Le conditionnement du second ordre pour la référenciation sociale

Nous décrivons dans ce chapitre une mini-séquence développementale, basée sur l'architecture Per-Ac, testée sur notre robot. Dans cette séquence nous voyons comment notre robot a développé la compétence de reconnaître des expressions faciales à travers un jeu d'imitations. Ensuite, en utilisant les expressions faciales comme outil de communication, nous décrivons l'émergence de capacité de référenciation sociale chez le robot. Finalement, sur la base des capacités acquises, dans le cadre de la référenciation, et en ajoutant un système de suivi visuel d'objets, nous avons testé la capacité de notre robot à acquérir un sens esthétique en évoluant dans un environnement centré sur les humains et à forte composante esthétique comme les musées. Le système de suivi visuel est présenté dans le chapitre 4, et les expériences aux musées dans le chapitre 3. Finalement, nous présentons un nouveau modèle de conditionnement qui permettra d'une part de simplifier la référenciation sociale et d'autre part d'apprendre des préférences esthétiques en autonomie.

5.1 Expressions faciales - Interaction di-adique

L'interaction homme/robot au travers des signaux émotionnels est de plus en plus étudiée. Bien que les solutions ingénieur ad-hoc exhibent des résultats impressionnants, elles ne permettent pas au système d'évoluer et de s'adapter à son environnement. Dans cette section nous présentons comment à travers un processus d'imitation en utilisant une architecture sensori-motrice simple notre robot a développée la capacité d'apprendre des expressions faciales. Un des principaux avantages de ce type d'architecture est sa versatilité et son adaptation au monde réel. Dans ce modèle l'apprentissage se fait en interaction avec un tuteur sans avoir recours à des stratégies Ad-hoc de reconnaissances. Cette capacité à reconnaître des expressions faciales est un premier schème essentiel au développement de capacité de référenciation sociale dans notre paradigme pour le développement du sens esthétique. Dans une perspective développementale nous nous sommes inspirés de la capacité du bébé à apprendre des expressions faciales sans signal de supervision (Gergely and Watson [1999]).

Plusieurs travaux de recherche se sont focalisées sur la construction de systèmes robustes pour la reconnaissance des expressions faciales. Cette capacité n'est quasiment jamais étudiée

du point de vue d'une construction autonome et développementale. Dans ces travaux, l'accent est mis sur la sélection des meilleurs ensembles de caractéristiques et les performances de reconnaissance. 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 méthodes utilisent des à priori forts et ils ont besoin d'accéder à toute la base d'apprentissage. Ces méthodes sont très performantes sur des bases de données déjà construites. Par contre elles sont rarement confrontés à des images réelles en milieu naturel. Ainsi, elles ne permettent pas une autonomie du robot et ses capacités de généralisation en sont amoindries. Nous utilisons, de notre coté, le même système pour apprendre des expressions faciales et des objets. La généricité du modèle nous permet d'acquérir des capacité de plus en plus complexe comme la référenciation sociale. Nous avons supposé que l'apprentissage des expressions pouvaient se passer de la détection de visage. Cette reconnaissance est vue comme une propriété émergente d'un mécanisme (jeu) de mirroring. Nous avons même démontré que la discrimination des visages résulte d'une détection de synchronisation, donc de la capacité à prédire ce que l'autre agent fait. (voir Boucenna et al. [2008] pour plus de détails).

Certaines méthodes sont basées sur l'analyse en composante principale (PCA) par exemple le LLE (Locally Linear Embedding) qui réalise une réduction des vecteurs d'entrés (Liang et al. [2005]). Des architectures neuronales ont également été développées pour la reconnaissance des expressions faciales, par exemple le réseau de Franco and Treves [2001]. Ces derniers utilisent un réseau de neurones multicouches utilisant une règle d'apprentissage supervisée classique. Cependant, l'architecte du réseau détermine le nombre de neurones associés aux différentes expressions selon leur complexité. D'autres méthodes sont basées sur des modèles du visages comme par exemple les modèles d'appartenance (Abboud et al. [2004]). Yu and Bhanu [2006] utilisent un support vecteur machine (SVM) pour catégoriser les expressions. Le but étant d'obtenir les meilleurs frontières de séparation. Wiskott [1997] utilise des ondelettes de gabor pour le codage des caractéristiques faciales appelées "jets". Ces caractéristiques sont insérées dans un graphe labellisé où les nœuds sont des "jets" et les liens entre les différents nœuds sont des distances (distances entre la bouche et un œil par exemple). La reconnaissance se traduit par une comparaison de graphe ("graphe matching"). Toutes ces techniques et beaucoup d'autres encore utilisent un apprentissage hors ligne ainsi qu'un grand nombre d'à priori pour améliorer les performances. Notons également que toutes ces méthodes accèdent à l'ensemble de la base d'apprentissage ce qui rend impossible la comparaison avec un processus développemental. Ces modèles ne peuvent pas mimer des processus biologiques par leurs manques d'adaptation aux différents environnements. Breazeal et al. [2005] a conçu Kismet, une tête robotique qui peut reconnaître des expressions faciales. Bien que Kismet apprenne à mimer les expressions à partir d'une interaction sociale, il a un fort a priori ce qu'est un visage et les caractéristiques visuelles sont prédéfinies. De plus, si un système comme celui de Breazeal est utilisé avec des objets ou des scènes, la tête expressive ne donne aucune réponse. Généraliser sur des vues jamais rencontrées parce qu'on y retrouve des caractéristiques visuelles commune est, pour nous, une capacité importante.

Nous considérerons dans nos développements que le robot est similaire à un jeune enfant et le partenaire humain interagissant avec lui correspond à un parent. De manière générale, le robot ne connaît rien sur son environnement. Il ne possède que des comportements réflexes (expressions faciales prototypiques). Il commence à apprendre grâce à l'interaction avec le monde

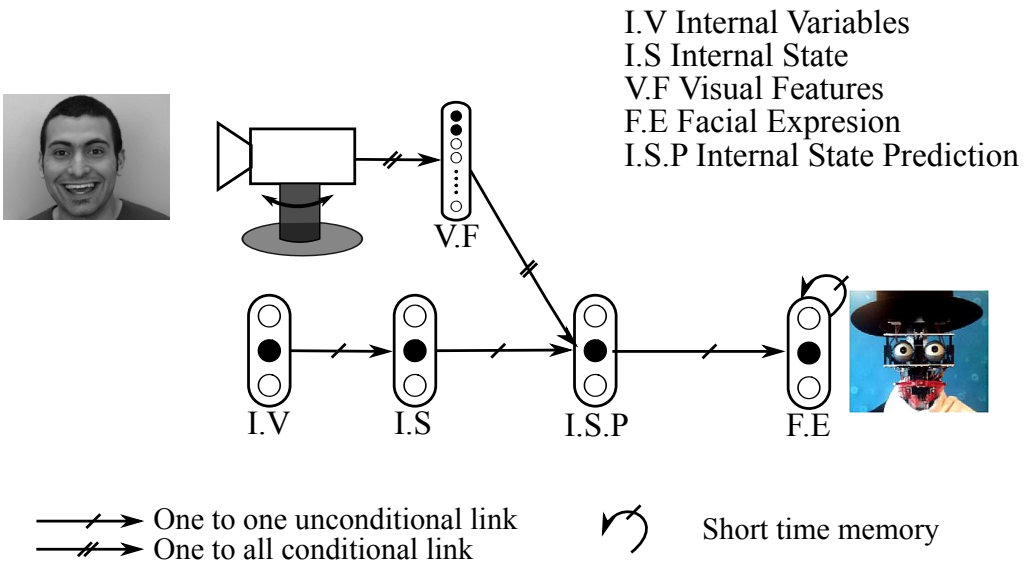


FIGURE 5.1 – Modèle utilisé pour l'apprentissage des expressions faciales. Le système associe les caractéristiques visuelles extraites du visage du professeur avec l'état interne du robot. L'état interne est lui contrôlé par les variables internes du robot (ex. La faim peut déclencher de la tristesse et la satiété de la joie).

extérieur : ici le partenaire humain constitue une grande partie de l'environnement. Le système est composé de deux agents interagissant dans un environnement neutre. Les deux agents peuvent afficher des expressions faciales en fonction de leur état interne (état "émotionnel") et perçoivent le visage du vis-à-vis. L'un des agents est un adulte capable de reproduire les expressions de l'autre. Le deuxième agent est un nouveau né sans connaissance du monde extérieur. Nous supposons que l'apprentissage des expressions faciales est possible si l'agent 2 produit des expressions et que l'agent 1 l'imite. Par conséquent, si l'agent 2 est doté de mécanismes sensori-moteurs alors il sera capable d'associer des sensations avec des actions. Sans langage, sans connaissance de l'autre, l'agent 2 (bébé) pourra reproduire des expressions vues sur le visage d'un expérimentateur. Pour que l'apprentissage se stabilise l'expression des états internes des deux agents doit être corrélée. L'état interne du robot n'étant pas contrôlable, il faut que l'agent tuteur imite ou résonne (Nadel et al. [2006]) aux expressions faciales du robot, permettant une corrélation explicite entre l'état interne du robot et l'entrée visuelle (expression du tuteur). Le tuteur n'est alors rien de plus qu'un miroir.

Nous avons développé au laboratoire un setup expérimental comme un outil de recherche en science cognitive mettant en jeu à la fois des expériences en psychologie et de la modélisation computationnelle (Nadel et al. [2006]). La tête robotique a été désigné pour être un système minimal permettant d'afficher quelques expressions faciales prototypiques (Izard [1971]; Ekman et al. [1972]; Ekman and Friesen [1978]; Plutchik [1980]). Elle a été conçue par Pierre Canet au centre émotion de la Pitié Salpétrière dans le cadre d'un projet avec l'équipe neurocybernétique du laboratoire ETIS. Dans l'expérience décrite ici seules cinq expressions seront utilisées : joie, tristesse, colère, surprise, plus un état neutre. La figure 5.2 montre les expressions utilisées.

L'avantage d'utiliser une tête robotique peut être justifie par :

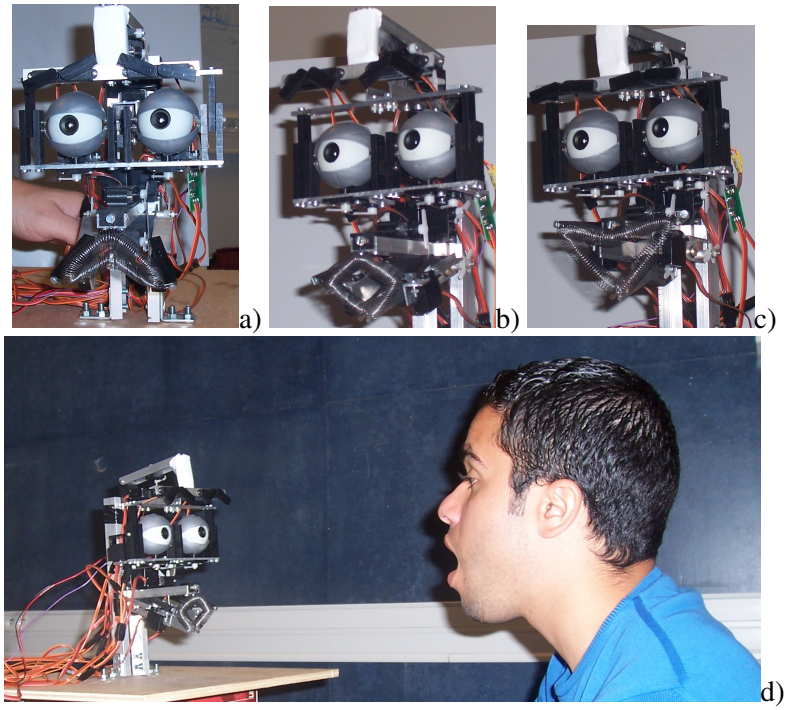


FIGURE 5.2 – Exemples d'expressions faciales que le robot est capable de reproduire : a) tristesse, b) surprise, c) joie. d) situation typique du robot et de l'humain dans le cas d'un jeu d'interaction.

- résonance émotionnelle : Les humains reproduisent "involontairement" les expressions faciales d'une tête de robot (Nadel et al. [2006])

- Le robot facilite la communication car l'homme et le robot interagissent dans le même environnement ("présence" du robot). L'aspect actif et communicatif est agréable pour un expérimentateur humain.

- Le robot a besoin de diriger son regard vers des objets d'intérêt. Il est pratique que ce dispositif puisse aussi permettre à l'humain de savoir dans quelle direction regarde le robot et ainsi servir de moyen de communication implicite.

Le scénario expérimental se divise en deux phases d'interaction toutes les deux basées sur l'imitation. En premier, le robot affiche des expressions faciales liées à son état interne. Le groupe I.V. représente les variables internes du robot (faim, satiété, douleur ...) qui modifient l'état interne du robot (joie, tristesse, surprise ...). Les états internes se manifestent alors sous forme d'expressions faciales affichées par la tête robotique (tristesse, joie, colère, surprise, et une expression neutre). L'humain mime de manière non intentionnelle les expressions faciales du robot (Nadel et al. [2006]).

Dans cette expérience, le changement d'état interne est simulé et le robot affiche les expressions dans l'ordre. Chaque expression faciale est présentée pendant 5 secondes pour accélérer l'apprentissage. Il est explicitement demandé au tuteur d'imiter les expressions faciales du robot. Cette phase dure entre 2 et 3 minutes. Elle permet au robot d'apprendre les expressions faciales

de son tuteur. Le groupe *VF* catégorise les vues locales de la scène en utilisant l'algorithme du SAW (cf chapitre.4). Le système visuel pour l'extraction des caractéristiques est le même que celui décrit dans le chapitre 4. Le groupe *ISP* associe les vue locales sur le visage du tuteur avec l'état interne du robot par un conditionnement pavlovien en utilisant l'algorithme des moindres carrés LMS (Widrow and Hoff [1988]) (cf chapitre.4). Les associations vues locales/ état interne sont intégrées dans le groupe *FE*. Ce groupe somme les activités des réponses de *ISP* et il est remis à zéro à chaque nouvelle image. Une fois l'apprentissage terminé le robot devient capable de reconnaître et d'imiter les expressions faciales de son tuteur. A la fin de procédure d'apprentissage, lors de la phase de test (restitution) l'humain produit une expression pendant quelques seconde et on test la capacité du robot à répondre par la même expression. Les résultats expérimentaux ne seront pas présentés dans ce manuscrit. Une description détaillée en est faite dans Boucenna [2011].

Cette expérience montre qu'un dispositif constitué d'une simple tête expressive est capable d'apprendre et reconnaître de manière autonome les expressions faciales par un jeu d'imitation. Cette procédure suppose la préexistence d'un adulte "empathique" (Decety and Jackson [2004]; Berthoz and Thirioux [2010]). La résonance émotionnelle est la clé de voûte de notre paradigme expérimentale.

L'approche proposée ici pour la reconnaissance des expressions faciales rencontre un certain nombre de contraintes liées à l'apprentissage en ligne et en interaction. L'apprentissage en ligne induit des problèmes dus aux temps de réaction de l'humain. Ce délai perturbe la phase d'apprentissage car lorsque le robot change d'expression faciale, les premières images capturées correspondent toujours à l'ancienne expression. A chaque fois que le robot exprime une nouvelle expression, les premières images apprises correspondent à l'expression précédente. Il est nécessaire d'intercaler une expression neutre pour éviter les mauvaises interprétations. Ceci peut se justifier notamment par le délai de reconnaissance de l'expression faciale par l'humain (150ms selon Simon Thorpe pour les objets Thorpe et al. [1996]). Un certain temps d'adaptation est indispensable.

5.2 Référenciation sociale - interaction tri-adique

Le jeune enfant acquiert très tôt, dans les six premiers mois, la capacité à discriminer et à comprendre les expressions émotionnelles d'autrui (Nelson [1987]). Ce premier schème de communication non-verbale est très utile au développement de nouvelles capacités d'interaction. Particulièrement dans la régulation du comportement face à des situations nouvelles. D'un point de vue développemental, la reconnaissance des expressions faciales est un premier schème. Son assimilation amène à l'adaptation à l'environnement et à la création d'un second schème qui est la référenciation sociale.

L'expression des émotions est une forme de communication non-verbale qui permet de communiquer ses propres émotions et de comprendre celle des autres. Les travaux en psychologie (Campos [1983]; Campos et al. [1989]; Izard [1990]; CAMPOS et al. [2013]) montrent l'utilité des expressions émotionnelles dans l'interaction de l'humain avec son environnement physique et social. Les émotions jouent un rôle prédominant dans l'adaptation de l'individu à son milieu social. Dans *"Les origines du caractère chez l'enfant"*, Wallon [1936] prête aux émotions une fonction de régulation du comportement et des poses corporelles : "C'est aux variations locales ou généralisées du tonus, nous dit l'auteur, que sont dus les états d'hypotonie, d'hypertonie, ou

de spasme d'où procèdent les émotions". A partir de l'age d'un an environ, le jeune enfant utilise sa capacité à communiquer par des expressions émotionnelles pour faire de **la référenciation sociale** (Social Referencing). La référenciation sociale est la capacité d'attribuer des valences émotionnelles à des situations (scène, objet, lieu...) a priori neutre. Campos et al. [1981] et Feinman1982 la définissent comme la capacité à utiliser des informations émotionnelles pour lever les ambiguïtés des situations nouvelles et adapter le comportement en conséquence. La fonction de régulation du comportement de la référenciation sociale est partagée par la majorité des chercheurs (CAMPOS et al. [2013]; Walden and Baxter [1989]; Feinman et al. [1992]; Rosen et al. [1992]; Mumme et al. [1996]; Baldwin and Moses [1996]). Les expériences de Klinnert [1984] ont montré une corrélation entre le comportement du bébé et l'expression faciale de sa mère. Klinnert a présenté des jouets à connotation négative à des enfants âgés de 12 à 18 mois, tout en demandant à leurs mères d'exprimer de la joie, de la peur ou un visage neutre. L'enfant s'approche des objets et les explore lorsque sa mère exprime de la joie. A l'inverse l'expression de peur, modifie le comportement de l'enfant qui s'éloigne de l'objet et s'approche de sa mère. La même année Gunnar and Stone [1984] ont utilisé des jouets avec des connotations différentes (plaisantes, ambigues, négatives), notamment grâce à un robot mécanique utilisé comme stimulus ambigu. Ils ont montré que des enfants d'un an réagissaient positivement à ces jouets lorsque leurs mères expriment un affect positif. Les observations de Sigman and Kasari [1994] ont aussi montré que les enfants se réfèrent aux adultes face à l'arrivée d'un stimulus ambigu comme un robot bruant. Ces expériences montrent que la communication émotionnelle permet d'affronter les situations nouvelles et ambiguës. Cette communication peut se faire au travers de différents canaux, expression faciale, tonalité de la voix, gestes. La maman parlant d'un individu de manière joyeuse permet à l'enfant d'accepter d'avantage cet individu (Feinman and Lewis [1983]). Les travaux de Feinman sont intéressants car ils montrent un lien entre le conditionnement au second ordre et la référenciation sociale.

FIGURE 5.3 – Référenciation sociale : falaise visuelle (Galinsky [2010]).

Le changement de comportement par référenciation sociale est bien démontré dans l'expérience de la falaise visuelle (fig.5.3) de Klinnert (Sorce et al. [1985]). Cette expérience montre l'effet de la communication émotionnelle sur le comportement de l'enfant face à des situations ambiguës et inconnues. La falaise visuelle est une table vitrée dont la moitié est recouverte d'un

motif à carreaux. Le même motif est placé sous la partie vitrée de sorte à être visible par l'enfant qui se balade sur la table. Ce dispositif donne une impression de chute. La mère fait face à son bébé installé sur la table du coté opaque. Les mères expriment différentes expressions faciales : joie, peur, colère, tristesse. Les observations montrent que la majeure partie des bébés (75%) traversent la falaise visuelle lorsque les mères expriment des affectes positifs (joie). A contrario, les bébés reculent ou s'arrêtent lorsque les mères expriment de la peur ou de la colère. Ceci montre que les enfants ont la capacité à comprendre les expressions de leurs mères et modifient leur comportement en conséquence. Ils s'approchent et traversent devant une expression positive, et reculent ou s'arrêtent devant des expressions négatives. Une autre expérience de Sorce and Emde [1981] découple la présence de la mère et son expressivité. Cette étude montre l'effet de l'interaction de la mère avec son enfant sur son comportement. Particulièrement sur le comportement de jeu et d'exploration en présence de perturbation extérieur. Au début, l'enfant joue dans une pièce où la mère est présente. En plus de la mère un individu étranger et un robot sont positionné en retrait dans un coin de la pièce. Après un certain temps l'individu étranger s'approche de l'enfant pour interagir avec lui et le robot se met en marche pour ajouter de la confusion à l'enfant. Les couples mères enfants sont divisés en deux groupes. Dans le premier groupe les mères étaient présentent mais non-disponible. Ces mères sont occupées à lire des magazines et ne changent pas de comportement face aux demandes de leurs enfants. Dans le second groupe les mères sont présentes et disponibles prêtes à intéragir avec leurs enfants. Les observations ont montré que les enfants dont la mère est disponible sont plus joyeux, leur exploration s'intensifie, et leur comportement change en fonction des expressions émotionnelles de leurs mères.

Ces expérience montrent le rôle de la communication émotionnelle dans l'adaptation du comportement de l'enfant. Elles mettent en valeur exclusivement des interactions mères/enfants. D'autres expériences comme celle de Dickstein and Parke [1988] montrent que l'enfant se réfère aussi à son père. Tandis que Hirshberg and Svejda [1990] montrent que les enfants se réfèrent indifféremment au père ou à la mère lorsque les deux sont présents. Par ailleurs, Klinnert et al. [1986] montre qu'en plus de la mère et du père, l'enfant se réfère aussi à un individu familier étranger à sa famille. Dans cette expérience l'enfant joue avec l'individu familier en présence de la mère. La mère est passive et n'interagit pas avec son enfant ni avec l'étranger. L'enfant joue d'avantage avec un jouet ambigu (robot) lorsque l'expérimentateur affiche une expression de joie que lorsqu'il affiche de la peur. Cela montre, d'une part que la seule présence de la mère n'influence pas spécialement le comportement de l'enfant. D'autre part, que l'enfant est capable de discriminer les expressions des individus autre que ceux de sa famille et les utilise pour adapter son comportement vis à vis des objets. D'autres recherches montrent que l'effet de la référenciation est orienté vers une situation spécifique sans modifier l'humeur globale de l'enfant. Hornik et al. [1987] et Stenberg and Hagekull [1997] montrent dans leur expérience comment un enfant ne se soucie plus d'un objet devant lequel la mère a exprimé de la colère mais continue à jouer avec les autres objets.

Les recherches sur la référenciation sociale s'accordent à y attribuer une fonction de régulation du comportement (CAMPOS et al. [2013]; Walden and Baxter [1989]; Feinman et al. [1992]; Rosen et al. [1992]; Mumme et al. [1996]; Baldwin and Moses [1996]). Par contre, deux théories se confrontent quant à l'origine de la capacité de référenciation sociale. La première théorie soutient que la référenciation est une capacité innée et biologiquement précablée, c'est le cas notamment de Campos [1983]. La seconde défend l'acquisition de la référenciation sociale au travers d'un apprentissage opérant lors d'interactions sociales (Gewirtz and Pelaez-Nogueras

[1992]). Dans les expériences de Gewirtz and Pelaez-Nogueras [1992] les signaux envoyés à l'enfant ne sont pas uniquement émotionnels, mais peuvent aussi être de simple geste corporels autre que des expressions émotionnelles. Cette expérience montre que d'un côté la référenciation sociale n'est pas innée, et d'un autre côté qu'elle peut être conditionnée par des messages autre que émotionnels. Nous défendons de notre coté que la référenciation sociale est un processus développemental. Elle permet à un enfant ou à un robot (Breazeal et al. [2005]; Thomaz et al. [2005]; Boucenna et al. [2010b]) de chercher l'information venant d'autrui et de l'utiliser pour réguler son comportement face aux situations nouvelles ou ambiguës Klinnert [1984].

Expérimentation robotique

Nous nous intéressons au laboratoire à comprendre comment un robot peut apprendre à développer des compétences complexes à partir d'architecture sensori-motrices simples qui à la base ne sont pas dédiées aux interactions sociales. Notamment dans le cadre de la référenciation sociale plusieurs expérience mettant en jeu des robots ont été mises en place Boucenna et al. [2010b]; Hasson et al. [2010]. Dans ces expériences, les situations sont définies au travers d'interactions émotionnelles et de manière non-verbale. En sachant que plusieurs modalités émotionnelles peuvent servir d'outil de communication, nous avons choisi d'utiliser les expressions faciales. Ceci est motivé par trois raisons. Premièrement, notre modèle d'expressions faciales, comme expliqué plus haut, permet un apprentissage rapide et autonome (Boucenna et al. [2010b]). Deuxièmement, le modèle est basé sur une architecture de conditionnement sensori-motrice simple ce qui facilite le chaînage des conditionnements. Finalement, les expressions faciales primaires sont universelles, reconnues dans toutes les cultures et elles ont le même sens dans des cultures différentes.

Nous expliquons ici une expérience de référenciation sociale utilisant les expressions faciales et un bras robotique. Cette expérience mené par Boucenna en 2010 (Boucenna et al. [2010b]), a été reprise durant cette thèse pour comprendre le fonctionnement du système. De plus l'explication permettra d'introduire le modèle de référenciation sociale utilisé dans le cadre du développement de l'esthétique artificielle.

Le setup expérimental est composé de la tête expressive décrite plus haut (tête expressive capable de reproduire les expressions faciales), d'un bras Katana à 6 degrès de liberté capable d'attraper ou d'éviter des objets, d'une caméra installée dans l'œil droit de la tête qui regarde le partenaire humain, et d'une seconde caméra tournée vers les objets et l'espace péricorporel du robot. Cette seconde caméra facilite l'apprentissage et nous évite de gérer les problèmes épineux du contrôle de l'action. L'humain et le robot communiquent au travers des expressions faciales. La figure 5.5 schématise le setup expérimental mis en place. Tandis que la figure 5.4 montre le modèle utilisé.

La tête robotique est tout d'abord entraînée par le jeu d'imitation à apprendre et reconnaître les expressions faciales (tristesse, joie, colère, surprise et visage neutre) tel qu'expliqué dans la section précédente. Le premier groupe *VF* permet de catégoriser les vues locales prises sur le visage en utilisant l'algorithme du *SAW*. Chacune de ses vues est associée à l'état interne du robot par le groupe *ISP*. Le groupe *ISP* est un *LMS* qui trouve l'erreur quadratique moyenne minimale entre l'état interne du robot et la reconnaissance d'expressions faciales. Ensuite, par un apprentissage visuo-moteur décrit dans de Rengervé et al. [2010], le bras Katana apprend à atteindre différentes positions dans son espace de travail. L'apprentissage visuo-moteur correspond à l'association entre la position visuelle de l'effecteur (extrémité du bras) et la proprioception des

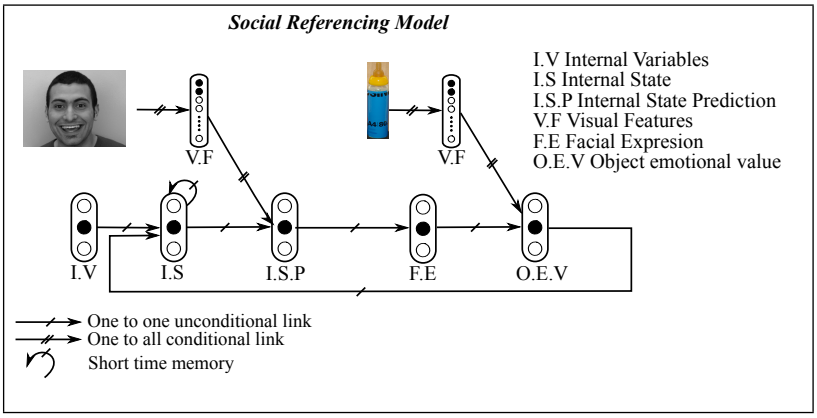


FIGURE 5.4 – Modèle de référenciation sociale. Les groupes *VF* sont des réseaux de neurones qui catégorisent les expressions faciales de l'humain ou l'objet à apprendre. Les groupes *ISP* and *OEV* sont basés sur des conditionnement pavlovien utilisant l'algorithme du LMS.

articulations du bras. Après l'apprentissage, le bras est capable de rejoindre plusieurs positions, crées sous la forme d'attracteurs, dans l'espace de travail (Andry et al. [2001]). Le contrôle du bras est réalisé par un système dynamique (contrôleur Yuragui) lissant la trajectoire (Fukuyori et al. [2008]). Ce système dynamique utilise aussi un signal de renforcement dans le but de donner beaucoup ou peu d'importance à certains attracteurs, par exemple une récompense peut être attribué si le bras suit la bonne direction, dans le cas contraire on le punie. Le renforcement peut bien entendue être émotionnel (e.g l'expression de joie est un signal positif tandis que l'expression de colère est considérer comme un signal négatif).

Puis, l'humain présente un objet au robot (tête plus bras) en affichant une expression faciale (joie, colère). Le robot associe alors son état émotionnelle (et donc l'expression du tuteur) avec l'objet. L'objet présenté est catégorisé par un second groupe *VF*, pour éviter les confusions entre les expressions de l'humain et les objets. L'association entre l'expression faciale et l'objet est effectuée par un second *LMS*. L'objet acquiert de ce fait une valence émotionnelle (positive ou négative). Il est intéressant de remarquer dans ce modèle que la partie modèle qui permet d'apprendre les expressions faciales (catégorisation plus association avec un signal émotionnel) est identique à celle qui permet d'apprendre à attribuer des valences aux objets. Ceci montre la généricité du système de perception-action. Finalement, le bras Katana atteint et prend l'objet si celui-ci est conditionné comme étant positif, ou sinon s'en éloigne si celui-ci est conditionné comme étant négatif. Ce qui est mis en évidence dans ce scénario, c'est que l'émotion joue un rôle non seulement pour essayer de réguler nos propres signaux internes mais également que l'émotion joue un rôle non négligeable pour l'adaptation du robot dans un environnement réelle. L'émotion a un rôle de régulation et d'adaptation de son comportement sur soi-même ainsi que sur l'environnement. Les résultats sont détaillés dans Boucenna et al. [2010b].

A ce stage de développement, le robot est capable d'utiliser l'expression faciale du partenaire sociale comme un signal de communication pour attribuer une valeur émotionnelle à des objets. Grâce à l'interaction avec un partenaire humain, le robot est capable de reconnaître et de comprendre les expressions de l'autre dans le but de désambiguer certaines situations comme la présence d'un nouvel objet dans l'espace de travail. Sans connaissance de ce qu'est l'autre et

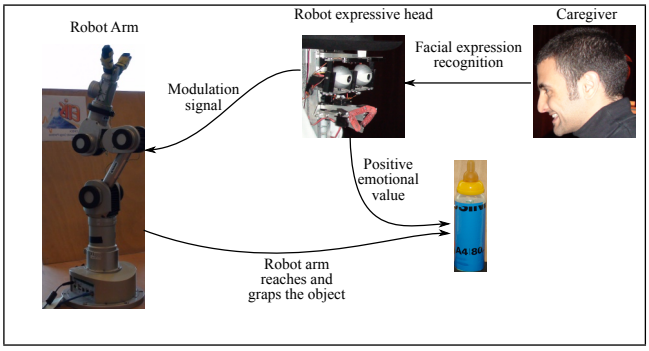


FIGURE 5.5 – Scénario expérimental de référenciation sociale pour la manipulation d'objets. Le tuteur exprime de la joie. La tête du robot reconnaît l'expression faciale et affiche la même expression que celle de l'humain. Le système associe une valence émotionnelle positive à l'objet (le biberon), et envoie un signal de renforcement au bras. Finallement, le bras atteint la position de l'objet et attrape l'objet appris comme étant positif.

ce que veux l'autre autrement dit sans théorie de l'esprit, le robot est capable de produire des comportements sociaux relativement complexes.

5.3 Chaîne de conditionnement

Dans une perspective d'apprentissage autonome, le robot doit être capable d'apprendre même en absence d'un professeur. Le développement d'appréciations basées sur un modèle de référenciation sociale permet d'associer une valence à un objet par le biais d'interactions expressives. Seulement un modèle de référenciation sociale comme le notre ne permet pas un apprentissage qui se maintient en autonomie sur le long terme. En effet avec un telle architecture les moments d'apprentissage sont contrôlés par la présence du tuteur, et seule l'expression faciale de l'humain peut attribuer des valences aux objets. Il existe alors une construction des préférences par un effet social, mais l'historique du robot n'a aucun impact sur cette construction.

Notre robot arrive à associer une valence émotionnelle avec une œuvre à partir des expressions faciales de son tuteur, ou bien dans le cas des expériences aux musées grâce à l'assignation des visiteurs au travers des boutons d'une souris. Ce stade est le premier niveau du sens esthétique. A ce niveau, le robot a un goût esthétique conformiste, c'est à dire qu'il est la moyenne des préférences des visiteurs avec lesquels il a interagit. Pour reprendre les termes des historiens de l'art, nous avons fait appel au *suffrage universel* pour éduquer le robot. De plus, le robot ne peut qu'apprécier ou déprécier un objet, il n'a pas la capacité à discriminer les situations intéressantes. Comme nous l'avons vu au chapitre 4, le robot pourrait détecter des ambiguïtés dans ses préférences et classer les objets selon quatre catégories : Les catégories positives, négatives et neutre associées contenant les oeuvres d'art appréciées, dépréciées ou sans intérêt, plus une nouvelle catégorie, la catégorie *objet intéressant*. Nous qualifions de situation intéressantes, un objet ambivalent qui serait fortement associé à une valence positive et à une valence négative en même temps. La détection des objets *"intéressants"* est le second stage de développement. A ce stade le robot n'est plus une simple copie du goût des humains qu'il a rencontré. Il se crée sa propre différence. Pat contre, deux (ou plusieurs) robot ayant accès au mêmes apprentissages

seront identiques et développeront les mêmes préférences esthétiques.

Comment prendre en compte le vécu du robot ? Des robots ayant après un premier apprentissage identique vécu des expériences différentes, auront-ils tendance à développer les mêmes préférences esthétiques ? Quel serait le modèle d'apprentissage qui permettrait d'intégrer l'historique du robot dans sa conduite esthétique ? Pour répondre à ces questions nous avons cherché à modifier notre modèle d'apprentissage par référenciation sociale de manière à permettre aux stimuli de propager leur valence émotionnelles sur des stimuli proches. On peut voir ce comportement comme celui d'un enfant qui apprécie les plats de sa mère qui se trouve faire des omelettes aux champignons. L'odeur de l'omelette aux champignons senti en passant devant un restaurant le mettra dans état émotionnel positive car l'omelette est associée aux plats de sa mère qui elle même est associée à une humeur positive.

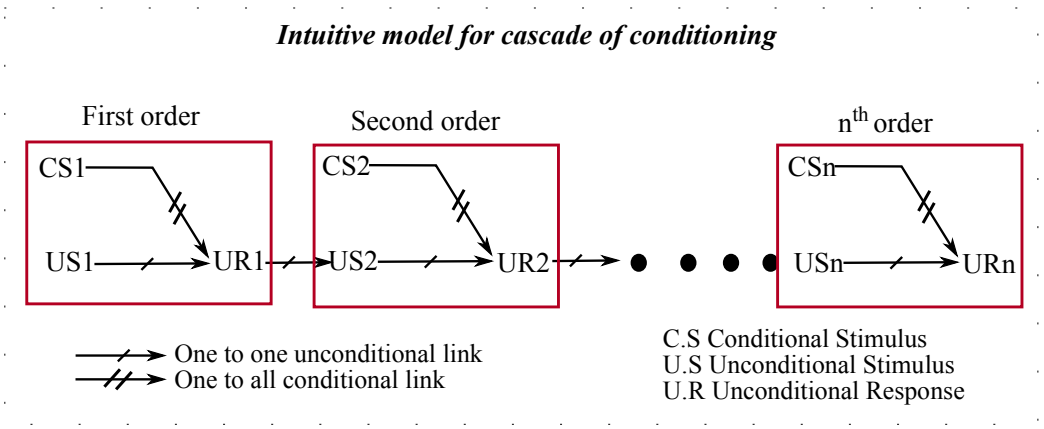


FIGURE 5.6 – Proposition intuitive (mais probablement incorrecte) d'une architecture composée d'une cascade de conditionnement.

Pour réussir à chaîner les conditionnements, une solution intuitive est de mettre en cascade des conditionnements tel que montré dans la figure fig.5.6. On peut imaginer par exemple que le premier bloc de conditionnement apprend à catégoriser les expressions faciales et à leur associer des valences émotionnelles, puis le second associe la vue d'une œuvre à la réponse des expressions faciales, le troisième associe la réponse de l'œuvre précédente à une œuvre plus proche et ainsi de suite. Cette mise en cascade des conditionnements atteint cependant rapidement ses limites. Tout d'abord, la demande en ressources computationnelles est proportionnelle au nombre de conditionnement chaînés, et ces conditionnements peuvent rapidement être saturés. De plus, l'implémentation contrôle le nombre de stimuli sur lesquels la propagation s'opère, et l'apprentissage ne peut être facilement arrêté avant la fin de la chaîne et ne démarre que si le tuteur est présent. Finalement, les stimuli ne modifient pas explicitement l'état émotionnel du robot. Pour dépasser ce problème et permettre l'implémentation de schemes Piagessiens, la solution est d'utiliser une structure ré-entrante.

Nous proposons une architecture dans laquelle la réponse d'un stimulus est utilisée comme signal pour modifier l'état "émotionnel" du robot. Le système agit alors en boucle fermée. De cette façon tout stimulus avec une appréciation forte peut alors faire office de tuteur, déclenchant l'apprentissage et propageant sa valeur sur ses proches voisins. Dans cette méthode aucune hiérarchie n'est a priori inscrite dans le modèle, tous les conditionnements sont au même plan.

Ainsi, un seul conditionnement avec des signaux réentrants bien choisis peut remplacer la cascade de conditionnement. Il existe par contre bel et bien une hiérarchie entre les stimuli, mais elle est une une propriété émergente du système résultant de l'ordre dans lequel sont fait les apprentissages. Ce type d'architecture de conditionnement s'apparente aux méthodes de conditionnement d'ordre supérieur inspirés des résultats observés par Pavlov (Pavlov [1927]).

Cependant avec une simple réetrance l' apprentissage ne peut se maintenir sur le long terme. La complexité des scènes visuelles fait que la segmentation des épisodes d'apprentissage n'est pas facile. La valence d'un objet contamine de proche en proche tous les objets que le robot rencontre. La propagation de proche en proche peut très vite diverger et le robot peut se retrouver bloqué dans le même état émotionnel (Comme un dépressif). Il est nécessaire d'avoir un contrôle des phases d'apprentissage par des mécanismes d'auto-evaluation et de meta-controle. Pour palier ce problème nous avons introduit un mécanisme de différenciation pour isoler les épisodes d'apprentissage tout en permettant malgré tout une certaine diffusion des informations pour avoir des conditionnements du second ordre.

Dans un premier temps, le professeur enclenche la première phase d'association entre un stimulus conditionnel et un stimulus inconditionnel (CS1-US). Ensuite le système transforme la réponse conditionnelle (CR) en un nouvel US pour conditionner un nouveau CS (CS2). Le modèle de conditionnement d'ordre supérieur avec cascade des conditionnement est montré dans la figure 5.9.

5.4 Conditionnement Pavlovien

Le conditionnement du second ordre est un champs de l'apprentissage associatif largement étudié. Il permet l'étude de l'émergence d'une relation entre deux événements qui n'ont jamais été présenté ensemble (Rizley and Rescorla [1972] ;Pavlov [1927]]). Une procédure de conditionnement secondaire se divise en trois phases. Dans une première phase, un stimulus conditionnel (CS1) est suivi d'un stimulus inconditionnel (US). [1] Dans une seconde phase, un stimulus conditionnel du second ordre (CS2) est présenté en même temps que le stimulus primaire (CS1). Finallement, en phase de test, le stimulus du second ordre (CS2) est présenté seul au sujet pendant que des données sont relevées (Jara et al. [2006]). Les résultats expérimentaux de Pavlov [1927] ont montré qu'un CS2 engendre la même réponse conditionnelle (CR) qu'un CS1, bien que le CS2 ne soit jamais directement associé à l'US.

Dans un conditionnement Pavlovien du premier ordre, l'apprentissage est acquis en couplant un stimulus conditionnel (CS ; ex. le son d'une cloche) et un stimulus intrinsèque inconditionnel (US ; ex. la nourriture ou un choc). Dans les conditionnements Pavlovien d'ordre supérieur, le CS est couplé à un stimulus ayant une valeur motivationnelle acquise et non intrinsèque (Gewirtz and Davis [2000]). Par exemple un animal peut apprendre dans un premier temps à associer de la nourriture avec un son (**conditionnement primaire**), ensuite l'animal peut associer une lumière avec le son (**conditionnement du second ordre**). Un stimulus conditionnel, après avoir acquit la capacité d'engendrer une réponse conditionnelle peut être utilisé comme stimulus inconditionnel. Ce phénomène est appelé **"Conditionnement d'ordre supérieur"**.

Dans une expérience, Pavlov conditionne un chien à saliver au rythme d'un métronome, en utilisant de la viande hachée comme stimulus inconditionnel. Ce conditionnement correspond à

[1] Selon les expériences les stimuli sont soit présentés l'un après l'autre soit en même temps. Dans nos travaux les stimuli sont toujours présentés les une après les autres.

un conditionnement primaire. Le métronome acquière des propriétés de renforcement et devient capable d'évoquer la réponse conditionnelle. Dans la phase suivante, Pavlov présente à l'animal un carré noir suivi du son du métronome mais sans présenter la nourriture. La vue du carré noir produit la salivation. De cette manière, un conditionnement du second ordre est établi. La nourriture est appelée *renforçateur primaire*. Le métronome en étant associé avec la nourriture acquière des propriétés de renforcement. Le métronome est vu comme un renforçateur secondaire. Dans la majeur partie des cas, le conditionnement d'ordre supérieur est assez faible. La réponse conditionnelle d'un conditionnement du second ordre est à peu près de moitié inférieur à celle d'un premier ordre. Le conditionnement du troisième ordre fut difficile à établir n'apparaissant que dans le cas où un renforcement négatif est utilisé comme renforçateur primaire comme un choc électrique par exemple. Le conditionnement du quatrième ordre semble impossible à montrer (Saurab [2015]).

Le conditionnement secondaire a été étudié chez les animaux en utilisant différents types d'US et de réponses. Par exemple, il a été démontré dans le conditionnement d'aversion (Cheatle and Rudy [1978] Rescorla [1976]), dans des expériences d'autoshaping (Rashotte et al. [1977] Rescorla and Cunningham [1979]) et dans des comportements sexuels (Crawford and Domjan [1995]). Plusieurs modèles théoriques ont été proposés pour expliquer le conditionnement du second ordre. La plus part d'entre eux découlent des théories d'apprentissage associatif. Ces modèles peuvent être rassemblés dans quatre types différents suivant la nature des associations qui peuvent avoir lieu durant un conditionnement du second ordre (Barnet et al. [1997]) comme montré dans la figure 5.7. Le premier type de modèle est le plus traditionnel, il repose sur un modèle stimulus-réponse (S-R), comme dans les modèles de Rescorla (Rescorla et al. [1972] ; Rizley and Rescorla [1972] ; Rashotte [1981]). Ce modèle propose une connexion directe entre le CS2 et la CR renforcé par la présence du CS1, du à la proximité temporel entre les stimuli lors de la phase d'apprentissage du second ordre. Le second type de modèle suggère l'existence d'une chaîne d'association CS2-CS1-US-CR dans laquelle CS2 évoque CS1 qui lui évoque l'US qui vient déclencher la CR (Hall [1996] ;Mackintosh [1974]). Un troisième type de modèle propose une association directe entre le CS2 et l'US. Selon ce modèle lors de la seconde phase la représentation de l'US peut être activée par le CS1 lors de la présence du CS2, et par cette association le CS2 peut évoquer la CR (Barnet and Miller [1996] ;Miller and Barnet [1993]). Le quatrième type de modèle considère l'existence d'une association entre CS2 et CS1 où le déclenchement de la CR est uniquement du à la représentation de CS1 évoquée par le CS2 (Miller and Escobar [2004]). Notre modèle s'approche du troisième modèle dans la figure 5.7.

Deux critères sont à considérer pour démontrer l'existence d'un conditionnement du second ordre. D'une part, la réponse conditionnelle engendrée par le stimulus du second ordre (CS2) doit résulter directement de l'association du CS2 avec un stimulus primaire CS1. D'autre part, les propriétés de renforcement du CS1 doivent venir de son association avec un stimulus inconditionnel (US) (Rizley and Rescorla [1972]). On peut ajouter un troisième critère qui est le renforcement des nouveaux apprentissages sans oublier ceux précédemment acquis (Baldassare et al. [2009]). Ce critère est notamment utilisé pour l'apprentissage des compétences d'une manière hiérarchique

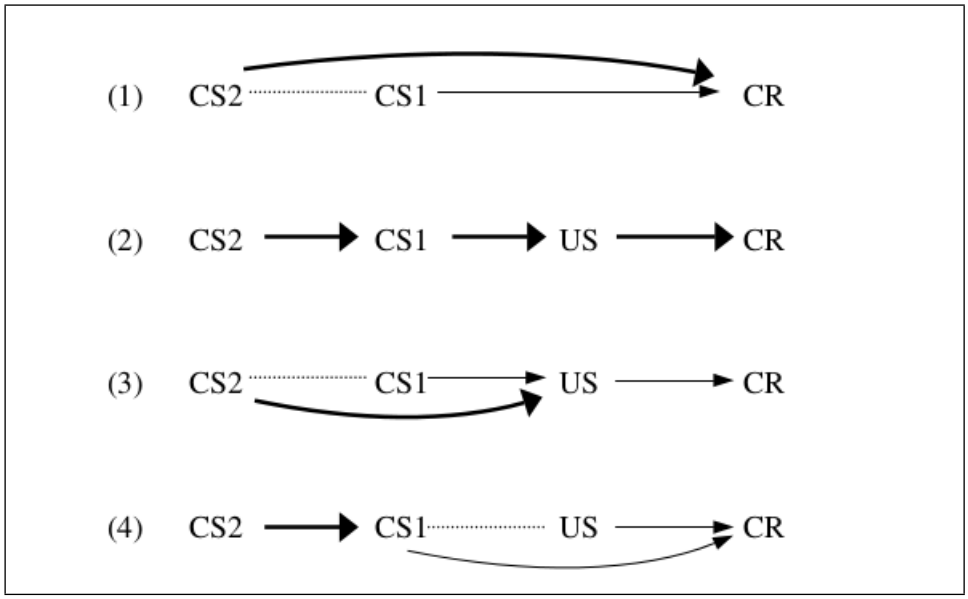


FIGURE 5.7 – Les différentes hypothèses sur les associations établies pendant un conditionnement du second ordre, inspiré de la théorie des apprentissages associatifs. CS1 représente le stimulus conditionnel du premier ordre, CS2 un stimulus conditionnel du second ordre, US le stimulus inconditionnel et CR la réponse conditionnelle. Extrait de Jara et al. [2006]

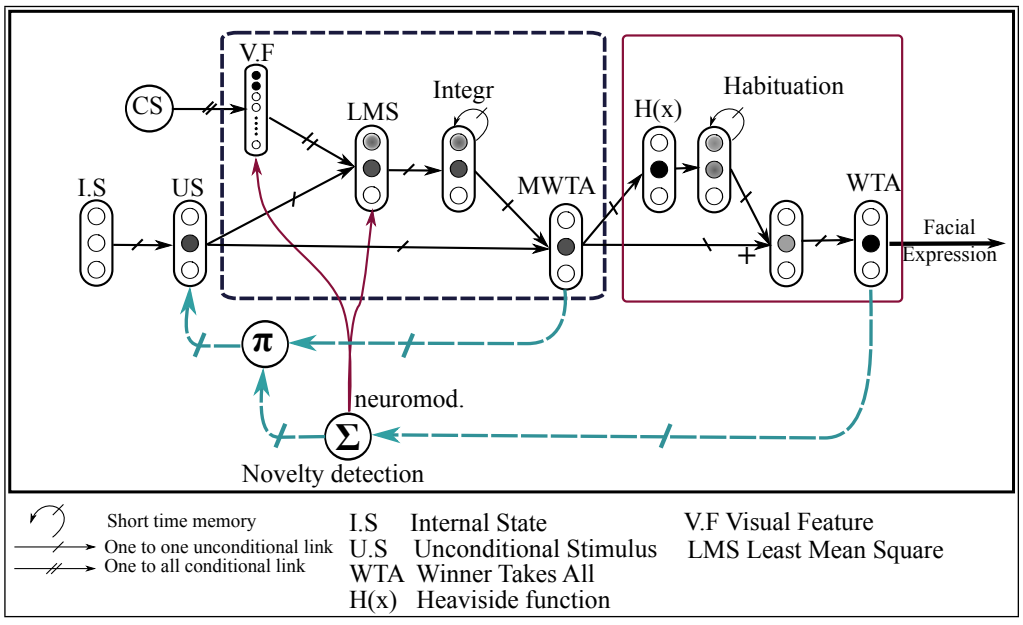


FIGURE 5.8 – Modèle de conditionnement aux ordres supérieurs. Les lignes en tirets bleus sont des connections réentrantes qui évitent une cascade de conditionnement comme dans la figure 5.6. Ce modèle est celui qui donne les meilleurs résultats sur base de donnée et sur camera.

5.5 Modèle computationnel

Nous avons testé plusieurs variantes de notre modèle de conditionnement d'ordre supérieur. Dans ce chapitre nous présenterons uniquement trois variantes. Un premier modèle testé uniquement sur une base de donnée d'image avec un scénario expérimental simple. Un second modèle testé aussi sur une base de données d'images mais cette fois-ci en essayant d'avoir un scénario proche des expérimentations en neurobiologie. Finalement, le dernier modèle présenté est le même que le second avec de très légère modification dans le système de contrôle des apprentissages pour tenir compte des aléa d'un test sur une caméra réelle.

Le modèle dans la figure 5.8 permet de faire des conditionnements au second-ordre ou à des ordres supérieurs. Le modèle est composé de deux blocs. Un bloc de reconnaissance visuelle avec un conditionnement par LMS et un second bloc d'habituation pour gérer les phases et temps d'apprentissage. Ces temps sont inspirés de la décharge des neurones dopaminergiques (Mannella et al. [2008] ;Moren and Balkenius [2000]).

Le premier bloc est le même que celui décrit plus haut. Il est utilisé dans l'apprentissage des expressions faciales ou bien des œuvres dans le musée. La premier groupe noté *IS* (Internal State) représente l'état interne du robot. La réponse de ce groupe est un signal réflexe impératif. Le groupe suivant *US* est une compétition entre la réponse du réseau de neurone et le signal réflexe. Le signal réflexe ayant un poids plus important, sa présence prime sur tous les autres. Ainsi, en présence d'un signal réflexe la réponse de l'état interne du robot est la seule à s'exprimer. Autrement, la réponse du réseau de neurone au temps *t-1* est réinjectée dans le réseau. Le groupe *CS* représente le stimulus conditionnel. Ici nous nous intéresserons uniquement à des stimuli visuels. Le groupe *VF* est un groupe de catégorisation utilisant l'algorithme du SAW. Ce groupe stocke les vues locales récupérées sur le stimulus conditionnel. Le groupe *LMS* permet d'associer chaque vue locale du stimulus conditionnel avec le stimulus inconditionnel résultat de la compétition entre le signal réflexe et la réponse du réseau de neurone (association CS-US). En sortie du LMS il existe une réponse conditionnelle par vue locale. Puis le groupe *integr* somme les réponses conditionnelles des vues locales du stimulus. Le traitement visuel pour obtenir les vues locales, la catégorisation, le conditionnement, et la somme des réponses, sont décrits dans le chapitre 4.

Le dernier groupe du premier bloc *MWTA* permet d'extraire la réponse maximale et l'argument de cette réponse. Ce groupe de neurones cherche tout d'abord l'activité maximale entre la réponse du stimulus conditionnel et le signal inconditionnel, ensuite un WTA (Winner Takes All) active la catégorie gagnante et inhibe les autres. Cette opération choisit la réponse conditionnelle (CR) qui s'exprime en sortie. Deux réponses sont possibles, soit le stimulus inconditionnel (US), soit la sortie intégrée (CR') du groupe *integr*. Si l'US est supérieur à la sortie de l'intégration (CR'), alors l'US est propagé sur le CS. Sinon si la sortie de l'intégration (CR') est supérieure à l'US, c'est alors cette CR qui devient un US et est propagée sur les prochains CS. En pratique le groupe *MWTA* compare la réponse conditionnelle au temps *t* avec la réponse conditionnelle au temps *t-1* qui est réinjectée en tant que US. Si le signal réflexe est présent, à cause de ses poids fort, il gagnera la compétition entre la CR_{t-1} et la CR_t et il sera propagé à l'itération suivante. De cette manière un renforçateur d'un ordre inférieur avec une CR forte ne peut être changé ou diminué par l'effet de propagation d'un conditionnement d'ordre supérieur avec une CR plus faible. Par contre deux renforçateurs du même ordre peuvent se propager les uns sur les autres.

Le second bloc (à gauche) est un système d'habituation qui contrôle les phases et durées d'apprentissage. Ce système agit comme les neurones dans le nucleus central de l'amygdale qui injectent des neuromodulateurs comme la dopamine à travers l'aire tegmentale ventrale. Le système d'habituation calcule l'écart entre la réponse conditionnelle et la moyenne des réponses conditionnelles passées. La moyenne est une mémoire à courte terme qui calcul une moyenne glissante dont l'horizon est contrôlé par le paramètre α. La catégorie ayant le plus grand écart à sa moyenne active l'apprentissage. De cette manière si la même catégorie est rencontrée, après un certain temps le système s'y habitue et ne la considère plus comme intéressante à propager. De plus, les CR de la même catégorie mais d'un ordre supérieur (donc avec une activité plus faible) ne sont pas propagée sur les CR d'ordre plus petit (dont la reconnaissance est plus forte). La vitesse d'apprentissage aussi est modulée par l'écart entre une CR et sa moyenne glissante. Le système d'habituation agit alors comme un contrôleur proportionnel, plus l'écart entre la CR et sa moyenne est grand plus l'apprentissage est rapide et vice versa. La propagation est une multiplication entre la CR et l'habituation. Ainsi, une réponse conditionnelle forte à laquelle le système n'est pas encore habitué se propage plus fortement, avec une vitesse d'apprentissage plus rapide et se propage plus longtemps. La CR n'est pas propagée lorsque l'écart entre la CR et sa moyenne est nul et la neuromodulation arrête l'apprentissage.

Le robot fait connaître son état émotionnel en affichant une expression faciale relative à son appréciation du stimulus. Cette expression faciale est contrôlée par la CR du stimulus ou bien par la sortie du processus d'habituation. Afficher la sortie du système d'habituation en expression faciale permet d'avoir un retour uniquement lorsque le robot apprend. Cette information est utile à l'expérimentateur pour valider et contrôler l'apprentissage du système. Autrement, l'affichage de la CR (sortie du *MWTA*) comme information pour les expressions faciales est plus informative dans le cadre d'une expérience grande échelle où l'appréciation des objets par le robot doit être observée. Le robot exprimera son appréciation qu'il soit ou non en phase d'apprentissage. Par contre, cette richesse ne permet pas de délimiter les phases d'apprentissage et les phases d'utilisation. Nous avons arbitrairement et pour la bonne conduite de nos expérience choisi d'afficher le résultat de l'habituation.

Nous avons testé un grand nombre de variantes de notre modèle pour obtenir un comportement cohérent. Nous entendons par comportement cohérent, la capacité à propager des valences sans détruire les apprentissages passés. Nous présentons ici trois variantes qui nous permettent d'expliquer l'utilité des différentes modifications introduites. Le tableau tab5.1 montre les différences majeures entre les variantes.

Le modèle le plus simple pour avoir de la propagation de valence serait l'utilisation d'un lien récurent qui réinjecte la CR comme étant un US. Ce modèle ne converge pas car le lien de retour est positif. L'apprentissage ne peut donc s'arrêter et les stimuli seront appris et désappris, d'où la nécessité d'un mécanisme d'habituation et de contrôle.

	Expression faciale	Rebouclage S^d
Modèle 0 5.9	Habituation ou CR_i	CR_i
Modèle 1 5.10	Habituation ou $Max_i(CR_i + S_i^d)$	habituation
Modèle 2 5.8	Habituation ou $Max(CR, S^d)$	$Max(CR, S^d) * H_x(habituation)$

TABLE 5.1 – Différences entre modèles de conditionnement

La première variante de notre modèle est montrée dans la figure fig5.9. Dans cette version, comme toutes les autres, l'état interne du robot a un poids plus fort que tous les autres signaux. Dans les illustrations, lorsqu'un poids n'est pas spécifié sur un lien inconditionnel un vers un,

sa valeur est par défaut égale à 1. Le groupe *U.S* est un WTA qui effectue une compétition entre le stimulus inconditionnel et la réponse conditionnelle réentrante. Si le stimulus inconditionnel (représenté par l'état interne *IS*) existe, il est propagé sans tenir compte des autres signaux. Autrement, la réponse conditionnelle au temps $t-1$ est propagée et ce quelque soit le stimulus conditionnel. Dans ce là, si un CS d'ordre supérieur est suivi d'un CS primaire, ce dernier verra sa reconnaissance altérée par la CR du CS d'ordre supérieur.

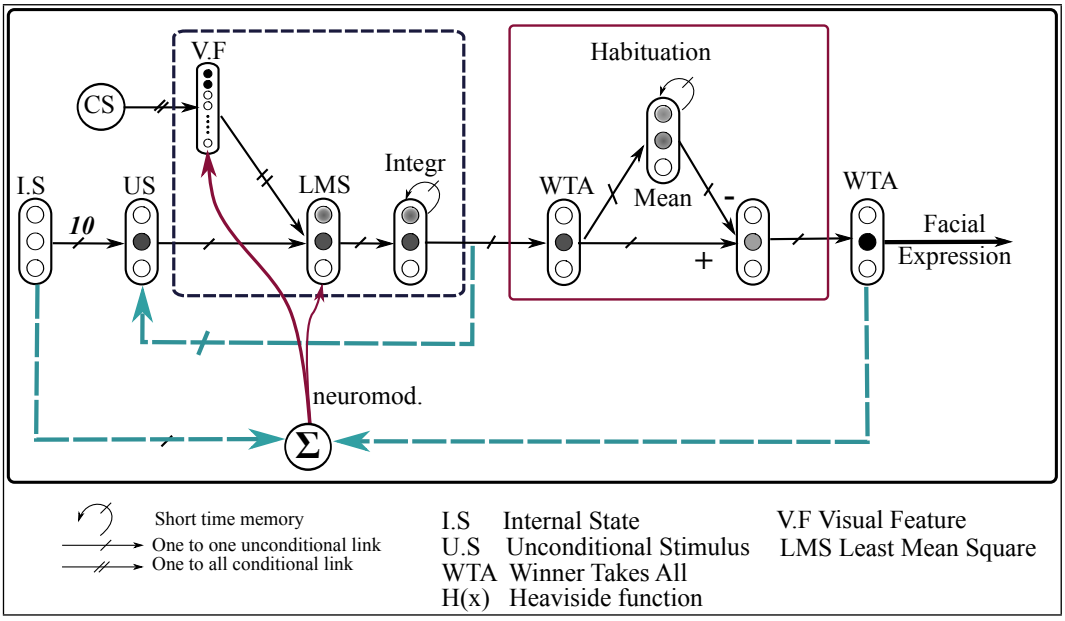


FIGURE 5.9 – Première variante du modèle de conditionnement d'ordre supérieur. Il n'y a aucune compétition entre la CR_{t-1} et la CR_t.

Le bloc d'habituation par contre n'est pas différent de celui utilisé dans la dernière version. La réponse conditionnelle est moyennée dans le temps grâce à une mémoire à court terme. Tant que la CR est supérieure à la moyenne des CR précédentes l'apprentissage continue. Autrement, le système arrête l'apprentissage considérant que le même stimulus lui est présenté pendant longtemps et qu'il ne sert à rien de continuer à apprendre. Les résultats de ce modèle seront présentées dans la prochaine section.
La seconde version est montrée dans la figure 5.10. La différence majeure de cette version avec la précédente est la compétition par un WTA entre le groupe *Integr* et le groupe *US*. Grâce à ce procédé, un stimulus conditionnel dont la réponse conditionnelle est forte n'est pas impacté par la réponse conditionnelle réentrante d'un stimulus plus faible. La réponse réentrante dans ce modèle est fournie par l'écart entre la moyenne glissante et la réponse conditionnelle.
Dans la dernière version expliqué ci-haut (figure 5.8), nous avons réutilisé le même modèle mais en modifiant le principe de compétition et le signal réentrant. Dans cette dernière version, la compétition s'effectuent par un opérateur de maximum entre les signaux. Cet opérateur nous évite de faire la somme entre la CR et l'US. En effet, le WTA comme expliqué dans le chapitre 2 calcul d'abord la somme des poids par les entrées avant la compétition. De cette manière la réponse conditionnelle a une valeur plus forte. Ce qui nous empêchait d'utiliser cette CR comme

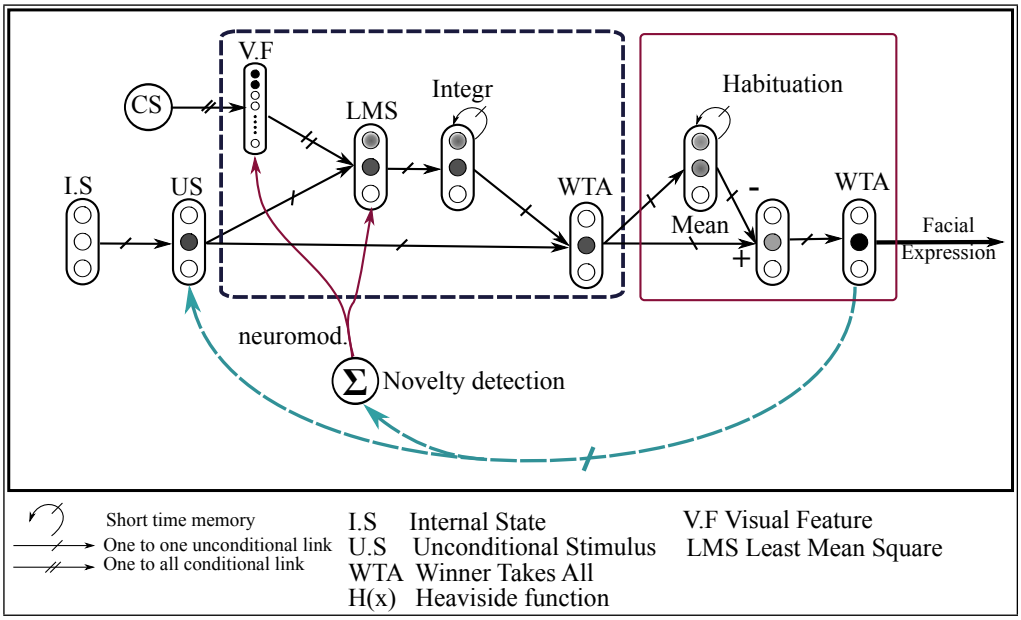


FIGURE 5.10 – Second modèle de conditionnement du second ordre. Les lignes en tirets bleus sont des liens réentrants. Ce modèle donne de bon résultats sur une base de donnée mais il est difficile à contrôler dans un scénario réel sur robot.

signal réentrant directement. Nous devions passer par l'écart avec la moyenne pour diminuer cette valeur. Grâce à l'opérateur du maximum, la CR réentrante est directement l'expression du résultat de la reconnaissance du stimulus.

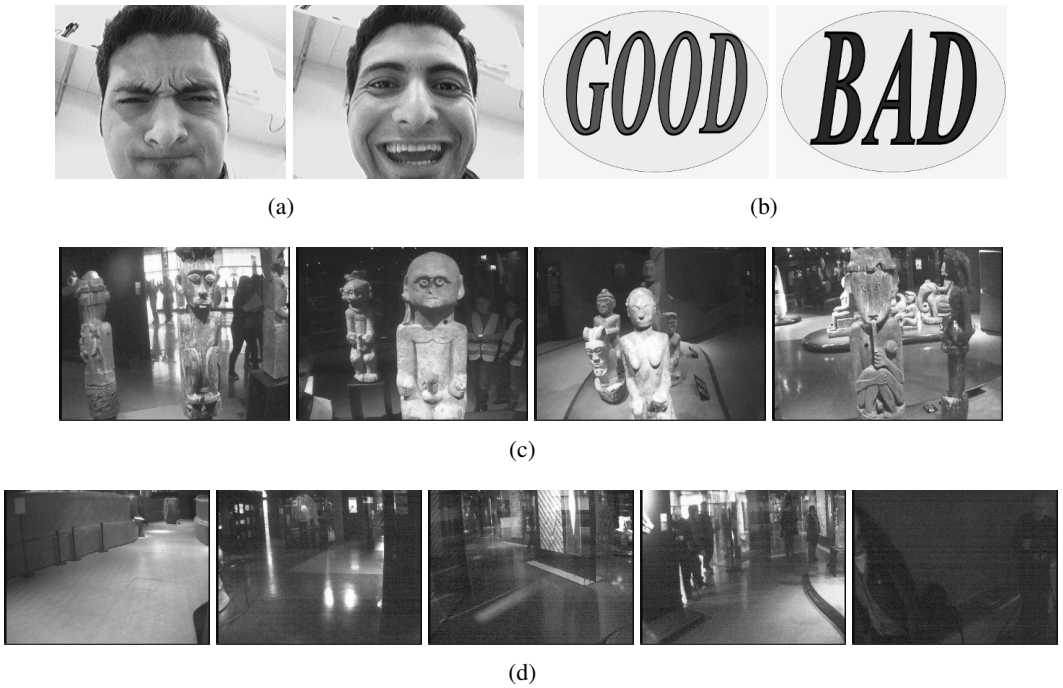
5.6 Expérience et résultat

Nous avons mis en place un scénario pour valider le modèle de conditionnement aux ordres supérieur. Ce scénario reprend les étapes classiques de conditionnement du second ordre telle que mise en place par Pavlov. Dans une première étape un stimulus conditionnel CS1 (expression faciale du professeur) est conditionné par l'état interne US (figure 5.12). Cette étape effectue un premier conditionnement primaire du premier ordre. L'état interne est simulé par les boutons d'une souris. Le premier bouton est vu comme une punition qui enclenche chez le robot une expression de tristesse. Le second bouton est vu comme une récompense qui déclenche une expression de joie. Par un jeu d'imitation comme celui décrit dans Boucenna et al. [2010a], le robot apprend à reconnaître les expressions faciales de son professeur. La figure 5.12 montre les performances d'apprentissage des expressions faciales. Dans une seconde étape le CS1 (expression faciale) est présenté avec un stimulus du second ordre CS2 (oeuvre du musée). La valence émotionnelle du CS1 est propagée sur le CS2. Le CS1 prédit alors l'US, permettant au CS2 d'acquérir le renforcement de l'US. On continue à propager la valence des CS d'ordre inférieur vers les CS d'ordre supérieur de proche en proche. Dans la dernière étape l'apprentissage est arrêté, on représente tous les CS et on collecte les réponses. Nous avons d'abord testé notre modèle sur une base donnée contenant des oeuvres du musée du Quai Branly. Ensuite nous avons testé le

modèle dans un cadre naturel en utilisant une caméra.

5.6.1 Sur base de donnée

La base de donnée est composé de quatre catégories d'images. Deux images d'expressions faciales (joie, colère) qui représentent les stimuli primaires (CS1) (fig.5.11(a)), deux images de synthèses utilisées comme stimluli du second ordre (fig.5.11(b)), quatre images d'œuvres du musée du Quai-Branly qui sont les stimuli d'ordres supérieurs (CSn) et des images du Quai-Branly ne contenant aucune œuvre qui représentent des stimuli neutre qui ne sont pas appris. Dans la phase de test seules les réponses des expressions faciales, des signes et des œuvres sont analysées. Les réponses des stimuli neutre ne sont pas importantes, leur réponse est un effet de généralisation du système.



(a) (b)

(c)

(d)

FIGURE 5.11 – Images de la base de donnée de test pour le conditionnement secondaire. Les expressions faciales 5.11(a) sont les CS1, les signes 5.11(b) et les oeuvres 5.11(c) sont des stimuli d'ordre supérieur (CSn), les images 5.11(d) sont des stimuli neutre.

Dans un premier temps nous avons testé la première version du modèle (figure 5.9) sur une base de donnée contenant uniquement des expressions faciales et des images d'œuvres. Nous avons alors comparé les résultats avec et sans système d'habituation. Ce modèle n'effectue aucune compétition entre la réponse du réseau de neurone et l'entrée du réseau. Ainsi soit le signal impératif, si il existe, est associé au stimulus. Sinon la réponse au temps $t-1$ réinjectée dans le système est associée.
Tout d'abord le système apprend les expressions faciales du tuteur. Ces expressions faciales sont présentées pendant trois itérations chacune avec 200 répétitions. Ici le signal réflexe est

toujours présent, il prime sur tous les signaux réentrant. Ceci revient à un apprentissage supervisé classique du premier ordre.
La figure fig.5.12 montre les résultats de ces associations. Les lignes en tirets bleues correspondent à la réponse au stimulus de colère et les lignes en pointillés rouges correspondent à la réponse au stimulus de joie. L'apprentissage est rapide et les valeurs des réponses se maintiennent sur le long terme.

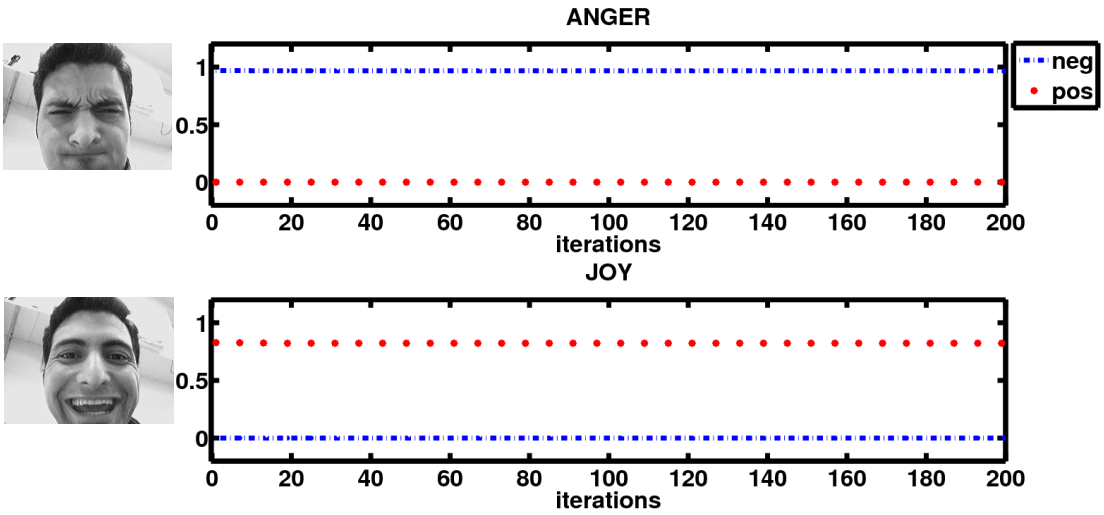


FIGURE 5.12 – Apprentissage des expressions faciales du tuteur. Les tirets bleus représente l'apprentissage d'une expression de colère, et les points rouges l'apprentissage d'une expression de joie. Chaque expression est présentée trois fois. L'opération est répétée 200 fois.

Par la suite les stimuli d'expression faciale sont suivi d'un stimuli représentant des œuvres du musée (O1, O2) en l'absence d'un signal inconditionnel réflexe. Chaque expression faciale est présentée pendant 3 itérations suivie d'une œuvre pendant trois itérations, et le tout est répété. Dans l'ordre : expression de colère, oeuvre "O1", expression de joie, oeuvre "O2" et on répète.
La figure fig.5.13 montre le résultat du conditionnement en absence du système d'habituation. Les réponses conditionnelles sont directement transformée en US. Au début le conditionnement a bien lieu, puis avec le temps, dès que la réponse à un stimulus de même ordre est plus grande notamment à cause du bruit, elle se propage sur tous les stimuli même sur les stimuli qui sont du même ordre. Ainsi, l'expression de joie et les deux œuvres donnent une réponse de colère.
L'utilisation du même scénario avec le système d'habituation en place permet une diffusion sur les stimulus du second ordre tout en gardant la réponse des stimuli du premier ordre en état. Ces résultat sont montré dans la figure fig.5.14.
Ces exemples simples permettent de voir l'utilité d'avoir un système de contrôle des apprentissages et de la propagation. Ces résultats bien que préliminaires, valident la capacité du système dans un cas simple de procéder à des conditionnement du second ordre. Néanmoins, le scénario expérimental dans lequel le modèle est testé reste simple et schématique. Pour simuler le comportement du système dans des conditions réelles et valider les conditionnements, nous avons mis en place un second scénario divisé en phases d'apprentissages et phases de tests avec une

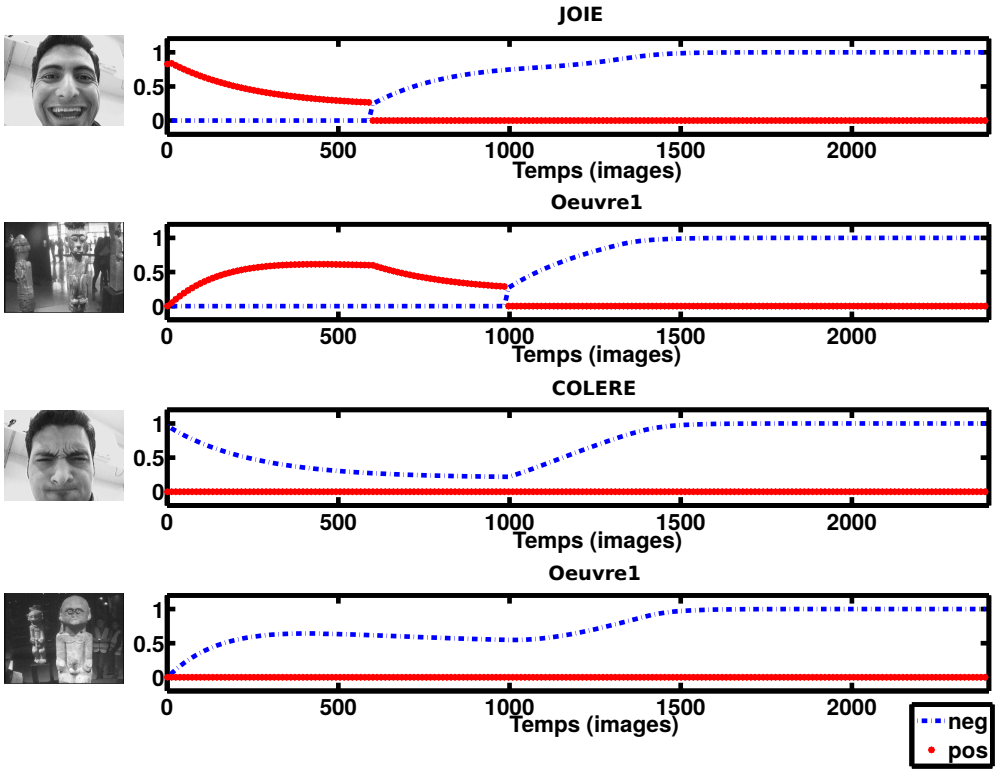


FIGURE 5.13 – Conditionnement du second ordre sans système d'habituation. Les réponses conditionnelles sont toujours propagées. Dans ce cas un stimulus avec une faible réponse peut endommager l'apprentissage d'un stimulus primaire comme le stimulus de joie dans ce cas là. Comme la propagation ne s'arrête pas, à un certain moment tout les stimuli sont associés à la même catégorie.

séquence de présentation des stimuli identique à celle utilisée dans la littérature.
La figure fig.5.15 montre l'ordre dans lequel sont présenté les stimulus et les signaux de conditionnement. Les stimulus du premier ordre noté *CS1* sont présentés en même temps qu'un signal inconditionnel réflexe *US*. Une fois l'association CS1-US effectuée, les CS1 sont présentés avec des stimuli du second ordre notés *CS2*. Les CS2 sont ensuite présentés avec les CS3, les CS3 avec les CS4. Entre les associations des CS d'ordres différents des stimuli neutres sont présentés. Ces stimuli neutres représentent une pause dans le système, ce sont des cas où la propagation peut avoir lieu ou pas et permettent uniquement un temps de déshabituation. Il est intéressant de noter dans notre scénario que, contrairement à la littérature, chaque stimulus peut avoir deux réponses différentes (joie, colère).
En faisant un parallèle avec l'expérience de Pavlov, l'expression faciale de joie peut être vue comme la salivation. L'appui sur un bouton de la souris simule ici la présentation de la nourriture. Le visage joyeux de l'expérimentateur représente le son du métronome et l'œuvre du musée représente le carré noir. Dans nos expériences le setup proposé par Pavlov est complexifié. On peut imaginer le second bouton de la souris comme la présence d'un serpent, l'expression de colère simule la colère ou la peur du chien, l'expression de colère de l'expérimentateur peut être remplacée par le son d'une cloche, et finalement l'œuvre remplacée par un cercle rouge. Ces

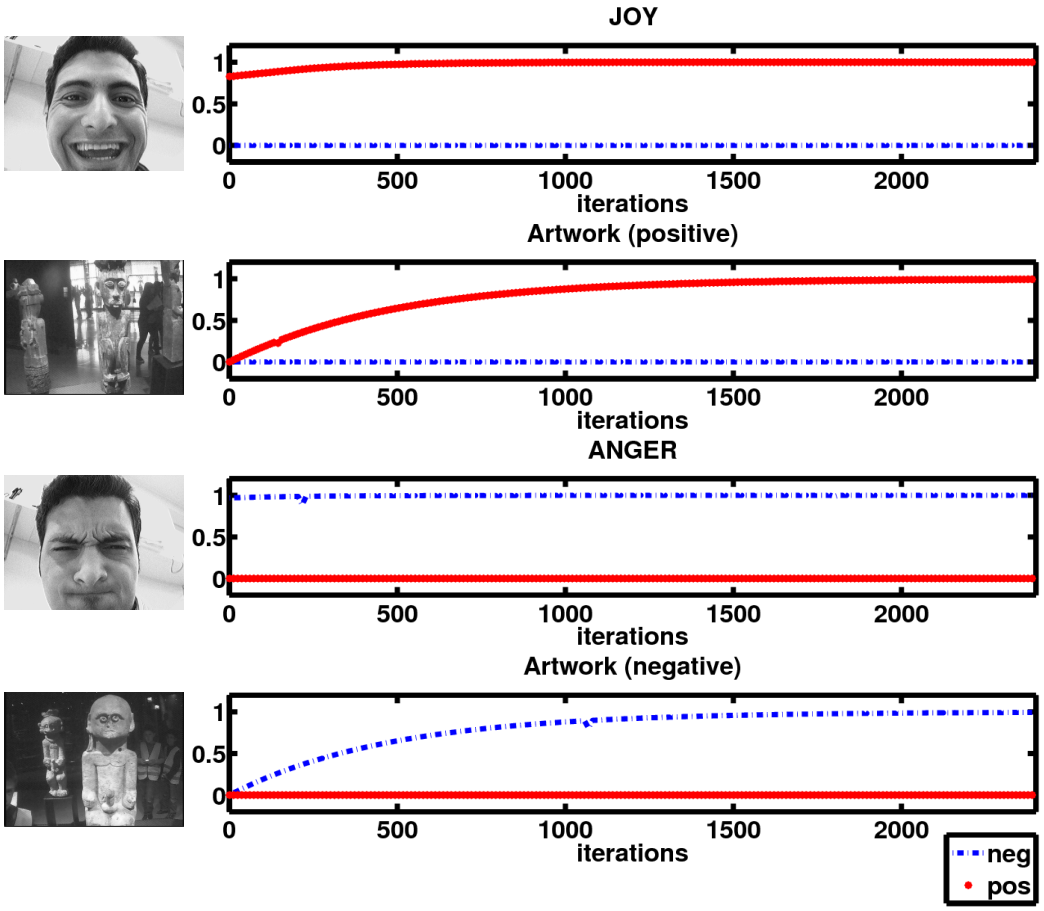


FIGURE 5.14 – Conditionnement au second ordre. Les lignes en pointillés rouges représente l'activité des réponses positives et les lignes en pointillés bleus les réponses négatives. Chaque stimulus est présenté trois fois dans l'ordre : Joie, Oeuvre_1, Colère, Oeuvre_2.

deux catégories (joie, colère) ne sont pas indépendantes, elles sont expérimentées ensembles durant le même scénario. Ceci amène un nouveau comportement que nous avons essayer d'éliminer. Celui d'empêcher un stimulus primaire de changer la valeur d'un autre stimulus primaire. Nous avons testé ce scénario avec le modèle de la première version (voir fig.5.9. Après avoir présenté tous les stimuli un certain nombre de fois, l'apprentissage est arrêté et les stimuli sont présentés les uns après les autres et on enregistre les réponses conditionnelles. Cette phase est appelée ici la phase de test dont les résultats sont montrés dans la figure fig.5.16. Dans cette figure les stimuli primaires et du second ordre sont bien associés, mais par la suite tous les autres stimuli sont associés au stimulus de joie.
Les résultats obtenus par la première version montre un effet des stimuli d'ordre supérieur sur les stimuli d'ordre inférieur. Le système d'habituation permet de contrôler les phases d'apprentissage mais ne contrôle pas dans ce cas là la catégorie propagée. Dans la seconde version du

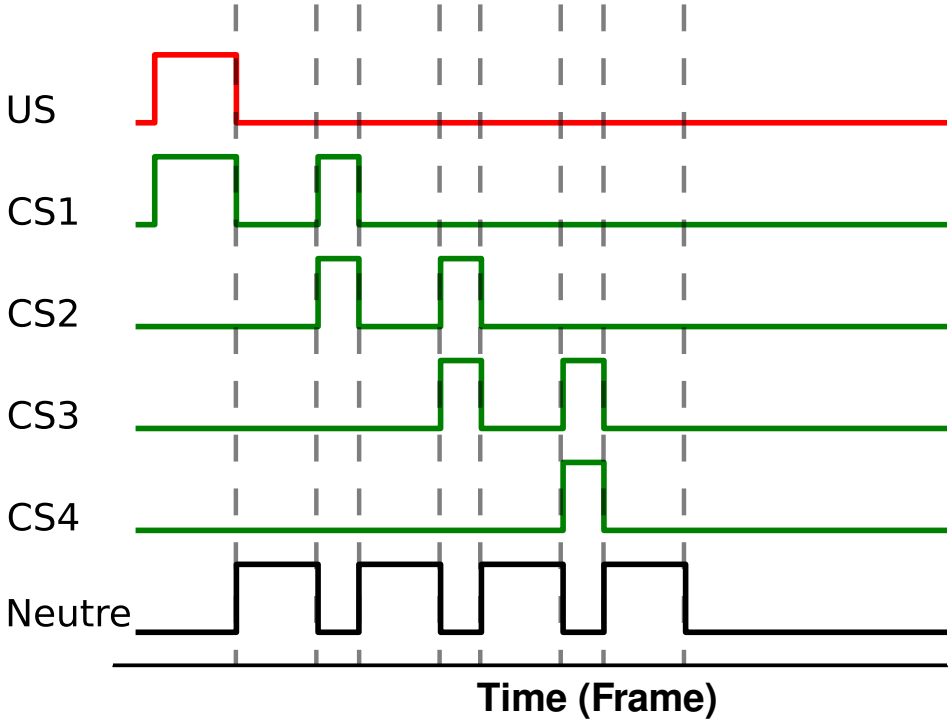


FIGURE 5.15 – Schéma de l'ordre des signaux présentés. Dans un premier temps l'US et le CS1 sont présentés ensemble pendant un certain temps. Ensuite le CS1 est présenté avec le CS2 pour diffuser le conditionnement de CS1 vers CS2. Cette séquence est ensuite répétée entre CS2 et CS3, puis CS3 et CS4. Entre deux présentation de CS le neutre est présenté seul pour permettre une deshabituation.

modèle montré dans la figure fig.5.10 nous proposons de propager uniquement la CR gagnante entre la réponse du stimulus actuel et celle du stimulus précédant. De la sorte, une réponse forte d'un stimulus ne sera pas altérée par la réponse d'un stimulus qui serait plus faible. Dans le modèle V1 nous avons introduit un groupe de compétition noté *WTA*. Ce groupe calcul la somme entre la CR et l'US sur chaque catégorie et active uniquement la catégorie gagnante. De plus dans ce modèle le résultat de l'habituation est propagé et non plus la CR. De cette façon, les stimuli suivant le stimulus primaire sont associés à des valeurs de plus en plus faibles dans le temps et à des vitesses d'apprentissage de plus en plus lentes. Ce comportement des valeurs est apprécié car il permet une différenciation des ordres de conditionnement. Les renforçateurs primaires ayant une valeur plus fortes que les renforçateurs d'ordre supérieur.
La base de donnée est rejouée en utilisant la seconde version du modèle dans l'ordre précisé dans 5.15. En premier lieu les expressions faciales sont associées à une catégorie émotionnelle issue du signal impératif simulé par les bouton d'une souris. Ce premier conditionnement, dit conditionnement primaire, attribue aux expressions faciales de l'humain une valeur de renforcement. Entre chaque conditionnement des stimuli neutres sont présentés. Une fois le signal impératif coupé, le visage en colère de l'expérimentateur est présenté avec le signe BAD, puis après une phase de neutre, le visage joyeux de l'expérimentateur est présenté avec le signe GOOD. Cette seconde phase effectue un conditionnement au second ordre des signaux (BAD et GOOD) grâce au renforcement des stimuli primaires. Pour le conditionnement au troisième ordre les si-

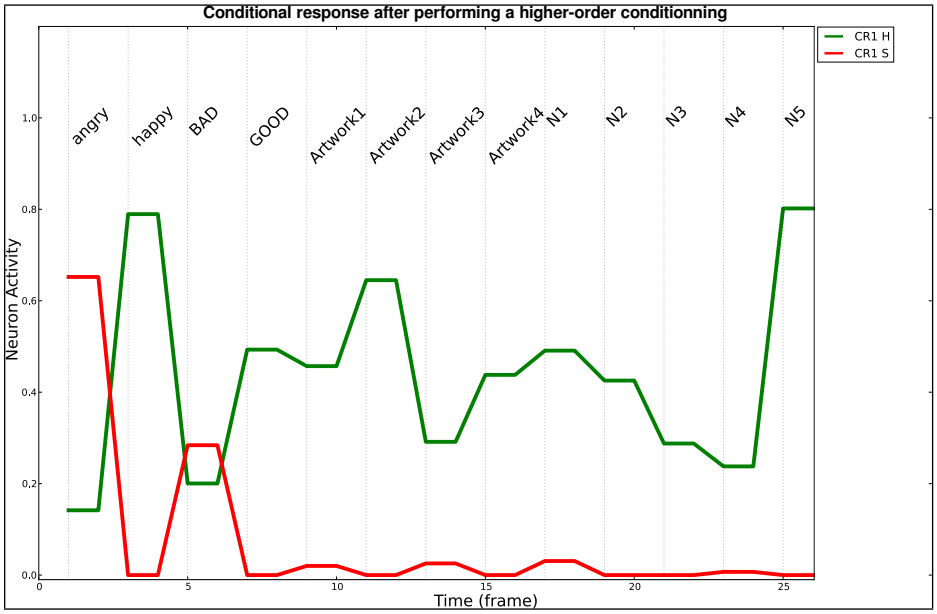


FIGURE 5.16 – Résultat de la phase de test du modèle V0 fig.5.9 sur la base de données 5.11 après un apprentissage dans l'ordre montré dans fig.5.15.

gnaux sont présentés avec les œuvres (œuvre-1 et œuvre-2). Finalement, un conditionnement au quatrième ordre est appris dans lequel les œuvres (œuvre-1 et œuvre-2) conditionnent d'autres œuvres du musée (œuvre-3 et œuvre-4).
La base de donnée est rejouée 20 fois et la figure 5.17 montre les résultats en fin d'apprentissage. Les réponses au valences positives sont affichées en vert, les réponses des valences négatives sont en rouge et finalement les moments d'apprentissage sont en bleu. Il est intéressant de noter l'arrêt des apprentissages lors de la présentation des stimuli neutres.
Une fois la phase apprentissage terminé, le réseau arrête d'apprendre et les stimuli sont présentés les uns après les autres au système. Les réponses conditionnelles sont enregistrées et affichées. Cette phase de test nous permet d'évaluer les valeurs de renforcement acquises par chaque stimulus. La figure 5.18 montre les résultats de la phase de test de la seconde version (modèle V1). Contrairement aux résultats observés avec le modèle V0 (5.16), le modèle V1 en comparant les réponses conditionnelles des stimuli évite au système de propager la même catégorie émotionnelle sur tous les autres stimuli. Sur la figure 5.18, l'expression de joie, le signe "GOOD" et l'œuvre 2 sont tous associés à la catégorie positive. Par ailleurs, les réponses des stimuli sont discriminables. Aucune hiérarchie n'est à priori écrite dans le modèle, elle émerge de l'ordre de présentation des stimuli. Les conditionnement d'ordre inférieur ont une réponse conditionnelle plus forte que les conditionnements d'ordre inférieur. Par effet de généralisation, les stimuli neutres donnent aussi lieu à des réponses conditionnelles. Ici nous ne nous intéressons pas à la réponse des stimuli neutre, mais il serait intéressant d'évaluer l'impact de la généralisation sur le système de conditionnement.
Le second modèle donne des résultats satisfaisant sur la base de donnée. Néanmoins, à cause des valeurs faibles des réponses conditionnelles ce modèle est plus difficile à contrôler sur ca-

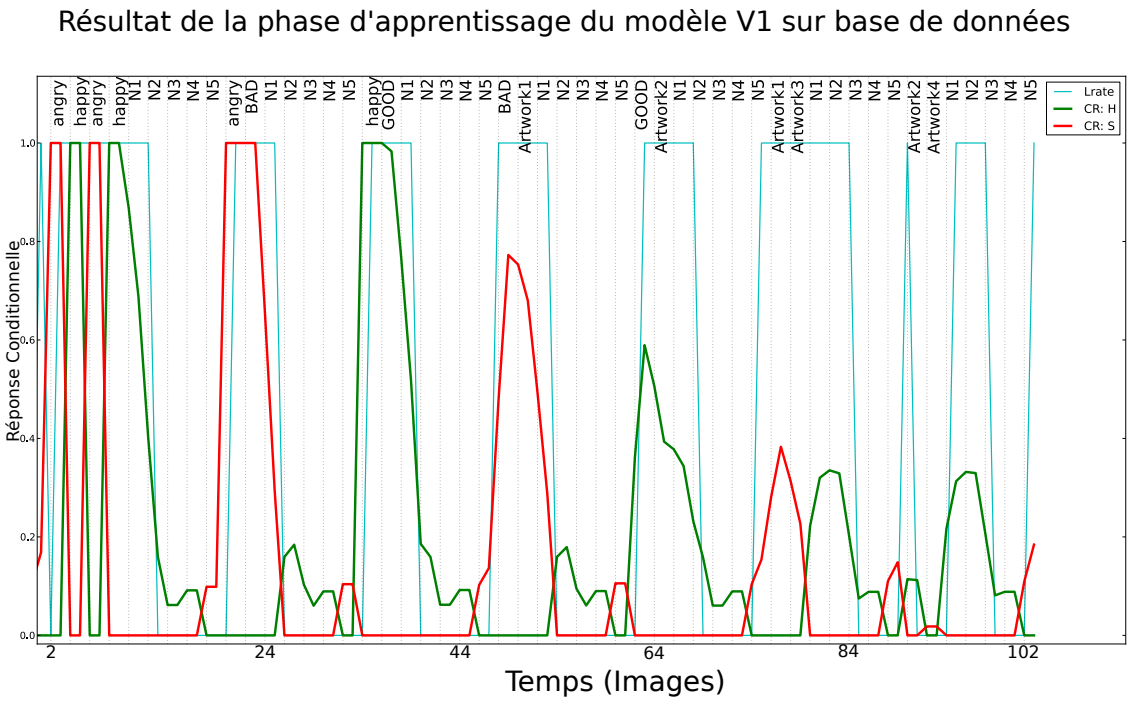


FIGURE 5.17 – Réponses conditionnelles en phase d'apprentissage sur base de données de la seconde version du modèle (figure 5.10). Les réponses des valences positives sont en vert, les réponses de valences négatives en rouge et les moments d'apprentissage en bleu. Les images présentées au système sont décrite en haut de la figure au moment où elles sont en entrée.

méra réelle. Sur base de données les temps de présentation des stimuli sont connus a priori. Les temps d'intégrations de l'habituation peuvent être facilement définis. Seulement, sur robot en interaction avec des humains, les stimuli ne sont pas tous présentés pendant les mêmes durées. Les temps d'habituation ne peuvent plus être défini de manière précises dès le début. De plus ici la diminution des réponses est fabriquée artificiellement en réinjectant l'écart entre la CR et l'habituation. Nous avons alors modifié le signal réentrant en utilisant directement la CR. L'habituation aussi est modifiée de sorte à avoir une entrée binaire. La CR est multipliée par l'habituation pour remettre à zéro la compétition des US. Le WTA est remplacé par un max suivi d'un WTA pour éviter d'avoir la somme des signaux.
Avant d'utiliser la dernière version du modèle de conditionnement d'ordres supérieurs sur une caméra réelle, nous l'avons testé sur la base de données. Le but est de savoir si les temps d'apprentissages sont contrôlés et si les associations s'effectuent correctement.
La figure 5.19 montre le résultat des activités de reconnaissance lors de la dernière phase d'apprentissage (après 20 présentations de la même base de données). L'apprentissage démarre lors de la présentation d'un stimulus ayant déjà acquis un renforcement et s'arrête une fois le stimulus suivant présenté. Une fois cette phase d'apprentissage terminée les stimuli sont présentés au système en phase de test et les réponses conditionnelles sont montrées dans la figure 5.20. Ce dernier modèle exhibe les mêmes résultats que le modèle précédent. Il n'y a pas de destruction d'apprentissage. Par ailleurs, les valeurs de reconnaissance étant assez forte, ce modèle se prête

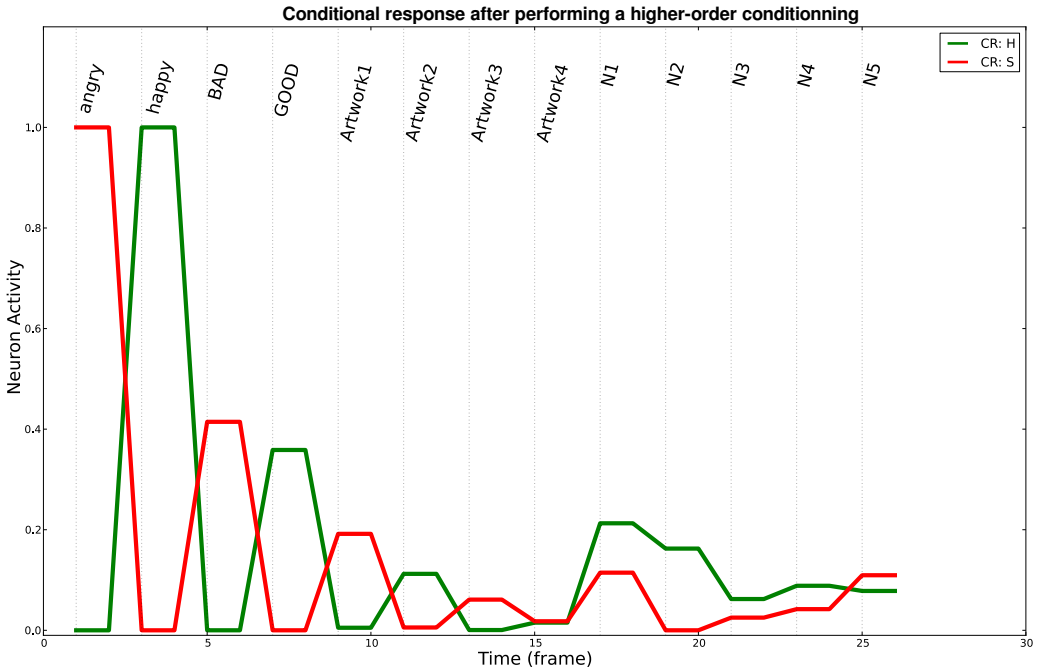


FIGURE 5.18 – Réponses conditionnelles en phase de test sur base de données de la seconde version du modèle (figure 5.10).

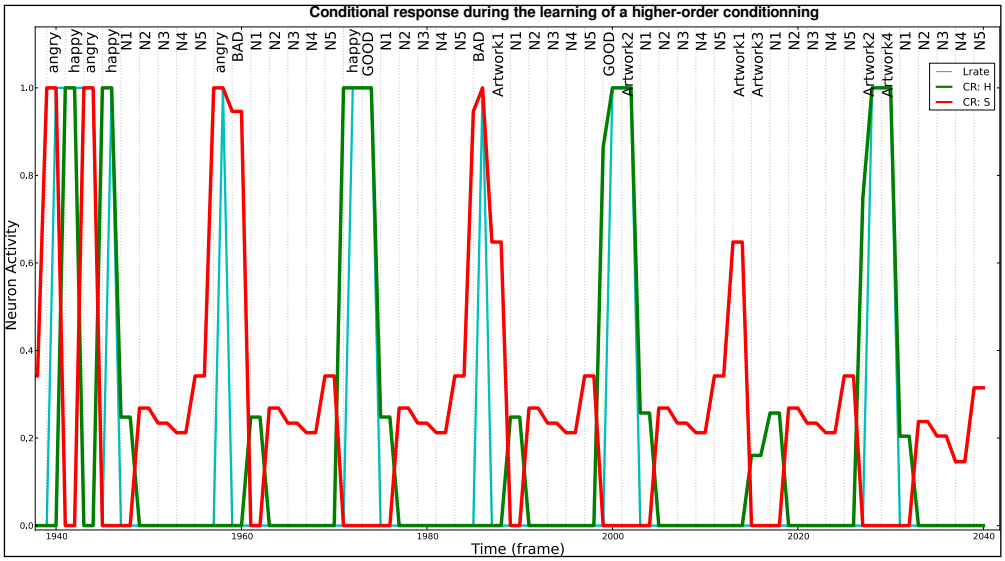


FIGURE 5.19 – Réponses conditionnelles en phase d'apprentissage sur base de données de la dernière version du modèle (figure 5.8).

mieux à un test en dynamique.

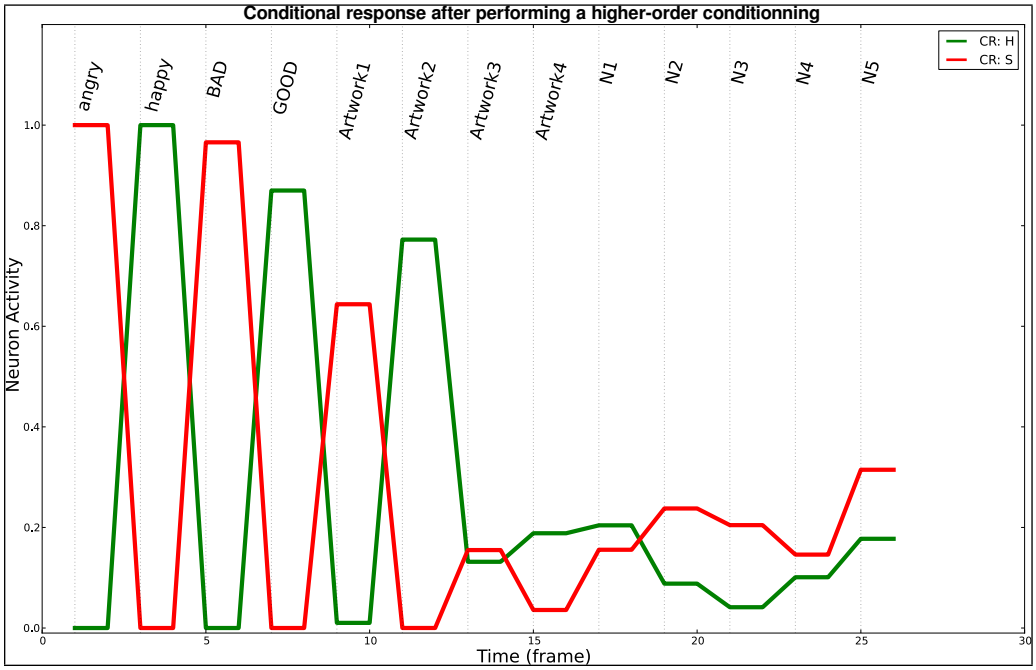


FIGURE 5.20 – Réponses conditionnelles en phase de test sur base de données de la dernière version du modèle (figure 5.8)

5.6.2 Expérience sur robot

Les tests sur base de données nous ont permis de choisir le modèle le plus approprié pour un test sur robot réel. Bien que l'utilisation des bases de données permettent d'éliminer de facto les modèles erronés, dans le cas de la seconde version du modèle il a fallu faire un test sur robot pour voir ses limites et changer de variante. Ceci nous a mené à ajouter une nouvelle variante au modèle présentée dans la figure 5.8 (noté *modèleVF* dans la suite). Après avoir validé le comportement du système avec le dernier modèle sur base de données (voir ci-dessus) nous l'avons testé sur robot réel. Dans la suite, seuls les résultats de la dernière version sont montrés. Le robot ici se résume à une tête expressive munie de servomoteurs et d'une caméra. La même tête expressive que celle décrite dans le chapitre 3 est utilisée. Tout d'abord un jeu d'imitation permet au robot d'apprendre les expressions faciales du robot (voir ci-dessus). Au début la tête expressive affiche des expressions faciales (joie, colère, expression neutre) et on demande au tuteur (l'expérimentateur) d'imiter le système pendant un certain temps. Le réseau de neurone associe l'expression du robot avec l'expression faciale du tuteur. Durant toute cette étape l'apprentissage est forcé par la présence d'un signal de conditionnement impératif. Une fois l'apprentissage terminé, le tuteur affiche des expressions faciales et il est imité par le système.
Contrairement au cas sur base de données, la catégorie des valences neutres est apprise dans l'expérience sur robot. Cette catégorie permet au robot de reconnaître l'expression faciale neutre du tuteur, d'afficher un visage neutre et de s'habituer à une nouvelle catégorie pour pouvoir enclencher l'apprentissage à nouveau en cas de changement.
Une fois l'apprentissage terminé, il est figé et le tuteur affiche des expressions faciales que

le robot imite. Les résultats de reconnaissance (réponses conditionnelles) sont enregistrés et affichés dans la figure 5.21.

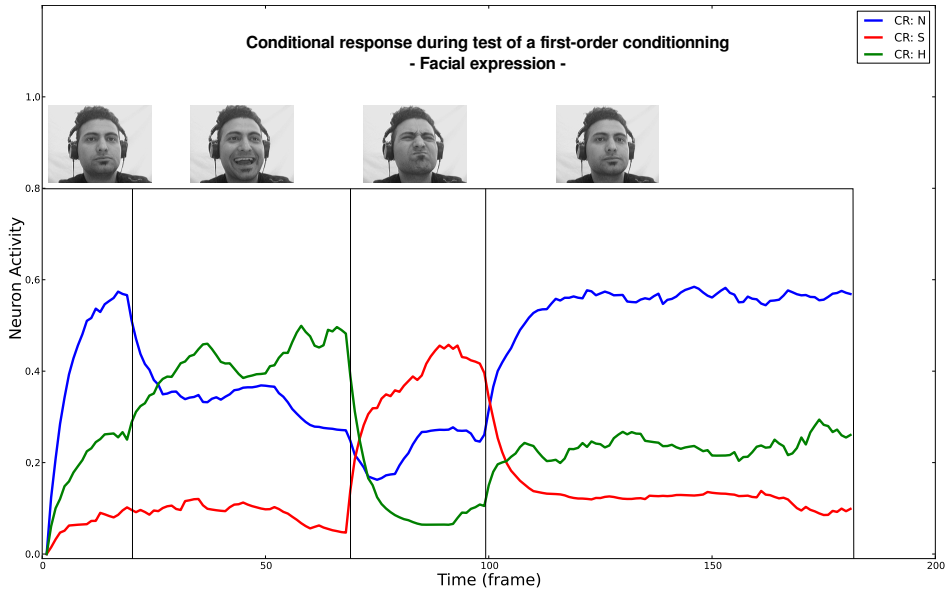


FIGURE 5.21 – Réponses conditionnelles après un conditionnement du premier ordre. Les images au dessus des courbes montrent l'expression faciale affichée par le "tuteur".

Dans cette figure les courbes représentent les activités des réponses conditionnelles et les expressions faciales affichées par le robot. Les images en haut de la courbe montre le visage du tuteur tel que perçut par le robot. Pour chaque expression du tuteur l'activité la plus forte de la CR est celle attendue. Néanmoins, comme des parties du visage sont communes à toutes les expressions, les CR des autres catégories s'expriment aussi. Nous ne considérons ici que la catégorie gagnante avec la plus forte activité.
Une fois que les expressions faciales du tuteur sont apprises, elle deviennent des renforçateurs primaires et peuvent être utilisées pour conditionner des stimuli présentés par la suite. Le tuteur fait une expression (joie ou colère) puis présente un objet à la caméra du robot. Entre deux présentation d'objet le tuteur doit afficher une expression neutre pour éloigner la présentation de deux stimuli les uns des autres et éviter que ces stimuli se conditionnent entre eux. La figure 5.22 montre les résultats en phase de test pour un stimulus du second ordre.
Pour la dernière étape nous avons cherché à mettre en évidence la capacité de notre système à établir un conditionnement au troisième ordre. Ici le tuteur montre à la caméra un objet ayant acquis un renforcement secondaire (CS2) puis le fait suivre d'un objet jamais appris (CS3). La figure 5.23 montre les courbes de test.

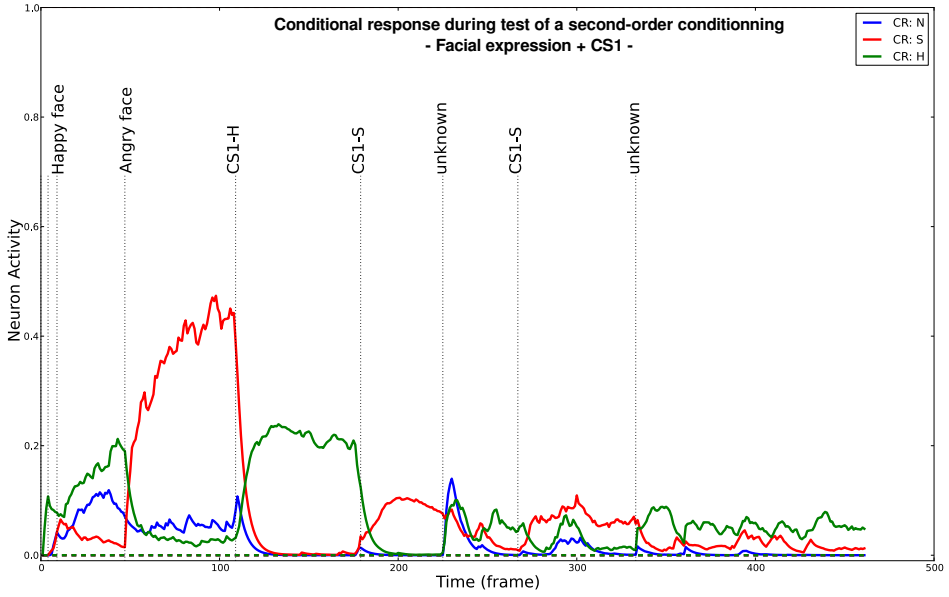


FIGURE 5.22 – Réponses conditionnelles en phase de test après un apprentissage au second ordre.

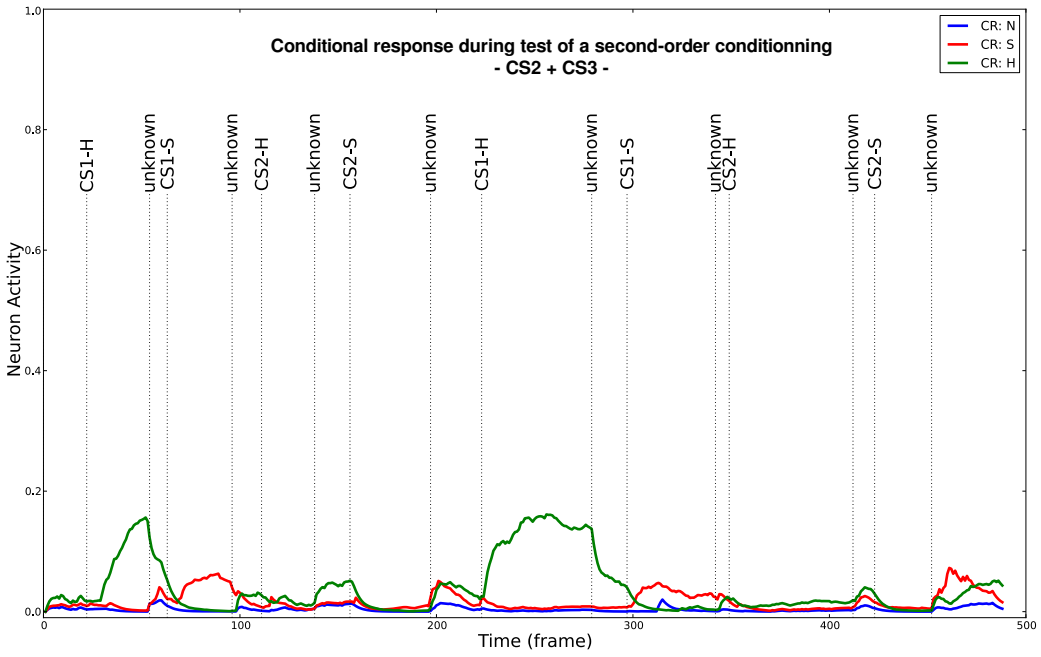


FIGURE 5.23 – Réponses conditionnelles en phase de test après un apprentissage au second ordre.

5.7 Conclusion et Discussion

Nous avons présenté dans ce chapitre un modèle de conditionnement aux ordres supérieurs. Le début du chapitre décrit comment à partir d'une simple architecture de conditionnement notre robot a développé la capacité à reconnaître les expressions faciales d'autrui et d'utiliser cette reconnaissance comme un outil de communication non verbale pour de la référenciation sociale. Dans le cadre de la référenciation sociale le robot peut attribuer des valences émotionnelles à des objets à priori neutres. Ces associations permettent au robot de développer ses appréciations esthétiques à travers des interactions sociales. Néanmoins, le modèle basé sur la cascade de conditionnements connaît quelques limites. Premièrement, dans le cas de plusieurs conditionnements en cascade, le système nécessite de fortes ressources computationnelles. Deuxièmement, chaîner les conditionnements en boucle ouverte suppose une hiérarchie entre ces derniers. Ainsi seule la présence de l'humain peut enclencher l'apprentissage. Cet apprentissage se termine en bout de chaîne. Ceci facilite le contrôle des apprentissages mais ne leur permet pas de se propager dans le temps. Finalement, avec un tel système l'historique du robot ne joue aucun rôle dans son développement. En effet, tant que les couples (expression de l'humain objet) présentés sont les mêmes, les réponses du système seront les mêmes. Ceci quelque soit l'ordre de présentation des couples. De plus en absence de l'humain le robot ne peut continuer à apprendre.
Pour permettre à notre robot d'apprendre en autonomie et de considérer son historique dans le développement de ses préférences esthétiques, nous avons proposé un modèle de conditionnement aux ordres supérieurs. Ce modèle s'inspire des travaux de Pavlov, de Rescorla et des décharges dopaminergique pour réguler l'apprentissage. De plus le lien entre la référenciation sociale et les conditionnements du second ordre a été observé et discuté par Feinman and Lewis [1983]. Avec les conditionnement d'ordre supérieur deux robots strictement identiques, ayant accès à la même base de données d'apprentissage en interaction avec leur entourage (ex. visiteurs du musée) exhiberont des comportements et des jugements différents vis-à-vis des œuvres d'art. En effet, leurs préférences esthétiques seront aussi dirigées par leur vécu, leur manière de se déplacer dans une scène et les trajectoires différentes qu'ils peuvent prendre. Cette question est centrale dans le développement de ce modèle. Nous considérons le conditionnement d'ordre supérieur et la liaison entre les stimuli comme le quatrième niveau (et le dernier dans cette thèse) du développement de l'esthétique artificielle.
Dans ce chapitre plusieurs modèles sont proposés et testés sur un scénario expérimental qui reprend les scenarii utilisés dans la littérature. Le dernier modèle proposé est testé sur un robot réel dans des conditions d'interaction dynamique. Ce modèle vient simplifier le modèle de référencement social par cascade de conditionnement. Il n'y a aucun à priori sur l'ordre des stimuli. On peut facilement imaginer que les expressions faciales soient d'un ordre supérieur et qu'une œuvre du musée acquièrent un renforcement primaire. Par contre pour avoir cette simplification il a été nécessaire d'ajouter au bloc de conditionnement un second bloc de contrôle de l'apprentissage. Ce second réseau de neurones évalue la pertinence d'une réponse conditionnelle et décide si il faut la propager ou bien arrêter l'apprentissage. Le réseau d'habituation propage une CR pendant un certain temps. Ce temps peut être vu comme le temps de décharge des neurones dopaminergiques. La dopamine a longtemps été vu comme un neuromodulateur qui enclenche l'apprentissage, mais d'autres travaux (Redgrave and Gurney [2006]) lui attribuent un rôle de liant entre deux événements. Lorsqu'un apprentissage est enclenché, il dure pendant un certain temps et associe tous les stimuli conditionnels qui apparaissent dans ce laps de temps avec le

stimulus inconditionnel.
Le modèle proposé ici est testé sur des scènes naturelles. Il est alors plus facile de l'intégrer dans les expériences sur l'esthétique artificielle. Néanmoins, le modèle dans l'état actuel connaît certaines limites notamment dues à la richesse des stimuli visuels qui ne facilitent pas son utilisation par un non-initié. Nous utilisons d'une part plusieurs stimuli, et d'autre part plusieurs catégories émotionnelles. Il faut veiller non seulement à garder les conditionnements primaires d'une même catégorie stable, c'est à dire veiller à ce qu'un stimulus avec une faible réponse ne vienne écraser l'apprentissage d'un autre. De plus, il faut aussi veiller à ce qu'une catégorie émotionnelle n'influence pas une autre.
Nous avons choisi pour ce modèle une association stimulus-réponse. D'autres modèles de l'amygdale introduisent aussi des associations stimulus-stimulus (Mannella et al. [2008]). Dans ces modèles là le CS2 évoque d'abord le CS1 qui lui évoque l'US pour finalement donner lieu à une réponse conditionnelle. Cette association stimulus-stimulus serait intéressante à explorer dans la suite des travaux.
Notre modèle a été testé sur un robot minimal (tête expressive plus caméra). Par la suite nous prévoyons deux autres expériences à mener sur un robot. Dans la première expérience, le modèle de conditionnement du second ordre sera utilisé pour apprendre des conditionnements dans le cadre d'un attention partagée. A ce jour les travaux menés sur l'attention partagée dans notre laboratoire utilisent deux caméras, une dirigée vers le visage du tuteur et une autre dirigée vers les objets à apprendre. Chacune de ces caméras est dotée d'une bloc de conditionnement PerAc. Le but serait alors d'utiliser une seule caméra capable de passer d'un stimulus à un autre et un bloc de conditionnement au second ordre pour propager l'expression du tuteur sur l'objet à apprendre. Dans la seconde expérience nous prévoyons de remplacer notre modèle d'apprentissage des préférences esthétiques par le modèle de conditionnement du second ordre et d'effectuer des tests au musée pour évaluer la capacité de notre modèle à tenir le passage à l'échelle et d'évaluer l'impact d'un tel apprentissage sur le comportement de notre robot Berenson.