ (2014). « Shaping inhibition : activity dependent structural plasticity of GABAergic synapses ». In : *Frontiers in cellular neuroscience* 8 (cf. p. 199, 202).
- FUJIYAMA, Fumino, Jaerin SOHN, Takashi NAKANO, Takahiro FURUTA, Kouichi C. NAKAMURA, Wakoto MATSUDA et Takeshi KANEKO (2011). « Exclusive and common targets of neostriatofugal projections of rat striosome neurons : a single neuron-tracing study using a viral vector ». In : *European Journal of Neuroscience* 33.4, p. 668-677 (cf. p. 133).
- FUKUYORI, Ipppei, Yutaka NAKAMURA, Yoshio MATSUMOTO et Hiroshi ISHIGURO (2008). « Flexible control mechanism for multi-dof robotic arm based on biological fluctuation ». In : *International Conference on Simulation of Adaptive Behavior*. Springer, p. 22-31 (cf. p. 113).

- GALLESE, Vittorio et Alvin GOLDMAN (1998). « Mirror neurons and the simulation theory of mind-reading ». In : *Trends in cognitive sciences* 2.12, p. 493-501 (cf. p. 100, 118).
- GALOUZ, Zinedine (2018). « Apprentissage par renforcement d'une trajectoire avec un bras de robot humanoïde ». Mém. de mast. Université de Cergy-Pontoise, Laboratoire ETIS (cf. p. 159).
- GAN, Zhenyu, Thomas WIESTNER, Michael A. WEISHAUPT, Nina M. WALDERN et C. DAVID REMY (2016). « Passive dynamics explain quadrupedal walking, trotting, and tölting ». In : *Journal of computational and nonlinear dynamics* 11.2 (cf. p. 190).
- GAUSSIER, Philippe, Cédric JOULAIN, Stéphane ZREHEN, Jean-Paul BANQUET et Arnaud REVEL (1997). « Visual navigation in an open environment without map ». In : *Intelligent Robots and Systems, 1997. IROS'97., Proceedings of the 1997 IEEE/RSJ International Conference on*. T. 2. IEEE, p. 545-550 (cf. p. 88, 113).
- GAUSSIER, Philippe, Sorin MOGA, Mathias QUOY et Jean-Paul BANQUET (1998). « From perception-action loops to imitation processes : A bottom-up approach of learning by imitation ». In : *Applied Artificial Intelligence* 12.7-8, p. 701-727 (cf. p. 3, 113).
- GAUSSIER, Philippe et Stéphane ZREHEN (1994). « Avoiding the world model trap : An acting robot does not need to be so smart! ». In : *Robotics and Computer-Integrated Manufacturing* 11.4, p. 279-286 (cf. p. 3, 88).
- (1995). « PerAc : A neural architecture to control artificial animals ». In : *Robotics and Autonomous Systems* 16.2-4. Moving the Frontiers between Robotics and Biology, p. 291-320. ISSN : 0921-8890. DOI : [http://dx.doi.org/10.1016/0921-8890\(95\)00052-6](http://dx.doi.org/10.1016/0921-8890(95)00052-6). URL : <http://www.sciencedirect.com/science/article/pii/S0921889095000526> (cf. p. 3, 88, 100, 113).
- GAUTHIER, Jean-Paul, Bastien BERRET et Frédéric JEAN (2010). « A biomechanical inactivation principle ». In : *Proceedings of the Steklov Institute of Mathematics* 268.1, p. 93-116 (cf. p. 50).
- GELLER, Tom (2008). « Overcoming the uncanny valley ». In : *IEEE computer graphics and applications* 28.4, p. 11-17 (cf. p. 15).
- GEORGIU, Ioanna, Cristina BECCHIO, Scott GLOVER et Umberto CASTIELLO (2007). « Different action patterns for cooperative and competitive behaviour ». In : *Cognition* 102.3, p. 415-433 (cf. p. 2, 51, 144, 156).
- GEORGOPOULOS, Apostolos P., Andrew B. SCHWARTZ et Ronald E. KETTNER (1986). « Neuronal population coding of movement direction ». In : *Science* 233.4771, p. 1416-1419 (cf. p. 58, 112).

- GERGELY, Anna, Anna B. COMPTON, Ruth C. NEWBERRY et  MIKLSI (2016). « Social interaction with an “Unidentified Moving Object” elicits A-not-B error in domestic dogs ». In : *PloS one* 11.4, e0151600 (cf. p. 112).
- GEYER, Hartmut et Hugh HERR (2010). « A muscle-reflex model that encodes principles of legged mechanics produces human walking dynamics and muscle activities ». In : *IEEE Transactions on neural systems and rehabilitation engineering* 18.3, p. 263-273 (cf. p. 190).
- GIRARD, Benot, Vincent CUZIN, Agns GUILLOT, Kevin N. GURNEY et Tony J PRES-COTT (2003). « A basal ganglia inspired model of action selection evaluated in a robotic survival task ». In : *Journal of integrative neuroscience* 2.02, p. 179-200 (cf. p. 59, 125).
- (2002). « Comparing a bioinspired robot action selection mechanism with winner-takes-al ». In : *From Animals to Animats 7. Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior*. Sous la dir. de The MIT Press CAMBRIDGE (cf. p. 59, 125).
- GLASS, Leon (2001). « Synchronization and rhythmic processes in physiology ». In : *Nature* 410.6825, p. 277 (cf. p. 175).
- GLIMCHER, Paul W. (2011). « Understanding dopamine and reinforcement learning : the dopamine reward prediction error hypothesis ». In : *Proceedings of the National Academy of Sciences* 108.Supplement 3, p. 15647-15654 (cf. p. 134).
- GRAND, Caroline, Ghiles MOSTAFAOUI, Syed Khursheed HASNAIN et Philippe GAUS-SIER (2013). « Combining Synchrony and shape detection to sustain the robot focus of attention on a selected human partner ». In : *Development and Learning and Epigenetic Robotics (ICDL), 2013 IEEE Third Joint International Conference on*. IEEE, p. 1-2 (cf. p. 91).
- GRAND, Caroline, Ghils MOSTAFAOUI, Syed Khursheed HASNAIN et Philippe GAUS-SIER (2014). « Synchrony detection as a reinforcement signal for learning : Application to human robot interaction ». In : *Procedia-Social and Behavioral Sciences* 126, p. 82-91 (cf. p. 3, 181, 185).
- GREEN, Robert D., Karl F. MACDORMAN, Chin-Chang HO et Sandosh VASUDEVAN (2008). « Sensitivity to the proportions of faces that vary in human likeness ». In : *Computers in Human Behavior* 24.5, p. 2456-2474 (cf. p. 15).
- GRILLNER, Sten, Peter WALLEN, Lennart BRODIN et Anders LANSNER (1991). « Neuronal network generating locomotor behavior in lamprey : circuitry, transmitters, membrane properties, and simulation ». In : *Annual review of neuroscience* 14.1, p. 169-199 (cf. p. 196).
- GROSSBERG, Stephen (1987). « Competitive learning : From interactive activation to adaptive resonance ». In : *Cognitive science* 11.1, p. 23-63 (cf. p. 58).

- GROSSBERG, Stephen et Daniel LEVINE (1975). « Some developmental and attentional biases in the contrast enhancement and short term memory of recurrent neural networks ». In : *Journal of Theoretical Biology* 53.2, p. 341-380 (cf. p. 91, 92).
- GROSSBERG, Stephen et David SOMERS (1991). « Synchronized oscillations during cooperative feature linking in a cortical model of visual perception ». In : *Neural Networks* 4.4, p. 453-466 (cf. p. 196).
- GU, Shixiang, Ethan HOLLY, Timothy LILICRAP et Sergey LEVINE (2017). « Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates ». In : *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE, p. 3389-3396 (cf. p. 11).
- GURNEY, Kevin, Tony J. PRESCOTT et Peter REDGRAVE (2001). « A computational model of action selection in the basal ganglia. I. A new functional anatomy ». In : *Biological cybernetics* 84.6, p. 401-410 (cf. p. 59).
- GUYTON, Arthur C. et John E. HALL (2006). « Golgi Tendon Reflex ». In : *Textbook of medical physiology 11th ed.* Sous la dir. d'Elsevier SAUNDERS (cf. p. 78).
- GWIAZDA, Jane, Sarah BRILL, Indra MOHINDRA et Richard HELD (1980). « Preferential looking acuity in infants from two to fifty-eight weeks of age. » In : *Optometry & Vision Science* 57.7, p. 428-432 (cf. p. 88).
- HAJNAL, Alen, Michael J. RICHARDSON, Steven J. HARRISON et Richard C. SCHMIDT (2012). « Location but not amount of stimulus occlusion influences the stability of visuomotor coordination ». In : *Experimental brain research* 221.3, p. 351-355 (cf. p. 177).
- HARI, Riitta et Miiamaaria V. KUJALA (2009). « Brain basis of human social interaction : from concepts to brain imaging ». In : *Physiological reviews* 89.2, p. 453-479 (cf. p. 1, 175).
- HARRIS, Christopher M. et Daniel M. WOLPERT (1998). « Signal-dependent noise determines motor planning ». In : *Nature* 394.6695, p. 780 (cf. p. 50).
- HASNAIN, Syed Khursheed (2014). « Synchronisation et coordination interpersonnelle dans l'interaction Homme-robot ». Thèse de doct. Cergy-Pontoise (cf. p. 187, 189).
- HASNAIN, Syed Khursheed, Ghiles MOSTAFAOUI, Caroline GRAND et Philippe GAUSSIER (2014). « Synchrony based side by side walking ». In : (cf. p. 3, 181).
- HASNAIN, Syed Khursheed, Ghiles MOSTAFAOUI, Robin SALESSE, Ludovic MARIN et Philippe GAUSSIER (2013). « Intuitive human robot interaction based on unintentional synchrony : a psycho-experimental study ». In : *2013 IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL)*. IEEE, p. 1-7 (cf. p. 3, 180, 185).

- HATZE, Herbert et Johan D. BUYS (1977). « Energy-optimal controls in the mammalian neuromuscular system ». In : *Biological cybernetics* 27.1, p. 9-20 (cf. p. 50).
- HECKMAN, C. J. et Roger M. ENOKA (2012). « Motor unit ». In : *Comprehensive physiology* 2.4, p. 2629-2682 (cf. p. 75).
- HEIDER, Fritz et Marianne SIMMEL (1944). « An experimental study of apparent behavior ». In : *The American journal of psychology* 57.2, p. 243-259 (cf. p. 17).
- HENNEMAN, Elwood, George SOMJEN et David O. CARPENTER (1965). « Functional significance of cell size in spinal motoneurons ». In : *Journal of neurophysiology* 28.3, p. 560-580 (cf. p. 78).
- HERCULANO-HOUZEL, Suzana (2009). « The human brain in numbers : a linearly scaled-up primate brain ». In : *Frontiers in human neuroscience* 3, p. 31 (cf. p. 60).
- HERSCH, Micha et Aude G. BILLARD (2006). « A biologically-inspired controller for reaching movements ». In : *The First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics, 2006. BioRob 2006*. IEEE, p. 1067-1072 (cf. p. 113).
- HERZFELD, David J. et Reza SHADMEHR (2014). « Motor variability is not noise, but grist for the learning mill ». In : *nature neuroscience* 17.2, p. 149 (cf. p. 86).
- HILL, Archibald Vivian (1938). « The heat of shortening and the dynamic constants of muscle ». In : *Proceedings of the Royal Society of London. Series B-Biological Sciences* 126.843, p. 136-195 (cf. p. 79).
- (1949). « The abrupt transition from rest to activity in muscle ». In : *Proceedings of the Royal Society of London. Series B-Biological Sciences* 136.884, p. 399-420 (cf. p. 79).
- (1950). « The series elastic component of muscle ». In : *Proceedings of the Royal Society of London. Series B, Biological Sciences*, p. 273-280 (cf. p. 79).
- (1970). *First and last experiments in muscle mechanics*. Cambridge University Press (cf. p. 79).
- HOGAN, Neville (1984a). « Adaptive control of mechanical impedance by coactivation of antagonist muscles ». In : *IEEE Transactions on automatic control* 29.8, p. 681-690 (cf. p. 49).
- (1984b). « Impedance control : An approach to manipulation ». In : *1984 American control conference*. IEEE, p. 304-313 (cf. p. 50).
- HORN, Berthold K. P. et Brian G. SCHUNCK (1981). « Determining optical flow ». In : *Artificial intelligence* 17.1-3, p. 185-203 (cf. p. 187).
- HOUK, James C. et James L. ADAMS (1995). « 13 A Model of How the Basal Ganglia Generate and Use Neural Signals That ». In : *Models of information processing in the basal ganglia*, p. 249 (cf. p. 54).

- HOUK, James C., Joel L. DAVIS et David G. BEISER (1995). *Models of information processing in the basal ganglia*. MIT press (cf. p. 133).
- IJSPEERT, Auke Jan (2008). « Central pattern generators for locomotion control in animals and robots : a review ». In : *Neural networks* 21.4, p. 642-653 (cf. p. 196).
- IJSPEERT, Auke Jan, Alessandro CRESPI, Dimitri RYCZKO et Jean-Marie CABELGUEN (2007). « From swimming to walking with a salamander robot driven by a spinal cord model ». In : *science* 315.5817, p. 1416-1420 (cf. p. 196).
- IJSPEERT, Auke Jan, Jun NAKANISHI et Stefan SCHAAAL (2002). « Learning rhythmic movements by demonstration using nonlinear oscillators ». In : *Proceedings of the ieee/rsj int. conference on intelligent robots and systems (iros2002)*. CONF, p. 958-963 (cf. p. 181).
- IOSSIFIDIS, Ioannis et G. SCHONER (2004). « Autonomous reaching and obstacle avoidance with the anthropomorphic arm of a robotic assistant using the attractor dynamics approach ». In : *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004*. T. 5. IEEE, p. 4295-4300 (cf. p. 113).
- ISSARTEL, Johann, Ludovic MARIN et Marielle CADOPI (2007). « Unintended interpersonal co-ordination : “can we march to the beat of our own drum?” ». In : *Neuroscience letters* 411.3, p. 174-179 (cf. p. 175, 180).
- ITO, Makoto et Kenji DOYA (2015). « Parallel representation of value-based and finite state-based strategies in the ventral and dorsal striatum ». In : *PLoS computational biology* 11.11, e1004540 (cf. p. 133).
- IVALDI, Serena, Olivier SIGAUD, Bastien BERRET et Francesco NORI (2012). « From humans to humanoids : the optimal control framework ». In : *Paladyn, Journal of Behavioral Robotics* 3.2, p. 75-91 (cf. p. 49).
- IZAWA, Jun et Reza SHADMEHR (2011). « Learning from sensory and reward prediction errors during motor adaptation ». In : *PLoS computational biology* 7.3, e1002012 (cf. p. 86).
- IZHIKEVICH, Eugene M. et Niraj S. DESAI (2003). « Relating stdp to bcm ». In : *Neural computation* 15.7, p. 1511-1523 (cf. p. 210).
- JACOB, Pierre et Marc JEANNEROD (2005). « The motor theory of social cognition : a critique ». In : *Trends in cognitive sciences* 9.1, p. 21-25 (cf. p. 51, 155).
- JENCKINS, Ward (2004). « The Polar Express : A Virtual Train Wreck ». In : *Ward-O-Matic blog*. URL : http://wardomatic.blogspot.com/2004/12/polar-express-virtual-train-wreck_18.html (cf. p. 15).
- JOEL, Daphna, Yael NIV et Eytan RUPPIN (2002). « Actor-critic models of the basal ganglia : New anatomical and computational perspectives ». In : *Neural networks* 15.4-6, p. 535-547 (cf. p. 134).

- JOHNSON, Jeffrey S., John P. SPENCER, Steven J. LUCK et Gregor SCHÖNER (2009). « A dynamic neural field model of visual working memory and change detection ». In : *Psychological science* 20.5, p. 568-577 (cf. p. 113).
- JOUAITI, Melanie et Patrick HENAFF (2018). « CPG-based Controllers can Trigger the Emergence of Social Synchrony in Human-Robot Interactions ». In : *2018 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*. IEEE, p. 71-76 (cf. p. 181, 189).
- KACEM, Linda (2011). « Connectivité anatomique des ganglions de la base : développements méthodologiques et application aux troubles moteurs ». Thèse de doct. (cf. p. 126).
- KALVERAM, Karl Theodor (1998). « A neural oscillator model learning given trajectories, or how an allo-imitation algorithm can be implemented into a motor controller ». In : *Motor control and human skill : A multi-disciplinary perspective*, p. 127-140 (cf. p. 181).
- KAMEN, Gary et Carlo J. DE LUCA (1989). « Unusual motor unit firing behavior in older adults ». In : *Brain research* 482.1, p. 136-140 (cf. p. 78).
- KASUGA, Tomofumi et Minoru HASHIMOTO (2005). « Human-robot handshaking using neural oscillators ». In : *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*. IEEE, p. 3802-3807 (cf. p. 180).
- KAY, Bruce A., Elliot L. SALTZMAN et J. A. Scott KELSO (1991). « Steady-state and perturbed rhythmical movements : A dynamical analysis. » In : *Journal of Experimental Psychology : Human Perception and Performance* 17.1, p. 183 (cf. p. 176).
- KELSO, JA Scott (1995). *Dynamic patterns : The self-organization of brain and behavior*. MIT press (cf. p. 176).
- KIEHN, Ole et Simon J. B. BUTT (2003). « Physiological, anatomical and genetic identification of CPG neurons in the developing mammalian spinal cord ». In : *Progress in neurobiology* 70.4, p. 347-361 (cf. p. 196).
- KILNER, James M., Yves PAULIGNAN et Sarah-Jayne BLAKEMORE (2003). « An interference effect of observed biological movement on action ». In : *Current biology* 13.6, p. 522-525 (cf. p. 179).
- KIM, Kangsoo, Arjun NAGENDRAN, Jeremy N. BAILENSON, Andrew RAIJ, Gerd BRUDER, Myungho LEE, Ryan SCHUBERT, Xin YAN et Gregory F. WELCH (2017). « a large-scale study of surrogate Physicality and gesturing on human-surrogate interactions in a Public space ». In : *Frontiers in Robotics and AI* 4, p. 32 (cf. p. 13).
- KNIGHT, Rob, Ulrich NEHMZOW et Colchester Co SQ (2002). « Walking robots-a survey and a research proposal ». In : *Univ. Essex, Essex, UK, Technical Report CSM-375* (cf. p. 12).

- KOBER, Jens et Jan PETERS (2010). « Imitation and reinforcement learning ». In : *IEEE Robotics & Automation Magazine* 17.2, p. 55-62 (cf. p. 54, 100, 112).
- KOHONEN, Teuvo (1982). « Self-organized formation of topologically correct feature maps ». In : *Biological cybernetics* 43.1, p. 59-69 (cf. p. 58, 90).
- KUGIUMUTZAKIS, Giannis (1999). « Genesis and development of early infant mimesis to facial and vocal models. » In : (cf. p. 100).
- KUKULKA, Carl G. et H. Peter CLAMANN (1981). « Comparison of the recruitment and discharge properties of motor units in human brachial biceps and adductor pollicis during isometric contractions ». In : *Brain research* 219.1, p. 45-55 (cf. p. 79).
- KUROKAWA, H., Chun Ying HO et S. MORI (1997). « A local connected neural oscillator network for sequential character segmentation ». In : *Neural Networks, 1997., International Conference on*. T. 2. IEEE, p. 838-843 (cf. p. 199).
- KUTTA, Wilhelm (1901). « Beitrag zur näherungsweise Integration totaler Differentialgleichungen ». In : (cf. p. 162).
- LANGEVIN, Paul (1908). « Sur la théorie du mouvement brownien ». français. In : SCIENCES, Académie des. *Comptes-Rendus de l'Académie des Sciences*. 146, p. 530-532. URL : <https://gallica.bnf.fr/ark:/12148/bpt6k3100t/f530> (cf. p. 113).
- LARGE, Edward W. (2000). « On synchronizing movements to music ». In : *Human movement science* 19.4, p. 527-566 (cf. p. 175).
- LE MASSON, Gwendal, Sylvie RENAUD-LE MASSON, Damien DEBAY et Thierry BAL (2002). « Feedback inhibition controls spike transfer in hybrid thalamic circuits ». In : *Nature* 417.6891, p. 854 (cf. p. 196).
- LEISMAN, Gerry, Orit BRAUN-BENJAMIN et Robert MELILLO (2014). « Cognitive-motor interactions of the basal ganglia in development ». In : *Frontiers in systems neuroscience* 8, p. 16 (cf. p. 64).
- LEMON, Neal et Denise MANAHAN-VAUGHAN (2006). « Dopamine D1/D5 receptors gate the acquisition of novel information through hippocampal long-term potentiation and long-term depression ». In : *Journal of Neuroscience* 26.29, p. 7723-7729 (cf. p. 129).
- LESUEUR-GRAND, Caroline (2017). « Le Robot mobile compagnon : De l'apprentissage interactif vers un modèle d'IHM intuitive. » Thèse de doct. Université de Cergy Pontoise (cf. p. 3, 181, 185, 187, 189).
- LEVIEN, R. B. et S. M. TAN (1993). « Double pendulum : An experiment in chaos ». In : *American Journal of Physics* 61.11, p. 1038-1044 (cf. p. 159).
- LEWKOWICZ, Daniel, Yvonne DELEVOYE-TURRELL, David BAILLY, Pierre ANDRY et Philippe GAUSSIER (2013). « Reading motor intention through mental imagery ». In : *Adaptive Behavior* 21.5, p. 315-327 (cf. p. 3, 51, 156).

- LIEN, Cheng-Chang, Yangling MU, Mariana VARGAS-CABALLERO et Mu-ming POO (2006). « Visual stimuli-induced LTD of GABAergic synapses mediated by presynaptic NMDA receptors ». In : *Nature neuroscience* 9.3, p. 372-380 (cf. p. 199, 202).
- LIU, Chengju, Yifei CHEN, Jiaqi ZHANG et Qijun CHEN (2009). « CPG driven locomotion control of quadruped robot ». In : *2009 IEEE International Conference on Systems, Man and Cybernetics*. IEEE, p. 2368-2373 (cf. p. 196).
- LUCAS, Bruce D., Takeo KANADE et al. (1981). « An iterative image registration technique with an application to stereo vision ». In : (cf. p. 187).
- LUNGARELLA, Max, Giorgio METTA, Rolf PFEIFER et Giulio SANDINI (2003). « Developmental robotics : a survey ». In : *Connection science* 15.4, p. 151-190 (cf. p. 3).
- MACDORMAN, Karl F., Robert D. GREEN, Chin-Chang HO et Clinton T. KOCH (2009). « Too real for comfort ? Uncanny responses to computer generated faces ». In : *Computers in human behavior* 25.3, p. 695-710 (cf. p. 15).
- MADIEU, Emmanuel (2015a). « Mémoire de Master 1 : Mesure de l'effet d'entraînement lors d'interactions non intentionnelles entre humain et robot ». Mém. de mast. Euromov, université de Montpellier (cf. p. 182).
- (2015b). « Mémoire de Master 2 : Effet d'une perturbation sur les synchronisations non-intentionnelles entre humain et robot ». Mém. de mast. Euromov, université de Montpellier (cf. p. 182).
- MADIEU, Emmanuel, Ghiles MOSTAFAOUI, R. N. SALESSE, Eva ANSERMIN, Nils BEAUSSÉ, Philippe GAUSSIER, B. G. BARDY et Ludovic MARIN (2017). *Frequency properties allowing human-robot unintentional motor synchronization ?* Joint Action Meeting 2017 (cf. p. 184).
- MALSBURG, Christoph von der et Joachim BUHMANN (1992). « Sensory segmentation with coupled neural oscillators ». In : *Biological cybernetics* 67.3, p. 233-242 (cf. p. 196, 200).
- MAPELLI, Jonathan, Daniela GANDOLFI, Antonietta VILELLA, Michele ZOLI et Albertino BIGIANI (2016). « Heterosynaptic GABAergic plasticity bidirectionally driven by the activity of pre-and postsynaptic NMDA receptors ». In : *Proceedings of the National Academy of Sciences* 113.35, p. 9898-9903 (cf. p. 202, 211).
- MARATOS, Olga (1973). « The Origin and Development of Imitation in the First Six Months of Life. » In : *ERIC* (cf. p. 2, 100).
- MARDER, Eve et Dirk BUCHER (2001). « Central pattern generators and the control of rhythmic movements ». In : *Current biology* 11.23, R986-R996 (cf. p. 196).
- MARIN, Ludovic, Johann ISSARTEL et Thierry CHAMINADE (2009). « Interpersonal motor coordination : From human-human to human-robot interactions ». In : *Interaction Studies* 10.3, p. 479-504 (cf. p. 178, 179).

- MARQUIS, Paul J., Nancy A. RUIZ, Mary S. LUNDY et Robert G. DILLARD (1984). « Retention of primitive reflexes and delayed motor development in very low birth weight infants. » In : *Journal of Developmental and Behavioral Pediatrics* (cf. p. 64).
- MARR, David et W. Thomas THACH (1991). « A theory of cerebellar cortex ». In : *From the Retina to the Neocortex*. Springer, p. 11-50 (cf. p. 61, 62).
- MARSH, Kerry L., Michael J. RICHARDSON et Richard C. SCHMIDT (2009). « Social connection through joint action and interpersonal coordination ». In : *Topics in Cognitive Science* 1.2, p. 320-339 (cf. p. 175).
- MATIGNON, Laetitia, Guillaume J. LAURENT et Nadine LE FORT-PIAT (2006). « Improving reinforcement learning speed for robot control ». In : *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, p. 3172-3177 (cf. p. 12).
- MCGEER, Tad (1990). « Passive walking with knees ». In : *Proceedings., IEEE International Conference on Robotics and Automation*. IEEE, p. 1640-1645 (cf. p. 2, 12).
- MCGEER, Tad et al. (1990). « Passive dynamic walking ». In : *I. J. Robot Res.* 9.2, p. 62-82 (cf. p. 190).
- MCNEILL, William Hardy (1997). *Keeping together in time*. Harvard University Press (cf. p. 175).
- MCPHEDRAN, Alexander M., Raymond B. WUERKER et Elwood HENNEMAN (1965). « Properties of motor units in a homogeneous red muscle (soleus) of the cat ». In : *Journal of neurophysiology* 28.1, p. 71-84 (cf. p. 78).
- MELILLO, Robert (2011). « Primitive reflexes and their relationship to delayed cortical maturation, underconnectivity and functional disconnection in childhood neurobehavioral disorders ». In : *Funct. Neurol. Rehabil. Ergon* 1, p. 279-314 (cf. p. 64).
- MELILLO, Robert et Gerry LEISMAN (2010). *Neurobehavioral disorders of childhood : An evolutionary perspective*. Springer Science & Business Media (cf. p. 64).
- MELNYK, Artem (2014). « Perfectionnement des algorithmes de contrôle-commande des robots manipulateur électriques en interaction physique avec leur environnement par une approche bio-inspirée ». Thèse de doct. (cf. p. 180).
- MELNYK, Artem, Patrick HENAFF, Viacheslav KHOMENKO et Volodymyr BORYSENKO (2014). « Sensor network architecture to measure characteristics of a handshake between humans ». In : *2014 IEEE 34th International Scientific Conference on Electronics and Nanotechnology (ELNANO)*. IEEE, p. 264-268 (cf. p. 180).
- MELNYK, Artem, Viacheslav KHOMENKO, Vladimir BORYSENKO et Patrick HENAFF (2013). « Physical human-robot interaction in the handshaking case : learning of rhythmicity using oscillators neuron ». In : (cf. p. 180).
- MELTZOFF, Andrew N. et M. Keith MOORE (1983). « Newborn infants imitate adult facial gestures ». In : *Child development*, p. 702-709 (cf. p. 100).

- MELTZOFF, Andrew N., M. Keith MOORE et al. (1977). « Imitation of facial and manual gestures by human neonates ». In : *Science* 198.4312, p. 75-78 (cf. p. 100).
- MICHALOWSKI, Marek P., Selma SABANOVIC et Hideki KOZIMA (2007). « A dancing robot for rhythmic social interaction ». In : *2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, p. 89-96 (cf. p. 180).
- MILLER, George A. et Kristina OLLS (1980). *The social foundations of language and thought : Essays in honor of Jerome S. Bruner*. WW Norton & Company Incorporated (cf. p. 175).
- MIROLLO, Renato E. et Steven H. STROGATZ (1990). « Synchronization of pulse-coupled biological oscillators ». In : *SIAM Journal on Applied Mathematics* 50.6, p. 1645-1662 (cf. p. 175).
- MIYAKE, Yoshihiro (2009). « Interpersonal synchronization of body motion and the Walk-Mate walking support robot ». In : *IEEE Transactions on Robotics* 25.3, p. 638-644 (cf. p. 189).
- MONTAGUE, P. Read, Peter DAYAN, Steven J. NOWLAN, A. POUGET et T. J. SEJNOWSKI (1993). « Using aperiodic reinforcement for directed self-organization during development ». In : *Advances in neural information processing systems*, p. 969-976 (cf. p. 59, 131).
- MORI, Masahiro (1970). « Bukimi no tani [the uncanny valley] ». In : *Energy* 7, p. 33-35 (cf. p. 15).
- MOUSTAFA, Ahmed A., Izhar BAR-GAD, Alon KORNGREEN et Hagai BERGMAN (2014). « Basal ganglia : physiological, behavioral, and computational studies ». In : *Frontiers in systems neuroscience* 8, p. 150 (cf. p. 126).
- MURRAY, G. K., J. VEIJOLA, K. MOILANEN, J. MIETTUNEN, D. C. GLAHLN, T. D. CANNON, P. B. JONES et M. ISOHANNI (2006). « Infant motor development is associated with adult cognitive categorisation in a longitudinal birth cohort study ». In : *Journal of Child Psychology and Psychiatry* 47.1, p. 25-29 (cf. p. 64).
- MURRAY, Graham K., Peter B. JONES, Diana KUH et Marcus RICHARDS (2007). « Infant developmental milestones and subsequent cognitive function ». In : *Annals of neurology* 62.2, p. 128-136 (cf. p. 64).
- NADEL, Jacqueline (2002). « When do infants expect ?-commentary of field, tm (1984). early interactions between infants and their post-partum depressed mothers. ibad, 7, 517-522 ». In : *Infant Behavior and Development* 1.25, p. 30-33 (cf. p. 1).
- NADEL, Jacqueline et George BUTTERWORTH (1996). « Présentation générale : l'imitation immédiate enfin réhabilitée ». In : *Enfance* 49.1, p. 5-7 (cf. p. 2, 100).

- NADEL, Jacqueline, Isabelle CARCHON, Claude KERVELLA, Daniel MARCELLI et Denis RÉSERBAT-PLANTEY (1999). « Expectancies for social contingency in 2-month-olds ». In : *Developmental science* 2.2, p. 164-173 (cf. p. 1).
- NADEL, Jacqueline et Caroline POTIER (2002). « Imitez, imitez, il en restera toujours quelque chose : le statut développemental de l'imitation dans le cas d'autisme ». In : *Enfance* 54.1, p. 76-85 (cf. p. 100).
- NADEL, Jacqueline, Ken PREPIN et Mako OKANDA (2005a). « Experiencing contingency and agency : First step toward self-understanding in making a mind ? » In : *Interaction studies* 6.3, p. 447-462 (cf. p. 188).
- NADEL, Jacqueline, Robert SOUSSIGNAN, Pierre CANET, Guillaume LIBERT et Priscille GÉRARDIN (2005b). « Two-month-old infants of depressed mothers show mild, delayed and persistent change in emotional state after non-contingent interaction ». In : *Infant behavior and Development* 28.4, p. 418-425 (cf. p. 188).
- NAGAI, Yukie, Akiko NAKATANI et Minoru ASADA (2010). « How a robot's attention shapes the way people teach ». In : *Proceedings of the 10th International Conference on Epigenetic Robotics*, p. 81-88 (cf. p. 51).
- NAKANO, Eri, Hiroshi IMAMIZU, Rieko OSU, Yoji UNO, Hiroaki GOMI, Toshinori YOSHIOKA et Mitsuo KAWATO (1999). « Quantitative examinations of internal representations for arm trajectory planning : minimum commanded torque change model ». In : *Journal of Neurophysiology* 81.5, p. 2140-2155 (cf. p. 49).
- NASSOUR, John, Patrick HÉNAFF, Fethi BENOUEZDOU et Gordon CHENG (2014). « Multi-layered multi-pattern CPG for adaptive locomotion of humanoid robots ». In : *Biological cybernetics* 108.3, p. 291-303 (cf. p. 196).
- NÉDA, Zoltán, Erzsébet RAVASZ, Tamás VICSEK, Yves BRECHET et Albert-László BARABÁSI (2000). « Physics of the rhythmic applause ». In : *Physical Review E* 61.6, p. 6987 (cf. p. 175).
- NISHII, Jun (1997). « A learning model for a neural oscillator to generate a locomotor pattern ». In : *Proc. of Joint Conference of Information Sciences*. T. 2. Citeseer, p. 153-156 (cf. p. 197, 199).
- (1999). « Learning model for coupled neural oscillators ». In : *Network : Computation in neural systems* 10.3, p. 213-226 (cf. p. 197, 199).
- NURZAMAN, Surya G., Yoshio MATSUMOTO, Yutaka NAKAMURA, Satoshi KOIZUMI et Hiroshi ISHIGURO (2009). « Yuragi-based adaptive searching behavior in mobile robot : From bacterial chemotaxis to levy walk ». In : *2008 IEEE International Conference on Robotics and Biomimetics*. IEEE, p. 806-811 (cf. p. 113).
- OULLIER, Olivier, Gonzalo C. DE GUZMAN, Kelly J. JANTZEN, Julien LAGARDE et J. A. SCOTT KELSO (2008). « Social coordination dynamics : Measuring human bonding ». In : *Social neuroscience* 3.2, p. 178-192 (cf. p. 1, 175).

- OZTOP, Erhan, David W. FRANKLIN, Thierry CHAMINADE et Gordon CHENG (2005). « Human–humanoid interaction : is a humanoid robot perceived as a human? » In : *International Journal of Humanoid Robotics* 2.04, p. 537-559 (cf. p. 179).
- PALADINI, C. A., P. CELADA et J. M. TEPPER (1999). « Striatal, pallidal, and pars reticulata evoked inhibition of nigrostriatal dopaminergic neurons is mediated by GABAA receptors in vivo ». In : *Neuroscience* 89.3, p. 799-812 (cf. p. 133).
- PEKONY, Sarah E., Jun IZAWA et Reza SHADMEHR (2015). « Reward-dependent modulation of movement variability ». In : *Journal of Neuroscience* 35.9, p. 4015-4024 (cf. p. 86).
- PELVIG, D. P., Henning PAKKENBERG, A. K. STARK et Bente PAKKENBERG (2008). « Neocortical glial cell numbers in human brains ». In : *Neurobiology of aging* 29.11, p. 1754-1762 (cf. p. 60).
- PEPER, C. Lieke E. et Peter J. BEEK (1998). « Are frequency-induced transitions in rhythmic coordination mediated by a drop in amplitude? » In : *Biological Cybernetics* 79.4, p. 291-300 (cf. p. 177).
- PFEUTY, Benjamin, Germán MATO, David GOLOMB et David HANSEL (2003). « Electrical synapses and synchrony : the role of intrinsic currents ». In : *Journal of Neuroscience* 23.15, p. 6280-6294 (cf. p. 189, 190).
- PHILLIPS-SILVER, Jessica, C. Athena AKTIPIS et Gregory A. BRYANT (2010). « The ecology of entrainment : Foundations of coordinated rhythmic movement ». In : *Music Perception : An Interdisciplinary Journal* 28.1, p. 3-14 (cf. p. 187).
- PIAGET, J. (1975). *L'équilibration des structures cognitives : problème central du développement*. Bibliothèque scientifique internationale. Etudes d'épistémologie génétique. Presses universitaires de France. URL : <https://books.google.fr/books?id=j3MQAQAIAAJ> (cf. p. 63).
- PIAGET, Jean (1936). « La naissance de l'intelligence chez l'enfant ». In : *Delachaux et Niestlé Neuchâtel, Parés* (cf. p. 63).
- (1937). « La construction du réel chez l'enfant. » In : (cf. p. 63).
- (1945). « La formation du symbole chez l'enfant (Imitation, jeu et rêve, image et représentation) Col ». In : *Actualités pédagogiques et psychologiques, Delachaux et Niestlé Ed. Neufchâtel* (cf. p. 2, 88, 100).
- (1948). *La naissance de l'intelligence chez l'enfant*. Delachaux et Niestlé Neuchâtel, Parés (cf. p. 63).
- PIKOVSKY, A. S., M. ROSENBLUM et J. KURTHS (2001). *Synchronization—A unified approach to nonlinear science* (cf. p. 176, 178).

- PIKOVSKY, Arkady, Jurgen KURTHS, Michael ROSENBLUM et Jürgen KURTHS (2003). *Synchronization : a universal concept in nonlinear sciences*. T. 12. Cambridge university press (cf. p. 176).
- PITTI, Alexandre, Ryuma NIIYAMA et Yasuo KUNIYOSHI (2010). « Creating and modulating rhythms by controlling the physics of the body ». In : *Autonomous Robots* 28.3, p. 317-329 (cf. p. 3).
- POLLICK, Frank E. (2009). « In search of the uncanny valley ». In : *International Conference on User Centric Media*. Springer, p. 69-78 (cf. p. 15).
- PONTRYAGIN, Lev Semyonovich, Vladimir Grigorevich BOLTYANSKY, Revaz Valerianovic GAMKRELIDZE et Evgenij F. MIŠČENKO (1962). Traduit du Russe vers l'Anglais. Sous la dir. de L.W. NEUSTADT. Trad. par K.N. TRIROGOFF. Interscience Publishers (cf. p. 49).
- PREPIN, Ken et Philippe GAUSSIER (2010). « How an agent can detect and use synchrony parameter of its own interaction with a human ? » In : *Development of Multimodal Interfaces : Active Listening and Synchrony*. Springer, p. 50-65 (cf. p. 3, 180).
- PREPIN, Ken et Arnaud REVEL (2007). « Human-machine interaction as a model of machine-machine interaction : how to make machines interact as humans do ». In : *Advanced Robotics* 21.15, p. 1709-1723 (cf. p. 3, 180, 187, 196).
- PRESTON, Stephanie D. et Frans B. M. DE WAAL (2002). « Empathy : Its ultimate and proximate bases ». In : *Behavioral and brain sciences* 25.1, p. 1-20 (cf. p. 100, 118).
- RAIBERT, Marc, Kevin BLANKESPOOR, Gabriel NELSON, Rob PLAYTER et T. B. TEAM (2008). « Bigdog, the rough-terrain quadruped robot ». In : *Proceedings of the 17th World Congress*. T. 17, p. 10822-10825 (cf. p. 2, 190).
- RASHDAN, Khalid (2010). « Le pointage communicatif à un an : Une capacité définitoire externe pour un objectif interne ». In : *La définition*, p. 20 (cf. p. 99).
- RAZAKARIVONY, Sébastien, Philippe GAUSSIER et Fathi Ben OUEZDOU (2010). « From force control and sensory-motor informations to mass discrimination ». In : *From Animals to Animats 11*. Springer, p. 221-231 (cf. p. 88, 113).
- REDGRAVE, Peter, Tony J. PRESCOTT et Kevin GURNEY (1999). « The basal ganglia : a vertebrate solution to the selection problem ? » In : *Neuroscience* 89.4, p. 1009-1023 (cf. p. 59).
- REIMANN, Hendrik, Ioannis IOSSIFIDIS et Gregor SCHÖNER (2011). « Autonomous movement generation for manipulators with multiple simultaneous constraints using the attractor dynamics approach ». In : *2011 IEEE International Conference on Robotics and Automation*. IEEE, p. 5470-5477 (cf. p. 113).
- RENGERVÉ, Antoine de, Sofiane BOUCENNA, Pierre ANDRY et Philippe GAUSSIER (oct. 2010). « Emergent Imitative Behavior on a Robotic Arm Based on Visuo-Motor As-

- sociative Memories ». In : *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'10)*. Taipei, Taiwan, p. 1754-1759 (cf. p. 88, 100, 113, 114).
- RENVERGÉ, Antoine (2013). « Apprentissage Interactif en Robotique, Autonome : vers de nouveaux types d'IHM ». Thèse de doct. ETIS, UMR8051 ENSEA - Université Cergy-Pontoise - CNRS (cf. p. 17, 80, 88, 95, 113, 114).
- REVEL, Arnaud et Pierre ANDRY (2009). « Emergence of structured interactions : From a theoretical model to pragmatic robotics ». In : *Neural networks* 22.2, p. 116-125 (cf. p. 3, 180).
- RHODES, Gillian et Leslie A. ZEBROWITZ (2002). *Facial attractiveness : Evolutionary, cognitive, and social perspectives*. Ablex (cf. p. 15).
- RIDLER, Khanum, Juha M. VEIJOLA, Päivikki TANSKANEN, Jouko MIETTUNEN, Xavier CHITNIS, John SUCKLING, Graham K. MURRAY, Marianne HAAPEA, Peter B. JONES, Matti K. ISOHANNI et al. (2006). « Fronto-cerebellar systems are associated with infant motor and adult executive functions in healthy adults but not in schizophrenia ». In : *Proceedings of the National Academy of Sciences* 103.42, p. 15651-15656 (cf. p. 64).
- RIGHETTI, Ludovic, Jonas BUCHLI et Auke Jan IJSPEERT (2006). « Dynamic hebbian learning in adaptive frequency oscillators ». In : *Physica D : Nonlinear Phenomena* 216.2, p. 269-281 (cf. p. 199).
- RIZZOLATTI, Giacomo, Luciano FADIGA, Léonardo FOGASSI et Vittorio GALLESE (1999). « Resonance behaviors and mirror neurons. » In : *Archives italiennes de biologie* 137.2, p. 85-100 (cf. p. 2, 118).
- RIZZOLATTI, Giacomo, Luciano FADIGA, Vittorio GALLESE et Leonardo FOGASSI (1996). « Premotor cortex and the recognition of motor actions ». In : *Cognitive brain research* 3.2, p. 131-141 (cf. p. 2).
- ROSENBAUM, David A. et Matthew J. JORGENSEN (1992). « Planning macroscopic aspects of manual control ». In : *Human Movement Science* 11.1-2, p. 61-69 (cf. p. 51).
- ROSENBLATT, Frank (1958). « The perceptron : a probabilistic model for information storage and organization in the brain. » In : *Psychological review* 65.6, p. 386 (cf. p. 70).
- ROUGIER, Nicolas P. (2009). « Implicit and explicit representations ». In : *Neural Networks* 22.2, p. 155-160 (cf. p. 113).
- ROWAT, Peter F. et Allen I. SELVERSTON (1993). « Modeling the gastric mill central pattern generator of the lobster with a relaxation-oscillator network ». In : *Journal of neurophysiology* 70.3, p. 1030-1053 (cf. p. 196).
- (1997). « Oscillatory mechanisms in pairs of neurons connected with fast inhibitory synapses ». In : *Journal of computational neuroscience* 4.2, p. 103-127 (cf. p. 187, 196).

- ROYER, Sébastien et Denis PARÉ (2003). « Conservation of total synaptic weight through balanced synaptic depression and potentiation ». In : *Nature* 422.6931, p. 518-522 (cf. p. 91, 93).
- RUGY, Aymar de, Olivier OULLIER et Jean-Jacques TEMPRADO (2008). « Stability of rhythmic visuo-motor tracking does not depend on relative velocity ». In : *Experimental brain research* 184.2, p. 269-273 (cf. p. 177).
- RUNGE, Carl (1895). « Über die numerische Auflösung von Differentialgleichungen ». In : *Mathematische Annalen* 46.2, p. 167-178 (cf. p. 162).
- SAYGIN, Ayse Pinar, Thierry CHAMINADE, Hiroshi ISHIGURO, Jon DRIVER et Chris FRITH (2011). « The thing that should not be : predictive coding and the uncanny valley in perceiving human and humanoid robot actions ». In : *Social cognitive and affective neuroscience* 7.4, p. 413-422 (cf. p. 16).
- SCARNATI, Eugenio, A. PROIA, E. CAMPANA et Claudio PACITTI (1986). « A microiontophoretic study on the nature of the putative synaptic neurotransmitter involved in the pedunclopontine-substantia nigra pars compacta excitatory pathway of the rat ». In : *Experimental brain research* 62.3, p. 470-478 (cf. p. 128).
- SCHAAL, Stefan et Christopher G. ATKESON (1998). « Constructive incremental learning from only local information ». In : *Neural computation* 10.8, p. 2047-2084 (cf. p. 53).
- SCHELER, Gabriele (2004). « Regulation of neuromodulator receptor efficacy—implications for whole-neuron and synaptic plasticity ». In : *Progress in Neurobiology* 72.6, p. 399-415 (cf. p. 129).
- SCHMIDT, Richard C., Claudia CARELLO et Michael T. TURVEY (1990). « Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. » In : *Journal of experimental psychology : human perception and performance* 16.2, p. 227 (cf. p. 176).
- SCHMIDT, Richard C. et Michael J. RICHARDSON (2008). « Dynamics of interpersonal coordination ». In : *Coordination : Neural, behavioral and social dynamics*. Springer, p. 281-308 (cf. p. 175, 176, 178, 184).
- SCHMIDT, Richard C., Michael J. RICHARDSON, Christine ARSENAULT et Bruno GALANTUCCI (2007). « Visual tracking and entrainment to an environmental rhythm. » In : *Journal of Experimental Psychology : Human Perception and Performance* 33.4, p. 860 (cf. p. 177).
- SCHÖNER, Gregor (1990). « A dynamic theory of coordination of discrete movement ». In : *Biological cybernetics* 63.4, p. 257-270 (cf. p. 176).
- SCHÖNER, Gregor, Michael DOSE et Christoph ENGELS (1995). « Dynamics of behavior : Theory and applications for autonomous robot architectures ». In : *Robotics and autonomous systems* 16.2-4, p. 213-245 (cf. p. 112).

- SCHULTZ, Wolfram, Peter DAYAN et P. Read MONTAGUE (1997). « A neural substrate of prediction and reward ». In : *Science* 275.5306, p. 1593-1599 (cf. p. 59, 131).
- SEO, Keehong, Soon-Jo CHUNG et Jean-Jacques E. SLOTINE (2010). « CPG-based control of a turtle-like underwater vehicle ». In : *Autonomous Robots* 28.3, p. 247-269 (cf. p. 181).
- SEO, Keehong et Jean-Jacques E. SLOTINE (2007). « Models for global synchronization in CPG-based locomotion ». In : *Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE, p. 281-286 (cf. p. 196).
- SHEN, Qiming, Hatice KOSE-BAGCI, Joe SAUNDERS et Kerstin DAUTENHAHN (2010). « The Impact of Participants' Beliefs on Motor Interference and Motor Coordination in Human-Humanoid Interactions ». In : *IEEE Transactions on Autonomous Mental Development* 3.1, p. 6-16 (cf. p. 179).
- SHOCKLEY, Kevin, Marie-Vee SANTANA et Carol A. FOWLER (2003). « Mutual interpersonal postural constraints are involved in cooperative conversation. » In : *Journal of Experimental Psychology : Human Perception and Performance* 29.2, p. 326 (cf. p. 175).
- SHORT, Martin W. et James H. CAURAUGH (1997). « Planning macroscopic aspects of manual control : End-state comfort and point-of-change effects ». In : *Acta Psychologica* 96.1-2, p. 133-147 (cf. p. 51).
- SIMON, J. Richard et James D. WOLF (1963). « Choice reaction time as a function of angular stimulus-response correspondence and age ». In : *Ergonomics* 6.1, p. 99-105 (cf. p. 111).
- SIMPLYMAYA (2016). « The Uncanny Valley and CGI - The Challenge in Creating Realistic Humans ». In : *Maya Blog*. URL : <https://simplymaya.com/sm-news-articles/75/the-uncanny-valley-and-cgi-the-challenge-in-creating-realistic-humans/> (cf. p. 15).
- STEINHAUS, H. (1956). « Sur la division des corps materiels en parties. Bull. Acad. Polon. Sci., C1. III vol IV : 801-804 ». In : (cf. p. 68).
- STEVENSON, Paul A. et Wolfram KUTSCH (1987). « A reconsideration of the central pattern generator concept for locust flight ». In : *Journal of Comparative Physiology A* 161.1, p. 115-129 (cf. p. 196).
- SUGAHARA, Atsushi (2010). « Generating circular motion of a human-like robotic arm using attractor selection model ». In : *Journal of robotics and mechatronics* 22.3, p. 315-321 (cf. p. 113).
- SUTTON, Richard S. (1988). « Learning to predict by the methods of temporal differences ». In : *Machine learning* 3.1, p. 9-44 (cf. p. 53, 59, 131).

- SUTTON, Richard S. et Andrew G. BARTO (1988). « Reinforcement Learning : An Introduction ». In : *IEEE Transactions on Neural Networks* 16, p. 285-286 (cf. p. 53).
- SUTTON, Richard S., Andrew G. BARTO et al. (1998). *Introduction to reinforcement learning*. T. 2. 4. MIT press Cambridge (cf. p. 55).
- TARDE, Gabriel (1890). *Les lois de l'imitation*. Sous la dir. de Félix ALCAN. Edition Kimé Paris 1993 (cf. p. 100).
- TAUBE, Jeffrey S., Jeremy P. GOODRIDGE, Edward J. GOLOB, Paul A. DUDCHENKO et Robert W. STACKMAN (1996). « Processing the head direction cell signal : a review and commentary ». In : *Brain research bulletin* 40.5, p. 477-484 (cf. p. 90).
- TAYLOR, Janet L., Markus AMANN, Jacques DUCHATEAU, Romain MEEUSEN et Charles L. RICE (2016). « Neural contributions to muscle fatigue : from the brain to the muscle and back again ». In : *Medicine and science in sports and exercise* 48.11, p. 2294 (cf. p. 154, 169).
- TEITELBAUM, Philip, Osnat B. TEITELBAUM, Joshua FRYMAN et Ralph MAURER (2002). « Infantile reflexes gone astray in autism ». In : *J. Dev. Learn. Disord* 6.15, p. 10-1073 (cf. p. 64).
- TEMPRADO, Jean-Jacques et Michel LAURENT (2004). « Attentional load associated with performing and stabilizing a between-persons coordination of rhythmic limb movements ». In : *Acta Psychologica* 115.1, p. 1-16 (cf. p. 175).
- TEPPER, J. M., L. P. MARTIN et D. R. ANDERSON (1995). « GABAA receptor-mediated inhibition of rat substantia nigra dopaminergic neurons by pars reticulata projection neurons ». In : *Journal of Neuroscience* 15.4, p. 3092-3103 (cf. p. 133).
- TEPPER, James M. et Christian R. LEE (2007). « GABAergic control of substantia nigra dopaminergic neurons ». In : *Progress in brain research* 160, p. 189-208 (cf. p. 133).
- THELEN, Esther, Gregor SCHÖNER, Christian SCHEIER et Linda B. SMITH (2001). « The dynamics of embodiment : A field theory of infant perseverative reaching ». In : *Behavioral and brain sciences* 24.01, p. 1-34 (cf. p. 112).
- THOMAZ, Andrea L. et Maya ÇAKMAK (2009). « Learning about objects with human teachers ». In : *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*. ACM, p. 15-22 (cf. p. 51).
- TODOROV, Emanuel (2006). « Optimal control theory ». In : *Bayesian brain : probabilistic approaches to neural coding*, p. 269-298 (cf. p. 49).
- TODOROV, Emanuel et Michael I. JORDAN (2002). « Optimal feedback control as a theory of motor coordination ». In : *Nature neuroscience* 5.11, p. 1226 (cf. p. 50).
- (2003). « A minimal intervention principle for coordinated movement ». In : *Advances in neural information processing systems*, p. 27-34 (cf. p. 50).

- TONDU, Bertrand (2017). « Has Uncanny Valley Feeling an Impact on Learning : A Preliminary Experiment Performed in a French Elementary School ». In : (cf. p. 15).
- TOPÁL, József, György GERGELY, Ágnes ERDŐHEGYI, Gergely CSIBRA et Ádám MIKLÓSI (2009). « Differential sensitivity to human communication in dogs, wolves, and human infants ». In : *Science* 325.5945, p. 1269-1272 (cf. p. 112).
- TOPÁL, József, Gergely GYÖRGY, Adam MIKLOSI, Agnes ERDOHEGYI et Gergely CSIBRA (oct. 2008). « Infants' Perseverative Search Errors Are Induced by Pragmatic Misinterpretation ». In : *Science (New York, N.Y.)* 321, p. 1831-4. DOI : [10.1126/science.1161437](https://doi.org/10.1126/science.1161437) (cf. p. 112).
- TRULLIER, Olivier, Sidney I. WIENER, Alain BERTHOZ et Jean-Arcady MEYER (1997). « Biologically based artificial navigation systems : Review and prospects ». In : *Progress in neurobiology* 51.5, p. 483-544 (cf. p. 88).
- ULZEN, Niek R. van, Claudine J. C. LAMOTH, Andreas DAFFERTSHOFER, Gün R SEMIN et Peter J. BEEK (2008). « Characteristics of instructed and uninstructed interpersonal coordination while walking side-by-side ». In : *Neuroscience letters* 432.2, p. 88-93 (cf. p. 175).
- UNO, Yoji, Mitsuo KAWATO et Rika SUZUKI (1989). « Formation and control of optimal trajectory in human multijoint arm movement ». In : *Biological cybernetics* 61.2, p. 89-101 (cf. p. 49).
- VALLAR, Giuseppe, Elie LOBEL, Gaspare GALATI, Alain BERTHOZ, Luigi PIZZAMIGLIO et Denis LE BIHAN (1999). « A fronto-parietal system for computing the egocentric spatial frame of reference in humans ». In : *Experimental Brain Research* 124.3, p. 281-286 (cf. p. 89).
- VAN DER POL, Balth et Jan VAN DER MARK (1928). « LXXII. The heartbeat considered as a relaxation oscillation, and an electrical model of the heart ». In : *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 6.38, p. 763-775 (cf. p. 196).
- VAN HAM, Ronald, Thomas G. SUGAR, Bram VANDERBORGHT, Kevin W. HOLLANDER et Dirk LEFEBER (2009). « Compliant actuator designs ». In : *IEEE Robotics & Automation Magazine* 16.3, p. 81-94 (cf. p. 19).
- VARELA, Francisco J., Evan THOMPSON et Eleanor ROSCH (1993). *L'inscription corporelle de l'esprit : sciences cognitives et expérience humaine* (cf. p. 12).
- VELICHKO, Andrei, Maksim BELYAEV et Petr BORISKOV (2019). « A Model of an Oscillatory Neural Network with Multilevel Neurons for Pattern Recognition and Computing ». In : *Electronics* 8.1, p. 75 (cf. p. 196).
- VIJAYAKUMAR, Sethu, Aaron D'SOUZA et Stefan SCHAAL (2005). « Incremental online learning in high dimensions ». In : *Neural computation* 17.12, p. 2602-2634 (cf. p. 53).

- VINYALS, Oriol, Igor BABUSCHKIN, Junyoung CHUNG, Michael MATHIEU, Max JADERBERG, Wojciech M. CZARNECKI, Andrew DUDZIK, Aja HUANG, Petko GEORGIEV, Richard POWELL et al. (2019). « AlphaStar : Mastering the real-time strategy game StarCraft II ». In : *DeepMind Blog* (cf. p. 11).
- VOOGD, Jan (2000). « The Purkinje cell ». In : *Neurological eponyms. Oxford University Press, Oxford*, p. 37-43 (cf. p. 60).
- WADA, Yasuhiro et Mitsuo KAWATO (2004). « A via-point time optimization algorithm for complex sequential trajectory formation ». In : *Neural Networks* 17.3, p. 353-364 (cf. p. 50).
- WAINER, Joshua, David J. FEIL-SEIFER, Dylan A. SHELL et Maja J. MATARIC (2006). « The role of physical embodiment in human-robot interaction ». In : *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, p. 117-122 (cf. p. 13).
- WALKER, Thomas J. (1969). « Acoustic synchrony : two mechanisms in the snowy tree cricket ». In : *Science* 166.3907, p. 891-894 (cf. p. 175).
- WALMSLEY, B., J. A. HODGSON et R. E. BURKE (1978). « Forces produced by medial gastrocnemius and soleus muscles during locomotion in freely moving cats. » In : *Journal of neurophysiology* 41.5, p. 1203-1216 (cf. p. 78).
- WANG, DeLiang (1995). « Emergent synchrony in locally coupled neural oscillators ». In : *IEEE transactions on neural networks* 6.4, p. 941-948 (cf. p. 196, 200).
- WANG, DeLiang, Joachim BUHMANN et Christoph von der MALSBERG (1990). « Pattern segmentation in associative memory ». In : *Neural Computation* 2.1, p. 94-106 (cf. p. 196, 200).
- WANG, DeLiang et David TERMAN (1995). « Locally excitatory globally inhibitory oscillator networks ». In : *IEEE transactions on neural networks* 6.1, p. 283-286 (cf. p. 196).
- WANG, Lipo (1996). « Oscillatory and chaotic dynamics in neural networks under varying operating conditions ». In : *IEEE Transactions on Neural Networks* 7.6, p. 1382-1388 (cf. p. 196).
- WANN, John P., Mark MON-WILLIAMS et Katherine RUSHTON (1998). « Postural control and co-ordination disorders : The swinging room revisited ». In : *Human Movement Science* 17.4-5, p. 491-513 (cf. p. 63).
- WATANABE, Tomio et Masashi OKUBO (1997). « Evaluation of the entrainment between a speaker's burst-pause of speech and respiration and a listener's respiration in face-to-face communication ». In : *Proceedings 6th IEEE International Workshop on Robot and Human Communication. RO-MAN'97 SENDAI*. IEEE, p. 392-397 (cf. p. 175).

- WATKINS, Christopher J. C. H. et Peter DAYAN (1992). « Q-learning ». In : *Machine learning* 8.3-4, p. 279-292 (cf. p. 54).
- WATKINS, Christopher John Cornish Hellaby (1989). « Learning from delayed rewards ». King's College, Cambridge (cf. p. 54).
- WEISS, Astrid et Christoph BARTNECK (2015). « Meta analysis of the usage of the god-speed questionnaire series ». In : *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, p. 381-388 (cf. p. 43).
- WIDROW, Bernard et Marcian E. HOFF (1960). *Adaptive switching circuits*. Rapp. tech. Stanford Univ Ca Stanford Electronics Labs (cf. p. 70).
- WILMORE, Jack H. et David L. COSTILL (1998). *Physiologie du sport et de l'exercice physique*. De Boeck (cf. p. 75).
- WILSON, Margaret et Thomas P. WILSON (2005). « An oscillator model of the timing of turn-taking ». In : *Psychonomic bulletin & review* 12.6, p. 957-968 (cf. p. 175).
- WINTERS, Jack M. (1990). « Hill-based muscle models : a systems engineering perspective ». In : *Multiple muscle systems*. Springer, p. 69-93 (cf. p. 80).
- WISSE, Martijn et Jan VAN FRANKENHUYZEN (2006). « Design and construction of mike ; a 2-d autonomous biped based on passive dynamic walking ». In : *Adaptive motion of animals and machines*. Springer, p. 143-154 (cf. p. 2, 12).
- WUERKER, Raymond B., Alexander M. MCPHEDRAN et Elwood HENNEMAN (1965). « Properties of motor units in a heterogeneous pale muscle (m. gastrocnemius) of the cat ». In : *Journal of neurophysiology* 28.1, p. 85-99 (cf. p. 78).
- YOSHIZAWA, Tomohiko, Makoto ITO et Kenji DOYA (2018). « Reward-predictive neural activities in striatal striosome compartments ». In : *Eneuro* 5.1 (cf. p. 133).
- YU, Junzhi, Ming WANG, Zongshuai SU, Min TAN et Jianwei ZHANG (2013). « Dynamic modeling of a CPG-governed multijoint robotic fish ». In : *Advanced Robotics* 27.4, p. 275-285 (cf. p. 196).
- ZHOU, F.M. (2016). « Chapter 15 - The Substantia Nigra Pars Reticulata ». In : *Handbook of Basal Ganglia Structure and Function, Second Edition*. Sous la dir. d'Heinz STEINER et Kuei Y. TSENG. T. 24. Handbook of Behavioral Neuroscience. Elsevier, p. 293-316. DOI : <https://doi.org/10.1016/B978-0-12-802206-1.00015-5>. URL : <http://www.sciencedirect.com/science/article/pii/B9780128022061000155> (cf. p. 128).
- ZIVOTOFSKY, Ari Z. et Jeffrey M. HAUSDORFF (2007). « The sensory feedback mechanisms enabling couples to walk synchronously : An initial investigation ». In : *Journal of neuroengineering and rehabilitation* 4.1, p. 28 (cf. p. 175).