Filière : Electronique, Energie Electrique et Automatique

Spécialité : M2 Systèmes Embarqués et Traitement de l'Information (SETI)

Stage de fin d'étude

Une tête expressive capable de reconnaître des expressions faciales secondaires

Auteur :

- Yassine OUZAR

Encadré par :

- M. Sofiane BOUCENNA

Année Universitaire : 2018 / 2019

Remerciement

En tout premier lieu, je remercie DIEU le tout puissant de m'avoir éclairé le chemin du savoir afin de terminer ce travail.

Je tiens à remercier tout particulièrement mon encadreur de stage Sofiane BOUCENNA pour son soutien qu'il m'a accordé, tout ce qu'il m'a apporté à travers les nombreuses discussions passionnantes et pour son amitié durant la période du stage.

Un grand merci également à tous les membres de l'équipe Neurocybernétique qui ont chacun contribué à l'aboutissement de ce travail. J'ai trouvé au sein de cette équipe un environnement de travail stimulant qui m'a permis de découvrir le domaine de la robotique développementale et l'interaction homme-robot.

Egalement, je tiens à exprimer ma sincère gratitude envers tous ceux qui m'ont aidé ou ont participé au bon déroulement de ce stage.

Enfin mes remerciements vont à mes chers parents, frères, sœurs, et amis pour leurs soutiens.

Yassine

« Je crois que dans une cinquantaine d'années il sera possible de programmer des ordinateurs, avec une capacité de mémoire d'à peu près 10^9, pour les faire si bien jouer au jeu de l'imitation qu'un interrogateur moyen n'aura pas plus de 70 pour cent de chances de procéder à l'identification exacte après cinq minutes d'interrogation »

Alan TURING

Résumé

Ce stage porte sur la reconnaissance des expressions faciales pour l'interaction homme-robot qui attire de plus en plus l'intention de la recherche en robotique et en intelligence artificielle. Notre objectif est de développer un modèle neuronale permettant à une tête robotique d'apprendre à reconnaître les expressions faciales secondaires en ligne et de manière autonome en interaction avec un partenaire humain. Nous proposons une architecture sensori-motrice (PerAc) qui associe les caractéristiques expressives extraites à des actions des servomoteurs agissant chacun comme muscle du visage.

En s'inspirant de la capacité du bébé à apprendre des expressions faciales sans signal de supervision, l'apprentissage basé sur des réseaux de neurones non supervisés apprend en ligne à reconnaître les groupements musculaires du visage à travers un jeu d'imitation entre le robot et un expérimentateur. Après une courte durée d'apprentissage, le robot sera capable de produire une multitude d'expressions faciales primaires et secondaires.

Abstract

This internship focuses on facial expressions recognition for human-robot interaction which is increasingly attracting the intention of research in robotics and artificial intelligence. Our aim is to develop a neural model that allows a robotic head to learn to recognize compound facial expressions online and in autonomous way interacting with a human partner. We propose a sensorimotor architecture (PerAc) that combines the extracted expressive characteristics with the actions of the servomotors, each acting as a facial muscle.

Inspired by the baby's ability to learn facial expressions without a supervision signal, learning based on unsupervised neural networks learns online to recognize facial muscle groups through an imitation game between the robot and an experimenter. After a short learning period, the robot will be able to produce a multitude of primary and secondary facial expressions.

Mots clés

Reconnaissance, Emotions, Tète expressive, Interaction homme-robot, Robotique développementale, Apprentissage non supervisé, Apprentissage en ligne, Expression faciale, Primitives motrices, Intensité expressive.

Table des matières

Table des figures

Introduction

Ce travail est motivé par la question de savoir comment un robot n'ayant au départ aucune connaissance priori du monde est capable d'apprendre de manière autonome à reconnaître et de produire une multitude d'expressions faciales avec une palette d'intensité expressives variée. En effet, l'acquisition d'une autonomie comportementale à savoir la capacité d'apprentissage et d'adaptation en ligne est indispensable pour que le robot puisse interagir émotionnellement de manière naturelle avec son partenaire humain et faire face à des perturbations imprédictibles. Pour cela, le robot doit posséder la capacité d'apprendre à reconnaître les expressions émotionnelles de l'autrui, à les associer à son état interne pour ensuite pouvoir communiquer lui aussi d'une façon non verbale, et ceci sans utiliser un signal de supervision externe.

Cette question est proche de la compréhension de la façon dont les bébés apprennent à reconnaître les expressions faciales de leurs entourage sans recevoir de signal d'enseignement explicite permettant d'associer, par exemple, un «visage heureux» à leur propre état émotionnel interne de bonheur. Nombreuses recherches en psychologie et en science cognitive ont montré que le développement de cette capacité chez l'enfant est lié et influencé par les interactions affectives avec son entourage (les parents,...) à travers l'imitation de ce qu'ils font.

En s'inspirant de cette propriété d'apprentissage et d'imitation chez les enfants ainsi que les principes de la robotique développementale, nous cherchons à vérifier si une architecture sensori-motrice basée sur un simple modèle de conditionnement classique peut développer la capacité du robot à apprendre en ligne, et à reconnaître n'importe quelle expression faciale, plutôt qu'un nombre limité d'expressions.

Le manuscrite est organisé comme suit: tout d'abord, un bref état de l'art sur la théorie de l'émotion et les systèmes de reconnaissances des expressions faciales. Le deuxième chapitre présente les réseaux et les architectures neuronales nécessaire pour le développement de notre modèle. Enfin nous présentons le travail réalisé durant le stage ainsi que les résultats obtenus.

1. Introduction

L'un des aspects les plus importants dans le domaine de la robotique sociale est la communication entre l'humain et le robot. Afin de faciliter les interactions naturelles et sociales entre les deux, il est nécessaire de doter les robots de capacités de reconnaissance des émotions qui peuvent fournir des informations importantes permettant d'améliorer la communication et de s'adapter à une situation donnée.

Les émotions sont multi componentielles, elles se manifestent sur différents canaux (verbal, vocal et non verbal). Bien que, la langue parlée joue un rôle si important mais nous oublions souvent l'importance des signaux non verbaux tels que les expressions faciales qui jouent un rôle majeur dans les interactions sociales [Adolphs, 1999]. D'après les travaux de Mehrabian [Mehrabian, 1968], la communication des émotions est à 7% verbale, à 38% vocale (tonalité de la voix) et 55% non verbale (expressions faciales, expressions corporelles). A cet égard, l'implémentation d'un système de reconnaissance des expressions faciales sur un robot social est indispensable pour renforcer l'empathie, l'attention ou la compréhension des compétences sociales dans une interaction homme-robot [Paiva et al. , 2004; Siegel et al., 2009].

Au cours des dernières années, plusieurs modèles de systèmes de reconnaissance des expressions faciales ont été proposés en développant des architectures permettant d'établir des relations empathiques avec les partenaires humains. La majorité des travaux se sont concentrés sur les émotions de base décrites par Ekman [Ekman, 1992]. Néanmoins, l'être humain peut exprimer une multitude d'expressions faciales autres que les 6 émotions primaires. D'autre part, les architectures proposées sont basées sur les solutions ingénieur ad-hoc qui utilisent des techniques d'apprentissage hors ligne ou basées sur un modèle. Malgré que les résultats obtenus sont impressionnants mais elles ne permettent pas au système d'évoluer et de s'adapter à son environnement.

Nous présentons dans ce chapitre, un état de l'art sur les émotions et la reconnaissance des expressions faciales pour l'interaction homme-robot.

2. Les émotions

2.1. Théorie des émotions

L'étude des phénomènes émotionnels remonte à plusieurs siècles mais la plupart des idées fondamentales dans la recherche sur les émotions modernes se trouvent dans des théories relativement «récentes». Au 17ème siècle, Spinoza définit un espace émotionnel dans lequel toutes les expériences émotionnelles peuvent être décrites selon 3 dimensions (axes): joie, désir et tristesse [Spinoza, 1677]. Une théorie contemporaine est celle de Descartes [Descartes, 1649], fondée sur l'idée que les émotions sont une combinaison de primitifs émotionnels. Malgré la nette distinction entre leurs définitions fondamentales, le désaccord majeur entre les deux théories réside dans le fait que Descartes établit une séparation nette entre l'esprit (cognition, cerveau rationnel) et le corps (émotion, instincts), tandis que Spinoza les unifie et voit les émotions comme les fondements de l'esprit. Plus tard, au 19ème siècle, Darwin introduit la théorie de l'évolution [Darwin, 1965]. À son avis, les émotions sont innées, universelles et font partie d'un patrimoine génétique. Il lie également les émotions au système nerveux.

William James associe l'émotion à une approche physiologique [James,1884], pour lui les émotions naquissent à partir d'une réponse physiologique et que chaque émotion est associée à un pattern dans le corps humain, un rythme cardiaque, un pattern de tensions. Les émotions sont secondaires et résultent de phénomènes physiologiques. Si on prends l'exemple d'un stimuli tel que la présence d'un serpent, la présence de cet animal provoque une augmentation du rythme cardiaque qui à son tour évoque l'émotion de la peur.

Cannon [Cannon, 1927] et Bard [Bard, 1928] ont étaient en désaccord avec la théorie de William James, disant que les émotions ne sont pas simplement des phénomènes physiologiques. Pour eux les réponses physiologiques ne peuvent pas expliquer l'expérience émotionnelle. L'argument est que la réponse physiologique comme le changement du rythme cardiaque est trop lente pour induire une expression émotionnelle aussi rapide et intense. Lors de la présence d'un stimuli externe une activation de certains patterns dans le cerveau implique une réponse physiologique et en même temps une réponse émotionnelle.

Plus tard Schacter et Singer [Schachter and Singer, 1962] ont proposé une théorie sur les émotions basée sur deux facteurs. L'évaluation de l'expérience physiologique définie et détermine l'expérience émotionnelle. Après qu'une réaction physiologique ait lieu vis-à-vis

d'un stimuli externe, l'interprétation de cette réponse physiologique est le facteur déterminant quant à l'émotion exprimée.

Certains des concepts fournis par Spinoza, Descartes et Darwin sont les fondements de théories et de modèles plus récents. Les théories de James-Lange et Cannon-Bard, toutes deux fondées sur une approche physiologique, ont eu une influence considérable sur la recherche de l'émotion. Il s'agit aujourd'hui d'un macrocosme regroupant des chercheurs de domaines aussi variés que la philosophie, la psychologie, la sociologie, l'éthologie, la neuro-imagerie, la neurophysiologie, la psychiatrie mais aussi l'intelligence artificielle et la robotique. des efforts considérables pour mener des recherches interdisciplinaires sur les émotions.

2.2. Les émotions fondamentales d'Ekman

Ekman est probablement l'un des partisans les plus importants de l'approche des émotions discrètes et fondamentales [Ekman et Friesen, 1971; Ekman et al., 1983; Matsumoto et Ekman, 2004]. Il soutient que les émotions correspondent à des modèles distincts d'activation dans le système nerveux autonome. En outre, les émotions distinctes sont régies par différents circuits neuronaux qui ont évolué dans le but de fonctions de survie. Dans sa théorie, Ekman distingue les émotions de base des émotions plus complexes, morales et prosociales. Les premières, acquises au cours de l'évolution, sont universelles, en ce sens qu'elles ont des propriétés communes à toutes les espèces. Dans la théorie d'Ekman, il existe sept émotions de base: initialement la colère, la peur, la tristesse, le dégoût, la surprise et la joie, auxquels le mépris était de plus en plus ajouté. Pour développer sa théorie des émotions discrètes, il a mené de nombreuses études interculturelles sur l'expression du visage. Selon lui, le fait que les individus soient capables de reconnaître les expressions de personnes de cultures avec lesquelles ils ont eu peu ou pas de contact (Nouvelle-Guinée par rapport à la culture occidentale) prouve que les émotions sont innées et universelles.

3. Les expressions faciales

3.1. Introduction

Les expressions faciales sont un aspect important du comportement et de la communication non verbale. Elles jouent un rôle majeur dans la recherche sur les émotions et dans les interactions sociales pour diverses raisons, elles sont visible, elles contiennent de nombreuses fonctionnalités utiles pour la reconnaissance des émotions, et il est plus facile de

collecter un grand ensemble de données de visages que d'autres moyens de communication humaine [Adolphs, 1999; de Gelder, 2009]. Pour cette raison, la conception d'un système méticuleux de reconnaissance des expressions faciales (FER) est essentielle pour concevoir un robot social.

3.2. Les expressions faciales primaires et secondaires

Ekman et Friesen [Ekman et Friesen, 1972] ont étudié la relation entre les émotions et l'expression faciale. Ensuite, ils ont identifié un sous-ensemble d'émotions en corrélation avec des expressions faciales spécifiques. Ces expressions sont appelées les émotions de base, à savoir le bonheur, la tristesse, la surprise, la colère, la peur et le dégoût. Néanmoins, l'être humain peut produire une multitude d'expressions faciales autres que les six émotions primaires d'Ekman. Dans la vraie vie, il y a une complexité de mélanges d'émotions. Nous sommes rarement dans une colère furieuse, dans une tristesse extrême ou dans une joie délirante, mais souvent dans un mélange de peur et de soulagement, d'amusement et de colère. Donc l'humain exprime non seulement les émotions primaires mais il peut également exprimer des expressions composées résultantes d'un mélange des expressions de base avec des niveaux d'intensités expressives variés. En effet, une simple expression peut exprimer plusieurs états selon son niveau d'intensité.

Les expressions secondaires ou composées peuvent correspondre à différents types d'émotions telles que la superposition de deux émotions, l'exagération d'une émotion ou le masquage de l'émotion ressentie par une autre émotion non ressentie. Par exemple, le sentiment de bonheur peut être classé dans une seule émotion (par exemple, la joie), mais le sentiment d'être heureux et surpris en même temps ne peut pas être classé ni dans la joie ni dans la surprise. Le fait de ressentir un événement joyeux et surprenant implique des comportements très différents de ceux observés lorsque nous sommes heureux mais pas surpris. De même, joyeusement surpris est différent de joyeusement dégoûté, bien que la joie soit un dénominateur commun à ces émotions [Du et al., 2014]. Du et Martinez [Du et Martinez, 2015] ont proposé des expressions faciales composées combinant deux catégories d'émotions de base et identifié 15 expressions composées produites de manière homogène dans toutes les cultures. Leurs théories indiquent que les émotions individuelles peuvent être mélangées ou fusionnées pour former de nouvelles émotions.

3.3. Les systèmes de reconnaissance des expressions faciales

Les systèmes automatiques de reconnaissance des expressions faciales consistent à classer le visage détecté dans l'image en émotion. La majorité des systèmes de reconnaissances des expressions faciales traite seulement les expressions faciales de base et elle se compose principalement de trois modules clés: détection de visage et extraction de caractéristiques, classification et reconnaissance et enfin la génération de l'émotion. La structure du système est illustrée à la figure 1.1. Tout d'abord, le module de détection de visage et extraction de caractéristiques segmente les régions de visage d'une séquence vidéo ou d'une image et localise les positions des traits du visage (sourcils, des yeux, du nez et de la bouche). Les positions peuvent être représentées par des points déterminés dotés de propriétés mathématiques spéciales (les minima locaux). Seulement les points importants des traits du visage sont extraits. Compte tenu des résultats de l'extraction de caractéristiques, le module de reconnaissance qui classe les visages d'entrée dans la classe correspondante d'expressions faciales (bonheur, peur,...). Enfin, le module de génération d'émotions artificielles peut contrôler le robot social pour imiter l'expression du visage en réponse à l'expression de l'utilisateur.

Dans la majorité des systèmes de reconnaissance des expressions faciales, l'accent est mis sur le choix des meilleures méthodes d'extraction de caractéristiques et également des algorithmes d'apprentissage et de classification adaptés. Leurs modèles s'inspirent des algorithmes de vision par ordinateur classiques divisés par étapes. Tout d'abord, le visage est détecté dans l'image, puis il est cadré. Finalement les expressions sont apprises en base de données hors ligne. Ces travaux utilisent le classificateur en cascade de Haar pour la détection du visage [Viola et Jones, 2001], quant à l'extraction de caractéristiques, de nombreuses approches traditionnelles ont été utilisé, telles que le modèle binaire local (LBP) [Liu et al., 2017], l'histogramme de gradient orienté (HOG) [Carcagni et al., 2015] et la transformation de caractéristiques visuelles invariante à l'échelle (SIFT) [Berretti et al., 2011]. Pour les méthodes traditionnelles d'apprentissage automatique, différentes techniques telles que la machine à vecteurs de support (SVM) [Jain et al., 2017] et le modèle de markov caché [Sandbach et al., 2012] ont été appliquées avec succès pour classifier l'expression faciale. Ces méthodes utilisent des à priori forts et ils ont besoin d'accéder à toute la base d'apprentissage. Bien que les résultats obtenus sont impressionnants sur des bases de données déjà construites (Cohn-Kanade, JAFFE…), mais elles sont rarement confrontés à des images réelles en

environnement naturel et elles sont basées sur des mécanismes ad-hoc non compatible de point de vu développemental et ne permettent pas une autonomie du robot.

En guise d'alternative, l'imitation du comportement humain a été utilisée pour des tâches d'apprentissage afin d'améliorer l'interaction homme-robot. L'imitation des émotions joue un rôle important dans le développement cognitif et a été étudiée les dernières années en robotique sociale [Breazeal et al., 2004 ; Ge et al., 2008]. Les informations visuelles et auditives sont utilisées pour imiter les expressions humaines en tant que moyen de développement des compétences sociales et de communication. Parmi les robots sociaux développés on trouve les robots Kismet [Breazeal et al., 2000], Einstein [Wu et al., 2009], et WE- 4RII [Zecca et al., 2007]... Ces robots imitent principalement les expressions faciales et le langage corporel, en modifiant la position de différents moteurs contrôlant les yeux et la bouche pour afficher l'émotion.

Notre approche se rapproche davantage de ces travaux où l'apprentissage est effectué à travers un jeu d'imitation avec le partenaire humain [Boucenna et al., 2010].

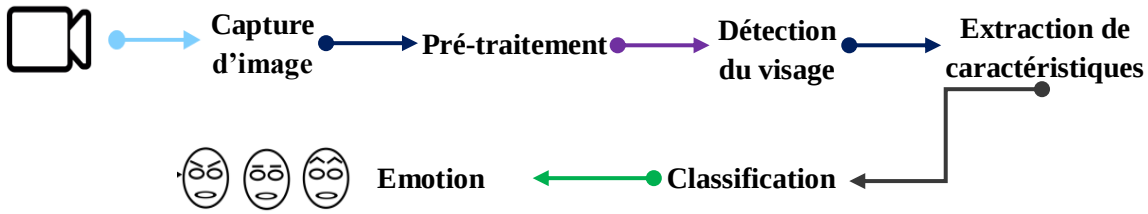

Figure 1.1 : Processus de la reconnaissance faciale des émotions.

1. Introduction

Le but de ce chapitre est de présenter les différents outils nécessaires pour le développement de l'architecture neuronale et à la compréhension des méthodes et des modèles utilisés tout au long de ces travaux. Les outils présentés se situent à différents niveaux d'abstraction que nous présenterons du plus élémentaire au plus complexe. Dans un premier temps nous présenterons le formalisme neuronal et les différents apprentissages et réseaux de neurones utilisés. Dans un second temps nous aborderons l'architecture PerAc qui a été développé pour décrire notre architecture de contrôle neuronale.

2. Les réseaux de neurones

2.1. Neurone biologique

Le cerveau humain contient environ 100 milliards de neurones interconnectés. Un neurone est constitué principalement par un corps cellulaire qui contient le noyau du neurone, un axone qui transmet l'influx nerveux (l'information), plusieurs dendrites qui reçoivent cet influx nerveux en provenance d'autres neurones et enfin plusieurs terminaisons neuronales, Le point contact entre l'axone d'un neurone et la dendrite d'un autre neurone est appelé une synapse, elle permet le déclenchement d'un potentiel d'action dans le neurone pour activer la communication avec un autre neurone [Hodgkin et Huxley, 1952].

La force d'un réseau de neurones réside dans la communication de ses neurones à travers des signaux électriques qu'on nomme "influx nerveux". Ces signaux se caractérisent par des fréquences qui jouent un rôle important au niveau de la propagation des signaux dans le réseau en question.

Au repos le neurone a un potentiel négatif, lorsqu'il est excité une différence de potentiel est générée par une concentration d'ions entre l'intérieur et l'extérieur du neurone. L'influx nerveux est envoyé sous forme de potentiel d'action et se propage le long de l'axone jusqu'aux terminaisons synaptiques. En arrivant au niveau des synapses, les potentiels d'action libèrent des neurotransmetteurs (médiateurs chimiques) dans la fente synaptique. Plus la fréquence de potentiel d'action est importante, plus le neurone produit les neurotransmetteurs.

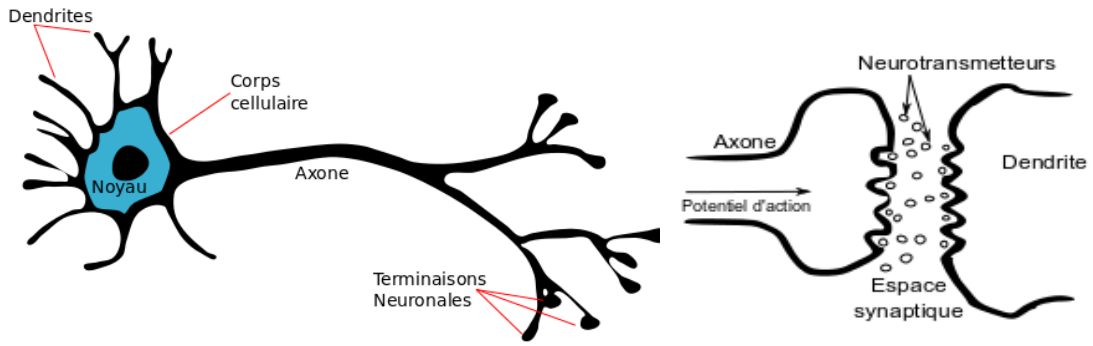

Figure 2.1: Schéma d'un neurone biologique et d'une synapse.

2.2. Neurone formel

Le neurone formel est une modélisation mathématique qui reprend les principes du fonctionnement du neurone biologique. McCulloch et Pitts [McCulloch et Pitts, 1943] ont développé un modèle du neurone formel (La figure 2.2 montre une description graphique du neurone formel). Ce modèle ne tient pas compte de l'aspect temporel des décharges du neurone, il code simplement une activité analogique comprise dans l'intervalle [0, 1]. Le neurone formelle de Mc. Culloch et Pitts se composent de :

- un vecteur de connexions W définissant les poids des synapses reliant le neurone émetteur et le neurone récepteur (efficacité des connexions synaptiques).
- un potentiel d'action Pot égal au produit scalaire du vecteur d'entrée par les poids:

$$Pot = W^T X \tag{2.1}$$

où X est le vecteur d'entrée, ou il s'agit simplement d'une somme pondérée des valeurs en entrée:

$$Pot = \sum_i w_i \, x_i \tag{2.2}$$

- un seuil d'activation θ au delà duquel la réponse du neurone est générée.
- une fonction d'activation f permettant le calcul de l'activité de sortie du neurone s = f (Pot − θ) où s représente la sortie.

Le neurone formel réalise d'abord une somme pondérée des entrées, à laquelle peut être ajoutée une constante appelée biais. Le résultat de la somme pondérée est comparée au seuil d'activation θ, Si la somme dépasse θ la sortie du neurone est 1, sinon elle vaut 0 dans le cas contraire.

Dans sa première version, le neurone formel était implémenté avec la fonction de Heaviside, puis il a été généralisé de différentes manières, en choisissant d'autres fonctions d'activations, comme la fonction sigmoïde, ReLU, softmax…

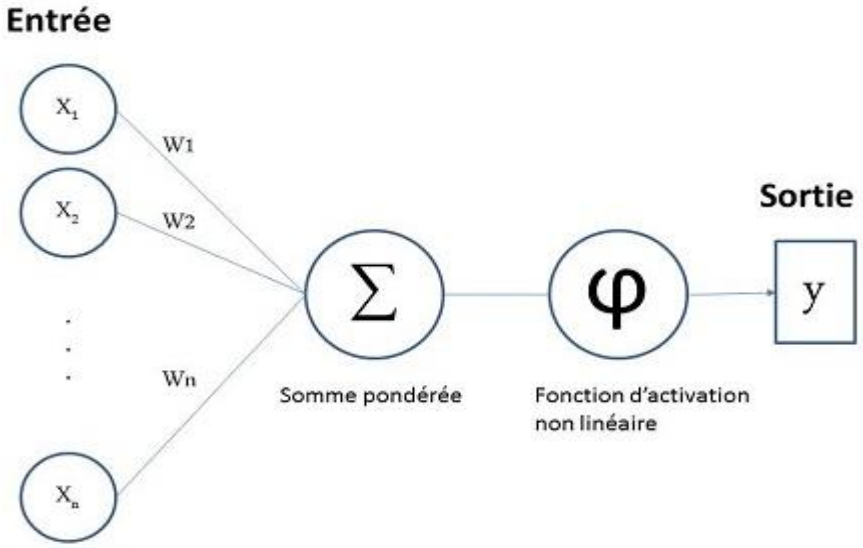

Figure 2.2: Schéma d'un neurone formel de Mc. Culloch et Pitts.

2.3. Apprentissage associatif hebbien

L'apprentissage hebbien est largement utilisé dans les domaines de la psychologie, de la neurologie et de la neurobiologie. C'est l'une des prémisses fondamentales de la neuroscience. La règle d'apprentissage de Hebb a été établie par Donald Hebb [Hebb, 1949], elle stipule que la modification de poids du réseau dépend de la co-activation répétée des neurones pré-synaptique et post-synaptique, c'est-à-dire, lorsque deux neurones sont excités conjointement, il se crée ou renforce un lien les unissant ce qui faisant croitre l'efficacité de la transmission d'information.

L'apprentissage hebbien peut être modélisé très simplement. La Figure 2.3 illustre son principe de fonctionnement. Il peut être décrit plus formellement de la manière suivante. Si deux neurones i et j sont reliés par une connexion (synapse) dont le poids est noté wij et que xi et xj représente le potentiel d'activation des neurones i et j (Le neurone pré-synaptique est défini par l'indice i et le neurone post-synaptique par l'indice j), la loi de Hebb décrivant l'évolution du poids wij dans le temps peut être modélisée par l'équation suivante:

$$\delta w_{ij} = \epsilon x_i y_i - \lambda w_{ij} - \lambda' w_{ij} x_i \qquad (2.3)$$

w_{ij} étant les poids synaptique en 2 neurones i et j ayant respectivement x_i et y_i comme activité, λ est le taux d'apprentissage, les termes λw_{ij} et $\lambda' w_{ij} x_i$ correspondent

respectivement à un oubli passif et un oubli actif. Les poids w_{ij} pouvant augmenter indéfiniment, ces oublis permettent une stabilisation de l'apprentissage.

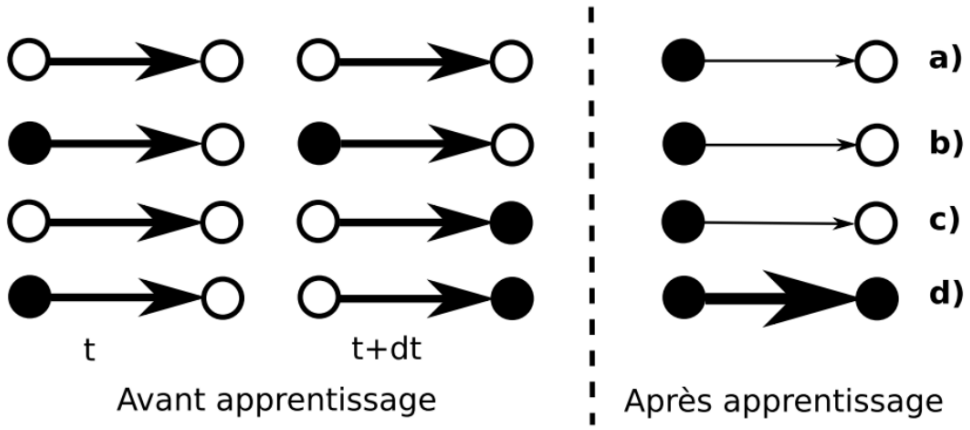

Figure 2.3: Principe de renforcement des poids pour la règle de Hebb.

2.4. Processus de catégorisation et classification

L'extraction des caractéristiques importantes de l'image (traits faciaux) et d'en réaliser une abstraction est nécessaire pour pouvoir reconnaître et reproduire les expressions faciales. Pour y arriver, la classification est l'une des manières efficaces permettant de catégoriser un sous ensemble des vues locales homogènes partageant les mêmes caractéristiques expressives. L'idée étant de comparer toute nouvelle vue locale à ces sous ensembles par des mesures de similarité ou de distance (euclidienne par exemple).

La classification peut être supervisée ou non supervisée et elle peut être en ligne ou hors ligne, considérant une approche incarnée de la robotique développementale, nous nous intéresserons principalement aux méthodes non supervisées et en ligne. Tout d'abord nous verrons la méthode des k-means qui est la méthode la plus simple qui illustre très bien la façon dont il est possible de classifier des données. Ensuite nous aborderons les cartes de Kohonen qui sont d'inspiration biologique et permettent de préserver la topologie des données. Puis nous présenterons l'algorithme du SAW (Selective adaptive winner) utilisé dans notre modèle, qui est proche de la règle de Kohonen et permet un apprentissage en ligne.

2.4.1. K-means

L'algorithme des k-moyennes (k-means) [MacQueen et al., 1967] est la méthode la plus simple et la plus utilisée qui illustre très bien la façon dont il est possible de classifier des données. Cette méthode propose de catégoriser un ensemble de données dans un nombre de classes fixées a priori (clusters). Il faut au préalable définir k clusters, une par centroïde, et les placer aléatoirement dans l'espace des données en les éloignant le plus possible les unes des

autres. Chaque nouvelle donnée est affiliée au centroïde le plus proche. La position des centroïdes est mise à jour en calculant le barycentre des données qui lui sont associées. Cette opération est répétée jusqu'à atteindre un état de convergence où l'erreur quadratique entre les données et les centroïdes est minimale.

2.4.2. La carte de kohonen

Des observations en neurobiologie ont montré que dans de nombreuses zones du cortex, des colonnes voisines ont tendance à réagir à des stimuli proches [Hubel et Wiesel, 1977 ; Knudsen et Konishi, 1979]. Ces observations ont mené Kohonen [Kohonen, 1989] à proposer un modèle de carte topologique auto-adaptative qui permet de coder des motifs présentés en entrée tout en conservant la topologie de l'espace d'entrée. L'algorithme de Kohonen propose une amélioration de la méthode des k-means en tirant profit des relations de voisinage pour réaliser une discrétisation dans un temps beaucoup plus court.

2.4.3. Self-Adaptative-Winner (SAW)

Notre système de reconnaissance des expressions faciales est basé sur un modèle d'apprentissage de catégorie inspiré de l'algorithme ART (Adaptive Resonance Theory) [Grossberg et Mingolla, 1985; Carpenter et Grossberg, 1987; Grossberg, 1988; Grossberg et Somers, 1991]. Ce modèle intitulé Selective Adaptative Winner "SAW" [Kanungo et al., 2002] est une variante du k-means couplé à un mécanisme de recrutement. Il permet une catégorisation en ligne et en temps réel. Sa règle d'apprentissage permet de catégoriser les caractéristiques visuelles VF de vecteurs d'entrées et de les classer en fonction d'une distance par rapport à des centroïdes comme le fait l'algorithme de k-means.

Chaque neurone du SAW peut représenter une catégorie. Lorsqu'il reçoit un signal de neuromodulation, un neurone peut être recruté pour représenter la catégorie correspondante à la vue locale perçue. La règle de mise à jour du SAW est décrite par l'équation suivante :

$$VF_j = net_j . H_{\max(\gamma, \overline{net} + \sigma_{net})}(net_j) \quad (2.4)$$

Avec net j l'activité du neurone j de la couche de sortie du SAW, $H_\theta(x)$ la fonction de Heaviside[1], γ la vigilance qui représente un seuil au-delà duquel le réseau effectue un recrutement. $\overline{net}$ σ_{net} et sont respectivement la moyenne et l'écart-type des neurones de sorties VF_j. Le calcul de l'activité des neurones de catégorisation est décrit par l'équation :

$$net_j = 1 - \frac{1}{N} \sum_{i=1}^{N} \left| W_{ij} - I_j \right| \quad (2.5)$$

[1] Fonction de Heaviside : $H_\theta(x) = \begin{cases} 1 & si\ \theta < x \\ 0 & sinon \end{cases}$

Le fonctionnement de ce réseau se base sur un apprentissage incrémental des catégories. Tous les poids synaptiques sont initialisés à 0 et l'apprentissage à proprement parlé est réalisé par l'adaptation des poids w_{ij} entre les neurones i de l'entrée et les neurones j du groupe de catégorisation. L'équation 2.6 décrit cette modification des poids w_{ij} :

$$\Delta w_{ij} = \delta_j^k \ (\ a_j\ (t) E_i] \tag{2.6}$$

Avec k = ArgMax (net_j), δ_j^k est le symbole de Kronecker [2] utilisé pour ne modifier les poids des liens que vers la catégorie gagnante. $a_j\ (t) = 1$ seulement quand le neurone j est recruté et 0 sinon.

La règle d'apprentissage permet à la fois un apprentissage en un coup et un moyennage des prototypes dans le temps. Cette règle mélange deux notions qui sont d'apprendre les choses nouvelles très rapidement et pouvoir s'adapter dans le temps.

L'évaluation des activités des catégories en fonction des vecteurs d'entrée se fait avec un mécanisme de comparaison entre les valeurs des neurones en entrée avec les poids appris pour chaque catégorie. La catégorie choisie est celle dont le vecteur se rapproche le plus. Si les valeurs de toutes les catégories sont en dessous de la vigilance γ alors une nouvelle catégorie est ajoutée (illustré par le recrutement d'un nouveau neurone) sinon les poids du neurone le plus actif sont moyennés.

[2] Fonction de Kronecker : $\delta_j^k = \begin{cases} x & si\ j = k \\ 0 & sinon \end{cases}$

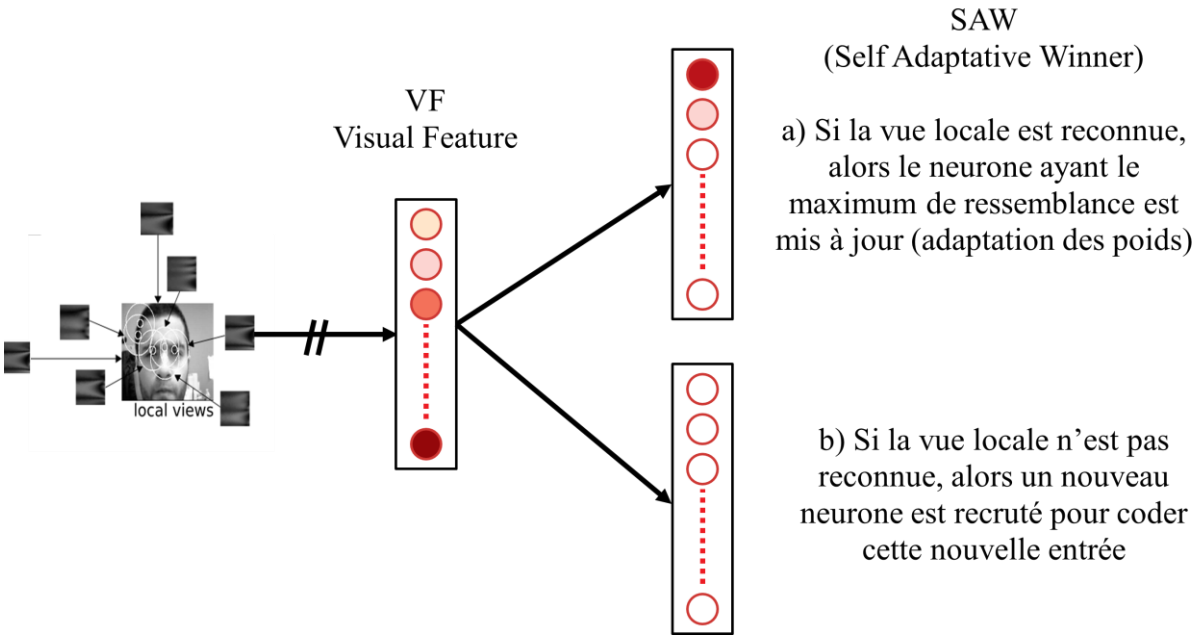

Figure 2.4: Principe de fonctionnement du SAW.

2.5. Conditionnement d'association

Le conditionnement est une forme d'apprentissage associatif observé chez l'homme et chez l'animal. Pavlov a introduit la notion de conditionnement pour ce que l'on appelle aujourd'hui conditionnement classique [Pavlov, 1927]. Dans son expérience, il a présenté un stimulus auditif juste avant de nourrir les chiens. Après quelques répétitions, les chiens commencent à saliver en réponse au stimulus, même sans nourriture. Pour généraliser, le concept consiste à associer un stimulus neutre conditionnel (par exemple un son) à un stimulus inconditionnel (par exemple plaisir ou douleur). En conséquence, la réponse réflexe inconditionnelle est également associée au stimulus conditionnel en tant que réponse conditionnelle.

L'un des algorithmes les plus utilisés pour modéliser un conditionnement classique est le principe de réduction de l'erreur quadratique moyenne (Least Mean Square (LMS)) proposé par Widrow et Hoff [Widrow et Hoff, 1960]. Cet algorithme consiste en la modification des poids synaptiques Wij connectant les stimulus conditionnels au groupe de conditionnement, jusqu'à trouver l'erreur quadratique moyenne minimale qui permet de converger des valeurs des sortie inconditionnelles Sj vers des valeurs de sorties désirées Sdj.
Où l'erreur à minimiser est définie par:

$$\xi = E\,[\,(S^d - W^T X)^2] \qquad (2.7)$$

L'apprentissage consiste ainsi à trouver le minimum global en fonction des poids synaptiques. La méthode utilisée pour atteindre à ce minimum est une descente des gradients où les poids sont modifiés comme suit:

$$\Delta w_{ij} = \lambda x_i \, (S_{dj} - S_j) \tag{2.8}$$

où w_{ij} est le lien entre le neurone de sortie Sj et le neurone d'entrée Xi, la sortie désirée Sdj et le taux d'apprentissage λ. Ici, Xi, Sj, et Sdj représentent respectivement le stimulus conditionnel, la réponse conditionnée et la réponse non conditionnée. Cette descente de gradient réduit l'erreur entre la sortie réelle et la sortie désirée en fonction des modèles d'entrée.

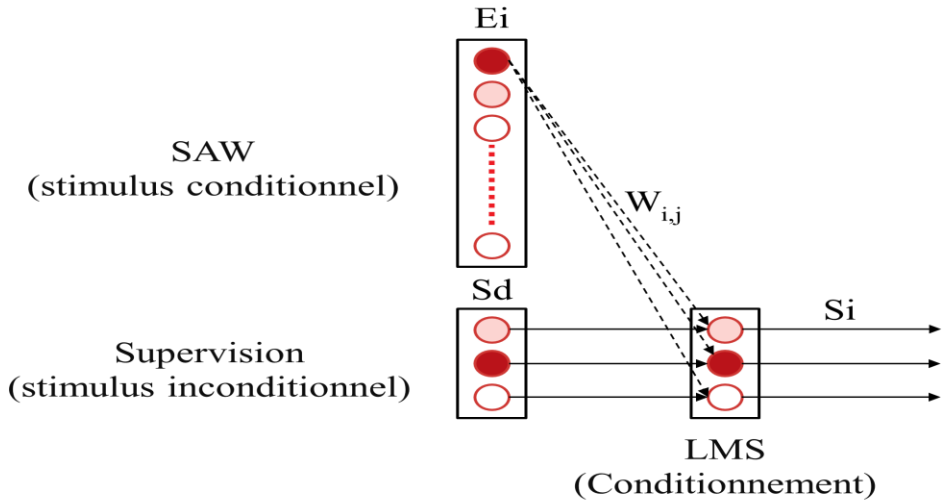

Figure 2.5: Principe de fonctionnement du LMS.

2.6. Compétition

Les mécanismes de compétition sont un type de réseaux de neurones qui se basent sur une comparaison des neurones les uns avec les autres, en fonction de l'interconnexion des voisins. Ainsi, avec des processus d'inhibition de voisinage une décision peut être prise sur la catégorie la mieux reconnue. Nous présentons ici deux mécanismes de compétition : le winner-take-all et les champs de neurones dynamiques.

2.6.1. Le Winner Takes all

L'idée de base du Winner Takes All (WTA) provient des modèles connexionnistes [Grossberg,1976, 1988; Kohonen, 1984; Feldman and Ballard, 1982]. Ils permettent de simuler les mécanismes de compétition existant entre neurones ou populations de neurones afin de prendre une décision. Le modèle courant utilise des groupes de neurones formels dont l'apprentissage est fixé par la règle de Hebb.

La compétition est donc le résultat d'une dynamique d'inhibition réalisée par le biais de liaisons inhibitrices latérales entre voisins. Après convergence, seul le neurone ayant la plus grande activité reste actif et inhibe tous les autres [Rumelhart and Zipser,1985; Lippman, 1987; Carpenter and Grossberg,1988].

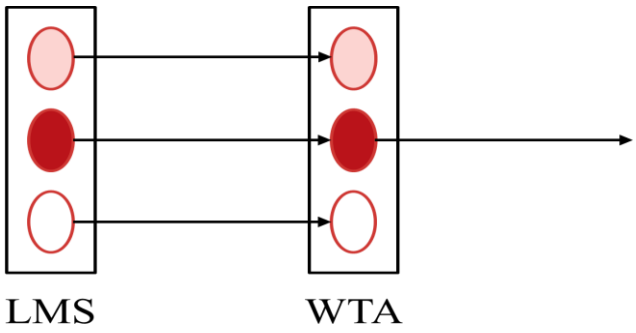

Figure 2.6: Principe de fonctionnement du WTA.

2.6.2. Champs de neurones dynamiques (Dynamic Neural Fields)

Le concept des champs de neurones est basé sur le fait que le cerveau utilise un codage de population [Taube et al., 1990; Bullmore et Sporns, 2009]. Contrairement au WTA, où l'information est codée par un seul neurone, dans un champ de neurones l'information est décrite par une population de neurones. Cette approche est issue des études sur l'anatomie du cortex. En effet, les neurones voisins ont généralement des activités similaires, selon le principe de redondance, chaque neurone possède une partie de l'information et l'information complète ne peut être obtenue qu'en combinant un grand nombre de neurones voisins.

La théorie des champs de neurones peut être attribuée à Beurle [Beurle, 1956] qui a observé des bulles d'activité dans le tissu cérébral. Wilson et Cowan [Wilson et Cowan, 1973] ont ensuite proposé le concept de neurones excitateurs et inhibiteurs pour modéliser les interactions au sein de populations de neurones. Aujourd'hui, l'un des modèles les plus couramment utilisés a été proposé par Amari [Amari, 1977]. Il est basé sur l'idée que les interactions excitatrices sont locales tandis que les interactions inhibitrices sont distales. Ceci est modélisé par un noyau d'interaction constitué d'un laplacien de gaussienne ou d'une différence de gaussienne. D'après les travaux d'Amari, [Schöner et al., 1995] ont proposé la théorie des champs neuronaux dynamiques (DNF) comme cadre pour le contrôle robotique.

Le modèle d'Amari [Amari, 1977] possède des propriétés intéressantes au sens de systèmes dynamiques (bifurcation, fusion, hystérésis, mémoire) en plus de la compétition que peut avoir un WTA.

3. Architecture PerAc

L'architecture PerAc (Perception-Action) est un modèle d'apprentissage sensori - moteur développé par l'équipe neurocybernétique du laboratoire ETIS pour le contrôle de robot interagissant avec leur environnement naturel [Gaussier et Zrehen, 1995, Gaussier et al., 1997, Gaussier et al., 2000, Gaussier, 2001, Giovannangeli et al., 2006, Giovannangeli et al., 2007, Giovannangeli et al., 2008]. Fondamentalement, PerAc permet un apprentissage en ligne des conditionnements entre des perceptions (exploration de l'environnement) et des actions à effectuer (mouvement des servomoteurs). Elle est constituée d'une première voie de bas niveau qui extrait des informations de base de l'entrée perçue (image, son…) afin de contrôler directement et approximativement les actions. La deuxième voie effectue la reconnaissance de la situation et permet d'apprendre le lien entre ce qui est reconnu dans le flux perceptif et le mouvement choisi. Cette voie permet de maintenir le comportement fourni par le système réflexe ou de l'éviter en cas de contradiction avec les contraintes de viabilité du robot.

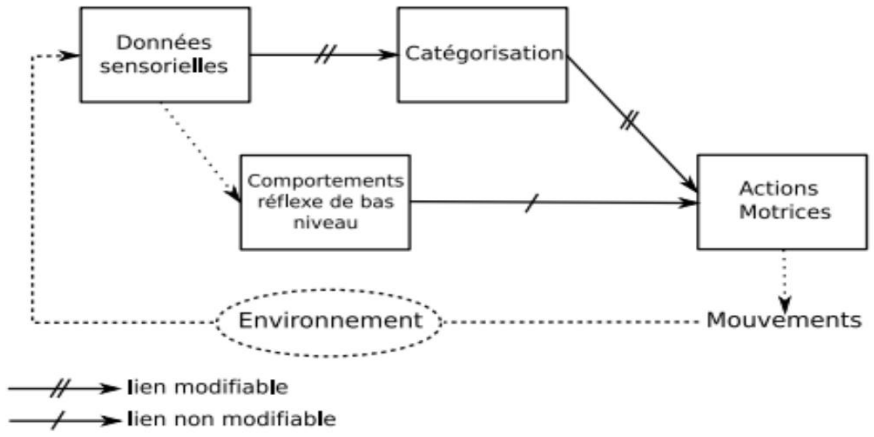

FIGURE 1.9 – Schéma de la l'architecture PerAC (Perception/Action)

Figure 2.7: Schéma de l'architecture PerAc.

4. Conclusion

Dans ce chapitre, nous avons présenté les outils essentiels pour le développement de l'architecture neuronale pour la reconnaissance des expressions faciales secondaires. Les réseaux de neurones qui ont été présentés ont des propriétés qui sont utiles pour la mise en place de notre modèle. En effet, les capacités d'apprentissage et d'adaptation de ces réseaux de neurones permettront au robot de pouvoir interagir correctement avec son partenaire dans un environnement non contrôlé et sans supervision externe. Le processus de catégorisation et classification permet un apprentissage en ligne et adaptation des poids. L'apprentissage par conditionnement et particulièrement le modèle de LMS entre en jeu pour l'association entre les caractéristiques visuelles du partenaire humain et l'état interne du robot. Enfin, le WTA et les champs de neurones dynamiques permettent de réaliser la compétition entre les différentes catégories et prendre une décision concernant l'expression faciale présentée.

Chapitre 3 : Système de reconnaissance des expressions faciales secondaires

1. Introduction

Les interactions émotionnelles en robotique autonome sont de plus en plus étudiées. Les architectures proposées se basent principalement sur des stratégies d'ingénierie ad-hoc qui exhibent des résultats impressionnants. Toutefois, elles ne permettent pas au système d'évoluer et de s'adapter à son environnement malgré l'utilisation des techniques d'apprentissage. En effet, il est nécessaire que le robot acquière une autonomie comportementale à savoir la capacité d'apprentissage et d'adaptation en ligne pour pouvoir agir et réagir dans un environnement naturel et faire face à des perturbations imprédictibles.

Dans ce chapitre nous montrons comment une tête robotique n'ayant au départ aucune connaissance du monde, peut développer la capacité d'apprentissage et la reconnaissance des expressions faciales à travers un jeu d'imitation et en utilisant une architecture sensori-motrice basée sur des réseaux de neurones.

2. Matériel utilisé

2.1. Outils de simulation de réseaux de neurones

Les travaux de l'équipe Neurocybernétique de l'ETIS utilisent en grande majorité deux outils de conception et simulation de réseaux de neurones : Coeos et Promethe [Lagarde et al., 2008a; Quoy et al., 2000]. Un autre outil développé et utilisé au sein de l'équipe fut employé dans le cadre de mon stage appelé Themis qui est une interface graphique permet de simplifier le lancement de plusieurs scripts à la fois. Ces outils sont développés dans l'équipe depuis une vingtaine d'années et évoluent au fur et à mesure des besoins de modélisation et de contrôle robotique. Durant mon stage, j'ai eu l'occasion d'utiliser ces outils pour développer mon architecture de contrôle neuronale, Ainsi, j'ai pu développer certaines fonctionnalités pour répondre à des problèmes particuliers pour lesquels aucune solution n'avait été implémentée.

2.1.1. Interface de conception : Coeos

Coeos est une interface graphique de conception de réseaux de neurones artificiels distribués. Elle permet de construire et de compiler des fichiers détaillant la structure du réseau de neurones et spécifiant la topologie des connexions entre les nœuds informatiques, afin d'être interprétées par la suite par le simulateur Promethe. Le logiciel permet de construire des réseaux complexes, tel un graphe orienté, à partir de "briques" de bases (ou

groupes) liées par des arêtes (Figure 3.1). Les groupes sont développés en amont et permettent de réaliser soit des calculs de types neuronaux avec des règles d'apprentissage (e.g. une catégorisation par un SAW) ou encore des fonctionnalités purement algorithmiques sans apprentissage qui seraient sous-optimisées sous forme neuronales. Certains groupes permettent de s'interfacer avec le matériel embarqué du robot, comme pour récupérer des informations de capteurs (un flux vidéo ou de capteurs ultrason). Les liens possibles offrent plusieurs types de connectivités : connexion plastique, réflexe/inconditionnel ou encore de neuromodulation, avec des topologies diverses comme un vers tous, un vers un ou un vers voisinage.

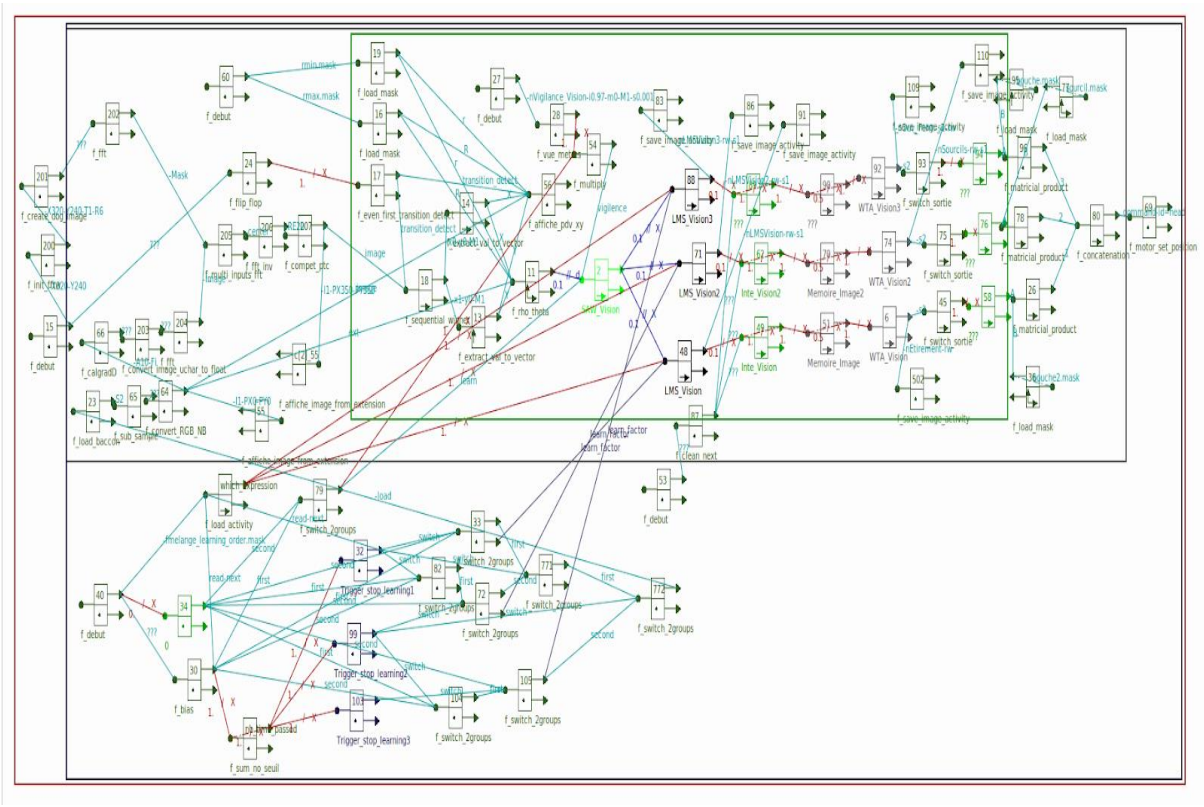

Figure 3.1: Capture d'écran de l'interface de Coeos montrant l'architecture neuronale de l'apprentissage des primitives motrices.

2.1.2. Simulateur temps-réel distribué : Promethe

Promethe est un simulateur de réseaux de neurones développé par les membres de l'équipe neurocybernétique du laboratoire ETIS, qui permet l'exécution distribuée de nœuds de calcul indépendants (groupes de neurones). Ces nœuds sont des fonctions écrites en langage C et implémentant diverses fonctionnalités: algorithmes neuronaux, opérations arithmétiques, algorithmes de traitement d'images, interfaces d'E / S, etc. La fonction

principale de ce simulateur est l'exécution des scripts conçus avec l'aide de Coeos et qui décrivent les réseaux de neurones et contiennent les fonctions associées aux différentes groupes de neurones (Figure 3.2).

Promethe planifie l'exécution des groupes (l'activité des neurones) et l'apprentissage (le poids des liens entre les groupes) en se basant sur des jetons transmis d'un nœud à ses successeurs une fois qu'il est exécuté. Chaque opération du flux de traitement d'informations peut être calculée dès que les informations des modules précédents sont mises à jour. Les modules indépendants sont exécutés en parallèle (c'est-à-dire dans des threads séparés).
On peut également concevoir différentes boucles s'exécutant avec des constantes de temps différentes par l'utilisation d'échelle de temps, qui fonctionnent comme des boucles *for* imbriquées. L'aspect d'exécution temps réel peut être géré par des groupes émettant des jetons d'exécution avec une période définie, et s'assurant que les contraintes temps réel sont bien respectées pour la mise à jour du réseau de neurones. Enfin, certains groupes algorithmiques peuvent gérer des aspects de communication réseau en recevant et transmettant des activités neuronales à d'autres parties de l'architecture.

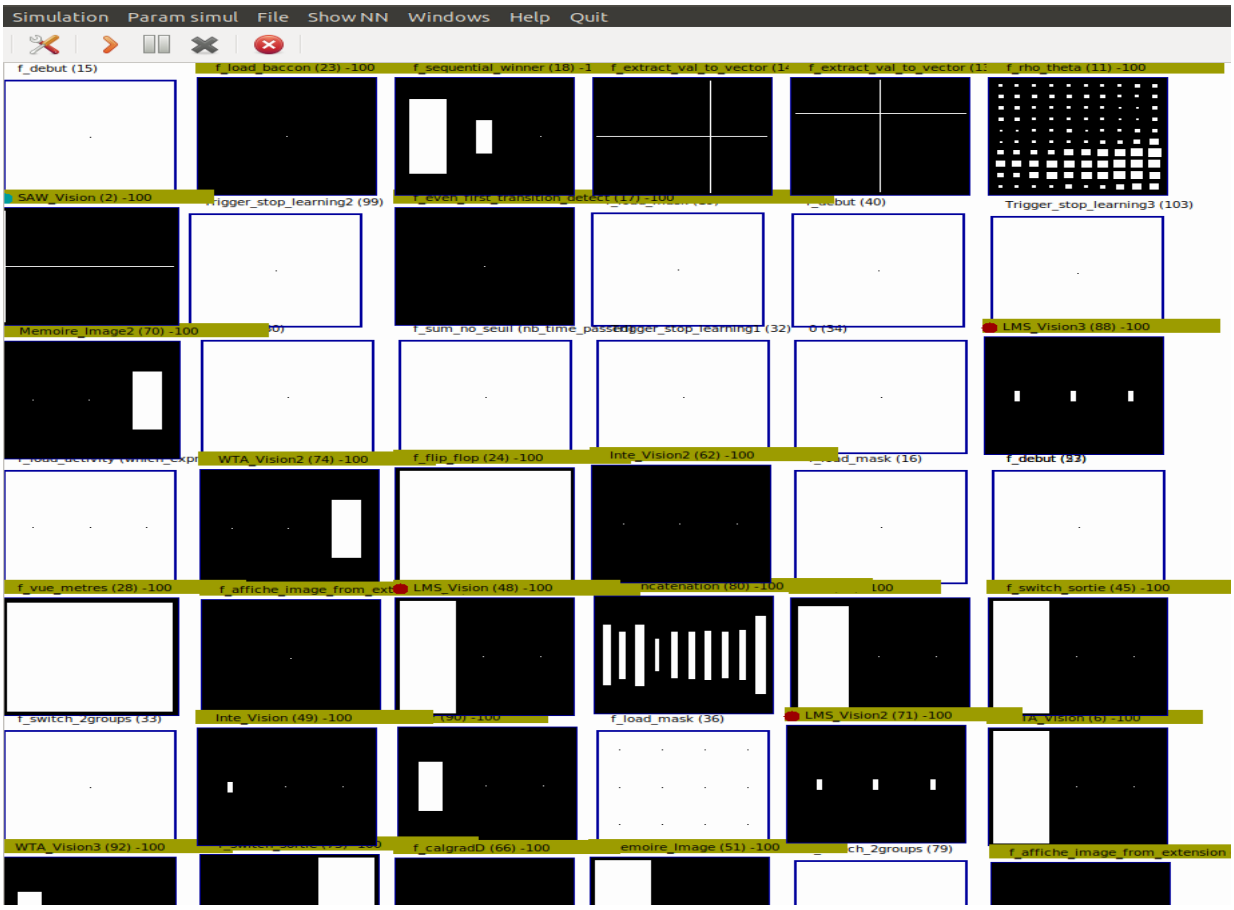

Figure 3.2 : Capture d'écran de l'interface de contrôle de Promethe.

2.1.3. Themis

Une interface graphique permettant de simplifier le lancement de plusieurs scripts en même temps. Il lit les fichiers ".net" et peut générer automatiquement la ligne de commande nécessaire pour exécuter les différents promethe. Il a aussi des raccourcis pour lancer coeos sur les scripts en question.

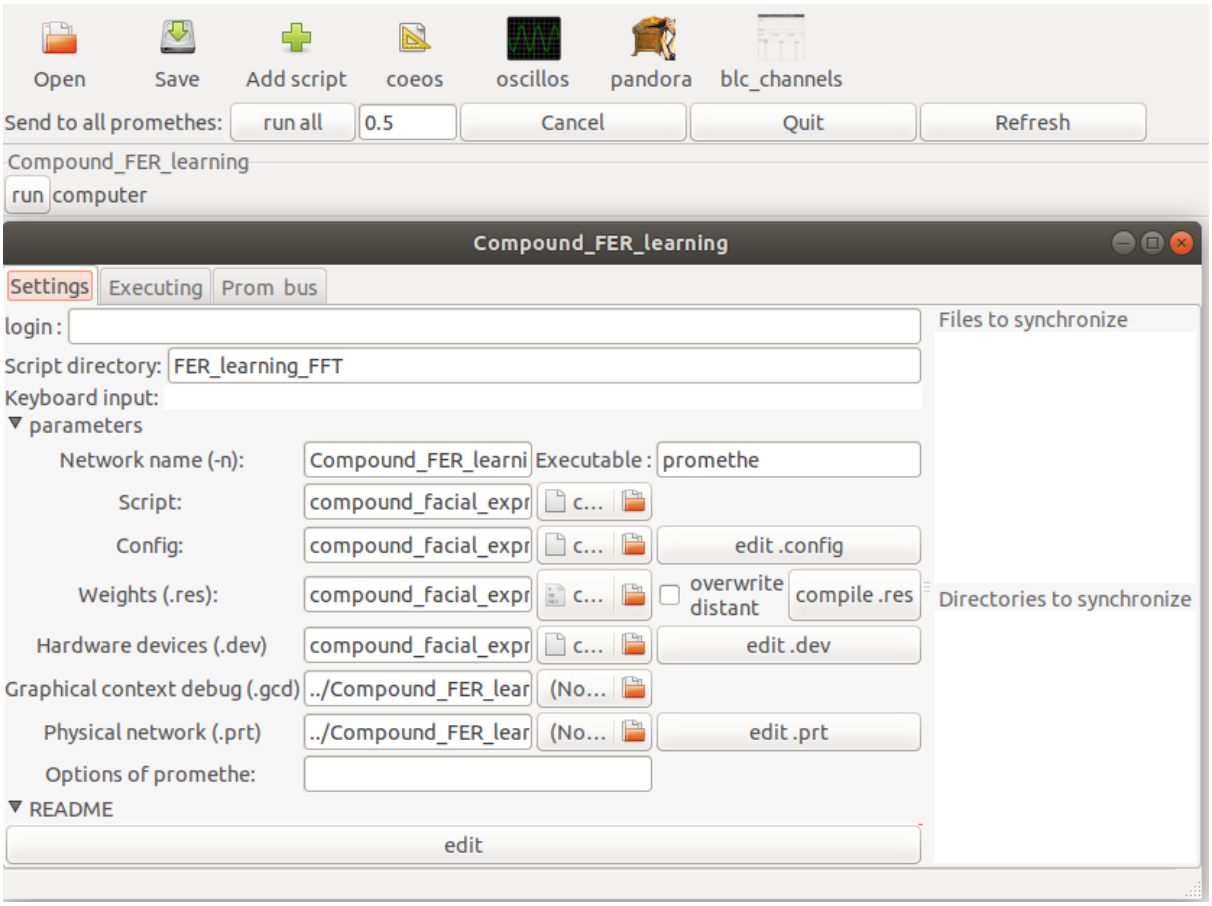

Figure 3.3 : Capture d'écran de l'interface de Themis.

2.2. La tête robotique

Notre système de reconnaissance d'expressions faciales pour l'interaction émotionnelle homme- robot est implémenté sur une tête robotique équipée de 10 servomoteurs agissant chacun comme un muscle du visage. Cette tête expressive présente les caractéristiques minimales nécessaire pour produire des expressions faciales émotionnelles humaines. La bouche peut se déformer, s'ouvrir et se fermer et les sourcils peuvent s'orienter, se plier et se hausser. Un œil est doté d'une caméra standard PAL permet à la tête de voir ce qui se trouve en face d'elle et fournir des images couleurs au robot.

Un module de contrôle Pololu est utilisé pour maintenir les servomoteurs dans une position donnée (contrôle en position) via une liaison USB. Tous les servomoteurs sont contrôlés en même temps ce qui permet une bonne dynamique expressive (changement d'expression en 200 à 400 ms).

La tête expressive est munie de 10 degrés de liberté (ddl) chacun actionné par un servomoteur. Il y quatre degrés de liberté dans les sourcils, un autre contrôle le tilt du front, cinq ddl contrôlent la bouche (deux pour une ouverture horizontale, deux pour le tilt et un pour la mâchoire). La tête expressive est capable d'exprimer 4 émotions basiques (joie, tristesse, colère, surprise) plus une expression neutre ([Izard, 1971] ; [Ekman et al., 1972] ; [Ekman and Friesen, 1978]). Ainsi que des expressions secondaires avec des niveaux d'intensités variées en jouant sur la position des servomoteurs. Ces expressions exprimées par la tête vont lui permettre d'interagir émotionnellement avec le sujet humain.

Ce dispositif n'a pas une réelle ressemblance humaine, mais il nous permet de distinguer les différentes expressions faciales exprimées par le robot.

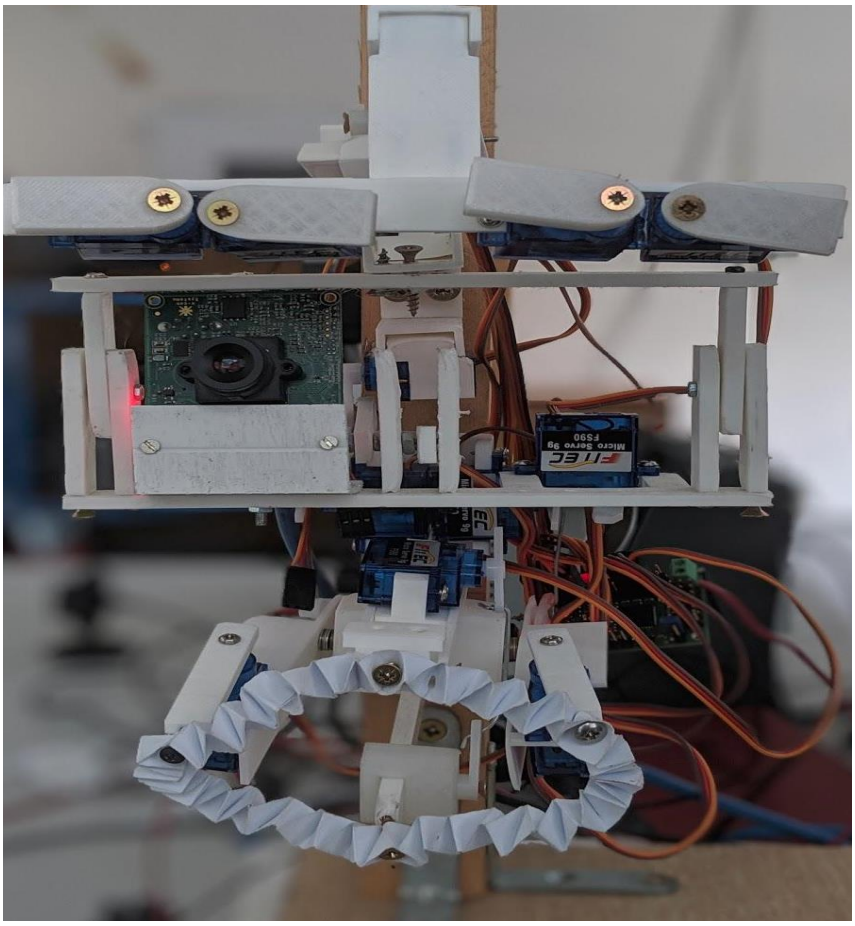

Figure 3.4: La tête robotique utilisé pour l'affichage des expressions faciales.

3. Modèle théorique

Nous considérons une interaction non verbale, à travers les expressions faciales émotionnelles entre deux agents sachant qu'un des deux agents est supposé être un adulte (agent 1) avec une parfaite reconnaissance et reproduction de l'expression émotionnelle des autres, par contre le second agent est considéré comme un nouveau né (agent 2) qui n'a pas de connaissances priori du monde extérieur et donc nécessite un apprentissage du rôle sociale des émotions [Boucenna, 2011]. Dans notre cas cet apprentissage se fait par un jeu d'imitation entre les deux agents. Nous supposons simplement que l'agent 2 est doté de mécanisme sensori-moteur lui permettant d'associer des sensations avec des actions. L'apprentissage des expressions faciales est possible si et seulement si l'agent 2 produit des expressions et que l'agent 1 l'imite, pendant cette phase d'apprentissage, les deux agents reçoivent un signal visuel (Vi vision de l'agent i). Cela peut être appris et reconnu par le groupe de neurones VFi (caractéristiques visuelles de l'agent i), VFi pouvant être le résultat d'un apprentissage non supervisé tel que SAW (Self Adaptative Winner) [Kanungo et al. [2002], réseau ART [Grossberg, 1987] ou une carte de Kohonen [Kohonen, 1989b]. Ainsi, la présence d'un visage exprimant une primitive motrice particulière déclenchera l'activation de plusieurs neurones correspondant à une certaine caractéristique faciale dans le groupe VFi :

$$VFi = c(Ai1.Vi) \tag{3.1}$$

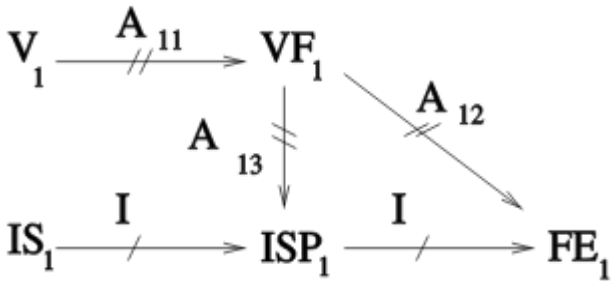

Figure 3.5: Représentation schématique d'un agent.

où c est un mécanisme de compétition, Ai1 représente la matrice de poids des neurones dans le groupe de reconnaissance de l'agent i permettant une quantification des caractéristiques (bouche, sourcils ...). Par ailleurs les deux agents possèdent un état interne (ISi) qui définit leurs état actuel (la faim, la peur, le plaisir…). La détection d'un état interne particulier déclenchera un état émotionnel ISPi (Internal State Prediction) qui dépendra de la

reconnaissance visuelle VFi (Visual Features) qui elle dépend à son tour du signal visuel Vi . FEi est le dernier groupe de neurones qui déclenchera l'expression faciale de l'agent i.

4. Protocole expérimental

Afin que la tête robotique puisse imiter les primitives motrices du sujet humain, un scénario d'apprentissage est nécessaire vue que notre système s'exécute en ligne. Notre scénario expérimental se divise en deux phases d'interaction basées sur un jeu d'imitation entre le robot et le sujet humain.

Durant la phase d'apprentissage, le robot produit des primitives motrices liées à son état interne (haussement de sourcil, ouverture de bouche, sourire,...) et en même temps, nous demandons à l'expérimentateur de l'imiter. Entre chaque mouvement moteur, le robot repasse à l'état neutre pour éviter les mauvaises interprétations par un observateur et donner le temps au sujet humain pour décontracter ses muscles du visage. Une procédure équivalente est utilisée en psychologie expérimentale pour éliminer le biais du set-up expérimental [Boucenna, 2011].

La tête robotique affiche les trois primitives motrices choisies arbitrairement et qui semblent être suffisantes pour exprimer une palette expressive variée. Cette séquence sera répétée trois fois pour améliorer la qualité de l'apprentissage. Une fois l'apprentissage fini (environ 3 minutes), le générateur des primitives motrices est stoppé. La phase de reconnaissance pourra commencer. Nous demandons alors à l'expérimentateur de produire une expression pendant quelques secondes et nous testons la capacité du robot à mimer à son tour l'expression faciale du partenaire humain.

5. Apprentissage des primitives motrices

5.1. Introduction

Afin d'obtenir un robot capable de produire n'importe quelle expression faciale, plutôt qu'un ensemble limité d'expressions comme les travaux précédents, nous proposons un modèle neuronale basé sur les réseaux de neurones non supervisés permettant l'apprentissage des groupements musculaires du visage se contractant ensemble pour produire une expression. Ceux sont ces groupes musculaires que nous appellerons primitives motrices qui seront apprises séparément par le robot. La tête du robot étant composée de dix servos moteurs agissant chacun comme un muscle du visage. Ceux-ci sont regroupés par

groupements musculaires ou groupe de servos moteurs fonctionnant ensemble. Trois primitives seront exploitées:

- Le haussement et le froncement des sourcils (P1): cinq servos moteurs contrôlent ce groupement musculaire. Cette primitive donne un air de surprise ou de colère au robot du moins pour la partie haute de la tête robotique.
- L'ouverture (du milieu) de la bouche (P2): un seul servo moteur contrôle cette primitive.
- La position des coins de la bouche (P3) est contrôlée par 4 servomoteurs. Ces servomoteurs sont capables d'exprimer un sourire ou une moue.

Dans le premier modèle, nous avons choisi d'utiliser trois neurones codant pour trois niveaux d'intensité différents par primitive motrices. Pour les sourcils correspondant à la primitive 1, trois neurones sont utilisés pour coder trois positions motrices: froncement, haussement et sourcils neutre. Pour la bouche (primitive 2), trois positions motrices sont également utilisées pour l'ouverture, la fermeture et la bouche neutre. Pour la primitive 3 correspondant à un sourire, une moue ou une bouche neutre, trois neurones sont également utilisés pour coder ces positions motrices [Boucenna, 2011].

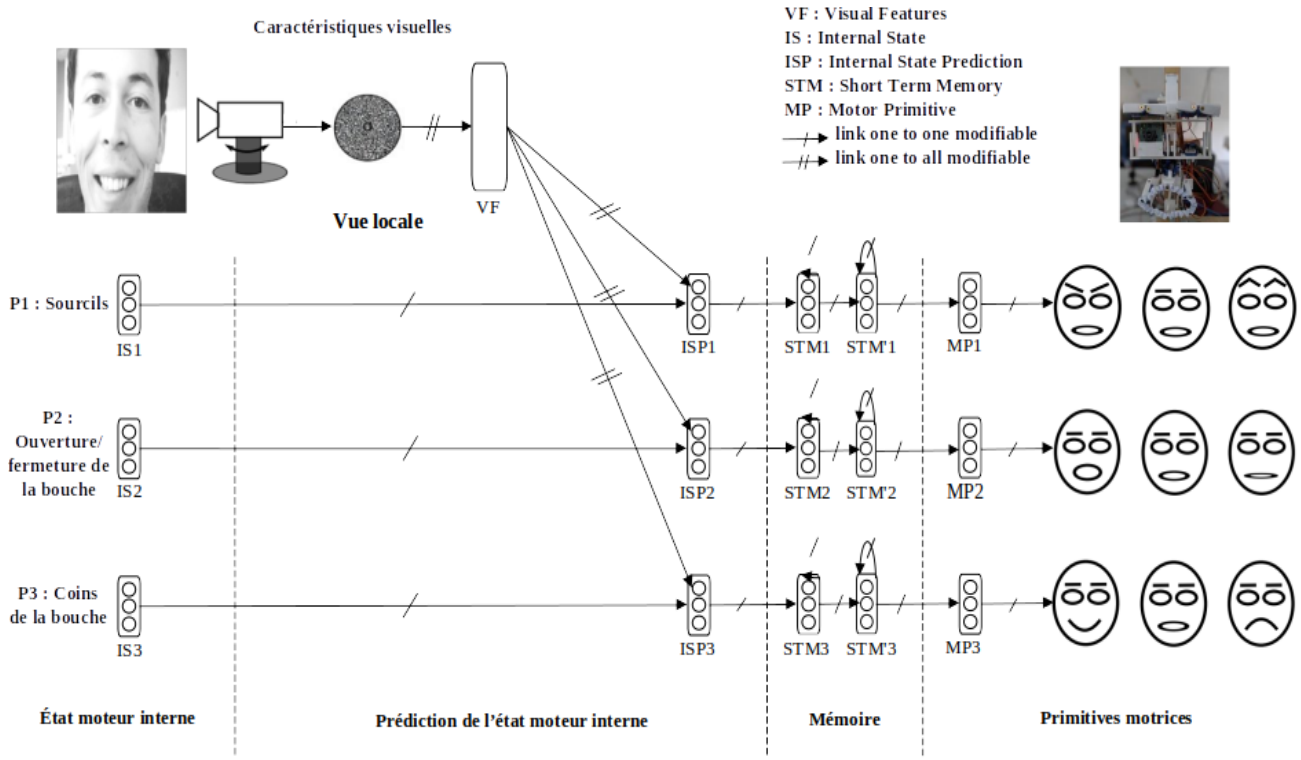

Figure 3.6: Modèle d'apprentissage des primitives motrices.

5.2. Traitement visuel bas niveau

Afin que l'apprentissage soit autonome, nous cherchons à éviter des mécanismes ad-hoc non compatible d'un point de vue développemental tels que la détection et le cadrage du visage utilisés dans la majorité des travaux de reconnaissance des expressions faciales. Pour cela, L'équipe Neurocybernetique a proposé un modèle biologiquement plausible de cellules visuelles qui a été utilisé dans plusieurs travaux notamment pour la navigation autonome et l'interaction homme-robot [Maillard et al., 2005 ; Giovannangeli et al., 2006; Giovannangeli et Gaussier, 2008 ; Boucenna, et al., 2014]. Ce modèle est basé sur l'exploration des points de focalisation de la scène. Une carte de saillance est obtenue par une méthode de détection de coins afin d'extraire des points d'intérêt qui sont les maxima locaux d'une convolution entre un filtre DoG (Difference of Gaussian) et l'image de gradient (obtenue par l'application d'un filtre de Canny-Derriche sur l'image en niveaux de gris)(Figure 3.7**)**. Notre système n'utilise pas l'image brute mais son gradient afin de minimiser l'impact des changements d'éclairage.

Les avantages d'utiliser ce mécanisme visuel par rapport aux autres méthodes :

- Il permet au système de focaliser plus sur les coins et les fins de lignes de l'image (sourcils, coins de bouche, etc).
- Un algorithme neuromimétique qui modélise bien le comportement de certaines cellules visuelles.
- Un coût computationnel réduit et très peu de points sont extraits.

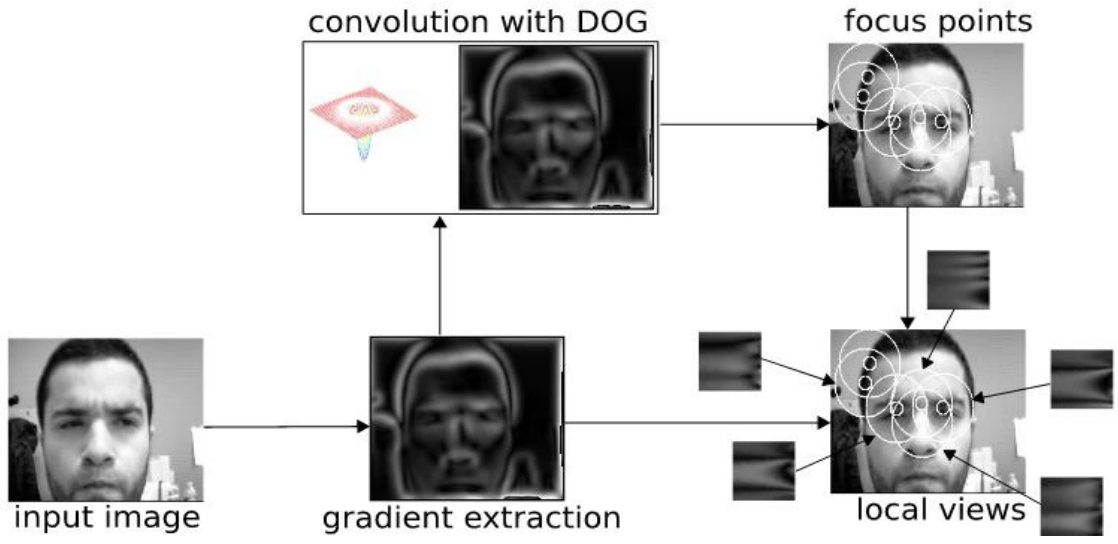

Figure 3.7: Processus visuel bas niveau.

Ensuite, Un mécanisme de compétition entre les points de focalisation permet de sélectionner les points les plus intenses (en termes de contraste et de rayon de courbure).

Enfin, les points de focalisation sélectionnés sont utilisés pour extraire une vue locale, à l'aide de la transformée log-polaire [Schwartz, 1980] qui permet d'augmenter la robustesse des vues locales extraites par rapport aux petites rotations et aux changements d'échelles. Cette transformation est centrée sur le point de focalisation et prend en compte une zone autour de celui-ci de taille 10x10 correspondant à un cercle de R pixels de rayon dans l'image d'entrée. Ce qui donne un vecteur signature de dimension 100 (Figure 3.8).

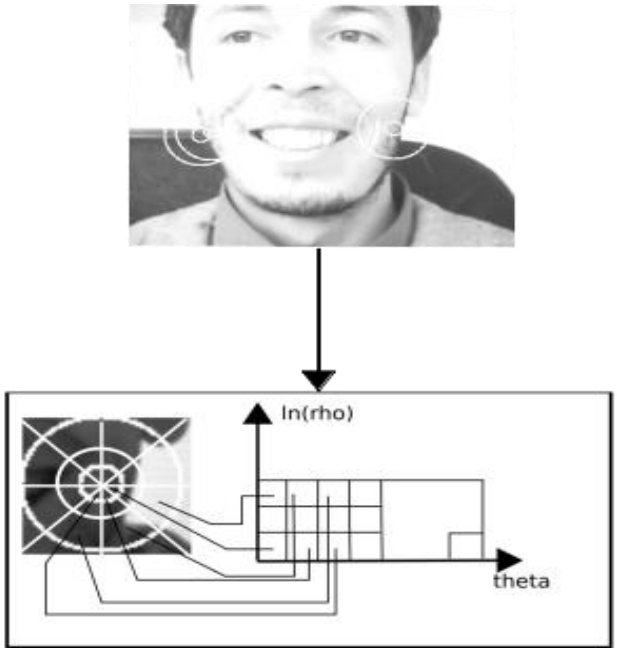

Figure 3.8: Extraction des vues locales par la transformée log-polaire.

5.3. Architecture de contrôle neuronale

5.3.1. Introduction

Dans une perspective développementale nous nous sommes inspirés de la capacité du bébé à apprendre des expressions faciales sans signal de supervision [Gergely et Watson, 1999, Andry et al., 2001].

Un modèle neuronal développé au laboratoire [Gaussier et al., 2004, Gaussier et al., 2007, Boucenna et al., 2008] permet au robot de reconnaître des expressions faciales basiques (Figure 3.9). Ces travaux ont été développés conjointement avec le groupe de Jacqueline Nadel du Centre Émotion de l'hopital Pitié-Salpêtrière qui s'intéresse par le rôle de l'imitation dans le développement du bébé et de l'enfant avec autism.

Dans cette section nous présentons en détails les différents processus d'apprentissage utilisés pour le développement de notre architecture sensori-motrice permettant à reconnaître les expressions faciales par l'apprentissage des primitives motrices.

5.3.2. Catégorisation

Les vues locales transformées en log-polaire sont apprises par recrutement de nouveaux neurones, cet apprentissage est réalisé par le groupe de neurone VF (caractéristiques visuelles) en utilisant l'algorithme de catégorisation SAW (Self Adaptative Winner) décrit dans le chapitre 2. Au début, le SAW ne dispose aucune vue locale apprise. Cependant, au fur et à mesure de l'exploration du monde, des nouveaux neurones sont recrutés codant des nouvelles caractéristiques visuelles. Chaque nouvelle entrée au SAW est comparée aux vues locales déjà apprises, si une caractéristique est partiellement reconnue (c-à-d l'activité du neurone codant cette nouvelle caractéristique est en dessus de la vigilance γ), alors le neurone le codant avec le maximum de ressemblance est moyenné pour ainsi généraliser les connaissances (adaptation des poids). Sinon, si la reconnaissance de cette nouvelle vignette est en dessous de la vigilance γ (trop différente de celles déjà apprises), une nouvelle catégorie est ajoutée en recrutant un nouveau neurone pour coder cette nouvelle vignette.

5.3.3. Conditionnement

Les vues locales apprises sont associées à une primitive motrice grâce au groupe de prédiction de l'état interne du robot ISP. L'association se fait par un simple mécanisme de

conditionnement utilisant la règle des moindres carrés (LMS) qui permet de modifier les poids de la sortie du SAW en minimisant l'erreur entre l'activité des neurones des vues locales apprises et le groupe de neurones de l'état interne du robot IS.

5.3.4. Mémorisation

Nous avons choisi d'analyser seulement 10 points de focalisations par image, et 20 images par échelle de temps pour permettre au système d'être réactif durant l'interaction avec le partenaire humain. Pour cela nous avons introduit deux mémoires à court terme STM.

La première mémoire STM1 permet d'accumuler tous les points de focalisation appartenant à la même image. Cependant, La seconde mémoire STM2 a été utilisée comme mémoire glissante sur les images consécutivement analysées. Ces mémoires rendent le contrôle des servomoteurs plus stable pour l'interaction [Boucenna, 2011].

5.3.5. Compétition

La prise de décision est réalisée par un mécanisme de compétition de type Winner-Takes- All qui active la catégorie gagnante ayant la plus grande activité et inhibe les autres. Celle-ci transmet son signal par la suite aux servomoteurs pour afficher la primitive motrice correspondante.

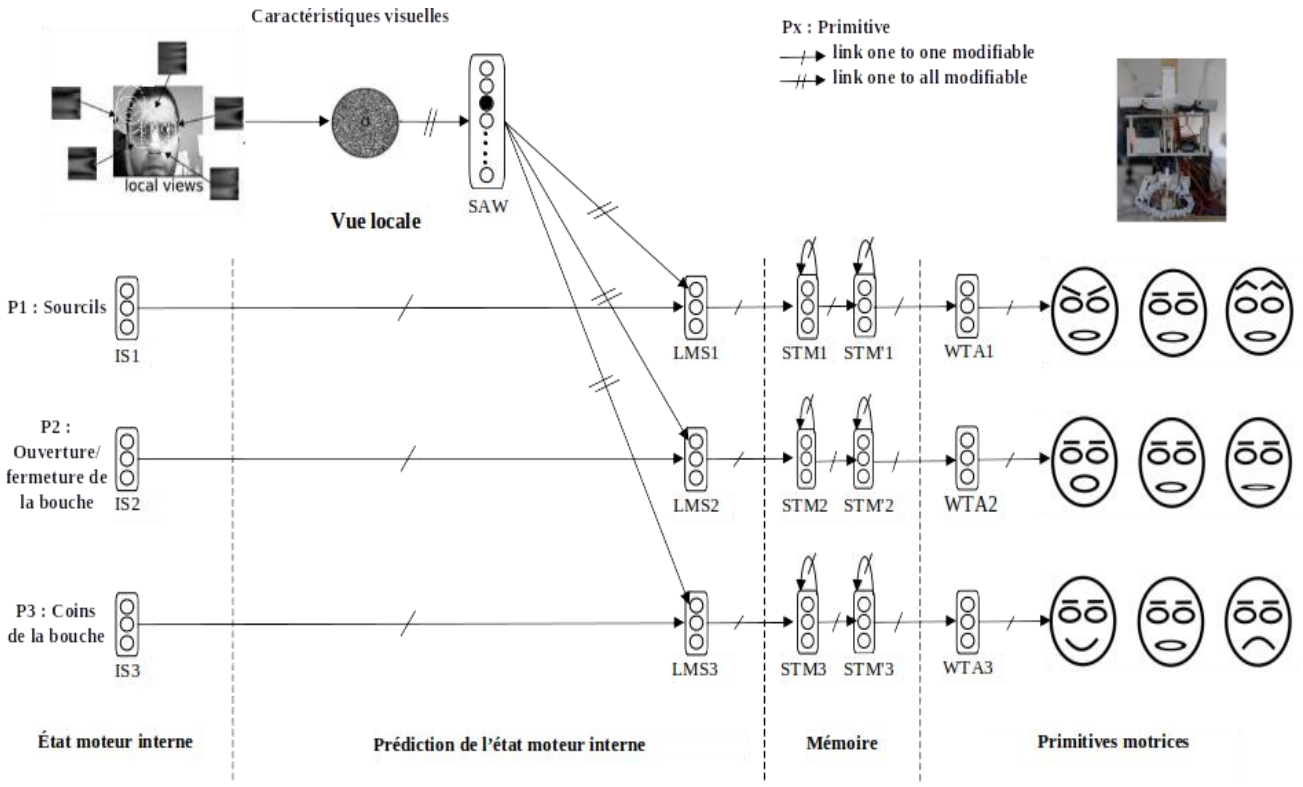

Figure 3.9: L'architecture neuronale pour l'apprentissage des primitives motrices.

5.4. Amélioration du modèle par l'utilisation des champs de neurones dynamiques

Le premier modèle de la reconnaissance et la production des primitives motrices utilise 3 neurones chacun codant le niveau d'intensité de plusieurs groupements musculaires. Cependant, l'être humain produit des expressions faciales avec des niveaux d'intensités différents. Afin d'améliorer la capacité du robot à produire n'importe quelle expression faciale avec une palette d'intensité expressive variée, nous remplaçons le WTA qui code l'intensité par un seul neurone par les champs de neurones dynamiques qui sont plus robustes et permettant de coder l'intensité par une population de neurones et obtenir plus de finesse vis à vis du contrôle moteur de la tête expressive.

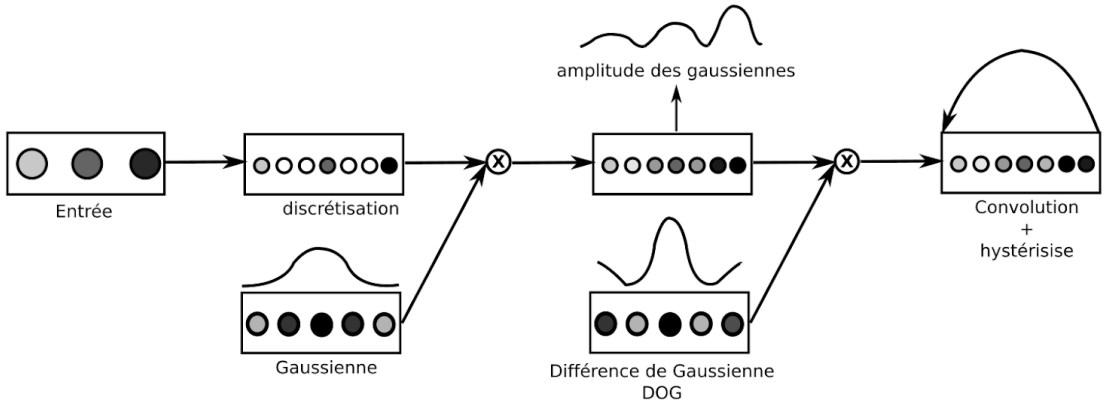

Figure 3.10: Modèle de champs de neurones dynamiques.

L'ajout du champ de neurone induit un contrôle plus précis de l'intensité motrice comparativement aux 3 neurones codant pour les 3 niveaux d'intensités. Les 3 neurones sont étalés sur une population de neurones (31 neurones codant chacun pour une intensité). Dans cette population, tous sont inactifs sauf les neurones extrêmes et le neurone central codant chacun pour les intensités extrêmes et intermédiaire. Une convolution avec une gaussienne est réalisée sur cette population permettant d'avoir un voisinage actif et un étalement des activités neuronales. Dans un second temps, une différence de gaussienne est réalisée, le but étant de mettre en compétition les neurones voisins. Enfin, une hystérésis temporelle est appliquée permettant ainsi une plus grande stabilité dans le temps (mémoire) [Boucenna, 2011].

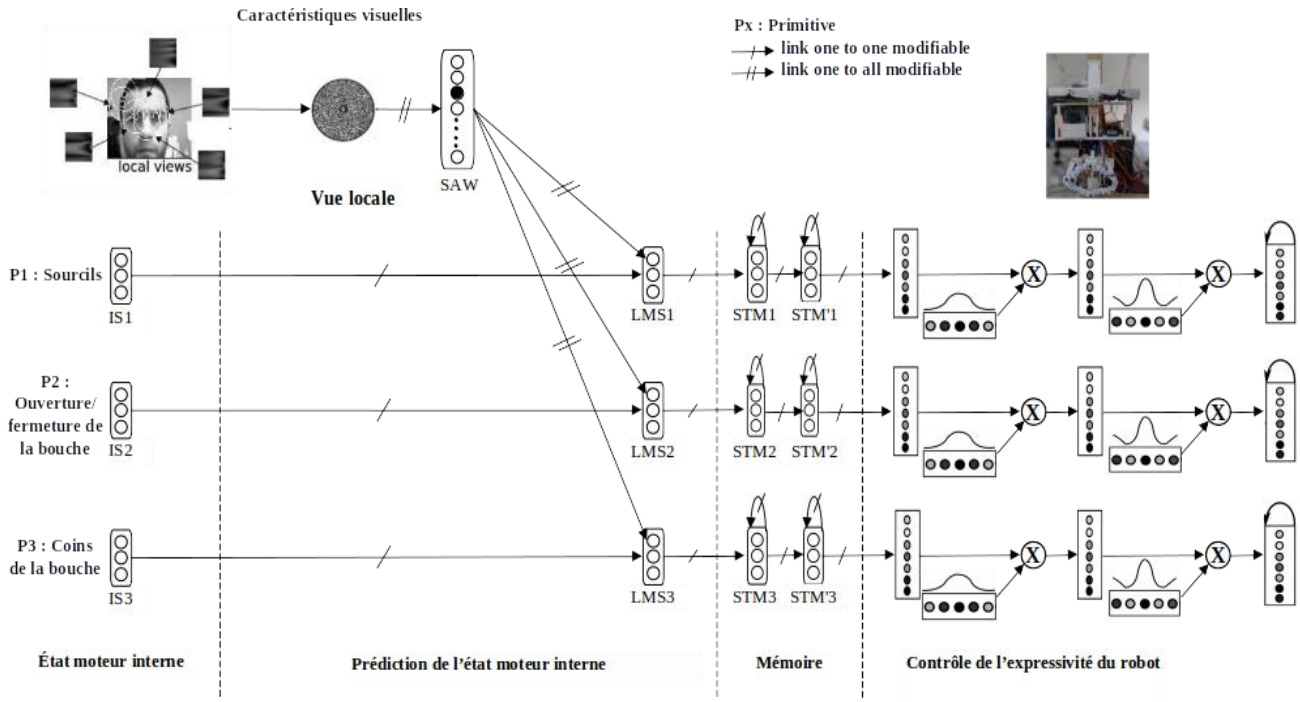

Figure 3.11: Reconnaissance des intensités motrices à l'aide des champs de neurones.

6. Performances et résultats

6.1. Apprentissage en ligne

Pour tester les performances de notre modèle, 51 individus (stagiaires, doctorants,...) qui ne sont pas experts du système ont interagi avec la tête du robot. Durant la phase d'apprentissage, les expérimentateurs doivent imiter le robot. Une fois l'apprentissage est terminé les rôles sont inversés, le partenaire fait une expression et le robot doit l'imiter.
Le taux de confusion de la reconnaissance en ligne de chaque primitive motrice est présenté dans les tableaux ci-dessous:

	Robot's mouth			
		Pout	Smile	Neutral
Human's mouth	Pout	74	0	26
	Smile	0	92	8
	Neutral	0	0	100

	Robot's mouth			
		Frowning	Raised	Neutral
Human's mouth	Frowning	78	0	22
	Raised	2	86	12
	Neutral	3	1	96

	Robot's mouth			
		Open	Closed	Intermediate
Human's mouth	Open	81	1	18
	Closed	0	68	32
	Intermediate	0	13	87

Figure 3.12: Tableaux montrant le taux de confusion pour les 3 primitives motrices apprises durant l'apprentissage en ligne.

Ces tableaux reflètent le taux de reconnaissance de notre système aux différentes primitives motrices affichées par les différents partenaires. Comme on peut le voir sur les tableaux, les résultats sont assez bons, quoique quelques fois notre système confonde la

primitive de fermeture de la bouche avec la position intermédiaire de la bouche, ainsi que la moue avec la bouche neutre. Cela est dû à la mauvaise expressivité des individus qui ont du mal à exprimer certaines primitives et également la succession des expressions affichées par la tête robotique sont trop rapide pour certains, ou pas claire pour d'autres. En plus du manque de ressemblance des mêmes émotions entre la première séquence et la deuxième ou la troisième fait que le SAW recrute trop de neurones ce qui ralenti un peu le processus de reconnaissance. Dans un cas idéal lorsque la procédure d'apprentissage est bien suivi par le sujet humain et il imite bien le robot, les performances de notre système sont très bonnes, il arrive à reconnaître les expressions affichées par l'utilisateur sans faute.

6.2. Apprentissage hors ligne

Après avoir passé les 51 sujets et testé la capacité du robot à reproduire les expressions faciales exprimées par les différentes personnes, nous passons maintenant à la validation de notre modèle. Pour cela, nous avons constitué notre base de données contenant 27540 images divisées sur l'ensemble des 51 sujets. Nous capturons 20 images pour chaque intensité de primitive motrice et pour chaque sujet la séquence d'interaction avec la tête robotique est répétée trois fois afin d'améliorer la qualité de l'apprentissage et collecter un grand nombre d'images. Les images de la base de données sont sauvegardées et annotées durant l'apprentissage en ligne.

Le script de l'apprentissage hors ligne est légèrement modifié par rapport au script de l'apprentissage en ligne. Après l'apprentissage des images de la base de données, chaque primitive motrice est corrélée avec un ensemble spécifique de caractéristiques visuelles associées à l'état interne du robot grâce au LMS.

L'ensemble des tests sont présentés dans les tableaux ci-dessous.

Test 1: Apprentissage: 51 sujets (540 images/sujet), 3 séquences ; Validation: même images

Human's mouth	Robot's mouth			
		Pout	Smile	Neutral
	Pout	58	0	42
	Smile	1	92	7
	Neutral	1	2	97

Human's mouth	Robot's mouth			
		Frowning	Raised	Neutral
	Frowning	66	4	30
	Raised	1	84	15
	Neutral	0	1	99

Human's mouth	Robot's mouth			
		Open	Closed	Intermediate
	Open	61	1	38
	Closed	1	73	26
	Intermediate	0	2	98

Figure 3.13: Tableaux montrant le taux de confusion pour les 3 primitives motrices apprises.

Cette figure montre qu'un temps très court d'apprentissage (30 minutes pour apprendre 3 groupements musculaires contenant chacun 3 intensités), le dispositif robotique est capable de reproduire les mimiques faciales de 51 individus appris durant la phase d'apprentissage. Les taux de réussite sont nettement supérieurs à 60% pour la plupart des intensités (haussement des sourcils, sourire, position neutre,...) par ailleurs les résultats sont moins performants pour la moue parce que les sujets ont du mal à l'exprimer. En moyenne, le taux de reconnaissance des intensités expressives des différents groupements moteurs est de 81%.

Test 2 : Apprentissage: 51 sujets (180 images/sujet), 1 séquence ; Validation: même sujets mais images différentes.

Human's mouth	Robot's mouth			
		Pout	Smile	Neutral
	Pout	24	10	66
	Smile	10	59	31
	Neutral	9	22	69

Human's mouth	Robot's mouth			
		Frowning	Raised	Neutral
	Frowning	32	0	68
	Raised	28	4	68
	Neutral	16	4	80

Human's mouth	Robot's mouth			
		Open	Closed	Intermediate
	Open	28	6	66
	Closed	3	47	50
	Intermediate	2	9	89

Figure 3.14: Tableaux montrant le taux de confusion pour les 3 primitives motrices apprises.

Cette figure montre la diminution du taux de reconnaissance par rapport au premier test à cause du manque de ressemblance des mêmes primitives entre les trois séquences. Pour chaque séquence, la majorité des sujets expriment différemment la même primitive. En moyenne, le taux de reconnaissance des intensités expressives de différentes primitives est de 48%.

Test 3 : Apprentissage: 26 sujets (180 images/sujet) ; Validation: 25 sujets différents.

Human's mouth	Robot's mouth			
		Pout	Smile	Neutral
	Pout	49	6	45
	Smile	4	53	43
	Neutral	2	4	94

Human's mouth	Robot's mouth			
		Frowning	Raised	Neutral
	Frowning	32	0	68
	Raised	28	4	68
	Neutral	16	4	80

Human's mouth	Robot's mouth			
		Open	Closed	Intermediate
	Open	44	20	36
	Closed	20	48	32
	Intermediate	24	24	52

Figure 3.15: Tableaux montrant le taux de confusion pour les 3 primitives motrices apprises.

Ici, l'apprentissage est fait sur les 26 premiers sujets (180 / sujet). Pour la reconnaissance, nous utilisons des images différentes des 25 autres sujets. La Figure 3.15 montre la diminution du taux de reconnaissance par rapport au premier test car les sujets expriment différemment les primitives motrices, ainsi que les individus constituant notre base d'images sont diversifiés (couleur de la peau, avec/sans barbe, forme du visage,...). En moyenne, le taux de reconnaissance des intensités expressives des différentes primitives est de 49%. On constate quand-même la capacité du robot à reproduire des expressions des sujets qu'il n'a jamais vu.

Test 4: Cross-validation

Human's mouth	Robot's mouth			
		Pout	Smile	Neutral
	Pout	**71%**	13%	16%
	Smile	4%	**84%**	12%
	Neutral	5%	6%	**89%**

Human's mouth	Robot's mouth			
		Frowning	Raised	Neutral
	Frowning	**66%**	7%	27%
	Raised	16%	**58%**	26%
	Neutral	6%	3%	**91%**

Human's mouth	Robot's mouth			
		Open	Closed	Intermediate
	Open	**69%**	12%	19%
	Closed	3%	**74%**	23%
	Intermediate	0%	7%	**93%**

Figure 3.16: Tableaux montrant le taux de confusion pour les 3 primitives motrices apprises.

Étant donné que notre base d'apprentissage ne contient pas suffisamment de personnes (51 personnes), nous avons utilisé le test de cross-validation, pour tous les sujets. Un individu est enlevé de la base d'apprentissage pour ensuite être testé en généralisation et ceci pour tous les individus constituant notre base d'images. Les résultats de ce test présentés dans la figure Figure 3.16 montre une grande capacité de généralisation. Le système arrive à reconnaitre en grande majorité les expressions des personnes n'ayant jamais interagi avec le robot.

Donc d'après ce dernier test nous constatons que plus nous augmentons le nombre de personnes constituant la base d'apprentissage, plus nous améliorons les performances du système.

7. Conclusion

Nous avons montré dans ce chapitre qu'une tête robotique dotée d'une architecture sensori-motrice basée sur des réseaux de neurones est capable d'apprendre les mouvements musculaires du visage humain à travers un jeu d'imitation avec un partenaire. Une durée d'apprentissage de 2 à 3 minutes est suffisante pour que le robot arrive à reconnaître les expressions faciales de son partenaire et à reproduire ses expressions. Maintenant le robot n'est plus limité à un ensemble limité d'expressions faciales mais il peut produire des expressions plus sophistiqués ou un mélange d'expressions primaires avec des niveaux d'intensité expressives variés.

Après l'analyse des images constituant notre base de données ainsi que les résultats obtenus, nous avons constaté que le robot confond entre certaines primitives car les individus qui ont interagi avec le dispositif ne sont pas familiers avec le fonctionnement de notre système et ils ont des difficultés à exprimer certaines primitives motrices.

Malgré les contraintes dues à la mauvaise expressivité et la grande variabilité entre les des expérimentateurs dans l'expression des différentes primitives, mais les résultats obtenus dans le test cross-validation montre que des capacités de généralisation apparaissent pour des personnes n'ayant jamais interagi avec le robot. Cependant les résultats pourraient être meilleurs si les individus ont bien exprimé les différentes primitives motrices. Autrement, le nettoyage de la base de données peut améliorer les résultats mais ça prend du temps.

Conclusion

Le travail réalisé durant ce stage fait partie des travaux de recherche du laboratoire ETIS dans le domaine de la robotique autonome et l'interaction homme robot. Le modèle développé est inspiré du développement cognitif d'un bébé et les principes de la robotique développementale, où le robot apprend à reconnaître et à produire les expressions faciales en ligne et de manière autonome en interaction avec son partenaire humain.

La première partie du stage a été consacré au développement du premier modèle neuronal pour la reconnaissance des expressions faciales par l'apprentissage des primitives motrices. Les résultats obtenus montrent qu'une tête expressive dotée d'une architecture sensori-motrice basée sur des réseaux de neurones est capable de produire une multitude d'expressions faciales. La propriété émergente de notre système réside dans la capacité du robot à produire des expressions faciales primaires comme la joie, la surprise ou la colère mais il peut également produire des expressions faciales secondaires (mélange d'expressions primaires).

La deuxième partie consiste à améliorer notre modèle en introduisant la notion d'intensité expressive par l'utilisation des champs de neurones dynamiques. L'ajout de ces derniers à notre modèle permet d'ajouter plus de finesse vis à vis le contrôle des servomoteurs. Maintenant le robot est capable de contrôler plus précisément la position des moteurs suivant l'intensité expressive du visage du partenaire humain.

La suite du stage sera consacrée au test des performances du deuxième modèle qui introduit la notion d'intensité expressive ainsi que la publication des travaux effectués dans un journal scientifique. D'autre part, en analysant notre base de données, nous pensons que les résultats pourraient être meilleurs si les images qui constituent notre bases de données sont plus significatives. Il est donc nécessaire de nettoyer la base de données et garder que les images avec une bonne expressivité des différentes primitives motrices.

Ce stage a été très enrichissant pour moi car il m'a permis de travailler dans le domaine de la reconnaissance des émotions et l'interaction homme-robot que je trouve très intéressant et passionnant, et il m'a permis également de découvrir la robotique développementale qui propose des solutions impressionnantes en s'inspirant du développement cognitif de l'être humain.

Conclusion

Durant ce stage j'ai appris à maîtriser de nouveaux outils tels que Promethe qui casse les barrières de la programmation classique, qui a pour but de faciliter l'implémentation de programmes complexes. J'ai eu aussi la chance de découvrir et travailler avec des nouveaux types de réseaux de neurones tels que le SAW et le WTA.

Autre que l'aspect technique, j'ai aussi appris ce qu'est le travail dans un cadre professionnel, la valeur de l'organisation, du sérieux, le sens de faire partie d'une équipe.

Bibliographie

Adolphs, R., Tranel, D., Hamann, S., Young, A.W., Calder, A.J., Phelps, E.A., Anderson, A., Lee, G.P., Damasio, A.R. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. Neuropsychologia 37 : 1111-1117

Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields.Biological Cybernetics, 27(2) :77–87.

Andry, P., Gaussier, P., Moga, S., Banquet, J., and Nadel, J. (2001). Learning and communication in imitation: An autonomous robot perspective. IEEE transactions on Systems, Man and Cybernetics, Part A, 31(5):431–444.

Bard, P. (1928). A diencephalic mechanism for the expression of rage with special reference to the central nervous system. American journal of psychology, 84:490–513.

Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., and Damasio, R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. 269(5227):1115–1118.

Berretti, S., Ben Amor, B., Daoudi, M., del Bimbo, A., (2011). "3D facial expression recognition using SIFT descriptors of automatically detected keypoints". Visual Computer, Springer Verlag, pp.1021-1036.

Beurle, R. L. (1956). Properties of a mass of cells capable of regenerating pulses.Philosophical Transactions of the Royal Society B : Biological Sciences, 240(669) :55–94.

Boucenna, S. (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. Thèses, Université de Cergy Pontoise.

Boucenna, S., Gaussier, P., & Andry, P. (2008). What should be taught first: the emotional expression or the face? In epirob.

Boucenna, S., Gaussier, P., Andry, P., and Hafemeister, L. (2010a). "Imitation as a communication tool for online facial expression learning and recognition". IROS, page 136.

Bibliographie

Boucenna, S., Gaussier, P., Andry, P., Hafemeister, L. (2014). A Robot Learns the Facial Expressions Recognition and Face/Non-face Discrimination Through an Imitation Game. International Journal of Social Robotics. Vol 6 No 4, Pages 633-652.

Breazeal, C., Bushbaum, D., Gray, J., Blumberg, B. (2004). "Learning from and about others: towards using imitation to bootstrap the social competence of robots". review Artificial Life, 11 (1-2), 31-62.

Breazeal, C., Edsinger, A., Fitzpatrick, P., Scassellati, B., Varchavskaia, P. (2000). "Social constraints on animate vision". IEEE Intelligent Systems and Their Applications 15 (4), 32-37.

Bullmore, E. and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nature reviews. Neuroscience, 10(3):186–98.

Cannon, W. (1927). The james-lange theory of emotions: A critical examination and an alternative theory. American journal of psychology, 39:106–124.

Carcagnì, P., Del Coco, M. , Leo, M., Distante, C., (2015). "Facial expression recognition and histograms of oriented gradients: a comprehensive study". SpringerPlus 4 (1), 645.

Carpenter, G. A. and Grossberg, S. (1987). Invariant pattern recognition and recall by an attentive self-organizing art architecture in a nonstationary world. Proceeding of Neural Network, 2 :737–745.

Carpenter, G. and Grossberg, S. (1988). The art of adaptive pattern recognition by a self-organizing neural network.Computer, 21(3) :77–88.

Darwin, C. (1965). The expression of emotion in man and animals. Chicago:University of Chicago Press (Originally Published in 1872).

De Gelder, B. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. Phil. Trans. R. Soc. B 364, 3475–3484.

Descartes, R. (1649). Les passions de l'ˆame.

Du, S., Martinez, A.M., (2015). “Compound facial expressions of emotion: from basic research to clinical applications” Dialogues in clinical neuroscience, vol. 17, no. 4, pp. 443.

Du, S., Tao, Y., Martinez, A.M., (2014). “Compound facial expressions of emotion” Proceedings of the National Academy of Sciences, vol. 111, no. 15, pp. E1454–E1462.

Ekman, P. (1982). Emotion in human face. Cambridge University Press.

Ekman, P. (1992). An argument for basic emotions. Cognition and Emotion, 6:169–200.

Ekman, P. (1992). An argument for basic emotions. Cognition and Emotion, 6:169–200.

Ekman, P. and Friesen, W. (1971). Constants across cultures in the face and emotion. Journal of Personality and Social Psychology, 17:124–129.

Ekman, P. and Friesen, W. (1978). Facial action coding system: A technique for the measurement of facial movement. Consulting Psychologists Press, Palo Alto, California.

Ekman, P., Friesen, W., and Ellsworth, P. (1972). Emotion in the human face: Guide-lines for research and an integration of findings. New York: Pergamon Press. Emery, N. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. Neuroscience and Biobehavioral Reviews, 24:581–604.

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. Science, 221(4616), 1208-1210.

Feldman, J. A. and Ballard, D. H. (1982). Connectionist models and their properties.Cognitive Science, 6 :205–254.

Gaussier, P. (2001). Toward a cognitive system algebra : A perception/action perspective. In European workshop on learning robots (EWRL), pages 88–100. Citeseer.

Gaussier, P. and Zrehen, S. (1995). PerAc: A neural architecture to control artificial animals.Robotics and Autonomous Systems, 16(2-4):291–320.

Gaussier, P., Boucenna, S., and Nadel, J. (2007). Emotional interactions as a way to structure learning. epirob, pages 193–194.

Gaussier, P., Joulain, C., Banquet, J.-P., Leprêtre, S., and Revel, A. (2000). The visual homing problem : an example of robotics/biology cross fertilization. Robotics and autonomous systems, 30(1) :155–180.

Gaussier, P., Joulain, C., Revel, A., Zrehen, S., Banquet, J.-P., Moga, S., and Quoy, M. (1997). Autonomous robot learning : what can we take for free ? In Industrial Electronics, 1997. ISIE'97., Proceedings of the IEEE International Symposium on, volume 1, pages SS1–SS6. IEEE.

Gaussier, P., Joulain, C., Zrehen, S., Banquet, J.-P., and Revel, A. (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, volume 2, pages 545–550.IEEE.

Ge, S. S., Wang, C., Hang, C.C. (2008). "Facial Expression Imitation in Human-Robot Interaction". In Proc. of the 17 IEEE International Symposium on Robot and Human Interactive Communication, Germany, pp. 213-218.

Gergely, G. and Watson, J. S. (1999). Early socio-emotional development : Contingency perception and the social-biofeedback model. Early social cognition : Understanding others in the first months of life, 60 :101–136.

Giovannangeli, C. and Gaussier, P. (2007). Orientation system in robots: Merging allothetic and idiothetic estimations. In 13th International Conference on Advanced Robotics, pages 349–354.

Giovannangeli, C. and Gaussier, P. (2008). Autonomous vision-based navigation: Goal-oriented action planning by transient states prediction, cognitive map building, and sensory-motor learning. In 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems,pages 676–683. IEEE.

Giovannangeli, C., Gaussier, P., and Banquet, J. P. (2006). Robustness of visual place cells in dynamic indoor and outdoor environment. International Journal of Advanced Robotic Systems, 3(2):115–124.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding, II : Feedback,expectation, olfaction, and illusions.Biological Cybernetics, 23 :187–202.

Grossberg, S. (1988). Nonlinear neural networks : Principles, mechanisms, and architectures.Neural Networks, 1(1) :17 – 61.

Grossberg, S. and Mingolla, E. (1985). Neural dynamics of form perception : boundary completion, illusory figures, and neon color spreading. Psychological review, 92(2) :173.

Grossberg, S. and Somers, D. (1991). Synchronized oscillations during cooperative feature linking in a cortical model of visual perception. Neural Networks, 4(4) :453–466.

Grossberg, S. E. and Morahan, P. S. (1971). Repression of interferon action : induce dedifferentiation of embryonic cells. Science, 171(3966) :77–79.

Hebb, D. (1949). The organization of behavior: A neuropsychological theory. LEA, Inc.

Hodgkin, A. and Huxley, A. (1952). “A quantitative description of membrane current and its application to conduction and excitation in nerve. The Journal of physiology, 117:500–544.

Hubel, D. H. and Wiesel, T. N. (1977). “Ferrier lecture : Functional architecture of macaque monkey visual cortex”. Proceedings of the Royal Society of London B : Biological Sciences, 198(1130) :1–59.

Izard, C. E. (1971). The face of emotion.

Jain, V., Aggarwal, P., Kumar, T., Taneja, V.,(2017)“Emotion Detection from Facial Expression using Support Vector Machine”. International Journal of Computer Applications (0975 –8887)Volume 167 –No.8.

James, W. (1884). What is an emotion. Mind, 9:188–205.

Knudsen, E. I. and Konishi, M. (1979). “Mechanisms of sound localization in the barn owl (tyto alba)”. Journal of Comparative Physiology, 133(1) :13–21.

Kohonen, T. (1984).Self-Organization and Associative Memory. Springer-Verlag, New York.

Lagarde, M., Andry, P., and Gaussier, P. (2008a). Distributed real time neural networks in interactive complex systems. InCSTST, pages 95–100.

Lippmann, R. P. (1987). An introduction to computing with neural nets. ASSP Magazine, IEEE, 4(2) :4–22.

Liu, Z. T., Wu, M., Cao, W. H., Chen, L. F., Xu, J. P., Zhang, R., Zhou, M. T., and J. W. Mao, (2017). “A facial expression emotion recognition basedhuman-robot interaction system,”IEEE/CAA J. of Autom. Sinica, vol. 4, no. 4,pp. 668−676.

MacQueen, J. et al. (1967). Some methods for classification and analysis of multivariate observations. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, volume 14, pages 281–297. Oakland, CA, USA.

Maillard, M., Gapenne, O., Hafemeister, L., and Gaussier, P. (2005). Perception as a dynamical sensori-motor attraction basin. In Advances in Artificial Life, pages 37–46. Springer.

Martinez, A.M., Du, S., (2012). “A model of the perception of facial expressions of emotion by humans: Research overview and perspectives”. Journal of Machine Learning Research, 1589-1608.

Matsumoto, D., Ekman, P. (2004). The Relationship Among Expressions, Labels, and Descriptions of Contempt. Journal of Personality and Social Psychology, 87(4), 529-540.

McCulloch, W. S. and Pitts, W. (1943). “A logical calculus of the ideas immanent in

nervous activity". The bulletin of mathematical biophysics, 5(4) :115–133.

Mehrabian, A., (1968).Communication without words.Psychology Today, 2 : 52–55.

P. Gaussier, K. Prepin, J. N. (2004). Toward a cognitive system algebra: Application to facial expression learning and imitation. In Embodied Artificial Intelligence, F. Iida, R. Pfeiter, L. Steels and Y. Kuniyoshi (Eds.) published by LNCS/LNAI series of Springer, pages 243–258.

Paiva, A., Dias, J., Sobral, D., Woods, S., Hall, L. (2004). Building empathic lifelike characters: the proximity factor. Workshop on Empathic Agents, AAMAS.

Pavlov, I. P. (1927). Conditioned reflexes. An Investigation of the physiological activity of the cerebral cortex.

Quoy, M., Moga, S., and Revel, A. (2000). Parallelization of neural networks using pvm. In LNCS, Proceedings of EuroPVM 2000, pages 289–296.

Rumelhart, D. E. and Zipser, D. (1985). Feature discovery by competitive learning. Cognitive science, 9(1) :75–112.

Sandbach,G., Zafeiriou,S., Pantic,M., Rueckert,D. (2012). "Recognition of 3D facial expressions dynamics". Image Vision Comput, 762–773.

Schachter, S. and Singer, J. (1962). Cognitive, social and physiological determinants of emotional state. Psychological Review, 69:379–399.

Schöner, G., Dose, M., Engels, C., Robotics, E., and Systems, A. (1995). Autonomous systems dynamics of behavior : theory and applications for autonomous robot architectures.Robotics and Autonomous Systems, 16(2-4) :213–245.

Schwartz, E. L. (1980). Computational anatomy and functional architecture of striate cortex : a spatial mapping approach to perceptual coding. Vision research, 20(8) :645–669.

Siegel, M., Breazeal, C., Norton, M.I. (2009)."Persuasive Robotics: The influence of robot gender on human behavior." Intelligent Robots and Systems. IROS 2009. IEEE/RSJ

International Conference on. 2009. 2563-2568.

Spinoza, B. (1677). Ethique, partie iii : Concernant la nature et l'origine des emotions.

Taube, J. S., Muller, R. U., and Ranck, J. B. (1990). Head-direction cells recorded from the postsubiculum in freely moving rats. i. description and quantitative analysis. The Journal of neuroscience, 10(2):420–435.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In CVPR, pages 511–518.

Widrow, B. and Hoff, M. (1988). Adaptive switching circuits. in 1960 ire wescon convention record, 1960. reprinted in. Neurocomputing.

Wilson, H. R. and Cowan, J. D. (1973). A mathematical theory of the functional dynamics ofcortical and thalamic nervous tissue.Kybernetik, 13(2) :55–80.

Wu, T., Butko, N. J., Ruvulo, P., Bartlett, M. S., and Movellan, J. R. (2009). Learning to make facial expressions. International Conference on Development and Learning, 0:1–6.

Zecca, M., Chaminade, T., Umilta, M.A., Itoh, K., Saito, M., Endo, N., Mizoguchi, Y., Blakemore, S., Frith, C., Gallese, V., Rizzolatti, G., Micera, S., Dario, P., Takanobu, H., Takanishi, A. (2007) " Emotional Expression Humanoid Robot WE-4RII Evaluation of the perception of facial emotional expressions by using fMRI ". The Proceedings of JSME annual Conference on Robotics and Mechatronics (Robomec).

Présentation du laboratoire ETIS

Mon stage a pris place au sein du laboratoire EIS (Équipes de traitement de l'information et systèmes) sous la direction de l'équipe Neurocybernétique. L'unité de recherche est commune au CNRS (UMR 8051), à l'ENSEA Cergy et à l'Université de Cergy-Pontoise. L'équipe "traitement des images" a été créée en 1980 au sein du laboratoire de recherche de l'École Nationale Supérieure de l'Electronique et de ses Applications (ENSEA), avec un effectif de trois personnes : un professeur des universités et deux maîtres-assistants.

Depuis 1988, l'équipe possède le label d'équipe d'accueil du Ministère de l'Enseignement Supérieur et de la Recherche (EA 1388). En 1990, l'équipe, devenue équipe "traitement des images et du signal", est reconnue jeune équipe CNRS pour deux ans. La création de l'Université de Cergy- Pontoise (UCP) donne en 1991 un nouvel essor à l'équipe.

La création du DEA "traitement des images et du signal" (DEA TIS), habilité la même année, marque l'engagement dans la formation doctorale. Le 1er janvier 1997, ETIS est devenue unité de recherche associée au CNRS (URA D2235, puis UPRES-A 8051). Depuis le 1er janvier 2002, le laboratoire ETIS, Equipes Traitement des Images et du Signal, est unité mixte de recherche du CNRS. Au 1er janvier 2009, ETIS accueille de nouveaux membres issus de l'équipe de recherche ECIME de l'ENSEA, et devient le laboratoire des Equipes Traitement de l'Information et Systèmes.

L'engagement d'ETIS dans la formation doctorale sur le site de Cergy s'est affirmé par notre appartenance à l'école doctorale multidisciplinaire Sciences et Ingénierie depuis sa création et par une forte implication dans la proposition des spécialités recherche du master SIGE (Systèmes Informatiques et Génie Electrique) de l'UCP en cohabilitation avec l'ENSEA.

ETIS est une entité de recherche commune à l'ENSEA et à l'Université de Cergy-Pontoise, reconnue par le CNRS (UMR 8051), et implantée dans les locaux de l'ENSEA et de l'UCP (St-Martin 1). Le laboratoire accueille des enseignants-chercheurs, chercheurs et personnels administratifs et techniques de ces trois établissements.

Les chercheurs du laboratoire relèvent de la section 7 du CNRS "Sciences de l'information : signaux, images, langues, automatique, robotique, interactions, systèmes intégrés matériel-logiciel".

Présentation de l'équipe cybernétique

L'équipe cybernétique travaille principalement sur de la robotique développementale et bio-inspirée d'où le nom de cette équipe. Les domaines les plus traités au sein de l'équipe sont l'interaction homme-robot, la perception active et multimodale et la navigation autonome.