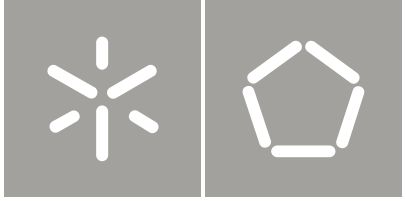
Universidade do Minho
Escola de Engenharia

Rui Manuel Gomes da Silva

Toward robots as socially aware partners in human-robot interactive tasks

Outubro de 2016



Universidade do Minho
Escola de Engenharia

Rui Manuel Gomes da Silva

Toward robots as socially aware partners in human-robot interactive tasks

Tese de Doutoramento
Programa Doutoral em Engenharia Eletrónica e de Computadores
Especialidade em Controlo, Automação e Robótica

Trabalho efetuado sob a orientação de
Professora Doutora Estela G. Silva Bicho Erlhagen
Professor Doutor Pedro Sérgio Oliveira Branco

Outubro de 2016

STATEMENT OF INTEGRITY

I hereby declare having conducted my thesis with integrity. I confirm that I have not used plagiarism or any form of falsification of results in the process of the thesis elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

University of Minho, 2016 / 10 / 21

Signature: Rui Manuel Gomes da Silva

(Rui Manuel Gomes da Silva)

This page was intentionally left blank.

Acknowledgements

This work is the conclusion of a path initiated long ago that would not be possible to undertake alone. I wish to thank all the people that I've come across and that in some way, made a contribution, no matter how small, to the success of this work.

First and foremost, I would like to express my deepest gratitude to my scientific advisor Prof. Estela Bicho, who during these years, even before I started my PhD, has been my mentor and nursed me to grow either scientifically and personally, through her support, confidence and knowledge. Also to my co-adviser Prof. Pedro Branco, who contributed with its expertise in one very important aspect of this work, the facial expressions, providing guidance throughout the work. During these years I had the opportunity to learn from them, which proved to be of the utmost importance.

My everyday work took place at the MARLab, in the University of Minho, where I had the privilege to work with a group of very talented people that together bring robots to 'life'. A special thank you is in order for Luís Louro, the person I worked with closely throughout my PhD, he was tirelessly supportive during the implementation of the robot's architecture and the robot experiences, and proved to be an invaluable work colleague and a true friend. To all the other people I met at the lab, namely (alphabetically), Angela Loureiro, António Ramiro, Carlos Faria, Carolina Vale, Emanuel Sousa, Eliana Costa e Silva, Flora Ferreira, Gianpaolo Gulleta, Miguel Cunhal, Miguel Sousa, Nzoji Hipólito, Paulo Vicente, Simão Antunes, Tiago Malheiro, Toni Machado and Weronika Wojtak. They created the most amazing work environment anyone can expect.

To my family, my parents and brother that throughout my life always supported and cared

for me, without them this would not be possible.

At last, but certainly not the least, to my wife Sílvia, for being part of my life and offering me her love and support during these years. In the most demanding times, she was the one that kept me going.

And finally to my lovely children Alexandre and Margarida, you are truly my biggest inspiration.

This work was supported by a Portuguese FCT (Fundação para a Ciência e a Tecnologia) grant SFRH/BD/48527/2008 financed by POPH-QREN 4.1-Advanced Training, co-funded by the European Social Fund and national funds from MEC (Ministério da Educação e Ciência). Also to Project JAST (Joint Action Science and Technology) (FP6-IST, Proj. n. 003747) financed by the European Commission and Project NETT (Neural Engineering Transformative Technologies) financed by the European Commission through its Marie Curie Training Network (grant agreement no 289146) for providing important resources for the development of this work.

Abstract

A major challenge in modern robotics is the design of socially intelligent robots that can cooperate with people in their daily tasks in a human-like way. Needless to say that non-verbal communication is an essential component for every day social interaction. We humans continuously monitor the actions and the facial expressions of our partners, interpret them effortlessly regarding their intentions and emotional states, and use these predictions to select adequate complementary behaviour. Thus, natural human-robot interaction or joint activity, requires that assistant robots are endowed with these two (high level) social cognitive skills.

The goal of this work was the design of a cognitive control architecture for socially intelligent robots, heavily inspired by recent experimental findings about the neurocognitive mechanisms underlying action understanding and emotion understanding in humans. The design of cognitive control architectures on these basis, will lead to more natural and efficient human-robot interaction/collaboration, since the team mates will become more predictable for each other.

Central to this approach, neuro-dynamics is used as a theoretical language to model cognition, emotional states, decision making and action. The robot control architecture is formalized by a coupled system of Dynamic Neural Fields (DNFs) representing a distributed network of local but connected neural populations with specific functionalities. Different pools of neurons encode relevant information about hand actions, facial actions, action goals, emotional states, task goals and context in the form of self-sustained activation patterns. These patterns are triggered by input from connected populations and evolve continuously in time under the influence of recurrent interactions. Ultimately, the DNF architecture implements a dynamic context-dependent mapping from observed hand and facial actions of the human onto adequate

complementary behaviours of the robot that take into account the inferred goal and inferred emotional state of the co-actor.

The dynamic control architecture has been validated in multiple scenarios of a joint assembly task in which an anthropomorphic robot - ARoS - and a human partner assemble a toy object from its components. The scenarios focus on the robot's capacity to understand the human's actions, and emotional states, detected errors and adapt its behaviour accordingly by adjusting its decisions and movements during the execution of the task. It is possible to observe how in the same conditions a different emotional state can trigger different a overt behaviour in the robot, which may include different complementary actions and/or different movements kinematics.

Resumo

Um dos desafios da robótica atual é o desenvolvimento de robôs socialmente inteligentes que consigam interagir e cooperar com humanos nas suas tarefas diárias de uma forma semelhante a estes. Na interação social diária, a comunicação não verbal desempenha um papel fundamental. Os humanos monitorizam constantemente as ações e expressões faciais, interpretando-as facilmente no que toca à sua intenção e estado emocional, e usam essa informação na seleção de comportamentos complementares adequados. Desta forma, uma interação humano-robô natural ou uma atividade conjunta, requer que os robôs estejam dotados destas duas competências sociais (de alto nível).

O objetivo deste trabalho passou pelo desenvolvimento de uma arquitetura de controlo para tornar robôs socialmente inteligentes, com inspiração em estudos recentes acerca dos mecanismos neuro-cognitivos subjacentes à interpretação de ações e emoções nos humanos. O projeto de arquiteturas de controlo cognitivo assente nesta base permitiu potenciar uma interação/colaboração humano-robô mais eficiente e natural, já que ambos os intervenientes se tornam assim mutuamente mais previsíveis.

A arquitetura de controlo para o robô é formalizada usando a teoria de Campos Dinâmicos Neuronais (CDNs), que representa uma rede distribuída de populações neuronais conectadas cada uma com objetivos específicos. Um aspeto importante nesta abordagem é a linguagem teórica usada para modelar cognição, estados emocionais, tomada de decisões e ações. Cada população neuronal codifica informação relevante acerca de ações levadas a cabo com as mãos, ações faciais, intenções subjacentes a ações, objetivos da tarefa e contexto na forma de padrões de ativação auto-sustentados. Estes padrões são desencadeados ao receber entradas de

populações conectadas e evoluem continuamente no tempo através da influência de interações recorrentes. Sinteticamente, a arquitetura baseada em CDNs implementa um mapeamento dinâmico, dependente do contexto, de ações observadas e expressões faciais na seleção de comportamentos complementares do robô que tem em consideração o objetivo e o estado emocional inferidos do parceiro.

A validação da arquitetura de controlo foi realizada em múltiplos cenários de uma tarefa conjunta de construção, onde um robô antropomórfico - ARoS - e um parceiro humano construíram um modelo de um brinquedo a partir dos seus componentes individuais. Os cenários focaram-se na capacidade do robô interpretar ações realizadas pelo humano, estados emocionais e erros detetados, adaptando o seu comportamento, e fazendo ajustes nas decisões tomadas e nos movimentos realizados durante a execução da tarefa. Foi possível observar que, nas mesmas condições, um estado emocional diferente pode desencadear no robô um comportamento diferente, o qual pode, por sua vez, incluir uma ação diferente e/ou uma cinemática de movimento diferente.

Contents

This page was intentionally left blank.

List of Abbreviations

AAM Active Appearance Model

AEFA Action Execution of Facial Actions sets

AEHA Action Execution of Hand Actions

AEL Action Execution Layer

ALICE Artificial Linguistic Internet Computer Entity

AOL Action Observation Layer

API Application Programming Interface

ARoS Anthropomorphic Robotic System

ASFA Action Simulation of Facial Actions sets

ASHA Action Simulation of Hand Actions

ASL Action Simulation Layer

AU Action Unit

CSGL Common Sub-Goals Layer

DNF Dynamic Neural Field

DOF Degree Of Freedom

EMFACS Emotional Facial Action Coding System

EMG Electromyography

EML Error Monitoring Layer

EMYS EMotive headY System

ESL Emotional State Layer

FACS Facial Action Coding System

FLASH Flexible LIREC Autonomous Social Helper

HSV Hue Saturation Value

HLS Hue Lightness Saturation

HCI Human-Computer Interaction

HRI Human-Robot Interaction

IL Intention Layer

JAST Joint Action Science and Technology

LIREC Living with Robots and interactive Companions

MEXI Machine with Emotionally eXtended Intelligence

NETT Neural Engineering Transformative Technologies

OML Object Memory Layer

OpenCV Open Source Computer Vision

PCA Principal Component Analysis

PFC Pre-Frontal Cortex

QoM Quantity of Movement

RGB Red Green Blue

SMI Silhouette Motion Image

This page was intentionally left blank.

List of Figures

This page was intentionally left blank.

List of Tables

This page was intentionally left blank.

Chapter 1

Introduction

Human-robot interaction aims to be in the future, as natural and fluent as human-human interaction. Robots must then be endowed with the necessary set of abilities that support this seamless interaction. One good example that can be used to explore human-robot interaction is in scenarios of collaborative joint activity.

1.1 Natural human-robot interaction and collaboration: the role of emotions

A major challenge in modern robotics is the design of socially intelligent robots that can interact or cooperate with people in their daily tasks in a human-like way. Needless to say that non-verbal communication is an essential component for every day social interactions. We humans continuously monitor the actions and the facial expressions of our partners, interpret them effortlessly regarding their intentions and emotional states, and use these predictions to select adequate complementary behaviour. Thus, natural human-robot interaction or joint activity, requires that assistant robots are endowed with these (high level) social cognitive skills.

There has been various kinds of interaction studies that have explored the role of emotion/affect in Human-Robot Interaction (HRI) (e.g. Breazeal, 2003a; b; Cañamero and Fredslund, 2000; Hegel et al., 2006; Kirby et al., 2010; Leite et al., 2010; Novikova and Watts, 2015;

Pereira et al. 2011). The results of such studies have clearly shown that endowing robots with –the recognition and display of human-like - emotions/affects critically contributes to making the HRI more natural and meaningful, from the perspective of the human interacting with the robot (e.g. Breazeal 2003a,b; Cañamero 2005; Hegel et al., 2006; Kędzierski et al. 2013; Leite et al. 2010; Pereira et al. 2011). Robots with infant-like abilities of interaction, such as Kismet (Breazeal, 2003b), have been used to demonstrate the ability of people to interpret and react appropriately to a robot's displays of emotions. Experiments with the robot Vikia (Bruce et al. 2002) demonstrated the effectiveness of an emotionally expressive graphical face for encouraging interactions with a robot.

However, in all these interaction experiments the robot and the human were not a team, in that the interactions did not involve joint action tasks. One exception goes to the work reported in Scheutz et al. (2006), where the robot and the human were both needed for the task and neither robot nor human could accomplish the task alone. Their results have shown that expressing affect and responding to human affect with affect expressions can significantly also improve team performance in a joint human-robot task. However, the human and the robot interacted solely based on 'natural' language, there was no physical interaction, and the robot was not making autonomous decisions, i.e. the robot always carried out human orders (see also Scheutz, 2011).

The work here reported aims to contribute to filling in this gap. Our approach is motivated by recent research in cognitive psychology and cognitive neuroscience that posits that various kinds of shared emotions can, not only motivate participants to engage and remain engaged in joint actions, but also facilitate processes that are central to the coordination of participants' individual actions within joint action, such as representing other participants' tasks, predicting their behaviour, detecting errors and correcting accordingly, monitoring their progress, adjusting movements and signalling (Michael, 2011; Rizzolatti and Sinigaglia, 2008).

There are a number of reasons why the inclusion of emotions in cognitive architectures may be necessary (Hudlicka, 2004), including: (a) research-motivated emulation of human decision-making and appraisal processes, to better understand their mechanisms; (b) enhancing agent and robot realism (e.g. for training and educational purposes, or assistive roles); (c) developing

user and operator models for improved adaptive Human-Computer Interaction (HCI).

In order to combine emotions into the decision making and complementary behaviour of an intelligent robot cooperating with a human partner our group relies on the development of control architectures for human-robot interaction that are strongly inspired by the neurocognitive mechanisms underlying joint action (Bekkering et al., 2009; Poljac et al., 2009; van Schie et al., 2008) and shared emotions in humans (Carr et al., 2003; Iacoboni et al., 2005; Wicker et al., 2003). Implementing a human-like interaction model in an autonomous assistive robot will greatly increase the user's acceptance to work with the artificial agent since the co-actors will become more predictable for each other (see also e.g. Fong et al., 2003; Kirby et al., 2010).

Humans have a remarkable ability to perform fluent organization of joint action, achieved by anticipating the motor intentions of others (Sebanz et al., 2006). An impressive range of experimental findings, about the underlying neurocognitive mechanisms, support the notion that a close perception-action linkage provides a basic mechanism for real-time social interactions (Newman-Norlund et al., 2007a; Wilson and Knoblich, 2005). A key idea is that action observation leads to an automatic activation of motor representations that are associated with the execution of the observed action. It has been advanced that this motor resonance system supports an action understanding capability (Blakemore and Decety (2001); Fogassi et al. (2005)). By internally simulating action consequences using his own motor repertoire the observer may predict the consequences of others' actions. Direct physiological evidence for such perception-action system came with the discovery of the so-called mirror neurons in the pre-motor cortex of the macaque monkey (for a review see Rizzolatti and Craighero (2004)). These neurons are a particular class of visuomotor neurons that are active during the observation of goal-direct actions, such as reaching, grasping holding or placing an object and communicative actions, and during execution of the same class of actions (Ferrari et al., 2003; Rizzolatti et al., 2001). More recently, Bekkering et al. (2009) have assessed, through neuroimaging and behavioural studies, the role of the human mirror neuron system while participants prepared to execute imitative or complementary actions. They have shown that the human mirror neuron system may be more active during the preparation of complementary than during imitative actions (Newman-Norlund et al., 2007a), suggesting that the it may be essential in dynamically

coupling action observation on to (complementary) action execution, and that this mapping is much more flexible and than previously thought (Poljac et al., 2009; van Schie et al., 2008).

There is also good evidence in neuroscience studies that also exists a facial expressions mirroring system. The work by Leslie et al. (2004) shows results that are consistent with the existence of a face mirroring system located in the right hemisphere (RH) part of the brain, which is also associated with emotional understanding (Ochsner and Gross, 2005). Specifically, the right hemisphere premotor cortex may play a role in both the generation and the perception of emotionally expressive faces, consistent with a motor theory of empathy (Leslie et al., 2004). van der Gaag et al. (2007) present a more in-depth study on the role of mirror neurons in the perception and production of emotional and neutral facial expressions. The understanding of other people from facial expressions is a combined effort of simulation processes within different systems, where the somatosensory, motor and limbic systems all play an important role. This process might reflect the translation of the motor program, emotions and somatosensory consequences of facial expressions, respectively (Keysers and Gazzola, 2006). The simulation processes in these individual systems have been previously described in the literature (Gallese et al., 1996; Keysers et al., 2004; Wicker et al., 2003). Specifically, and at a neuronal level, premotor mirror neurons might resonate the facial movement and its implied intention (Carr et al., 2003; Iacoboni et al., 2005), insula mirror neurons might process the emotional content (Wicker et al., 2003), and somatosensory neurons might resonate proprioceptive information contained in the observed facial movement (Keysers et al., 2004). This process is coherent with current theories of facial expression understanding (Adolphs, 2006; Carr et al., 2003; Leslie et al., 2004), pointing out that different brain systems collaborate during the reading of facial expressions, where the amount and pattern of activation is different depending on the expression being observed.

Current works that take a neuro/bio inspired approach for the integration of emotions into architectures for artificial intelligence focus on more low level aspects of emotions. The work by Talanov et al. (2015) explore how to produce basic emotions by simulating neuromodulators in the human brain, and applying it to computational environments for decision making. Lowe et al. (2007) explore how a dynamical systems perspective can be combined with an approach

that views emotions as attentional dispositions.

In previous works from our group, a cognitive control architecture for human-robot joint action was developed, that integrates action simulation, goal inference, error detection and complementary action selection (Bicho et al., 2011b,a), based on the neurocognitive mechanisms underlying human joint action (Bekkering et al., 2009). For the design and implementation, a neurodynamics approach based on the theoretical framework of Dynamic Neural Fields (DNFs) was used (Erlhagen and Bicho, 2006, 2014; Schöner, 2008). The robot is able to successfully collaborate with a human partner in joint tasks, but pays attention only to hand actions and to the task itself.

1.2 Socially interactive robots

Fong et al. (2003) describes socially interactive robots as being capable of peer-to-peer human-robot interaction, where social interaction is a key component.

This section will present various types of such robots. Virtual agents, medical therapy, entertainment, robot heads, humanoid robots and android robots

1.2.1 Entertainment

The robots in this category are commercially available products, used as toys, usually with a pet appearance or miniaturized humanoid (See Figure 1.1).

One example is the robot Nao (Beck et al., 2010), which is able to express emotions by using it's body pose and movements (Miskam et al., 2014). It is able to recognize human emotions through some non-verbal cues, such as postures, gestures and movements of the body.

Aibo (Fujita and Kitano, 1998) was developed as a robotic dog that should be able to live with people like a real dog would. The robot presents an autonomous behaviour generated by different "instincts" (e.g. love, search, movement, recharge and sleep). Each instinct drives a different behaviour sequence (Fujita, 2004). Although the robot does not express emotions, they are perceived as such in the form of changes in it's behaviour. The dog can recognise

(a) Nao (Image taken from Miskam et al., 2014).

(b) Aibo (Image taken from www.sony-aibo.com).

(c) Qrio (Image adapted from Tanaka et al., 2004).

Figure 1.1: Examples of entertainment robots.

the emotional state of it's owner from affective cues in the owner's speech and respond with appropriate actions (Jones and Deeming, 2008).

Qrio has an emotionally grounded control architecture (EGO-Architecture) composed by five parts (perception, memory, behaviour control, internal model and motion control). The most interesting part to analyse is the internal model part, where an internal state is defined and composed by a set of variables that describe the robot's needs. The goal of the robot is to maintain its "needs" satisfied, and accomplishes this by selecting specific behaviours, based on time and external stimuli. The satisfaction values are used to determine an emotional state that will influence the behaviour selection (Tanaka et al., 2004).

1.2.2 Robotic heads

Robots which are composed only by a robotic head and neck.

Kismet

Kismet (Breazeal, 2002). Its models of emotion interacted closely with its cognitive system

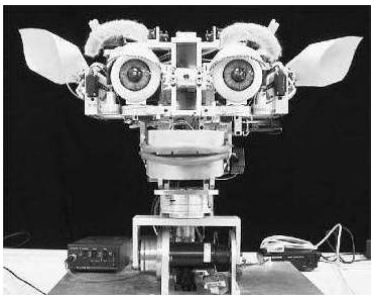


Figure 1.2: Kismet (Image taken from Breazeal, 2002).

to influence behaviour and goal arbitration (Breazeal, 2003b) using a process of behavioural homoeostasis, which drive the robot's affective state to a mildly aroused, and slightly positive state.

One of its purposes was to use emotive communication signals to regulate and negotiate its interactions with people (Breazeal, 1998a,b). It uses emotive displays to regulate the intensity of playful interactions, socially negotiating its interaction with people via its emotive responses to have humans help it achieve its goals, and satiate its drives, and maintain a suitable learning environment.

In order to express its affective states, Kismet is able to generate several facial expressions with corresponding body posture. Kismet's facial expressions are generated using an interpolation-based technique over a three-dimensional space (arousal, valence, stance). The basic facial postures are designed according to the componential model of facial expressions (Smith and Scott, 1997), whereby individual facial features move to convey affective information. In addition to facial expressions, Kismet is able to use a vocalization system to generate a set of emotive utterances corresponding to joy, sorrow, disgust, fear, anger, surprise, and interest (Breazeal, 2003c).

At the perception level, Kismet is able to determine the human affective state through audio and image processing. Extracts low level features from speech such as pitch and energy, and analyses motion and proximity from vision among others to determine a human affective state

(e.g. Neutral, approval, attention).

MEXI

MEXI (Machine with Emotionally eXtended Intelligence) is able to recognize emotions of the human interacting with it, and react in an emotional way by expressing its own emotions using its facial expressions and speech. It has an internal architecture based on emotions and drives, and integrates mechanisms of emotion regulation. MEXI is designed to be emotionally competent, hence it not only recognizes others' emotions, it represents its own emotions internally, reacts to recognized emotions and regulates its own (Esau et al., 2008).

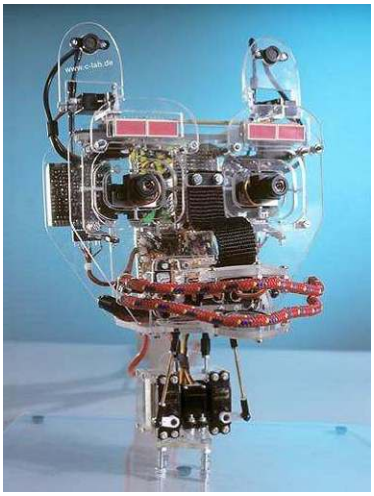


Figure 1.3: MEXI (Image taken from Esau et al., 2006).

The control architecture is comprised of four main blocks, namely: perception, reaction, action control and emotion engine.

At the perception level, inputs from vision and speech are acquired and processed. The vision system is responsible for the detection of objects and faces, and the robot is able to track them using its eyes and head movements. It recognizes speech using a commercially available software (ViaVoice). Emotion recognition is achieved by combining facial expressions detection (Esau et al., 2007) and the prosody of human speech (Austermann et al., 2005).

The reaction level is responsible for direct responses to the sensory stimulus, by generating appropriate head movements, facial expressions and speech. Within the reaction level, the

behaviour system is responsible for generating the robot's movements. The speech output is generated by a chatbot (ALICE[1]), the contents of the answers are influenced by the emotion engine.

The action control performs low level control of motors and speech production.

The internal state represents the strength of its emotions and drives, taking into account the current perceptions, it determines its actions by following two main goals: (a) Feel positive emotions and avoid negative ones; and (b) Maintain its drives at a comfortable level (Esau et al., 2008).

The emotion engine represents four basic emotions: anger, happiness, sadness and fear. Drives are what motivates behaviour (in humans e.g. thirst, hunger), MEXI has a drive to communicate with the human or play with a toy. A dynamic model is used to control emotions and drives, allowing adequate control of the robot's behaviour (Esau and Kleinjohann, 2011).

This way and according to the drive that is active, and the emotion represented, an adequate behaviour is selected in such a way that its two main goals are accomplished.

1.2.3 Humanoid robots

Similar to the Robot Heads, the difference is that the research in this area not only focuses on the interaction but also on bipedal walking or motion control in general. Nevertheless, there is also research in the area of socially interacting robots and in the development of emotion-based control architectures.

FLASH

FLASH (Flexible LIREC Autonomous Social Helper) is a prototype social robot built in the scope of the LIREC (Living with Robots and interactive Companions) project (European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement nº 215554).

It is a mobile-dexterous-social robot with two main purposes, (a) to serve as an integration

[1] http://www.alicebot.org

platform of diverse technologies developed in LIREC; (b) enable experimental verification of these technologies in social environments through HRI experiments.

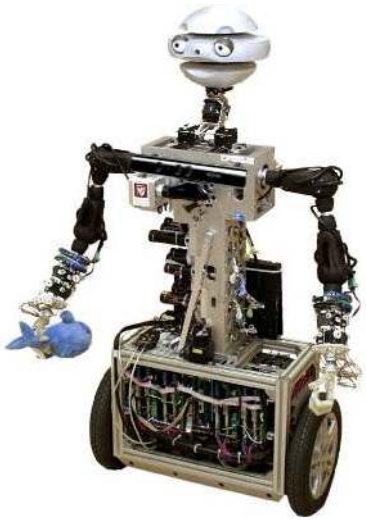


Figure 1.4: FLASH (Image taken from Kędzierski et al. 2013).

FLASH is mounted on a balancing mobile platform, it is equipped with an expressive head (known as EMYS) and two dexterous hands WANDA (Kędzierski and Janiak, 2012; Kędzierski et al., 2013). It is able to communicate with people in verbal and non-verbal ways. Its head serves to express emotions, by using facial expressions. Its hands are intended primarily to perform simple gesticulation tasks, expressing emotions, but they may also be able to execute basic object grasping.

The head EMYS (EMotive headY System) - is a "turtle" type head which can express 7 basic facial expressions representing basic emotions like anger, disgust, fear, joy, sadness and surprise (Kędzierski et al., 2013), by using the Facial Action Coding System (FACS). It is mounted on a movable neck for purposes of objects searching, turning toward a user, and gazing.

Kobian

Kobian is a whole-body humanoid robot with interaction and bipedal walking abilities (See Figure 1.5), this robot is a combination of the bipedal robot Wabian and the upper body humanoid robot WE-4RII (Zecca et al. 2008).

Figure 1.5: Kobian (Image taken from Endo and Takanishi, 2011).

Kobian has a mental model that allows the dynamical change of the emotional state based on external stimuli. The robot's behaviour is affected by it's emotional state, which is displayed with both facial expressions and whole-body patterns.

The robot is able to express six basic emotions (happiness, anger, disgust, fear, sadness and surprise) and a neutral state using facial expressions Trovato et al. (2012), it's whole body (Zecca et al., 2008) and even differentiate two emotions (happy and sad) by it's gait (Destephe et al., 2013).

Although very advanced in it's emotion expression capabilities, Kobian lacks perception abilities, it is unable to detect and recognize human emotions. It's vision system features only an object tracker for interaction.

Barthoc

Barthoc (see Figure 1.6) was created to explore communication with humans.

It is able to recognize objects and grasp them, as well as execute instructions by communicating with a human.

At the perception level, the robot is able to detect human faces by using vision and track them. Additionally it can use voice detection to localize a person. The detection of human emotions is performed by analysing speech from the person interacting with it. During communication

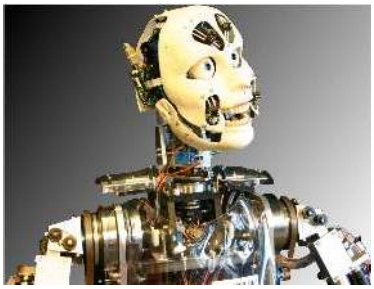


Figure 1.6: Barthoc robot with no skin (Image taken from Spexard et al. 2007).

it is able to display emotions through speech by using its dialogue system and perform facial expressions (Spexard et al., 2007).

1.2.4 Discussion

Today's robots are still far from being social, i.e. from being able to interact with humans in a natural and human-like way (Breazeal, 2003a; Fong et al., 2003). Needless to say that non-verbal communication is an essential component for every day social interaction. For example, we humans continuously monitor the actions and the facial expressions of our partners, interpret them effortlessly in terms of their intention and emotional/mental state and use these predictions to select adequate complementary behaviour (Blair, 2003; Frith and Wolpert, 2004). Thus, two important (high level) cognitive skills which are missing and which would greatly increase the acceptance by human users are the capacity of understanding human actions and the capacity of understanding/interpreting facial expressions.

The work here presented aims to advance towards more fluent and natural human-robot joint action. First one must understand what is joint action, and what it involves, in order for this knowledge to be applied in a human-robot scenario.

Most of the presented robots, built to be social companions that have the ability to recognize and express human-like emotions, are more focused on natural interaction and communication.

Both Kismet and MEXI have control architectures based in emotions and drives where the goal is to interact with the human, in a communicative basis to satiate their drives, by reading and expressing emotions. Even though Kismet has some object manipulation capabilities with

small objects, MEXI has only a head which uses to communicate with the human. Kobian implements also a similar behaviour with a need model, much similar to drives that influence its behaviour towards a goal.

Affective social robots are beginning to emerge in the home consumer market. One proof of this is the recent robot Pepper by Aldebaran Robotics[2], a mobile humanoid robot that has voice recognition technology for human-robot communication, as well as touch sensing and emotion recognition capabilities using facial expressions and body language. Although its behaviour is controlled using proprietary algorithms and not much information is available.

Additionally, other types of social robots exist, such as androids or medical therapy oriented robots.

The main focus of the development of android robots is on the human-like appearance (e.g. Geminoid (Sakamoto et al., 2007), Saya (Hashimoto et al., 2009) and Repliee (Minato et al., 2004)). It is however, an aspect that may prove to be useful in the future of socially interactive robots.

Robots developed for medical therapy are designed to be commercially available products for hospital environments, and the development of a complex and stable architecture that creates a truly socially interactive robot is expensive and difficult to achieve.

Our robot aims to read emotional cues from humans and apply them in a joint cooperation task, adapting its behaviour to the perceived emotion in order to make the interaction during the task execution more fluid. It has the advantage of being an anthropomorphic human sized robot, with the ability to, in the context of the presented joint task, manipulate the same objects that the human does and physically collaborate towards a common goal.

1.3 Automatic analysis of emotions and social signals

The human face is one of the most prominent mean of communicating and understanding somebody's affective state and intentions on the basis of the shown facial expression (Keltner and Ekman, 2000). Given the significant role of the face in our emotional and social lives,

[2] http://www.aldebaran.com

it is not surprising that the potential benefits from efforts to automate the analysis of facial signals are varied and numerous (Pantic and Bartlett, 2007; Zeng et al., 2009). As far as natural interfaces between humans and machines are concerned, facial expressions provide a way to communicate basic information about level of engagement, interest, puzzlement, and other emotions and social signals to the machine (Pantic et al., 2008). Where the user is looking (i.e., gaze tracking) can be effectively used to inform the machine about the user's current focus of attention. Also, combining facial expression detection with facial expression interpretation in terms of labels like "joyful", "curious" and "bored" could inform the machine on the type of the feedback/ change needed and it could be employed as a tool for monitoring human reactions during web-based lectures, automated tutoring sessions, or robot-based sessions.

Because of its practical importance and the theoretical interest of cognitive and medical scientists (Cohn, 2006; Ekman and Friesen, 1978), machine analysis of facial expressions, facial affect and social signals, attracted the interest of many researchers. For exhaustive surveys of the related work, refer to: Pantic and Rothkrantz (2000); Samal and Iyengar (1992) for overviews of early works, Pantic and Bartlett (2007) for a survey of techniques for detecting facial muscle actions, Gunes and Pantic (2010a); Zeng et al. (2009) for surveys of audiovisual (facial and vocal) methods for affect recognition in terms of either discrete emotion categories like happiness, anger, fatigue, etc., or affect dimensions like valence, arousal, expectation, etc., and Vinciarelli et al. (2012) for a survey on social signals processing.

Most facial expressions analysers developed so far, target human facial affect analysis and attempt to recognize a small set of prototypic emotional facial expressions like happiness and anger. Automatic detection of the six basic emotions (happiness, sadness, anger, disgust, fear and surprise) in posed, controlled displays can be done with reasonably high accuracy. Also detecting these facial expressions in the less constrained environments of human-computer interaction has been recently explored (Gunes and Pantic, 2010b; Koelstra et al., 2010; Nicolaou et al., 2011). Whilst the state of the art machine analysis of facial expressions is fairly advanced, it does suffer from a number of limitations that need to be addressed if it is to be used with freely moving subjects in a real-world environment. In particular, published techniques are still unable to handle natural scenarios typified by incomplete information due to occlusions, large

and sudden changes in head pose, and other temporal dynamics occurring in natural facial behaviour (Zeng et al., 2009).

Body movement and posture are also predictors of affective states but they have been largely neglected because of a lack of a commonly-accepted set of emotional descriptors. Yet, they are accurate predictors of human affect (Ambady and Rosenthal, 1992). In fact, perception of emotion has been shown to be often biased toward the emotion expressed by the body when facial and body expressions are incongruent (Meeren et al., 2005), and this perception has been shown to be robust even when cartoon-like characters or point-light displays are used (Pollick et al., 2001). Yet, as in facial studies, most studies have focused on acted basic emotions and stereotypical body movement (e.g. dance, for an overview of the state of the art, see Gunes and Pantic (2010a)). Natural expressions are more subtle than basic and stereotypical expressions, and approaches that rely on acted and often exaggerated behaviours typically fail to generalise to the complexity of expressive behaviour found in real-world settings.

Finally, agreement between humans rating affective behaviour is greater when multiple modalities are combined (Ambady and Rosenthal, 1992), and the dynamics of human behaviour is crucial to their rating (Ekman and Rosenberg, 2005). When it comes to fusion of multi-sensorial signals, past research has shown that this problem needs to be approached as the general classifier fusion problem, where correlations between input data streams (visual, audio, biophysical, etc.) are modelled while the requirement of synchronisation of these streams is relaxed (Zeng et al., 2009). Past research has also indicated that the prediction of the input in one data stream based on the input in other data streams may be a more robust and effective approach to multi-sensorial signal interpretation than is the case with standard multimodal data fusion (Petridis et al., 2010). However, in most of the published studies on multimodal analysis of human affective behaviour, the input from each modality is modelled independently and combined only at the final stage, which implements classifier fusion (i.e. decision-level data fusion). Although some attempts to make use of correlation between multiple data streams have been made, it remains unclear how to model the observed multimodal data on multiple time scales and how to model temporal correlations within and between different modalities.

1.4 User states

In order for human-robot interaction to be as natural and fluent as human-human interaction, it is essential for the robot to perceive and understand the human's current interaction and affective state, as well as the human social signals, which are composed of multiple behavioural cues (Vinciarelli et al., 2009).

A large part of research that aims to recognize the human state, is focused in recognizing emotions (Cowie et al., 2001), usually, the six basic emotions defined by Ekman (1992). The defined emotions, happiness/joy, anger, disgust, surprise, sadness and fear are described as being universal across cultures (Ekman, 1971), which made these emotional states to be the most widely used reference across studies in multiple fields of study such as psychology, human-computer interaction or neuroscience.

Affective computing aims to develop computer interfaces that automatically detect and react to a user expressed emotion. While traditional theory on emotion is often oriented for the discrete basic emotions, these do not occur as often in interaction with a machine (Russell and Barrett, 1999). Scherer et al. (2012) suggest the use of a set of labels defined by performing human perception tests of realistic interactions, and having uninformed participants to report what they perceive. The labels concern the user's attitude towards the interaction or define the current interaction state: (a) interested; (b) uninterested; (c) surprised; (d) embarrassment; (e) impatient; (f) stressed; (g) negative; (h) positive; (i) disagreement.

D'Mello and Calvo (2013) question the use of basic emotions in affective computing, and analysis on five studies is performed. The studies tracked basic and non-basic emotions, after generalizing across tasks, interfaces, and methodologies what the results show is that engagement, boredom, confusion, and frustration occur at five times the rate of basic emotions.

Steininger et al. (2002) define as user states the following labels: (a) joy/gratification (being successful); (b) anger/irritation; (c) helplessness; (d) pondering/reflecting; (e) surprise; (f) neutral. These are applied to the classification of user states in video analysis by trained observers in the *SmartKom* project, which aims to develop an intelligent computer-user interface that allows almost natural communication with an adaptive and self-explanatory machine, by

using natural speech, gestures and facial expressions.

Adelhardt et al. (2003) presented a method based on neural networks for combining facial expressions, gestures and voice into a user state recognition system. It uses three independent networks to classify each of the inputs, face, sound and gestures and associates a label with each one. The fusion of modalities is then performed by training a new network with all inputs to produce a classification, instead of using the networks already created. However, in this process the facial expressions module is discarded because of inconsistencies introduced in the classifier. For instance, the production of sound "Ah!" by a confused user can be classified incorrectly as joy.

More examples include the study by Shi et al. (2003) where three user states were defined to classify the gestures made by a human: (a) determined; (b) negative; (c) hesitant; Or Asteriadis et al. (2008) that defines user states based in eye gaze and head pose: (a) Frustrated/struggling to read; (b) distracted; (c) tired/sleepy; (d) not paying attention; (e) attentive; (f) full of interest.

Although the basic emotions are argued not to be an ideal set of user states for human-machine interaction, they are less susceptible to subjective interpretation, and because they are investigated across different fields (Ortony and Turner, 1990; Russell, 1994) offer a solid support in literature, including neuroscience (Hamann, 2012; Vytal and Hamann, 2010). Hence, the set of emotions comprised by the basic emotions (happyness, surprise, anger, sadness, fear and disgust) offer solid reasons to choose them over other emotional states to be used in this work. The focus is on – free floating – basic emotions that function as rapid appraisals of situations in relation to goals, actions and their consequences (Oatley and Johnson-Laird (1987), for a recent review see Oatley and Johnson-Laird (2014)).

1.5 Objectives and contributions of this thesis

This thesis extends the previous developed cognitive architecture for human-robot joint action (Bicho et al., 2011a) by endowing the robot with the ability to detect and interpret facial expressions, in order to infer the human emotional state. From the integration of reading

motor intentions and reading facial expressions into the robot's control architecture, the robot is endowed with the required high level cognitive skills to be a more intelligent and socially aware partner.

The information acquisition method used to provide the robot with the necessary information is a vision system comprised of several modules, each responsible with the processing of different types of information. A module was created to extract information about objects (type, position, orientation), task state and gesture recognition. Another part of the system is responsible for the analysis and extraction of information from a human face and classify it in accordance with FACS. The developed vision system is also capable of classifying human movement (hands, body and head).

The acquired information was integrated into the existing cognitive control architecture for joint action by using the theoretical framework of DNFs. The architecture capabilities were extended with the ability of understanding emotional facial expressions and human movement.

The results illustrate how the human emotional state influences various aspects of the robot behaviour. It is shown how it influences the decisions that the robot makes, e.g. the same goal directed hand action in the same context but with a different emotional state bias the robot decisions. How the emotional state can have a role in the robot's error handling capabilities, specifically, how the same error is treated in different ways. Also, how the robot can use its emotional expressive capabilities to deal with a human partner persisting in error. And finally, how the human's emotional state can influence the time it takes for the team to complete the joint construction task.

The work developed during this thesis provided numerous contributions to the Human-Robot Interaction field. The main contributions include:

- A humanoid robotic platform for HRI, that was used in studies on Human-Robot Joint action (Bicho et al., 2011a);
- The participation on the European project JAST, considered one of the ICT "success stories" (Bicho et al., 2012), was possible with the vision system developed for this work;
- An approach towards creating socially intelligent robots by endowing them with the ability

to classify emotional states in a human-robot joint construction task (Silva et al., 2016).

1.6 Structure of the thesis

The remainder of the thesis is organized as follows: Chapter 2 will introduce the core biological background used to support this work. How in the scope of joint cooperation, our brain handles action understanding and intention detection, using the mirror neuron system, and also how the same neural mechanism of mirror neurons exist for emotion recognition using facial expressions.

In Chapter 3, the used theoretical framework is presented, this chapter will focus on how Dynamic Neural Fields are used for behaviour representation and decision making, and how this approach can be applied to cognitive robotics.

Chapter 4 contains a description of the developed cognitive architecture that controls the robot ARoS. An overview of the previous work in this robot is presented and in further detail, how the existing cognitive architecture was extended to cope with human emotion recognition and its integration in human-robot joint action.

Chapter 5 describes the robotic setup used in this work, namely the task used for human-robot joint action and a brief description of the robot.

Chapter 6 offers details on how the implementation of several features of the robot's vision system was accomplished. The robot's vision system is able to extract information of the objects it can manipulate, identify hand gestures and recognize emotional facial expressions.

Chapter 7 presents the experimental results accomplished with the robot in a joint construction task, where the influence of the partners emotions in the robot's behaviour is explored.

The thesis concludes in Chapter 8 with a discussion of the presented work and also some developments for future work.

This page was intentionally left blank.

Chapter 2

Background

Human-Robot Interaction (HRI) and joint action can greatly benefit from the recognition of human face expressions, since robots are becoming more part of our daily lives, the ability of understanding natural non-verbal human language will be an inevitable step in robotics research.

One of the most expressive means of communication is our face. Even unconsciously we make and interpret facial expressions seamlessly, and engage in social interaction using this information to better understand and make decisions during the interaction itself.

2.1 The role of emotions as a coordinating smoother in joint action

Humans have a remarkable ability to perform fluent organization of joint action, achieved by anticipating the motor intentions of others (Sebanz et al. 2006). In everyday social interactions, humans continuously monitor the actions of their partners and interpret them effortlessly in terms of their outcomes. These predictions are used to select adequate complementary behaviours, and this can happen without the need for explicit verbal communication.

To achieve a useful and efficient human-robot collaboration, both partners are required to coordinate their actions and decisions in a shared task. The robot must then be endowed with some cognitive capacities such as action understanding, error detection and complementary

action selection.

Classical accounts of joint action involve the notion of shared intentions, which require that the participants have common knowledge of a complex, interconnected structure of intentions and other mental states (see Bratman, 1992, 1993, 1997, 2009; Gilbert, 1990; Tuomela, 2005).

Individual participants' actions within a joint action make sense in light of each other and in light of a shared intention to which they are committed. An example of this is a group playing music, each individual must be aware of the intentions of others so that its own actions make sense.

The classical accounts of joint action are able to explain complex actions involving rational deliberation and planning, and thus presupposes that participants in a joint action are capable of rational deliberation and planning. These accounts are suitable to explain complex interactions but fail to explain more simple cooperative tasks, presenting therefore a limitation. They can't explain joint action performed by non-human animals or young children, since these lack the previously mentioned cognitive skills required (Michael, 2011).

In recent years there have been proposals of minimalist accounts of joint action that either assume no shared intention is necessary or the participants do not require common knowledge of each other's intentions or other mental states, or the relations among the various participants' mental states.

Tollefsen (2005) suggests a minimalist account for joint action where joint attention is sufficient as a substitute for common knowledge of an interconnected structure of intentions. Joint attention involves two individuals attending to the same object or event, and being mutually aware that the other is also attending to the same object or event.

Another model of a minimalist joint action states that the participants do not even require to represent each other as intentional agents. The proposed model by (Vesper et al., 2010) relies in three aspects that play a role in joint action:

- Representations: An agent represents its own task and the goal state, but not necessarily any other agents' task;
- Processes: Prediction (sensory consequences of himself or other agent) and monitoring (identify errors);

- Coordination smoothers: exaggeration of one's movements to make them easier for the other participant to interpret; giving signals, such as nods; and synchronization, which makes partners in a joint action more similar and thus more easily predictable for each other.

Even though there are multiple accounts that attempt to explain human-human joint action, none of the above described accounts addresses the potential role of emotions as coordinating factors in joint actions. Michael (2011) aims to fill this gap by showing how emotions, more specifically shared emotions, can facilitate coordination in joint actions.

Before discussing how shared emotions can be used to facilitate coordination in joint action, it is important to define the necessary conditions for a shared emotion. Supposing we have an interaction between two agents A and B, two necessary conditions must be met in order for a shared emotion to be considered as such:

- A expresses his affective state verbally or otherwise (facial expressions, posture, ...);
- B perceives this affective state.

When these two conditions are satisfied we are in the presence of a shared emotion. Michael (2011) discusses in more detail different varieties of shared emotions: (a) emotion detection; (b) emotion/mood contagion; (c) empathy; (d) rapport.

From the various forms of shared emotions, the one that is most helpful in a human-robot joint action scenario is the emotion detection, it occurs as a result of agent B perceiving agent A's emotional expression, as a consequence, B detects A's emotional state. Emotion detection can be used to facilitate coordination in joint action in three aspects:

- Facilitate prediction;
- Facilitate monitoring;
- Serve as signalling function.

Emotion detection can facilitate prediction in a joint action, lets imagine agent A expresses an emotion in response to a certain action performed by agent B, if agent B detects this emotion and agent A is aware of this detection, then, agent A can predict that the decisions made by agent B will take into account its emotional state.

Regarding monitoring, a person's emotional expressions can transmit information about how

she appraises her progress toward the goal of her own task, or the group's progress toward the global goal of a joint action.

Emotion detection can also serve as signalling function, a positive emotional expression such as a smile may signal approval of another participant's action or proposed action, or the continued presence of rapport within the group.

Emotional mechanisms can contribute to fast adaptation (allowing to have faster or slower reactions), to resolve the choice among multiple conflicting goals, and through their external manifestations, to signal relevant events to others (Cañamero, 2001).

In a context of human-robot joint action, the use of emotions can be beneficial for the team, since the robot can harness more information about its partner (ex.: facial expressions, gesture velocity and body movement), it is expected that the decisions that the robot makes take into account not only the performed actions, but also the state of the partner, resulting in better decisions. From the human perspective, collaborating with a robot that has emotion recognition abilities could contribute to a less rigid interaction, making it more human-like and fluent.

Autonomous agents can benefit from the inclusion of emotions in their architectures as far as adaptation is concerned (Cañamero, 2001). If the robot is able to interpret the user facial expression in terms of its meaning, in the context of a joint task, it can select more appropriate behaviours and actions to improve the interaction with the human.

In the context of this work the facial expressions displayed by the human, might not be actual manifestations of an emotion, if the robot interprets a facial expression as sadness, it does not mean that the human is actually feeling sad, this cannot be measured accurately by the robot's sensing capabilities, the human can be only displaying a communicational signal to work with the robot. However, the interpretation that the robot makes of this expression, in the context of the task, will have consequences in its behaviour and decisions.

In order to integrate emotion recognition into the robot architecture, in a way that is biologically plausible, one must first investigate how humans understand emotions.

2.2 Neuro-cognitive mechanisms underlying the understanding of actions and emotions

Experimental evidence from behavioural and neurophysiological studies that investigate action and perception in social contexts have shown that when we observe others' actions, the corresponding motor representations in our motor system become activated (Rizzolatti and Craighero, 2004; Wilson and Knoblich, 2005).

Although the capacity that humans possess to understand an action could merely involve visual analysis of that action, it has been argued that we actually map this visual information onto our own motor representation in our nervous system (Rizzolatti et al., 2001). There are two hypotheses that attempt to explain how action understanding works:

- The "visual hypothesis" - is based on a visual analysis of the different elements involved in an action, and the motor system is not involved;
- The "direct matching hypothesis" - there is a mapping of the observed goal-directed action onto our own motor representation of the same action.

Rizzolatti et al. (2001) and Ferrari et al. (2003) present experimental evidence that an action observation/execution matching system does exist in primates, supported by the discovery of mirror neurons.

Mirror neurons are a particular class of visuo-motor neurons originally discovered in area F5, a sector of the ventral premotor cortex of monkeys (Gallese et al., 1996; Rizzolatti et al., 1996). Area F5 is characterized by the presence of neurons that code goal-related motor acts such as hand and mouth grasping (Murata et al., 1997; Rizzolatti et al., 1988, 2000).

Some of the cells present in area F5 are motor neurons, others also respond to visual stimuli, some of them are activated by the presentation of three-dimensional objects, while others - mirror neurons - require action observation for their activation (Rizzolatti et al., 2001). The main functional characteristic of mirror neurons is that they become active both when the monkey makes a particular action (for example, when grasping an object or holding it), and when it observes another individual (monkey or human) making a similar action.

A detailed review and discussion regarding the anatomical and functional organization of the pre-motor and parietal areas of monkeys and humans, and also, how the mirror neuron mechanism is involved in understanding the action and intention of others in imitative behavior can be found in Rizzolatti et al. (2014).

Mirror neurons link action observation and execution, this neural mechanism allows for a direct matching between the visual description of an action and its execution. Experiments have shown that action observation is related to activation of cortical areas that are involved in motor control in humans (Ferrari et al., 2003; Rizzolatti et al., 2001).

A goal-directed action is understood when its observation causes the motor system of the observer to "resonate". For instance, when we observe a hand grasping an object to place it somewhere else, the same population of neurons that control the execution of "grasping to place" movements becomes active in the observer's motor areas (Rizzolatti et al., 2001).

It has been suggested that the functional role of the automatic action resonance mechanism contributes to understanding the actions of other individuals during social interactions (Rizzolatti and Craighero, 2004; Wilson and Knoblich, 2005). The observer performs an internal motor simulation to predict the consequences of perceived actions using knowledge of his/her own actions and motor intentions.

The control architecture for human-robot joint action implemented in the robot ARoS (Anthropomorphic Robotic System), presented in Chapter 4, relies on neurocognitive principles to endow the robot with some capacities present in biological nervous systems (Silva et al., 2016). These capacities include memory, decision making, prediction and action understanding, which are essential when a robot is interacting with another agent in the execution of collaborative tasks (Erlhagen and Bicho, 2006).

One of the basis for this architecture is the existence in the human brain of so-called mirror neurons in the pre-motor cortex (Ferrari et al., 2003; Rizzolatti et al., 2001). These authors have reported how these neurons in a monkey brain responded in the same way to performing an action and seeing another individual perform the same action (in this case, grasp an object). Mirror neurons act as a link between action observation and action execution, which may explain our ability to understand the motor behaviour of others.

The robot's architecture allows it to understand the actions of its partner by internally simulating those actions, and because the robot knows how to execute them in a motor level, it has the ability to understand those actions (Fogassi et al., 2005; Gallese et al., 2004; Rizzolatti et al., 2001). This is plausible with the mirror neurons theory that correlates action execution and action observation.

In the same way mirror neurons allow humans to use motor primitives not only to execute actions in a task, but also understand those actions when they are performed by another individual, exhibiting motor resonance (Newman-Norlund et al., 2007b; Rizzolatti et al., 1996), there is also evidence the same happens with facial expressions associated with emotional states.

Imaging studies have shown that some areas in the brain that exhibit activity while observing facial expressions associated to emotions, are also active during imitation of the same facial expressions. This shows that there are common areas that are active during the observation and execution of facial expressions associated to emotions. In particular, results suggest that the right hemisphere premotor cortex may play a role in both the generation and the perception of emotionally expressive faces, consistent with a motor theory of empathy (Leslie et al., 2004).

There is evidence that patients with *Möbius Syndrome*, who are congenitally incapable of moving their facial muscles, seem to have difficulties in appreciating emotions conveyed by the faces of others (Cole, 1999, 2001). This points to the existence of an emotional resonance system in the brain associated with facial expressions (Ochsner and Gross, 2005), which in these patients is impaired.

Ekman (1971) identified seven emotions that are recognized universally across cultures. Studies that investigate the brain responses to emotional facial expressions use as stimulus, images of emotions or other stimulus that lead to the emotions considered universal. It was shown that each of the universal emotions is associated with a different brain activation pattern (see Calder and Young, 2005; Carr et al., 2003; Kesler/West et al., 2001; Leslie et al., 2004; Neta and Whalen, 2011; Sato et al., 2004; Schulte-Rüther et al., 2007; van der Gaag et al., 2007).

However, the representation of emotions is not consensual, there are different approaches that attempt to understand the nature of the basic units of emotion and whether these units

are essentially dimensional or discrete (Russell, 2009).

Discrete emotion theories defend that there exists a limited number of distinct emotions each with specific characteristic properties, as opposed to a continuum of emotional states (Barrett et al., 2007). One view of discrete emotion theory, proposes a limited set of basic emotions (happiness, sadness, anger, fear, disgust, and surprise (Ekman, 1971)) that are universal (across cultures) and have unique physiological and neural profiles that distinguish them from one another (Ekman, 1992; Ekman and Cordaro, 2011; Panksepp, 2007).

The dimensional theory, conceptualizes emotions as arising from combinations of more fundamental dimensions, such as emotional arousal (strength or intensity of emotion) and emotional valence (degree of pleasantness or unpleasantness), in combination with cognitive processes, such as appraisal and attribution of meaning (Barrett, 1998; Gable and Harmon-Jones, 2010).

Dimensional theories of emotion rely in the mapping of emotions in two dimensions: valence and arousal, the measure of these two dimensions is carried out measuring physiological signals using more or less invasive methods. For example, facial electromyography can be used to evaluate valence (Hazlett, 2006). Moreover, research in psychophysiology provides evidence that affective arousal has a range of somatic and physiological correlates such as heart rate, skin moistness, temperature, pupil diameter or respiration velocity (Cacioppo et al., 2000). Facial analysis using vision was used and correlated with facial activity measured using Electromyography (EMG) sensors to assess the user difficulty in completing a task (Branco et al., 2005, 2006).

It is however possible to represent the basic emotions in a valence-arousal space used by dimensional models.

As depicted in Figure 2.1, instances of basic emotions can be represented in a dimensional framework, for example, happiness induced by a beautiful sunset, in terms of variation along affective dimensions (arousal and valence).

Although less intrusive sensors and wearable computers offer possibilities for less invasive physiological sensing (Mann, 1997), the approach followed in this work was a completely non intrusive approach (by using vision), and the discrete approach to emotions favours this possibility.

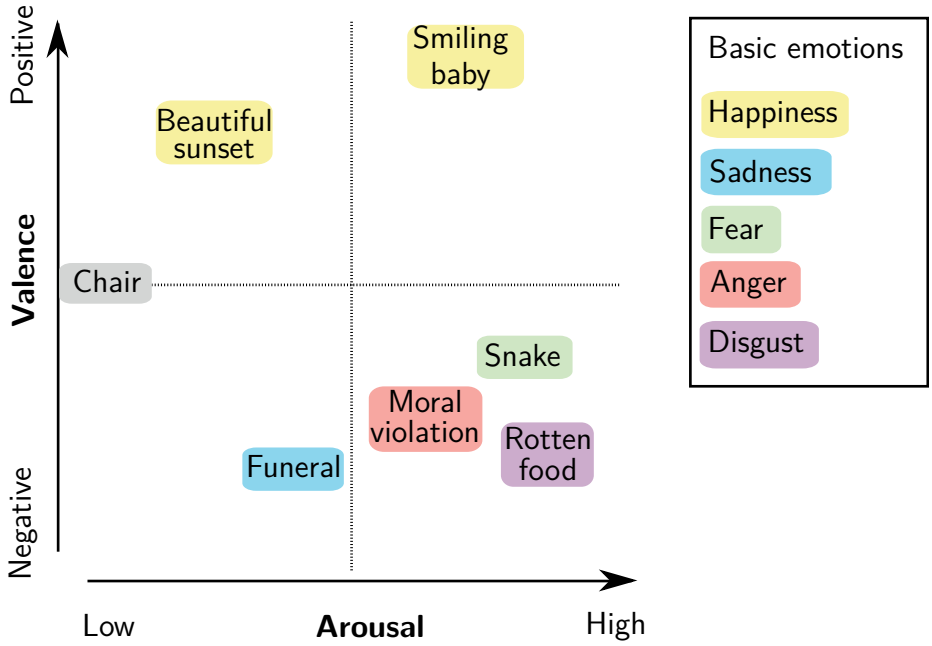


Figure 2.1: Representation of basic emotions within a dimensional framework (Image adapted from Hamann (2012)).

Typically, in neuroscience related works that study emotions, the discrete universal emotions suggested by Ekman (1971) are used, although, not all the emotions are used in a study, only a part of them, to control complexity. Usually, brain responses to emotional face expressions are compared to the brain responses of neutral facial expressions. Kesler/West et al. (2001) presented a comparison of brain activations between images of anger, fear, sadness, happiness, neutral and scrambled images and shows that there is a brain activation pattern associated with neutral faces and faces associated to emotions.

Figure 2.2 shows the brain areas responsible for each of the tested emotional states. Each coloured area in the image depicts the comparison between the emotional state tested and the neutral response, except the neutral areas, where these are the comparison between the brain response to a neutral face and the brain response to a scrambled image. The areas marked in green are associated with face processing.

Kesler/West et al. (2001) have shown that the activation of the amygdala during neutral face processing suggests that it might be responsible for the presence of faces per se. This is

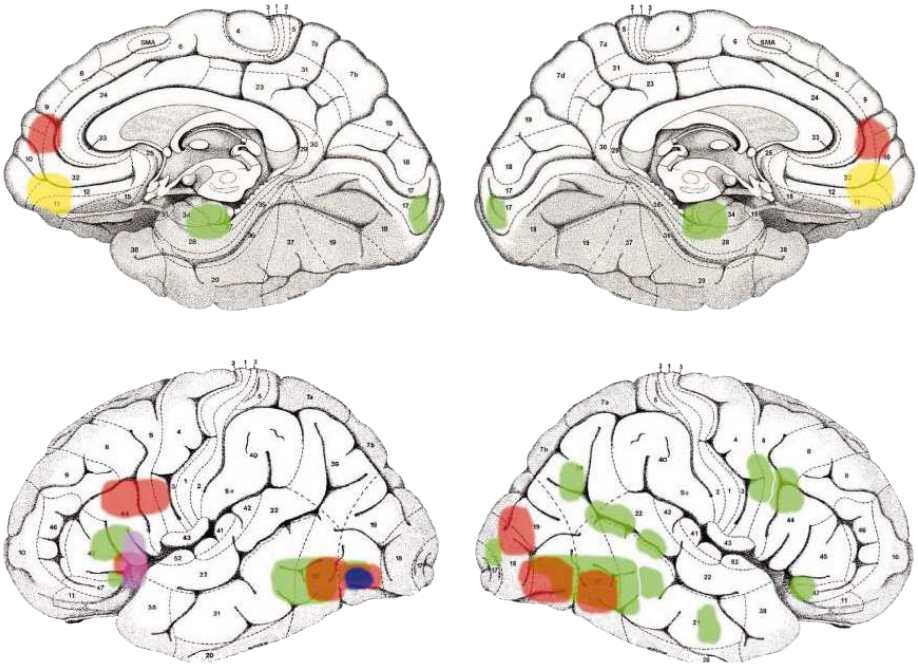


Figure 2.2: Brain areas associated with different emotional facial expressions. Green - neutral versus scrambled; red - angry versus neutral; purple - fear versus neutral; yellow - happy versus neutral; blue - sadness versus neutral (Kesler/West et al. 2001).

consistent with single cell recording studies in monkeys that have shown face-selective neurons in the amygdala (Leonard et al., 1985; Rolls, 1984).

van der Gaag et al. (2007) presents a more in-depth study on the role of mirror neurons in the perception and production of emotional and neutral facial expressions. There is a differential processing of neutral and emotional facial expressions in the brain, and shows that the brain discriminates between neutral and emotional facial movements within specific brain regions in a congruent manner. The understanding of other people from facial expressions is a combined effort of simulation processes within different systems, where the somatosensory, motor and limbic systems all play an important role (van der Gaag et al., 2007).

The simulation processes in these individual systems have been previously described in literature (Gallese et al., 1996; Keysers et al., 2004; Wicker et al., 2003). Specifically, and at a neuronal level, premotor mirror neurons might resonate the facial movement and its implied emotion (Carr et al., 2003; Iacoboni et al., 2005), insula mirror neurons might process the emotional content (Wicker et al., 2003), and somatosensory neurons might resonate proprioceptive information contained in the observed facial movement (Keysers et al., 2004). This process is coherent with current theories of facial expression understanding (Adolphs, 2006; Carr et al., 2003; Leslie et al., 2004), pointing out that different brain systems collaborate during the reading of facial expressions, where the amount and pattern of activation is different depending on the expression being observed (more details in van der Gaag et al., 2007).

Studies in social psychology show that the perception of another person's behaviour increases the likelihood of engaging in that behaviour (Bargh et al., 1996). People have an unconscious tendency to mimic the postures, behaviours, and facial expressions of the person with whom they are interacting. There is good evidence, provided by neuroscience through measures of electromyographic (EMG) activity, supporting that people imitate facial expressions of others seamlessly and unconsciously (Dimberg and Thunberg, 1998; Dimberg et al., 2000).

The work by Leslie et al. (2004) shows results that are consistent with the existence of a face mirroring system located in the right hemisphere (RH) part of the brain, which is also associated with emotional understanding (Ochsner and Gross, 2005). Specifically, the right hemisphere premotor cortex may play a role in both the generation and the perception of emotionally

expressive faces, consistent with a motor theory of empathy (Leslie et al., 2004).

van der Gaag et al. (2007) presents a more in-depth study on the role of mirror neurons in the perception and production of emotional and neutral facial expressions. There is a differential processing of neutral and emotional facial expressions in the brain, and shows that the brain discriminates between neutral and emotional facial movements within specific brain regions in a congruent manner. The understanding of other people from facial expressions is a combined effort of simulation processes within different systems, where the somatosensory, motor and limbic systems all play an important role (van der Gaag et al., 2007).

The simulation processes in these individual systems have been previously described in literature (Gallese et al., 1996; Keysers et al., 2004; Wicker et al., 2003). Specifically, and at a neuronal level, premotor mirror neurons might resonate the facial movement and its implied intention (Carr et al., 2003; Iacoboni et al., 2005), insula mirror neurons might process the emotional content (Wicker et al., 2003), and somatosensory neurons might resonate proprioceptive information contained in the observed facial movement (Keysers et al., 2004). This process is coherent with current theories of facial expression understanding (Adolphs, 2006; Carr et al., 2003; Leslie et al., 2004), pointing out that different brain systems collaborate during the reading of facial expressions, where the amount and pattern of activation is different depending on the expression being observed (more details in van der Gaag et al. (2007)).

Since the current robot's architecture is biologically inspired, having in its foundations the mirror system, the face categorization system implemented in the robot is also based in mirror neurons, implemented using as a design tool the dynamical neural field theory, making it a coherent choice with the existence of a face mirroring system, in the context of the developed previous work.

Chapter 3

Theoretical framework of Dynamic Neural Fields

Dynamic Neural Fields (DNFs) provide a theoretical framework to endow artificial agents with cognitive capacities like memory, decision making or prediction (Erlhagen and Bicho, 2006; Schöner, 2008). DNFs are based on dynamic representations that are consistent with fundamental principles of cortical information processing, implementing the idea that task-relevant information about action goals, action primitives or context is encoded by means of activation patterns of local populations of neurons.

DNFs can be used to build multi-layered models where each layer is formalized by one or more DNFs. The basic units present in these models are local neural populations with strong recurrent interactions that cause non-trivial dynamic behaviour of the population activity. One important property that can be observed, is that population activity initiated by time-dependent external signals may become self-sustained in the absence of any external input. This property of the population dynamics behaves like an attractor state and is thought to be essential for organizing goal-directed behaviour in complex dynamic situations, they allow the nervous system to compensate for temporally missing sensory information or to anticipate future environmental inputs.

In Chapter 4, a DNF based architecture for joint action will be presented, which is built as

a complex dynamic system in which activation patterns of neural populations in the various layers can appear and disappear continuously in time as a consequence of input from connected populations and external sources to the network (e.g., vision, speech) and as defined by a field dynamics.

3.1 Behaviour representation

The problem of knowledge representation in robotic systems can be addressed using two distinct approaches. One approach follows a purely behaviouristic method, where the need for knowledge representations can be discarded (Brooks, 1991). This approach can be applied to a mobile robot by creating a simple set of rules that map input from sensory information to output as motor actuation. Bicho et al. (2000) demonstrated how this approach can be used, to endow with complex behaviours a robot with low computational power.

The other approach requires the use of some kind of knowledge representations to implement processes that are part of more complex behaviours Bicho et al. (2010, 2011a), for example memory or decision making. Memory is important for a system to be able to cope with the absence of sensory information, while decision making becomes important in cases where the sensory input provides ambiguous information.

Although, the simplicity and robustness inherent to the behaviour-based approach, makes it appealing to robotics applications. In order to merge the advantages of the two approaches, Engels and Schöner (1995) proposed a theoretical framework for creating control architectures that takes advantage of memory and representations, grounded on principles underlying the behaviour-based approach. The proposed framework uses artificial neuronal representations where the behaviour of each of the neurons is controlled using a dynamical system.

3.2 The dynamic approach to cognitive robotics

The DNF approach is based in the work of Amari (1977), which presented an equation to model cortical activations in neuronal tissue with lateral inhibition. Later, several works used

this approach to model characteristics from human cognitive behaviour (Erlhagen and Schöner, 2002; Erlhagen et al., 1999; Spencer and Schöner, 2003).

Schöner et al. (1995), proposed an approach to plan and control the movement of an autonomous mobile robot, by representing the navigational space around the robot with a dynamic neural field. This proposal provided a base for a large number of applications in the field of mobile robotics (Bicho, 2000; Bicho and Schöner, 1997; Bicho et al., 1998, 2000; Machado et al., 2013; Monteiro and Bicho, 2002, 2008; Monteiro et al., 2004; Soares et al., 2007; Sousa et al., 2012).

A dynamical neural field approach was used to synthesize higher level cognitive behaviours, such as, action understanding and imitation (Erlhagen and Bicho, 2006; Erlhagen et al., 2006a,b).

Also, Erlhagen and Bicho (2006) used a dynamical neural field approach to synthesize higher level cognitive behaviours, such as, action understanding and imitation.

Based on evidence on the anatomical fact that the largest contribution to cortical cells comes from neighbouring excitatory cells, it has been suggested that the basic mechanism for cortical information processing is in the form of recurrent interactions in populations of neurons (Douglas et al., 1995). The recurrent interactions allow the system to amplify input values, while at the same time exhibits some immunity to noise. It is also possible that internal self-stabilized states can emerge, allowing the compensation of temporary absence of external stimuli.

The equation presented by Amari (1977) offers a way to model neuronal activation, and presents characteristics like memory, where a self-stabilized peak is able to be maintained without the need for external input.

Amari (1977) proposed that when a sufficiently large number of neurons interact with each other, an approximately homogeneous network along the cortical surface is formed, and can be approximated to a continuous field of neuronal activation.

$$\tau \frac{\delta\, u(x,t)}{\delta\, t} = -u(x,t) + S(x,t) + h + \int w(x-x') f\left[u(x',t)\right] dx' \tag{3.1}$$

Equation 3.1 describes the activity $u(x,t)$ of a neuron at location x and time instant t (for mathematical details see Erlhagen and Bicho, 2014).

For simplicity, Equation 3.1 will be broken into parts, the first part is defined by:

$$\tau \frac{\delta\, u(x,t)}{\delta\, t} = -u(x,t) + S(x,t) \tag{3.2}$$

Equation 3.2 represents a dynamic linear system that relaxes at the input level, and acts as a low-pass filter, $u(x,t) = S(x,t)$. The parameter $\tau > 0$ defines the time scale.

$$\tau \frac{\delta\, u(x,t)}{\delta\, t} = -u(x,t) + S(x,t) + h \tag{3.3}$$

In Equation 3.2 the term h is added, which is the resting level of the field dynamics, shown in Equation 3.3. The resting level determines the global rate of inhibition and excitation. If $h < 0$ the system becomes immune to inputs with lower amplitudes, for example, noise induced in the system, and requires that $S(x,t) > h$, in order for an input to be reflected in the field output. This way, the system relaxes towards $u(x,t) = S(x,t) + h$.

Equation 3.3 is however, unable to generate interactions between neurons.

$$\int w(x - x') f\left[u(x',t)\right] dx' \tag{3.4}$$

These interactions are what makes possible to select one activation peak when several inputs are present (decision making), or to maintain an activation peak when the input removed (memory). The integral term shown in Equation 3.4 achieves this, allowing to describe the process of interaction as a weighted sum of the neighbouring neurons activations in the same layer.

Each neuron within a layer can either inhibit or excite other neurons, depending on their distance $(x - x')$. Typically, a neuron excites its closest neighbours and inhibits those that are further away.

The intra-field interactions are implemented by using a kernel of lateral inhibition type described by Equation 3.5:

$$w(\Delta x) = A \exp\left(\frac{-\Delta x^2}{2\sigma^2}\right) - w_{\text{inhib}} \tag{3.5}$$

where $A > 0$ describes the amplitude and $\sigma > 0$ the standard deviation of the Gaussian. The inhibition ($w_{\text{inhib}} > 0$) is assumed to be constant, only sufficiently activated neurons contribute to interaction.

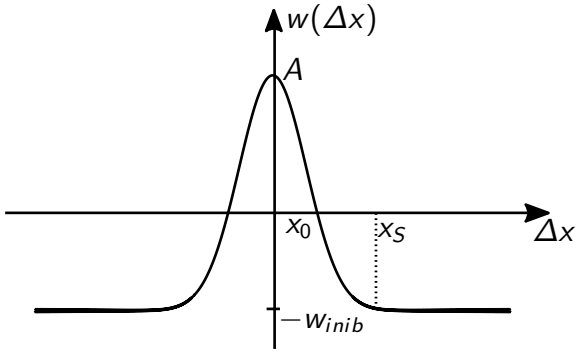


Figure 3.1: Symmetric synaptic weight function $w(\Delta x)$, $\Delta x = x - x'$, of center-surround type. The synaptic weights are positive ('excitatory') for two cells x and x' that are closer to each other than the distance x_0 and are negative ('inhibitory') for larger distances. For $\Delta x > x_S$, inhibition strength is constant ($-w_{\text{inhib}}$).

The threshold $f(u)$ is a sigmoidal function with slope parameter β and threshold u_0, described in Equation 3.6:

$$f(u) = \frac{1}{1 + \exp[-\beta(u - u_0)]} \tag{3.6}$$

Function $f(u)$ gives the firing rate of a neuron with activation value of u and it is responsible for preventing neurons with activation bellow threshold u_0 interfere in the field, by limiting activation and preventing field destabilization.

The model parameters are adjusted to ensure that the field dynamics is bi-stable (Amari 1977), allowing the attractor state of a self-stabilized activation pattern to coexist with a stable homogeneous activation distribution, that represents the absence of specific information (resting level - h). When the input ($S(x, t)$), to a local population is sufficiently strong, the homogeneous state loses stability and a localized pattern in the dynamic field evolves, however, weaker external signals lead to a sub-threshold, input-driven activation pattern in which the contribution of the interactions is negligible.

To represent and memorize simultaneously the location of several objects, and multiple common subgoals, the spatial ranges of the lateral interactions in the field are adapted to avoid a direct competition between different populations, enabling this field to support a multi-peak solution. The updating of the memorized information is performed by defining a proper dynamics

for the inhibition parameter, h, of the population dynamics (Bicho et al., 2000).

The summed input from connected fields u_l is given as $S_i(x,t) = k\sum_l S_l(x,t)$. The parameter k scales the total input to a certain population relative to the threshold for triggering a self-sustained pattern. This guarantees that the inter-field couplings are weak compared to the recurrent interactions that dominate the field dynamics (for details see Erlhagen and Bicho (2006)). The scaling also ensures that missing or delayed input from one or more connected populations will lead to a subthreshold activity distribution only. The input from each connected field u_l is modelled by a Gaussian function described in Equation 3.7:

$$S_l(x,t) = \sum_m \sum_j a_{mj} c_l(t) exp\left(\frac{-(x-x_m)^2}{2\sigma^2}\right) \tag{3.7}$$

where $c_l(t)$ is a function that signals the presence or absence of a self-stabilized activation peak in u_l, and a_{mj} is the inter-field synaptic connection between subpopulation j in u_l to subpopulation m in u_i. Inputs from external sources (e.g. vision) are also modelled as Gaussians.

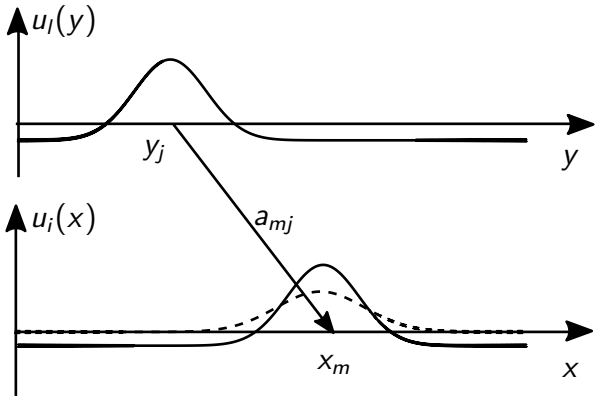


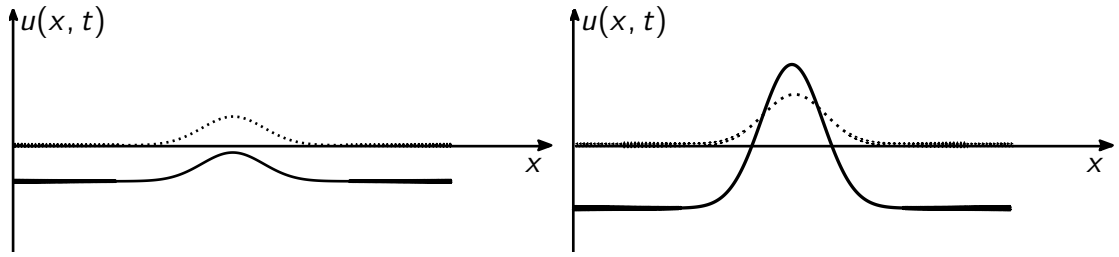
Figure 3.2: Schematic view of the input from a population j in layer u_l that appears to be activated beyond threshold level to a target population m in u_i. The solid line in each graphic represents the field activation and in the bottom graphic in a dashed line the input from connected field u_l.

As an example, Figure 3.2 shows the input from a connected population u_j in layer l connected to a target population m in layer u_i, modelled by a gaussian function. This input is applied whenever the activation in population j is above the threshold for a self-stabilized activation peak.

The Amari neural field model has a set of properties that allows to implement several behaviours useful in intelligent agents, such as memory, forgetting and decision making (Bicho, 2000; Bicho et al., 2000).

Detection

This property makes it possible for the output field to produce only one decision (stable peak), when the input is strong enough, presenting this way an immunity to noise at the entrance. The noise immunity is achieved by using the resting level parameter tuned to $h < 0$, this way, the interaction will produce results, only when the input is greater than the resting level ($S(x,t) > h$) and is present during a sufficient amount of time. When these conditions are met, the activated neurons will excite the neighbours producing a sustained peak and inhibit the rest of the field.



(a) No detection. (b) Detection.

Figure 3.3: Example of an input to a field with and without detection. In 3.3a the input (dashed line) is not strong enough to form an activation, while in 3.3b the input caused an activation peak.

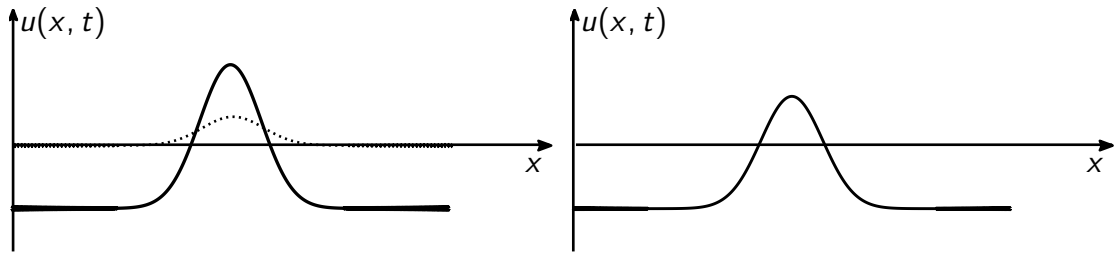
Figure 3.3 shows two examples of inputs and their consequence on the field. Figure 3.3a shows an input $S(x,t)$ (dashed line) that is not strong enough to form an activation peak, hence, the field (solid line) converges to $u(x,t) = h + S(x,t)$. In Figure 3.3b the presented input $S(x,t)$ is sufficiently strong for the field activity $u(x,t)$ (solid line) to go above the zero threshold and form a self-sustained activation peak.

Memory

Another important cognitive capacity is memory. Transposing this to dynamic neural fields, memory would be equivalent to a field to maintain an activation peak even after the input that originated is removed. This capacity is achieved by tuning the kernel parameters (w) and the resting level (See Erlhagen and Bicho, 2006, for a more in depth analysis).

The critical value for the resting level is given by:

$$-h < W_{\max} = \max_{x} \int w(\Delta x)dx \tag{3.8}$$



(a) Input present. (b) Input absent.

Figure 3.4: Example of a self-sustained activation peak that persists even when the input is removed.

The implementation of a working memory function is achieved through the existence of self-stabilized activation patterns. Figure 3.4 depicts an example of memory implemented in dynamic neural fields. Figure 3.4a shows an input (dashed line) triggering a self-sustained peak in the field activity (solid line), next Figure 3.4b shows the activation (solid line) being maintained active even when the input was removed.

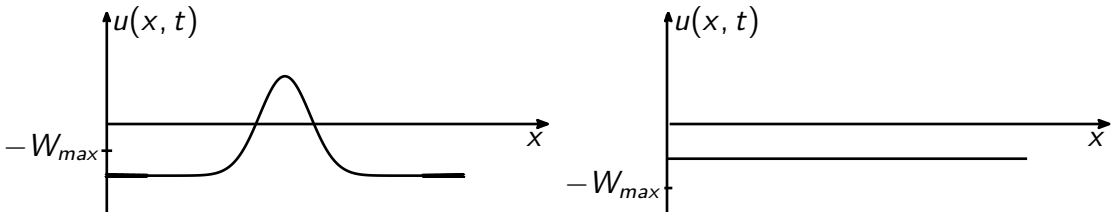
Forgetting

Forgetting is as important as memory. In case of occlusion of objects, its important to retain a representation of it even if the input is not there. However, this representation should not remain indefinitely, since the locations of objects in the world is dynamic, its important to

have the ability to forget, after some time, in order not to take the risk of relying in incorrect information. If an object is occluded and after the occlusion the object is no longer in the same location, the robot must forget its former representation, and be ready to create an updated representation of objects present in the environment.

The critical value for the resting level is given by:

$$h < -W_{\max} = -\max_{x} \int w(\Delta x)dx \tag{3.9}$$



(a) (b)

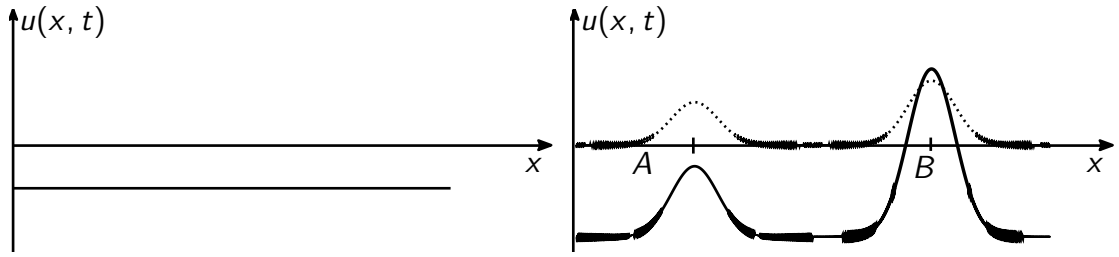
Figure 3.5: Example of a self-sustained activation peak that persists in the absence of input but tends to disappear after a certain time.

Figure 3.5 depicts an example of forgetting. Figure 3.5a shows a self-sustained peak in the field activity (solid line), but as time goes by, the activation ceases to exist (Figure 3.5b).

Decision making

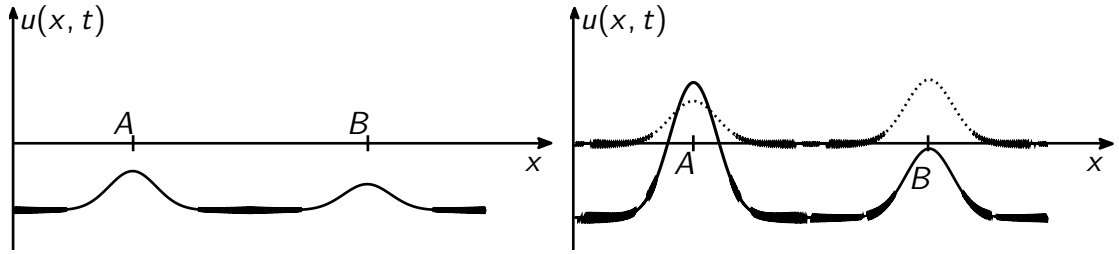
The process of decision making is important to handle multiple activation peaks at the field input, when at the output only one peak is expected. The Amari field model implements this process in the equation kernel. The gaussian kernel enables activated neurons to excite neighbours and inhibit the rest of the field, which produces a competition process between the multiple inputs where the strongest prevail over the weakest (see Figure 3.6).

However, if there are pre-activations at the input, these will also affect the decision making process. Additionally, inputs at locations with pre-existing activations will have greater chances of winning the competition process even if they are weaker than others (see Figure 3.7).



(a) Initial condition. (b) Decision.

Figure 3.6: Decision making with two inputs of different strengths without any preshape in the initial conditions.



(a) Initial condition. (b) Decision.

Figure 3.7: Decision making with two inputs of different strengths with a preshape in the initial conditions that favour position A.

The existence of a single, self-stabilized pattern of activation in a dynamic field is closely linked to decision making.

This page was intentionally left blank.

Chapter 4

Emotion aware robot control architecture for human-robot joint action

The developed architecture, at its core, is heavily inspired by experimental and theoretical findings about the neurocognitive mechanisms underlying joint action in humans (Bekkering et al., 2009; Sebanz et al., 2006). It endows the robot with some cognitive and social capacities present in biological nervous systems like memory, decision making, prediction, action understanding and goal inference (Erlhagen and Bicho, 2006). Previous work, on the cognitive control architecture for human-robot joint action (Bicho et al., 2011b,a), served as a basis for this work.

4.1 Cognitive architecture for human-robot joint action

The architecture implements a flexible action planning and decision formation in cooperative human-robot interactions, that take into account the inferred goal of the partner and other task constraints. This is supported by experimental evidence on the notion that a close perception-action linkage provides a basic mechanism for real-time social interactions (Newman-Norlund et al., 2007b; Wilson and Knoblich, 2005). Action observation leads to an automatic activation

of motor representations that are associated with the execution of the observed action, this resonance of motor structures supports an action understanding capacity (Blakemore and Decety, 2001; Fogassi et al., 2005; Rizzolatti et al., 2001). However, neuroimaging and behavioural studies provide evidence that goal and context representations may link an observed action to a different but functionally related motor response (Newman-Norlund et al., 2007a; van Schie et al., 2008), demonstrating that the mapping between action observation and action execution is much more flexible than previously thought.

A neurodynamics approach based on the theoretical framework of Dynamic Neural Fields (DNFs) (Erlhagen and Bicho, 2006, 2014; Schöner, 2008) was used for design and implementation.

Bicho et al. (2010, 2011a) provides more details about the functional role of the different layers and discusses the flow of information between layers with respect to experimental findings that have inspired this work. Each layer of the architecture is composed of one or more than one dynamical neural field with multiple neural populations that represent relevant information.

The architecture implements a context dependent mapping between an observed action and an executed action (Erlhagen et al., 2006a; Poljac et al., 2009; van Schie et al., 2008). The mapping occurs at the level of abstract motor primitives defined as whole object-directed motor acts like reaching, grasping, placing, attaching or plugging, which encode the motor act in terms of an observable end state or goal rather than in terms of a detailed description of the movement kinematics (Rizzolatti and Craighero, 2004; Schaal, 1999).

Experimental results in several human-robot scenarios (Bicho et al., 2011b,a, 2012) have demonstrated that the robot is able to participate in cooperation tasks with a human. However the interaction was done only through the detected hand goal directed actions. The robot only takes into account the human's hands regarding the motor intentions, this poses a limitation to the interaction.

To achieve a more natural and fluent interaction, the robot must be able to harness more information about his partner. The natural evolution for the architecture developed for action understanding in joint tasks is to take into account the human's emotional state. Because it may convey important information about the underlying intention, and because it must also

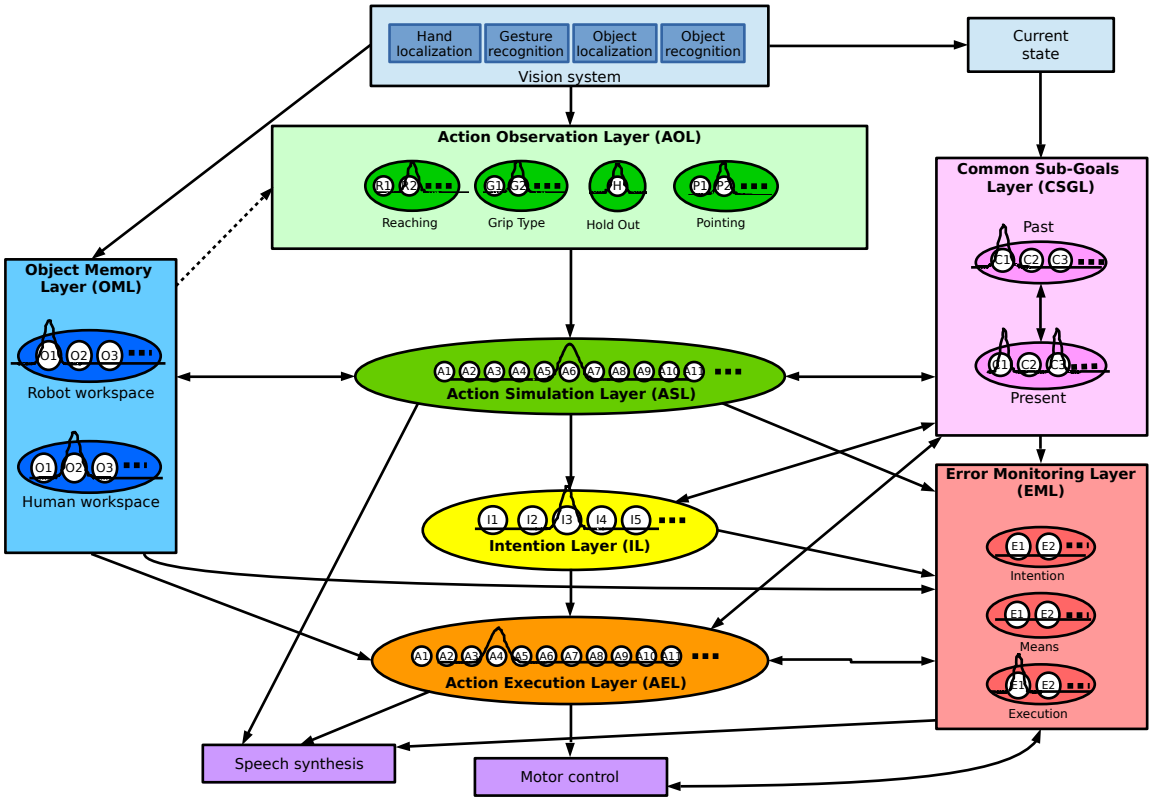


Figure 4.1: Schematic view of the cognitive architecture for joint action. It implements a flexible mapping from observed hand and facial actions (AOL) onto complementary actions taking into account the inferred goal (IL), detected errors (EML), contextual cues (OML) and shared task knowledge (CSGL). The goal inference capacity is based on motor simulation (ASL).

interfere in the selection of a more appropriate complementary action.

Figure 4.1 presents a sketch of the multi-layered robot control architecture, it reflects neurocognitive mechanisms that are believed to support human joint action (Bekkering et al., 2009) and emotional facial expressions (Carr et al., 2003; Iacoboni et al., 2005; Wicker et al., 2003). Each layer contains several neural populations encoding information relevant for the joint assembly task presented in Chapter 5. Every population can receive input from multiple connected populations that may be located in different layers.

Ultimately, the architecture implements a context-dependent mapping between observed action and executed action (Erlhagen et al., 2006a; Poljac et al., 2009; van Schie et al., 2008).

The fundamental idea is that the mapping takes place on the level of abstract motor primitives defined as whole object-directed motor acts like reaching, grasping, placing, attaching or plugging. These primitives encode the motor act in terms of an observable end state or goal rather than in terms of a detailed description of the movement kinematics (Rizzolatti and Craighero, 2004; Schaal, 1999). Also, there is evidence of premotor mirror neurons that might resonate the facial movement and its implied intention (Carr et al., 2003; Iacoboni et al., 2005).

4.2 Combining intention and emotional state inference in a control architecture for human-robot joint action

Figure 4.2 presents a sketch of the multi-layered robot control architecture, modified to cope with the extra information extracted from the human that will be incorporated into the robot's decision making mechanisms.

The cognitive architecture used in this work has its core in the work presented by Bicho et al. (2011b,a), where only hand actions have been considered. In the work here reported, additional layers have been added to reflect the extra information - e.g. observed facial actions - used by the robot in its distributed decision making process. That is, the inferred partner's emotional state, inferred goal and selection of an adequate complementary behaviour. The latter includes selection of an appropriate goal-directed hand-action and facial action sets to be performed and

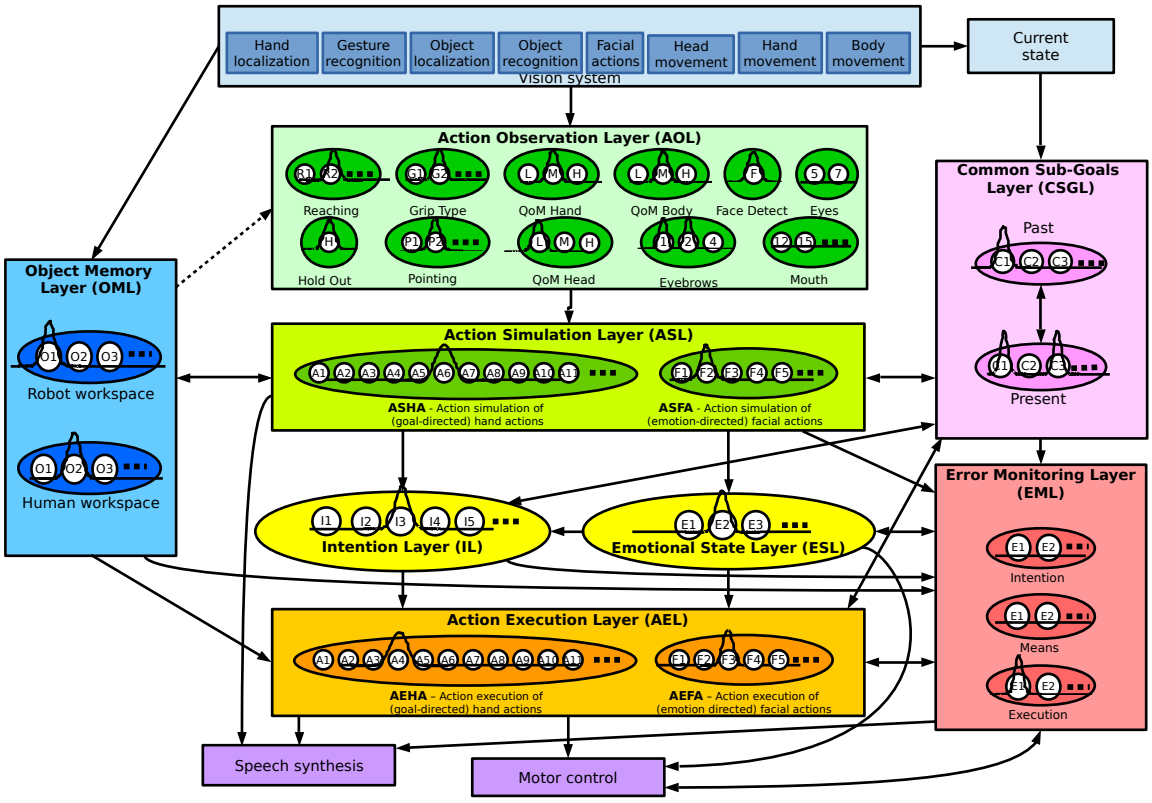


Figure 4.2: Schematic view of the emotion aware cognitive architecture for joint action. It implements a flexible mapping from observed hand and facial actions (AOL) onto complementary actions and emotional expressive faces (AEL) taking into account the inferred goal (IL), the inferred emotional state of the partner (ESL), detected errors (EML), contextual cues (OML) and shared task knowledge (CSGL). The goal and emotional state inference capacity is based on motor simulation (ASL).

displayed by the robot.

Bicho et al. (2010) provides details on the schema of the hand-coded connections for the concrete assembly task, offers an overview about the functional role of the different layers and discusses the flow of information between layers with respect to experimental findings that have inspired this work.

An observed hand movement that is recognized by the vision system as a particular primitive (e.g., top grip or side grip) is represented in the Action Observation Layer (AOL). This layer incorporates also neural populations that code facial actions identified by the vision system (eyes, eyebrows and mouth), as well as a qualitative quantification of the movement of the hand, head and body.

Additionally, there is a DNF with a single neuronal population that represents the presence of a face in the vision system, this population exhibits activation whenever a human face is detected. The presence of a face is encoded in this manner and represents face-selective neurons that exist in the human brain that were shown to exhibit activation when a face is present (Kesler/West et al. 2001; Leonard et al. 1985; Rolls, 1984), even in a neutral emotional state, exhibiting no emotion at all. Moreover, this encoding will prove useful to subsequent layers of the architecture to handle vision errors, since, if Action Units (AUs) are being detected without the presence of a face, this information will be discarded, and allows us to be aware that a human is present even when not expressing any emotion at all.

The Action Simulation Layer (ASL) implements the idea that by automatically matching the co-actor's hand and facial actions onto its own sensorimotor representations without executing them, the robot may simulate the ongoing action and facial expression and their consequences. ASL consists of two DNFs layers. One DNF with neural populations representing entire chains of hand action primitives that are in the motor repertoire of the robot (e.g., reaching-grasping-placing or reaching-grasping-holding out) - named Action Simulation of Hand Actions (ASHA) layer. The other DNF with neural populations representing facial action sets (e.g., lift eyebrows - open mouth - express surprise) - named the Action Simulation of Facial Actions sets (ASFA) layer.

In case of goal-directed hand actions, the chains are linked to neural representations of specific

goals or end states (e.g., attach wheel to base) which are represented by neural populations in the Intention Layer (IL). Facial action sets are linked to specific emotional states represented in the Emotional State Layer (ESL). This layer has influence in the IL, since an emotional state can play a role in identifying an intention.

If a chain (in ASL) is activated by observation of its first motor act, the observer may be able to predict future motor behaviour and the consequences of the whole action sequence before its complete execution, effectively inferring the partner's motor intention ahead of time. However, in some situations the observation of the first motor act per see, might not be enough if the motor act being observed is part of multiple chains. Likewise, a single facial action unit may by part of several different facial expressions. In order to disambiguate, additional contextual information is required to be integrated into the inference process (Erlhagen et al., 2007).

The Object Memory Layer (OML) that represents the robot's memorized knowledge about the location of the different objects in the two working areas, plays a key role.

Another important source of information, vital to the success of the task is the shared task knowledge about the possible sequences of sub-tasks (e.g. assembly steps in a joint assembly task). This information is provided by the Common Sub-Goals Layer (CSGL), which contains neural populations representing the subgoals of the task (e.g. individual assembly steps) that are currently available for the team. The connections between these populations encode subsequent assembly steps (for an example of how these connections could have established through learning by demonstration and tutor's feedback see (Sousa et al., 2015)).

In case of an assembly task, the sugoals are continuously updated in accordance with the assembly plan based on visual feedback about the state of the construction and the inferred goal of the co-actor (represented in the IL). Neurophysiological evidence suggests that in sequential tasks distinct populations in Pre-Frontal Cortex (PFC) represent already achieved subgoals and subgoals that have still to be accomplished (Genovesio et al., 2006). In line with this finding, CSGL contains two connected DNF layers with population representations of past and future events. The connections linking the neural populations in one DNF to the other DNF encode the different serial order of sub-goals of the task (see Sousa et al. (2015) for how these can be learned by tutors demonstration and feedback).

The Action Execution Layer (AEL) contains populations representing the same goal-directed action sequences and facial actions sets that are present in the ASL. Each population in AEL integrates inputs from the IL, ESL, OML and CSGL to select among all possible actions the most appropriate complementary behaviour. Specifically, the ESL (representing the inferred co-actor's emotional state) contributes to the selected emotional state to be expressed by the robot. The mapping from ESL to AEL implements some aspects of shared emotions in joint action (Michael, 2011). For example, if the human is in a positive state (Happy) the robot expresses also a Happy expression. This effect is known as emotion contagion and occurs when one person's perception of another person's emotional expression can have effects that are relevant to an interaction, if the perceiver thereby enters into an affective state of the same type (Michael, 2011).

In fact, one important way in which emotion contagion can function as a coordination smoother within joint action is by means of alignment. A key benefit of alignment is manifested by the likelihood of the increase in the participants' motivation to act jointly, since people tend to find other people with similar moods to be warmer and more cooperative, and prefer to interact with them (Locke and Horowitz, 1990).

The implemented context-sensitive mapping from observed actions on to-be executed complementary actions guarantees a fluent team performance if no errors occur (Bekkering et al., 2009). However, if an unexpected or erroneous behaviour of the partner occurs, neural populations in the Error Monitoring Layer (EML) are sensitive to a mismatch on the goal level, on the level of action means to achieve a valid sub-goal, and on the level of motor execution. This allows the robot to detect errors in human's intention and/or action means to achieve a sub-goal, and execution errors (e.g. a piece the robots was moving falls down), and thus allows the robot to efficiently cope with such situations.

The ESL is implemented as a dynamical neural field with different neural populations, where each of the populations represent an emotional state (Anger, Disgust, Fear, Happiness, Neutral, Sadness and Surprise).

Figure 4.3 depicts the implementation of the ESL, with the distribution of the different neural populations and an example of an emotional state being active. Here the 'State' axis attempts

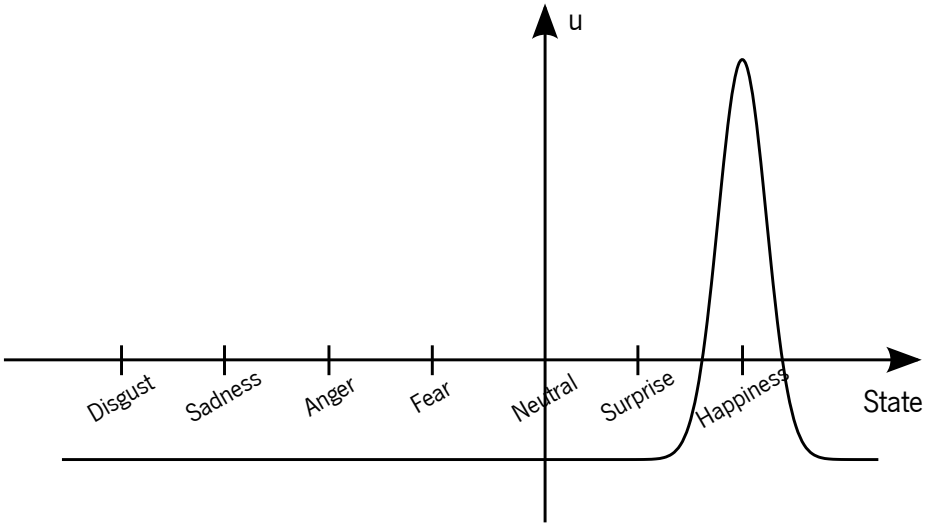


Figure 4.3: Emotional State Layer (ESL).

to represent the emotional valence associated with each emotion as being positive or negative according with the literature (for a review see: Hamann, 2012). This representation allows to define an order in which to place neural populations that represent each of the emotional states.

The ESL also plays a role in influencing the EML, implementing some aspects of shared emotions in joint action. Michael (2011) talks about the various types of shared emotions present in joint action tasks. One of the types of shared emotions used in our work is the emotion detection, which can facilitate prediction and monitoring of the partner's actions, and can also act as a signalling function.

For example, a positive emotional expression, such as a smile, may signal approval of another participant's action or proposed action (Michael, 2011). This way in our joint task, if the human partner is in a positive (e.g. Happy) emotional state, this might mean she/he is committed and engaged in the task, and thus it is not the probable that partner will make errors. In this situation, the processing of the DNFs detecting errors in action means and intention are disabled, since this allows to decrease the computational efforts of the robot's decision making processes, and hence the time it takes to select a complementary action is accelerated. In addition, if the human is in a positive emotional state, it means that she/he is comfortable with the robot and therefore one can increase the robot's movement velocity. Altogether this allows for the joint

task to be completed in less time.

Reversely, if the robot infers the human is in a negative emotional state (e.g. Sad), then it might be that the human is (also) not fully committed in the task and hence can be more prone to errors. The detection of a negative emotional state is used as a signal to activate the full processing of the EML. This is consistent with the modelling study by Grecucci et al. (2007), who proposed a computation model of action resonance and its modulation by emotional stimulation, based on the assumption that aversive emotional states enhances the processing of events. This way, the robot is fully alert to all types of errors that can occur during the execution of the task, being able to anticipate them, and act before they occur. This is fundamental for efficient team behaviour.

Through direct connections to the AEL, population activity in the EML may bias the robot's planning and decision process by inhibiting the representations of complementary actions normally linked to the inferred goal and exciting the representations of a corrective response. In order to efficiently communicate detected errors to the human partner a corrective response may consist of a manual gesture like pointing or a verbal comment to attract the partners' attention (Bicho et al., 2010).

Finally, it is important to highlight the connections from the ESL to both the AEL and Motor control. These connections implement the idea that perceived emotions play an important role not only in an early stage, during decision making and action preparation (AEL) of a complementary action, but also latter may affect the execution at the kinematics level (Motor control). This is motivated by recent studies in neuroscience by Ferri et al. (2010b,a), who have investigated the link between emotion perception and action planning & execution within a social context. In summary, they have demonstrated that assisting an actor with a fearful expression requires more smooth/slow movements, compared to assisting an actor with a positive emotional (e.g. happy) state.

In the different layers of the architecture subpopulations – encoding different hand action chains (ASHA), facial action sets (ASFA), goals (IL), complementary goal directed hand actions (Action Execution of Hand Actions (AEHA)), complementary facial actions (Action Execution of Facial Actions sets (AEFA)) and detected errors (EML) – interact through lateral

inhibition. These inhibitory interactions lead to the suppression of activity below resting level in competing neural pools whenever a certain subpopulation becomes activated above threshold. The population for which the summed input from connected populations is highest wins the competition process.

This page was intentionally left blank.

Chapter 5

The robotic setup

To test the dynamic neural field architecture for human-robot collaboration a joint assembly paradigm was chosen in which, the team has to construct a 'toy vehicle' from parts that are initially distributed on a table (see Figure 5.1).

The robotic setup used in this work was composed of an anthropomorphic robot performing a the joint construction of the toy vehicle with a human.

5.1 Joint construction task

The toy vehicle is a mockup of a mobile robot, composed of three sections. The lower section consists of a round platform with an axle on which two wheels have to be attached and then fixed with a nut (see Figure 5.2).

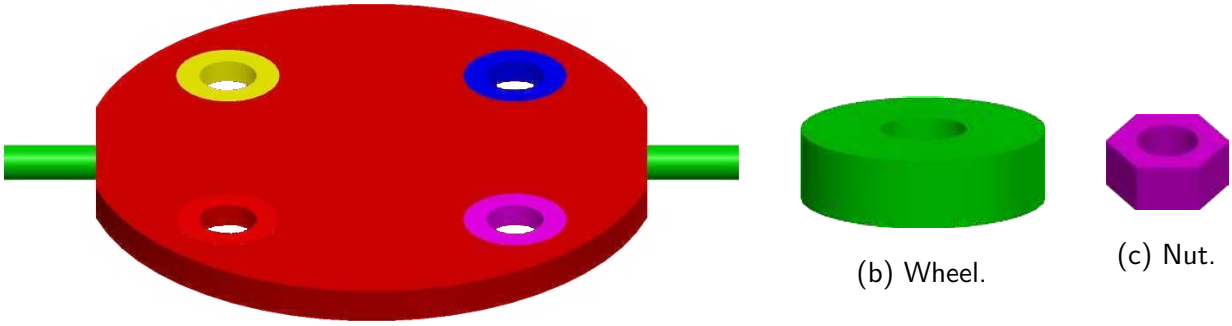
In the middle section, four columns that differ in their colour have to be plugged into specific holes in the platform (see Figure 5.3).

Finally, at the top section, the placing of another round object on top of the columns finishes the task (see Figure 5.4).

The parts have been designed to facilitate the workload of the vision and the motor system of the robot.

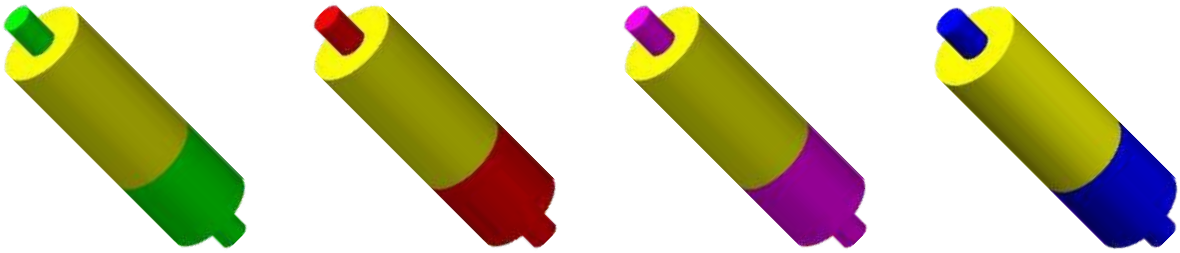
The working areas of the human and the robot do not overlap, the spatial distribution of the parts on the table obliges the team to coordinate handing-over sequences. It is assumed that

Figure 5.1: Anthropomorphic robot ARoS and the scenario for the joint construction task.



(a) Base. (b) Wheel. (c) Nut.

Figure 5.2: Objects that are part of the lower section of the toy vehicle.



(a) Column 1. (b) Column 2. (c) Column 3. (d) Column 4.

Figure 5.3: Objects that are part of the middle section of the toy vehicle.

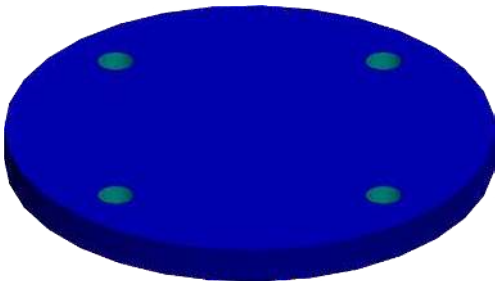


Figure 5.4: Object part of the top section of the toy vehicle: Top Floor.

each team mate is responsible to assemble one side of the toy, although, some assembly steps may require that one actor helps the other by holding still a part in a certain position. Both the human and the robot perform the same assembly actions.

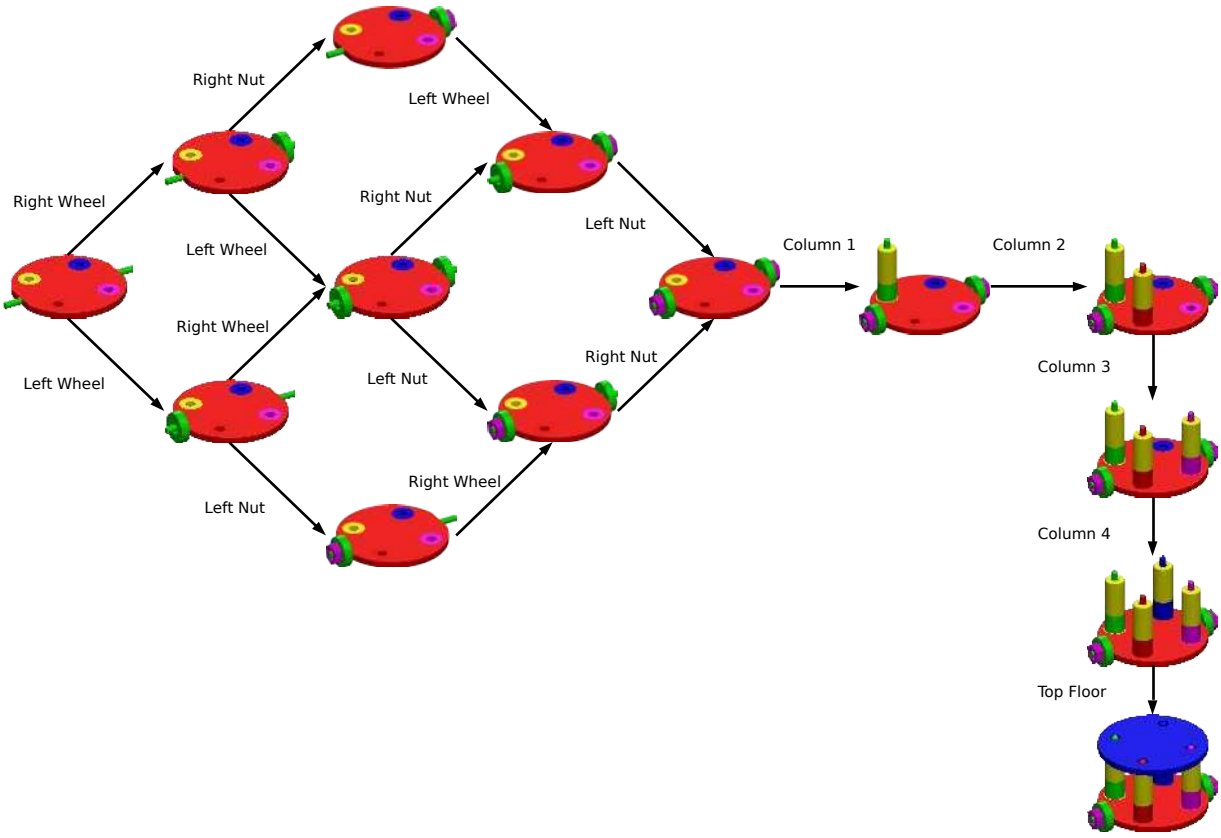


Figure 5.5: Construction plan for the Toy Vehicle.

It is assumed that both partners know the construction plan (see Figure 5.5) and keep track of the subtasks which have been already completed by the team. Since the desired end state does not uniquely define the logical order of the construction, at each stage of the construction the execution of several subtasks may be simultaneously possible.

The main challenge for the team is thus to efficiently coordinate in space and time the decision about actions to be performed by each of the team mates. The task is complex enough to show the impact of goal inference, emotional state inference, action understanding and error

monitoring on complementary action selection.

5.2 Anthropomorphic robot: ARoS

The robot ARoS (Anthropomorphic Robotic System) used in the experiments has been built in our lab (Silva et al., 2008). The robot consists of a stationary torus, on which a 7 Degrees Of Freedom (DOFs) AMTEC arm (*Schunk GmbH*) with a 3-fingers dexterous gripper (*Barrett Technology Inc.*), a stereo camera rig mounted on a pan-tilt unit, a PS3Eye camera with an adapted lens, are mounted.

The robot's body was designed from scratch to hold all components and to give the robot its anthropomorphic form. An anthropometric study (Drillis and Contini, 1966) was used to determine the measures of the body relatively to the arm, in order to have correct anthropometric proportions.

In addition, the robot has a monitor located on the chest, which is used to produce expressive faces in order to improve interaction with the human. The expressive faces the robot is able to produce, are performed using the same facial action primitives (Action Units) that can be recognized by the vision system. A speech synthesizer (Microsoft Speech SDK 5.1) allows the robot to communicate the result of its reasoning to the human user.

5.2.1 Manipulation: Arm & Hand

The execution of a complimentary hand action resulting from a robot's decision is translated into a fluent, smooth and collision-free arm trajectory. The robot has a motor repertoire of goal-directed movements that can consist of a simple pointing towards an object or more complex movements such as grasping an object with a certain grip.

For the control of the arm-hand system a global planning method in posture space was applied that allows us to integrate optimization principles derived from experiments with humans (Costa e Silva et al., 2015).

The generation of complete temporal motor behaviours of the robotic arm and hand was

achieved using an approach that draws inspiration from the posture model by Meulenbroek et al. (2001); Rosenbaum et al. (2001). It enables the generation of different types of human-like movements such as reaching, grasping and manipulation of objects, and also incorporates an obstacle avoidance mechanism. It was first proposed to be used in 2D movements and was later expanded to 3D workspaces (Vaughan et al. 2006).

The planning of movements for the arm is performed in joint space, instead of cartesian space, where the overall problem is divided into two sub-problems: end posture selection and trajectory selection. The planning system first selects a goal posture for the arm that satisfies two conditions: (a) The object is grasped without any collision with the robot's body or other object; (b) The displacement costs from the beginning to the end are minimized. This selection process is formalized as a non-linear constraint optimization problem (Costa e Silva et al. 2011), and numerically solved taking into account the information about the object: type, position and orientation (provided by the vision system) and how the object will be grasped (grip type and hand orientation relative to the object). Afterwards, the trajectory is generated by computing a sequence of position, velocity and acceleration in time from the initial posture to the final posture of the movement, for each of the 10 joints that compose the system arm and hand.

The trajectory selection applies the minimum jerk principle (Flash and Hogan, 1985), generating joint movements that follow a bell-shaped velocity profile. This trajectory defines the direct movement required to perform the action specified, without checking for possible collisions with intermediate obstacles or the object to be grasped. The collision detection is performed by internally simulating the movement execution (from start to end) with direct kinematics. When no collisions are anticipated, the movement is executed, otherwise an alternative trajectory must be found. To plan a trajectory that enables the movement to be executed with no collisions, the system selects a suitable bounce posture for this purpose.

The bounce posture is a back-and-forth movement superimposed on the previously selected direct movement. It is determined by solving a constrained optimization problem, similar to the one applied for the end posture. Acting as a subgoal for the movement, the bounce posture modifies the resulting path, but preserves the desired initial and end postures, enabling a collision free movement.

The generated movement is composed of a sequence of joint positions and time interval, these are sent to the low level controllers present in the arm and hand, which guarantee the execution of the multiple iterations to perform the movement.

The planning system generates collision free robot motion that is perceived by the human user as smooth and goal-directed.

5.2.2 Vision: Neck & Eyes

The vision system is composed of two independent systems, that provide distinct information. The first system (see Figure 5.6) is a stereo camera rig mounted on a pan-tilt unit and provides information about objects (type, position and orientation), hands (position, velocity, and classification of static hand gestures such as grasping, and communicative gestures like pointing) and the state of the construction task.

(a) Pan-Tilt unit.

(b) Stereo camera rig.

Figure 5.6: Stereo vision system (not to scale).

The second system is composed of a single camera (Figure 5.7a) with an adapted lens (Figure 5.7b). This camera, even though its not an industrial camera, offers a high performance acquisition (60 fps), unlike the majority of webcams on the market at the time. The used lens, replaces the camera's stock lens and enables us to adjust the camera field of view and create an image focused on the human face that maximizes the information that can be extracted from the face.

The second system provides information about the facial expressions and head movements

(a) PSEye camera.

(b) Vari-focal m12 lens, 2.8-12mm.

Figure 5.7: PSEye camera and lens (not to scale).

of the human interacting with the robot.

5.2.3 Expression of emotional states

Equipped on the chest is a VGA LCD monitor. This monitor is used by the robot to display an emotive (cartoon like) face to help in the interaction with the human. The displayed face is selected by the robot during the interaction, being the high level cognitive architecture responsible for the selection of the adequate face to display (see Figure 5.8).

The images were designed to display expressions of the emotions recognized by the robot using a prototypical set of Action Units (AUs) associated with each emotion.

(a) Sadness (AUs 1+15).

(b) Anger (AUs 4+7).

(c) Fear (AUs 2+5+26).

(d) Neutral (None).

(e) Surprise (AUs 1+2+26).

(f) Happiness (AU 12).

Figure 5.8: Images displayed by the chest monitor. Each image displays an emotion and the corresponding AUs used.

This page was intentionally left blank.

Chapter 6

Implementation details

This chapter presents details of implementation on several aspects of the robotic system. The scope of this work encompassed the design and implementation of a vision system capable of acquiring all the required information for the success of the joint cooperation task. The robot is equipped with two independent vision systems that provide information on different parts of the task.

The first vision system is a stereo system composed of two cameras fixed on an aluminium frame, which is mounted on a controlled pan-tilt unit. This system provides information about the objects that are part of the task, such as, 3D spatial position and orientation, task status, gesture recognition and movement quantification of the human hands and body.

The second system is composed of a small camera with a custom lens and is responsible for detecting and analysing the human's face, providing a description of the facial movements in the form of Action Units (AUs). Also, the system is able to quantify the head movement.

All the information generated by these systems is incorporated into the high level cognitive architecture, that processes it and generates the robot behaviour.

6.1 Object information

The object information is provided by the stereo system, which combines a image processing algorithm with stereo data to determine all the information required. The processing applied to

images from the stereo cameras identifies the objects that are part of the construction task (see Chapter 5, Figures 5.2, 5.3 and 5.4). The objects were designed with well defined colours so they could be more easily detected by a colour based search algorithm.

The image captured by the cameras is in Red Green Blue (RGB) colour space. However, because the search algorithm will be colour based, the use of the Hue Saturation Value (HSV) colour space or Hue Lightness Saturation (HLS) is more appropriated, the fact that they use a colour codification that is more intuitive and similar to the how the human vision processes colour, makes them are more appropriated to a colour processing application (Li et al., 2002). These colour spaces offer the advantage of coding the luminosity and chromatic information of colour separately, making the colour definition insensitive to illumination.

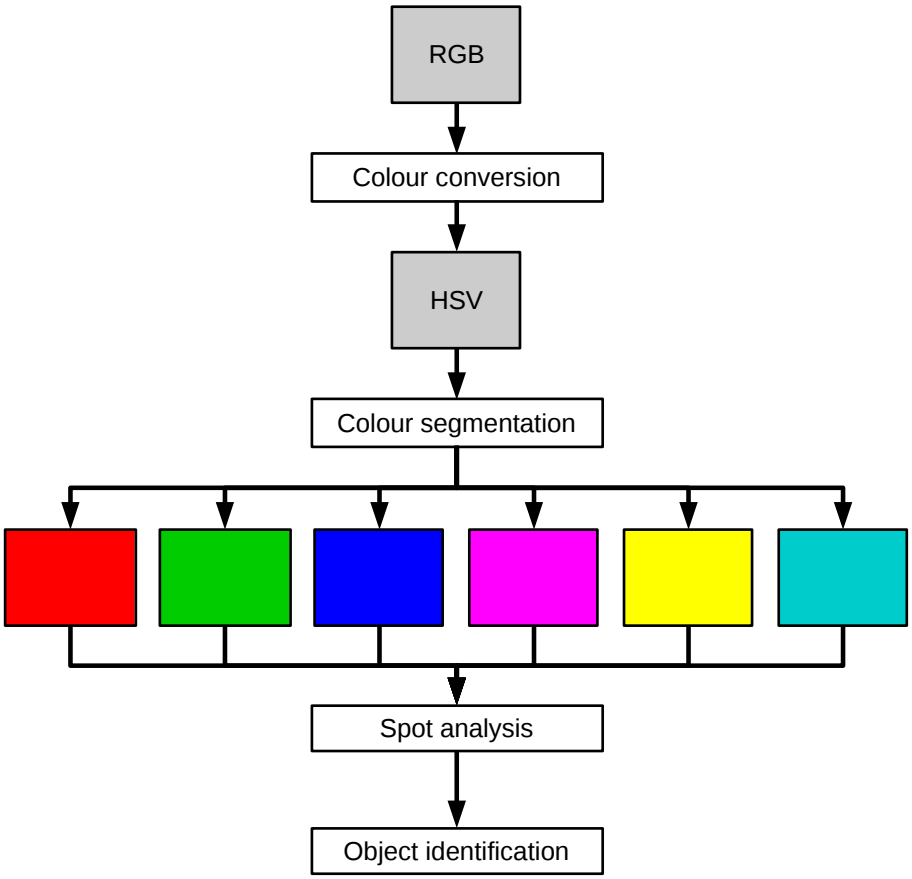
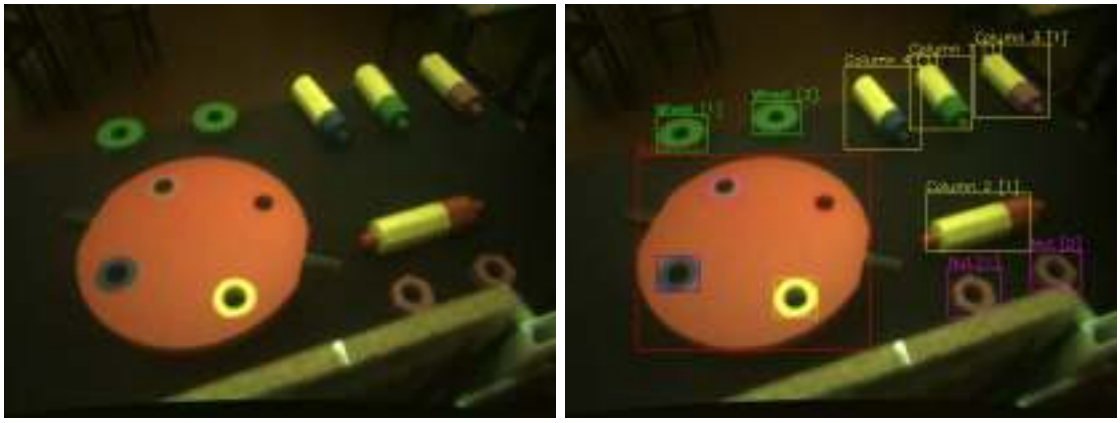


Figure 6.1: Illustration of the steps involved in the image processing algorithm.

Figure 6.1 illustrates the order of steps involved in the image processing algorithm responsible

for identifying the objects. First, the image is converted from the RGB colour space to HSV, next a colour segmentation is applied in order to create six binary images that contain spots of the colours that the algorithm is looking for (red, green, blue, magenta, yellow and cyan). Each of the image associated with a search colour is processed to remove noise and create regions associated with every spot detected. After the spots are identified, the object identification step will combine the information of spots and attempt to match the spots or sets of spots to objects.



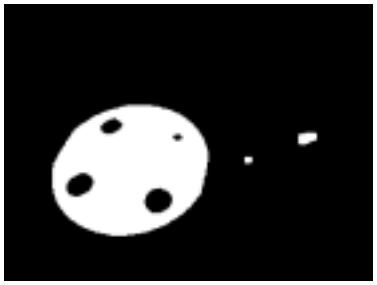
(a) Original. (b) Processed.

Figure 6.2: Images taken from the robot camera with the processing applied.

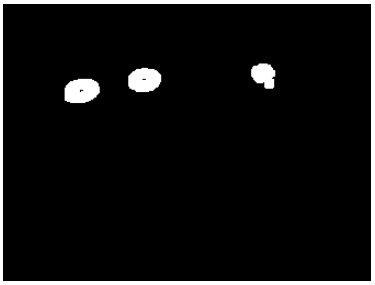
The match of spots to objects is performed by following a set of properties that define each object. A large red spot with holes is identified as a Base, the combination of each hole position with spots of another colour, identifies the hole in the base. Similar to the Base, a large blue spot is identified as the Top Floor. A yellow spot which has another spot of a different colour contiguous in the image is identified as the correspondent column. All green and magenta spots are treated as wheels and nuts, respectively.

After the identification of objects, the information of each one will be filled in various steps. The spots that originated the objects are used as a mask that is combined with the stereo information, to extract the coordinates of each object in the camera reference frame.

All the pixels (n) in the region of the object are used to calculate the object coordinates,



(a) Red.



(b) Green.

(c) Blue.

(d) Magenta.

(e) Yellow.

Figure 6.3: Binary images resulting from the colour segmentation.

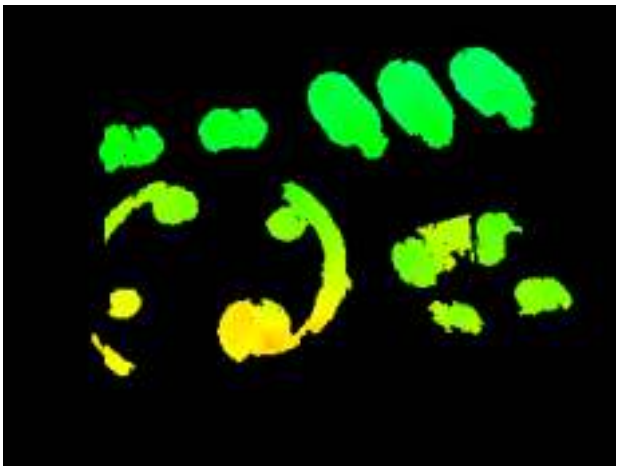


Figure 6.4: Result of the stereo computation.

The x, y and z coordinates of the object are given by:

$$x = \frac{\sum_{i=0}^{n} m_i\, v_i\, x_i}{\sum_{i=0}^{n} m_i\, v_i} \qquad y = \frac{\sum_{i=0}^{n} m_i\, v_i\, y_i}{\sum_{i=0}^{n} m_i\, v_i} \qquad z = \frac{\sum_{i=0}^{n} m_i\, v_i\, z_i}{\sum_{i=0}^{n} m_i\, v_i} \tag{6.1}$$

where, n is the number of pixels in the region of the image that contains the object. m defines the mask of the object, if the pixel i is part of the mask $m_i = 1$, otherwise $m_i = 0$. v is created by the stereo computation, if the stereo process was able to calculate valid coordinates to the pixel i, $v_i = 1$ otherwise $v_i = 0$. x_i, y_i and z_i are the coordinates in the camera reference frame for pixel i.

The coordinates are transformed from the camera reference frame to the robot's reference frame using transformation matrices defined by the robot's dimensions and the pan-tilt position (see Silva et al., 2008, for more details).

Each object, depending on its type can have sub-objects associated to it. For example, the base has sub-objects that identify the holes, and the columns have sub-objects that identify the primary and secondary colours. Using the coordinates calculated for these sub-objects, it is possible to calculate the orientation of these objects, since it is a vital information for the robot to be able to manipulate columns and figure out how the base is positioned in the table.

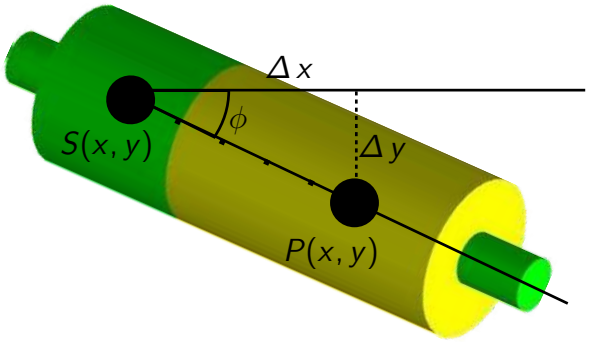


Figure 6.5: Orientation of a column, on the Z axis.

Figure 6.5 shows how the orientation is calculated from the x and y coordinates associated

to the primary and secondary colours present in the column.

$$\phi = \arctan\left(\frac{\Delta y}{\Delta x}\right) \tag{6.2}$$

The method for calculating the orientation of the base is similar, but because the base has several holes, each one different, any combination of two holes that are detected can be used for calculating the orientation, even if all the features of the base cannot be detected by the vision system.

6.2 Gesture recognition

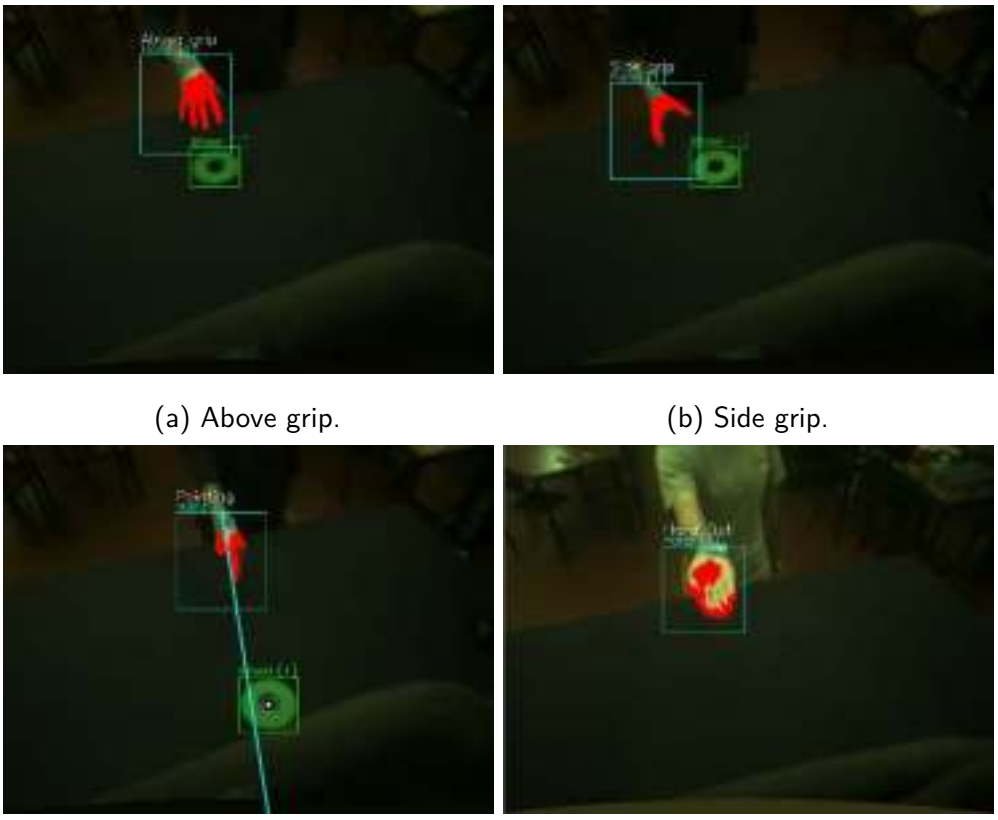
The first step in gesture recognition is to detect and segment the hands, separating in the image the environment from the hand that needs to be processed. Several approaches are used to address this issue, colour detection, shape, 3D models and movement based (Rautaray and Agrawal, 2015). The colour based methods, rely in the definition of a model of skin colour to detect the hands, this however poses some problems due to the variations of skin tones. Some methods try to mitigate these issues by applying compensations (Sigal et al., 2004), but fail under some circumstances, such as rapid variations in lighting.

During the interaction, as a means to ease the detection of the human hands, the humans interacting with the robot wear a coloured bracelet on the wrists. Making it easier to detect and dodging the problems that skin colour based models present. Applying a colour based search, a region of the hand is extracted. For the gesture recognition, only the region of the image that contains the hands is processed, in order to make the process faster and more efficient.

The used method for classifying the gesture was implemented by Cunhal (2014). It creates a database for desired gestures, where each gesture is defined using vectors of moment invariants (Hu, 1962), through image examples. In runtime, the image of the hand is processed to determine the vectors of moment invariants and then compared with the gestures in the database.

Currently, four gestures are supported: above grip, side grip, pointing and hand out[1]. Figure 6.6 shows an example of the gesture recognition.

[1] **!!! Get and image of hand-out gesture !!!**



(a) Above grip. (b) Side grip.

(c) Pointing. (d) Hand out.

Figure 6.6: Recognized gestures.

The detection of the 'hand out' gesture is performed based on a spatial location. If the human's hand is detected in a pre-defined location close to the robot and the gesture classification does not match any of the other three detected gestures, then this is classified as a 'hand out' gesture.

However, classifying the gestures is not enough, the vision system is also able to detect which object is the gesture being applied to, when possible. If the gesture is being recognized but the vision system detects no object, this information alone can still be relevant for the high level cognitive architecture.

When the vision system detects a gesture and an object, in the cases of above grip and side grip the system determines which object is being grasped, by searching for objects in the region of the hand. In the case of pointing, the system makes an estimation of where the human is pointing, by creating a line using the points in the bracelet and the fingertip, and then calculating the distance from this line to all objects around it. It then selects the closest object as the target being pointed to. The line is created by two points, the coordinates of the wrist $P1(x, y)$ and the coordinates of the fingertip $P2(x, y)$. If the coordinates of an object are defined by point $P0(x, y)$, the distance from the line to the object is given by:

$$distance(P1, P2, PO) = \frac{|(y_2 - y_1)x_0 - (x_2 - x_1)y_0 + x_2y_1 - y_2x_1|}{\sqrt{(y_2 - y_1)^2 + (x_2 - x_1)^2}} \tag{6.3}$$

6.3 Quantification of human movement

The movement of the human quantified by the vision system comprises the movement of the hand, the movement of the body and the movement of the head.

6.3.1 Hand

The hand movement is measured by tracking each of the detected hands position. The tracking is achieved by combining a labelling algorithm (Chang et al., 2004) with an occlusion handling method (Senior et al., 2006). This combination ensures a stable tracking over time, while dealing with occasional occlusions that might occur in real-time.

For each processed frame, the position of each hand is taken and used to calculate the speed relatively to the previous frame.

$$v_i = \frac{\left\| \vec{pos}_i - \vec{pos}_{i-1} \right\|}{dt} \tag{6.4}$$

where pos is the cartesian position at a given time.

Additionally, and to prevent large variations in speed, the determined value is filtered using an infinite impulse response low-pass filter of the form:

$$v_{filt,i} = v_{filt,i-1} + \frac{dt}{smooth} \cdot (v_i - v_{filt,i-1}) \tag{6.5}$$

where $v_{filt,i}$ is the current value filtered, v_i is the current speed measured, $v_{filt,i-1}$ is the previous value of speed filtered, dt is the time elapsed since the previous sample and $smooth$ is a smoothing factor to attenuate fast variations in the speed.

6.3.2 Body

One of the most widely used methods to measure movement in vision is by background subtraction (see Piccardi, 2004, for a review). By taking a reference image of the background and subtract it from the image being analysed, it is possible to create a segmented image of the subject in study.

The Quantity of Movement (QoM) is measured from the amount of detected motion, computed with a technique based on Silhouette Motion Images (SMIs), as reported by Castellano et al. (2007), which uses a structured environment with a uniform dark background to achieve silhouette extraction. Our system implements the movement measurement by making use of the stereo vision to remove the background and foreground of the image.

Taking advantage of the spatial cloud of points created by the stereo computation, a mask is created with only the pixels that are within a pre-defined spatial range, that can be thought as a cube containing the space where the human is located.

$$SMI(t) = \sum_{i=1}^{n} (S(i) - S(i-1)) \tag{6.6}$$

where $S(t)$ is the silhouette image at each instance computed by removing the background and foreground, n controls the amount of images that will be used as history, and SMI is the sum of all detected movement from instant '$t - n$' to instant 't'.

The QoM is computed as a ratio between the area (number of pixels) of the SMI and the area of the silhouette, as given by:

$$QoM = \frac{Area(SMI(t))}{Area(S(i))} \tag{6.7}$$

6.3.3 Head

Using the information available of the head position and orientation, these two measures are accumulated over time and an estimation is created of the current speed of the head. The instant linear ($\vec{v}$) and rotational $\vec{\omega}$ speeds at instant $t = i$ are given by:

$$\vec{v_i}(x, y, z) = \frac{\vec{p_i}(x, y, z) - \vec{p_{i-1}}(x, y, z)}{dt} \qquad \vec{\omega_i}(\theta, \psi, \phi) = \frac{\vec{\alpha_i}(\theta, \psi, \phi) - \vec{\alpha_{i-1}}(\theta, \psi, \phi)}{dt} \tag{6.8}$$

where $\vec{p}$ is the cartesian position at a given instant and $\vec{\alpha}$ is the angular position at a given instant. The instant values are then smoothed by a low-pass filter implemented using the method presented by Equation 6.5.

6.4 Face detection

In order to use information extracted from faces, an automated analysis system would have to detect a human face in an image. This would be the first step in autonomously extracting usable information from a human face. Only when this problem is addressed and solved, resulting in a sub-image or a region where the face is located, the face analysis can leap to the next step in the process.

Even though there is an extensive research effort in creating facial analysis algorithms (Pantic and Rothkrantz, 2000), there is no system that can be deployed in an unconstrained environment and is able to tackle with the inherent variability in imaging parameters such as sensor noise, viewing distance, and light conditions. The only known system that seems to work well, dealing

with all these challenges is the human visual system. There is some research that attempts to understand the strategies employed by the human visual system, and try to use them to create machine-based algorithms that address this issue (Sinha et al., 2006).

The detection and tracking of the human face is accomplished by using an available commercial product *faceAPI* by *Seeing Machines*. This Application Programming Interface (API) enables the creation of a tracking engine that continuously detects and tracks a human face in a video in real-time. The available data returned by this engine comprises the position of some facial characteristics and the position and orientation of the head in the camera reference frame.

6.5 Face analysis

The analysis performed on the face can be decomposed into two steps, the acquisition of data (position of facial points) and the data processing to implement the face coding (this last step will be further discussed in the next section).

The followed approach to analyse the face was to segment it into several parts in an attempt to create a coding system for each expression. Ekman (1971) studied the common and different aspects of face expressions in different cultures, when expressing a small set of basic emotions.

Ekman et al. (2002) developed a coding system for faces named Facial Action Coding System (FACS). This system emerged from the need to create methodologies to measure facial behaviour validly and reliably, to document the research being done when investigating the universality of emotions. It defines, based on facial muscles, measures for all types of visible facial behaviour, identifying the possible independent movements of the facial musculature, referenced as AUs. To each AU is assigned a simple numeric code, this way each face expression can be described by the different presence of AUs.

In order to accomplish a reliable and valid coding system for face expressions, the coding system created by Ekman et al. (2002) was used. In this system it is also defined, for each Action Unit, a scale of intensity ranging from A (faint presence) to E (strong presence). Since most of AUs happen in a very small time-scale, and some are very subtle, this introduces a limitation in the used hardware and software, and for this reason, the intensity scale of AUs will

be ignored during this work, while focusing on the presence or not of each Action Unit.

FACS defines all the possible visible movements in the human face, the number of Action Units and possible combinations is very large. This motivated the appearance of a number of derivatives of FACS. For instance, to account for the differences in some aspects in the face of children, Oster (2004) developed Baby FACS, a specialized version of FACS for infants and young children. Friesen and Ekman (1982) developed Emotional Facial Action Coding System (EMFACS), an abbreviated version of FACS, also following the work by Darwin (1872), which identifies only those AU that are theoretically or empirically related to emotion, the implemented coding process focuses only on the identification of a smaller set of AUs or AU combinations, greatly reducing the number of AUs to search for.

Emotion	**AUs**
Anger	4, 5 and/or 7, 22, 23, 24
Contempt	Unilateral 12, Unilateral 14
Disgust	9 and/or 10, 25 or 26
Fear	1, 2, 4, 5, 7, 20, 25 or 26
Happiness/Joy	6, 12
Sadness	1, 4, 15, 17
Surprise	1, 2, 5, 25 or 26

Table 6.1: Description of AU combinations associated with emotions according to EMFACS.

These AU combinations, although they are not rigidly defined and valid for every person, they help us to narrow the relevant information to extract from the face and the most important AUs to focus on.

Table 6.2: Description of Action Units and appearance changes caused by each AU in the face, according to Ekman et al. (2002).

Code	**Name**	**Description**
1	Inner Brow Raiser	Pulls the inner portion of the eyebrows upwards.

Table 6.2: (continued).

Code	Name	Description
2	Outer Brow Raiser	Pulls the lateral (outer) portion of the eyebrows upwards.
4	Brow Lowerer	Lowers the entire eyebrow.
5	Upper Lid Raiser	Widens the eye aperture.
6	Cheek Raiser	Lift the cheeks and compresses the eye socket.
7	Lid Tightener	Tightens eyelids, narrows eye aperture.
9	Nose Wrinkler	Pulls the skin along the sides of the nose upwards, causing wrinkles to appear along the sides of the nose.
10	Upper Lip Raiser	Raises the upper lip. Centre of upper lip is drawn straight up, the outer portions of upper lip are drawn up but not as high as the center.
12	Lip Corner Puller	Pulls the corners of the lips back and upward (obliquely).
14	Dimpler	Tightens the corners of the mouth, pulling the corners somewhat inwards, and narrowing the lip corners.
15	Lip Corner Depressor	Pulls the corners of the lips down.
17	Chin Raiser	Pushes the chin boss upward.
20	Lip Stretcher	Pulls the lips back laterally, the main movement is horizontal.
22	Lip Funneler	Lips funnel outwards taking on the shape of a funnel.
23	Lip Tightener	Tightens the lips, making the red parts of the lips appear more narrow.

Table 6.2: (continued).

Code	Name	Description
24	Lip Presser	Presses the lips together, without pushing up the chin boss.
25	Lips Part	The lips part, which may expose the inner mucosal area of the lips.
26	Jaw Drop	The mandible is lowered by relaxation.
51 / 52	Head Turn Left/Right	Codes the orientation of the head left to right on a vertical axis.
53 / 54	Head Up/Down	Codes the orientation of the head when oriented up or down.
55 / 56	Head Tilt Left/Right	Codes the orientation of the head when leaned to the left or right.

Table 6.2 shows a description of each AU referred in Table 6.1, it describes the most relevant changes that occur in the face when each AU is present. Having this set of AUs, and taking into account that their detection must be performed by an automated system, the focus of attention turned to the AUs that are feasible for being detected, eliminating others that may be either too subtle or too similar to others to be effectively detected in a proper time frame.

After the analysis of these AUs a smaller subset from those referred in Table 6.2 was created, with the AUs considered to offer a better confidence for automated detection: AU 1, 2, 4, 5, 12, 15, 20, 23, 26, 51, 52, 53, 54, 55, 56.

6.5.1 Action Unit detection

The detection of AUs is an automated coding system that uses data gathered from the face and applies FACS to generate the description of a facial expression.

It requires first that in the face which is being analysed to have a good tracking of stable facial landmarks. These landmarks are points in the face that are identifiable even if they move

within the face, as in the case of facial expressions. Zhao et al. (2011) defines the stable facial landmarks to generally include the nose tip, the inner eye corners, the outer eye corners, and the mouth corners, which are not only characterized by their own properties (texture and shape), but are also characterized by their global structure resulting from the morphology of the face.

Branco (2006) employs a modular method for analysing the face, after the face is detected by a face detection module, a combination of Active Appearance Model (AAM) and Principal Component Analysis (PCA) is used to extract features from the face and classify its expression. The work here presented relies on the *faceAPI* to detect and track facial features which are then processed by a set of pre-defined rules.

The result of the continuous detection and tracking of *faceAPI* engine, generates a set of points in the face and the coordinates for each point in several coordinate frames. To each point in the face, there are associated coordinates in these four different coordinate frames:

- Image;
- Camera;
- Head;
- Face texture.

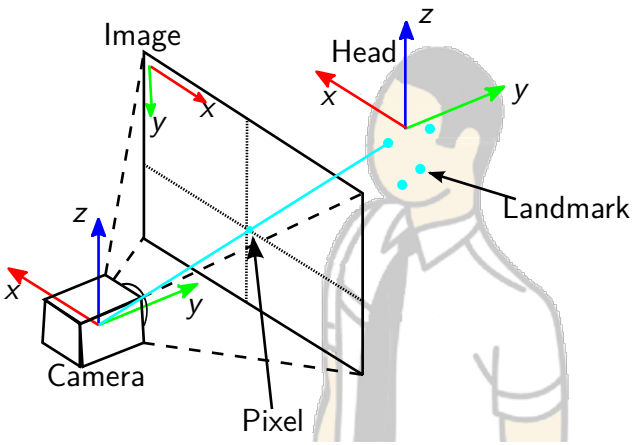


Figure 6.7: Different coordinate frames in *faceAPI* (Image adapted from documentation on *faceAPI*).

Figure 6.7 shows how all of the coordinate frames are integrated into the detection engine. The 'Image' coordinate frame is associated with the acquired image from the camera, each

point is defined by one pair of coordinates (x, y) that identify where its located in the image, the units are in pixels. The 'Camera' coordinate frame is a 3D axis where the origin is fixed in the used camera, each point is defined by three coordinates (x, y, z), measured in meters. The 'Head' frame is similar to the Camera frame but the origin of the axis is fixed in the head, the coordinates are also measured in meters. The 'Face texture' frame is a normalized plane placed in the face and each point is defined by a pair of coordinates (x, y) that range from -0,5 to 0,5 (not shown in Figure 6.7).

After running some comparison tests using the engine, two coordinate frames were candidates to be chosen to work with the data, the 'Head' frame and the 'Face Texture' frame. Although their accuracy is identical, the head coordinate frame was chosen because it maps directly the dimensions of the detected face in real-world units, rather than an abstract coordinate frame used by the 'Face Texture' frame.

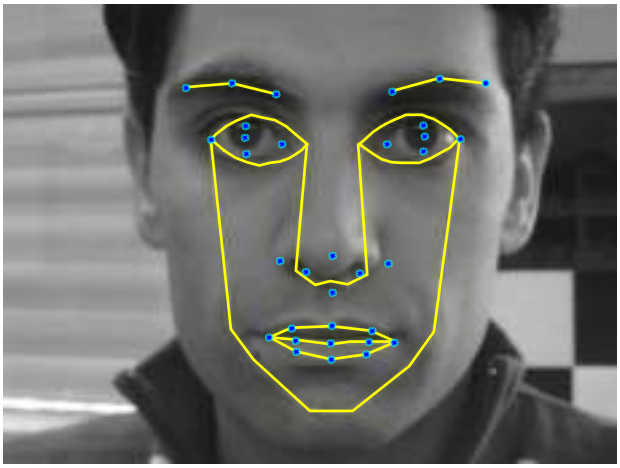


Figure 6.8: Detected points by *faceAPI* engine.

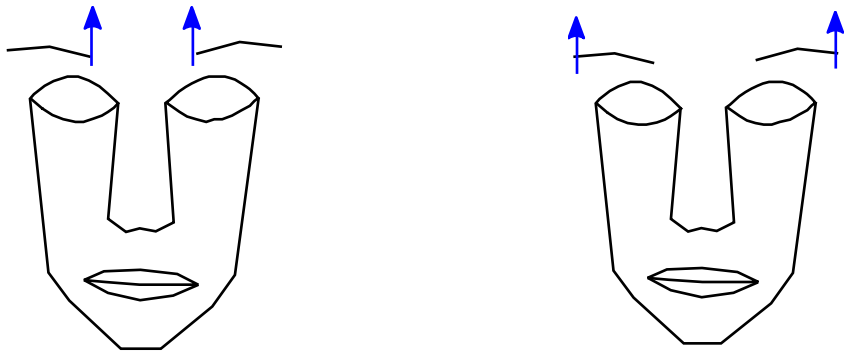
Figure 6.8 depicts all the tracked points in a face by the *faceAPI* engine. These points focus in three different aspects of the face, the eyes, the eyebrows and the mouth. From these areas, the ones that offer more accuracy and robustness in detection are the eyes and eyebrows. Although the mouth area has many points to track the lips, their response to movement is not as reliable as the eyebrows. Each point tracked in the face, designated as Face Landmark within *faceAPI* is marked as a blue dot, while the face mask that aggregates all these points is marked

with yellow lines. Additionally, the engine also determines the head position and orientation, making it possible to analyse the head movements.

Schyns et al. (2009) reports in a study with volunteers, where each expression was shown on 10 different faces, five male, five female, while brain-imaging equipment monitored how quickly different parts of the brain interpreted them, that all of the universal expressions have distinctive characteristics that the brain can easily distinguish between. Where "happy" requires the smiling mouth and the eyes, "surprised" the open mouth, "anger" the frowned eyes and the corners of the nose, "disgust" the wrinkles around the nose and the mouth, "sadness" the eyebrows and the corners of the mouth and "fearful" the wide-opened eyes (Schyns et al., 2009).

With the faceAPI data acquisition system, and taking into consideration some studies (e.g. EMFACS(Friesen and Ekman, 1982) and Schyns et al. (2009)), the focus was to implement only a small subset of AUs into the analysis system, rather than all the discussed AUs.

The determination if an AU is present or not, requires to overlap the definition of each AU with the points obtained by the *faceAPI* engine. Using the face mask generated by *faceAPI*, shown in Figure 6.8, and based on the description provided by Ekman et al. (2002) for each AU, a set of rules was created to map the data acquired into AUs. The movement associated with each AU is mapped on top of the face mask.



(a) AU 1. (b) AU 2.

Figure 6.9: Facial movement produced by AU 1 and AU 2.

Figure 6.9a defines the rule to detect AU 1. This AU codes the upward movement of the inner part of the eyebrow. Whenever the points associated with the inner part of the eyebrow

are above a certain threshold, the AU 1 is considered to be present.

In Figure 6.9b one can see the movement associated with AU 2. The detection rule is identical to AU 1, but the movement is associated with the outer part of the eyebrow. Usually AU 1 and AU 2 are present simultaneously, since it is generally difficult to perform these two movements separately.

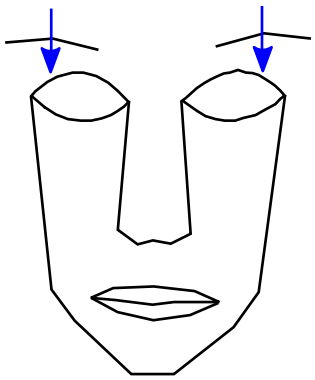


Figure 6.10: Movement produced by AU 4.

When the eyebrows are pushed down, this movement is coded by AU 4, shown in Figure 6.10. Although the upward movement of the eyebrow is coded by two separate AUs for the inner and outer parts, the downward movement is coded by a single AU. When the eyebrows are closer to the eyes than in the relaxed position, AU 4 is considered to be present.

(a) AU 12. (b) AU 15.

Figure 6.11: Facial movement produced by AU 12 and AU 15.

AU 12 codes an oblique movement upwards from the lip corners, producing a smile, shown in Figure 6.11a. This AU is detected comparing the position of the lip corners with the upper

point of the lips, if the lip corners are above, AU 12 is considered present.

AU 15 codes the opposite movement of AU 12. This AU presents an oblique movement downwards from the lip corners, as shown in Figure 6.11b. The detection is similar to AU 23, but in this case is considered when the lip corners are below the lower point of the lips.

(a) AU 20. (b) AU 23.

Figure 6.12: Facial movement produced by AU 20 and AU 23.

The horizontal movement of the lips is coded by two AUs. AU 20, shown in Figure 6.12a, codes an horizontal movement from the corner lips as they move further apart opposite direction, stretching the lips. While AU 23, shown in Figure 6.12b, codes the opposite movement of the corner lips, as these move to compress the lips. For the detection of AU 20 and 23, the distance between lip corners is calculated, and two limits are defined. When the upper limit is passed, AU 20 is present, if the lower limit is passed, AU 23 is present.

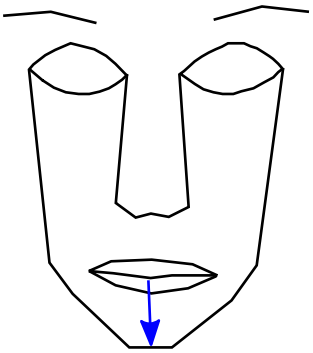


Figure 6.13: Movement produced by AU 26.

AU 26, shown in Figure 6.13, is produced by relaxing the jaw. This causes the mouth to

open, the detection is carried out by computing the distance between the lower point of the lower lip to the middle point of the upper lip.

6.5.2 Implementation

The implementation of AU coding relies in the comparison of the current data on the face being analysed with a reference face of the person interacting with the robot. The comparison with a reference face, attempts to solve the problem of the variability present in faces and the initial tracking of the *faceAPI* engine. Each AU has a rule for detection, where the parameters for them were tuned from the tests performed on the Cohn-Kanade database (Kanade et al. 2000; Lucey et al., 2010). The parameters code the ratio of change from the current expression to the reference neutral expression.

The initial conditions for the tracking engine vary in the sense that the mask fitting to the face is not always the same. Sometimes, the face mask generated is slightly larger or smaller than the tracked face, and even the mask orientation varies. This variation associated with the natural variability present in faces makes it impossible to adjust coding parameters robust enough for general use.

When a human starts interacting with the vision system, it is asked to present a neutral face. The neutral face is used by the system to establish the reference data required to code the AUs present in the face. According to the rules presented before, the position of each point is compared with the position in the reference frame. Each AU has detection parameters associated that determine a ratio of variability accepted to code each one.

Figure 6.14 shows an example of coding for AU 1 and 2. The reference data is shown in dashed line, while the solid line is the current face being analysed. For AU 1 the rate of variation of outer points of the eyebrows between the reference face and current is measured, if this variation exceeds the detection parameter, the AU is coded as present, otherwise coded as absent. In the same manner, for AU 2 the rate of variation is measured using the inner points of the eyebrow.

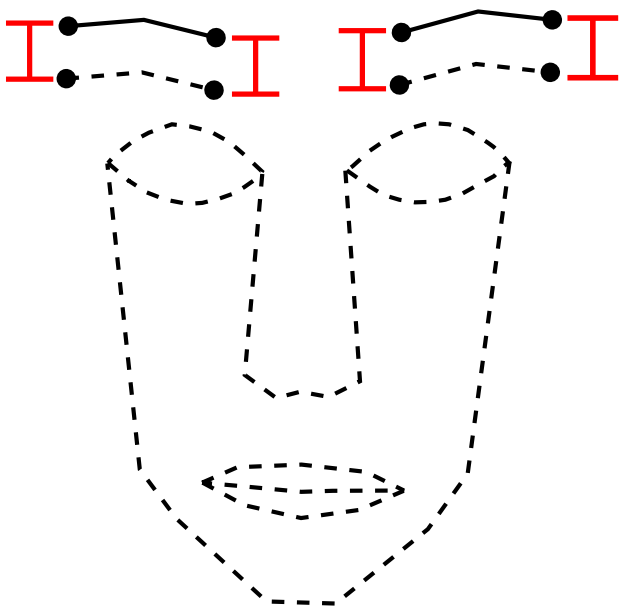


Figure 6.14: Example of AU coding (Dashed - reference; Solid - current).

This page was intentionally left blank.

Chapter 7

Results

To validate the dynamic neural field architecture for human-robot collaboration, real-time human-robot experiments in scenarios of the joint construction task described in Chapter 5, were designed and conducted.

For better understanding the construction task was divided in three logic stages, lower section (wheels and nuts), middle section (columns) and top section (Top Floor).

The focus is on showing and explaining how decision making and error detection are affected by the human partner's emotional state. In all cases, the initial spatial distribution of parts forces both actors to demand and hand-over parts. There is no verbal communication from the human to the robot. This obliges the robot to continuously monitor and interpret the actions of its co-worker. Both the human and the robot can manipulate the parts (e.g. plug a wheel on the axle).

The robot uses speech to communicate to the human partner the outcome of the goal inference and decision making processes implemented in the dynamic neural field model. As our studies with naive users show, this basic form of verbal communication facilitates natural and fluent interaction with the robot (Bicho et al., 2010).

To validate the high level cognitive control architecture, five different experiments were designed. Each experiment will address a specific feature with different scenarios, in order to better understand how the partner's emotional state can affect the robot's behaviour.

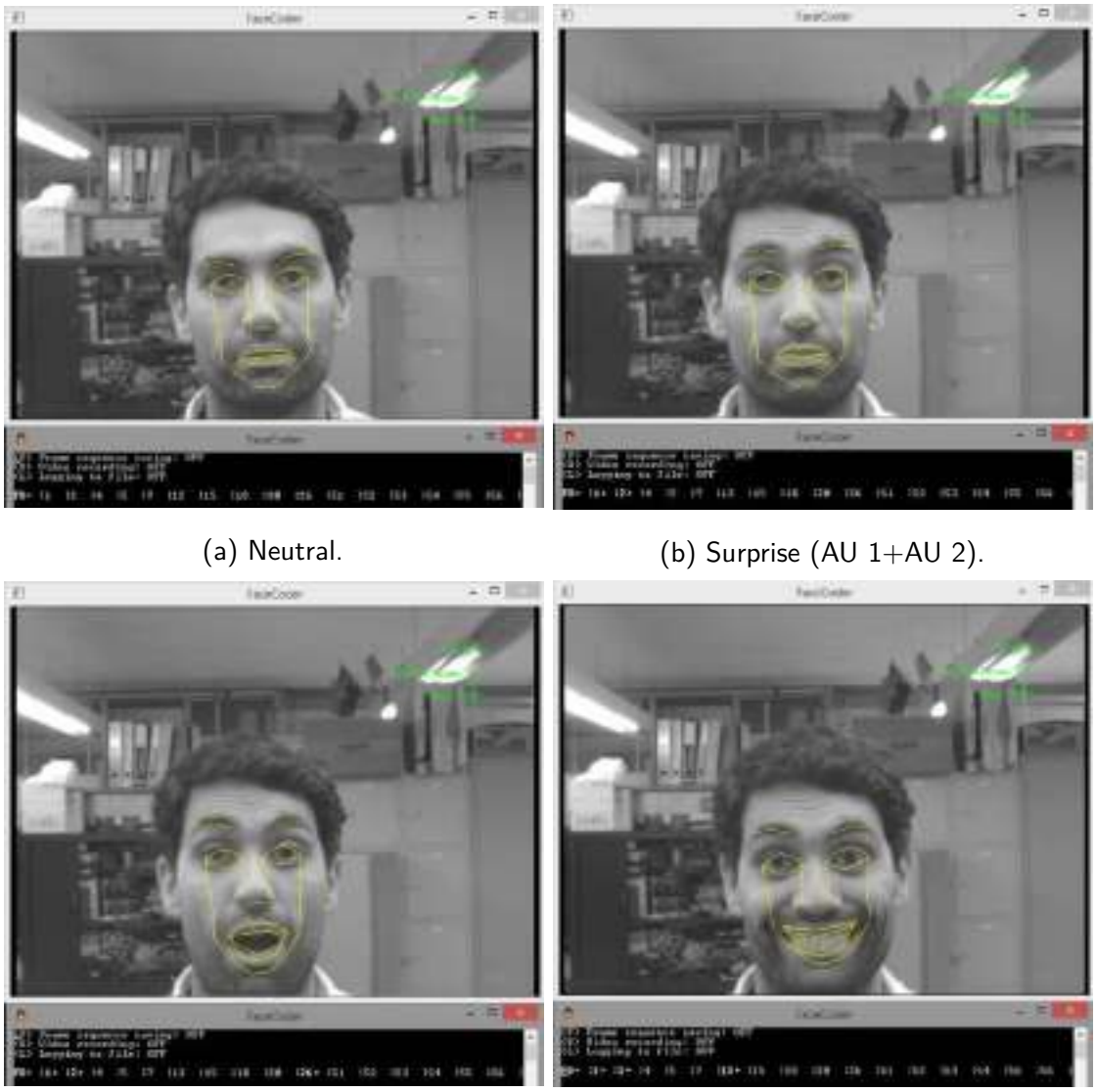
Experiment 1 explores how the robot's decisions can be influenced by the partner's emotional state. Experiment 2 shows how the inferred user's emotional state can influence how the robot detects and handles errors during task execution. Experiment 3 shows how the robot, by expressing emotional facial expressions, deals with a human persisting in an error. Experiment 4 presents a comparison of the influence of the user's emotional state in the time the task takes to be performed. Finally, experiment 5 shows the dynamic nature of the architecture in a longer interaction, where the robot adjusts its behaviour in real time, in response to the change of the human emotional state.

The graphics presented for each scenario, show the time evolution of fields activity in some layers of the control architecture. To show the evolution in all layers of the architecture would be impractical, hence, only key layers for each scenario will be presented.

The main contribution of this work is the integration of emotions into the robot's cognitive architecture. Hence, before presenting the interaction results, details on how the information acquired by the vision system regarding the human face is handled, will be presented. Figure 7.1 presents snapshots of the analysis performed by the system developed for this robot. A dedicated camera placed on the robot acquires an image of the face, which is then processed by combining the library *faceAPI* from *Seeing Machines* with some post-processing algorithms implemented using the Open Source Computer Vision (OpenCV) library. The system is continuously processing (at 60fps) and coding the face according to the Facial Action Coding System (FACS) (Ekman et al., 2002), resulting in a real-time description of the face with Action Units (AUs).

The entry point in the architecture for the information provided by the vision system is the Action Observation Layer (AOL). Three Dynamic Neural Fields (DNFs) in this layer are responsible for representing information about detected facial muscle movements that are associated to the eyes, eyebrows and mouth.

Figure 7.2 shows the time evolution of the DNFs involved in the processing of this visual information and the simulation and inference of the user's emotional state (layers AOL, Action Simulation of Facial Actions sets (ASFA) and Emotional State Layer (ESL) respectively). On top, and regarding AOL, one can see a DNF, $u_{\text{AOL\_FaceDetect}}(x, t)$, that codes the presence or not of a human face and the three DNFs responsible for representing AUs related to the eyebrows



(a) Neutral.

(b) Surprise (AU 1+AU 2).

(c) Surprise (AU 1+AU 2+AU 26).

(d) Happy (AU 1+AU 2+AU 12).

Figure 7.1: Face analysis by the vision system.

($u_{\text{AOL\_Eyebrows}}(x,t)$), mouth ($u_{\text{AOL\_Mouth}}(x,t)$) and eyes ($u_{\text{AOL\_Eyes}}(x,t)$). These fields provide input $S_{\text{ASFA}}(x,t)$ to the DNF $u_{\text{ASFA}}(x,t)$ that contains neural populations that respond or not to the presence of the several AUs detected. The field activity $u_{\text{ASFA}}(x,t)$ provides the input to the DNF in ESL. $u_{\text{ESL}}(x,t)$, which depending on the initial active populations and other dynamic factors, such as time, produces an activation at the correspondent inferred emotional state.

The example presented in Figure 7.1, starts with a facial expression where no AUs are present. Hence from times T1 to T2, the activity in ($u_{\text{AOL\_Eyebrows}}(x,t)$), mouth ($u_{\text{AOL\_Mouth}}(x,t)$) and eyes ($u_{\text{AOL\_Eyes}}(x,t)$) code absence of AUs, while the bump of activity in field ($u_{\text{AOL\_Face}}(x,t)$) represents the presence of the human face. During this time interval only this input arrives to $u_{\text{ASFA}}(x,t)$ which produces a pattern of activation that represents solely 'face detected' and thus activity in $u_{\text{ESL}}(x,t)$ produces a bump of activity centred at the emotional state 'Neutral'.

Next, from times T2 to T3, the human raises its eyebrows (see Figure 7.1b) producing an activation in $u_{\text{AOL\_Eyebrows}}(x,t)$ representing the detection of AUs 1 and 2. As a consequence of the spread of field activation from AOL to ASFA, a bump of activity in $u_{\text{ASFA}}(x,t)$ emerges centred in the population 'Raise eyebrows', which in turn leads to a bump of activity in $u_{\text{ESL}}(x,t)$ representing an inferred emotional state of 'Surprise'.

Afterward, from times T3 to T4, the human then opens the mouth by dropping its jaw, getting coded by the vision system as AUs 1+2+26 (Figure 7.1c). This gives rises to several inputs, $S_{\text{ASFA}}(x,t)$, competing for a decision in $u_{\text{ASFA}}(x,t)$. The population representing "raise eyebrows & mouth open" wins the competition. However, the inferred emotional state, represented in $u_{\text{ESL}}(x,t)$, remains as 'Surprise'. This demonstrates the ability to detect the same emotional state in more than one way.

At last, in the time interval T4-T5, the human smiles maintaining the eyebrows raised, the resulting expression is coded with AUs 1+2+12 (Figure 7.1d). The disappearance of AU 26 and presence of AU 12 changes the competition in $u_{\text{ASFA}}(x,t)$, and ultimately, the winning population in this field then triggers in $u_{\text{ESL}}(x,t)$ a different inferred emotional, i.e. 'Happy'.

Next, the scenarios addressing several aspects of human-robot joint action, will be presented.

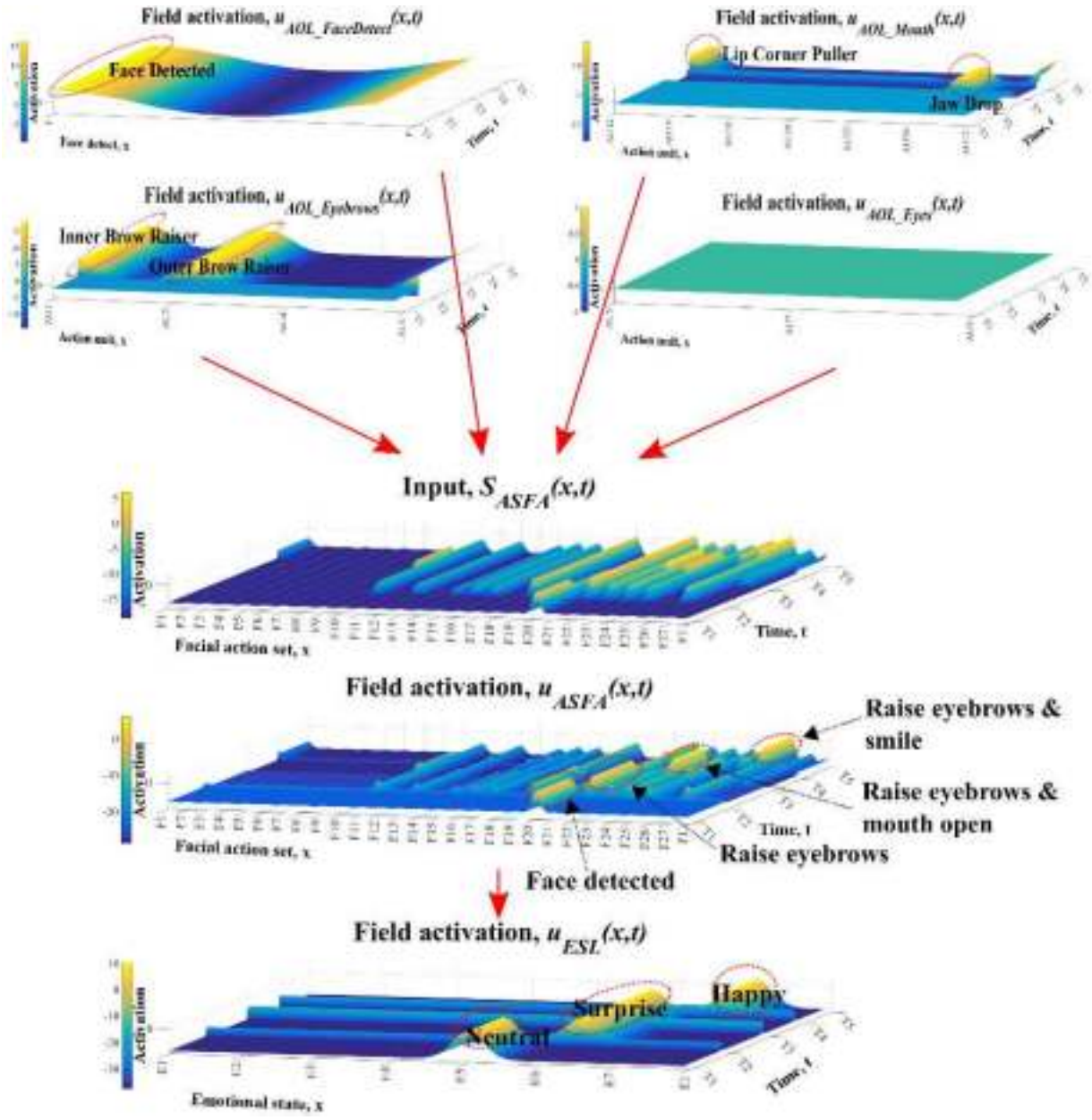


Figure 7.2: Field activities in layers AOL, ASFA and ESL, in response to the information provided by the vision system regarding the user's facial expressions depicted in Figure 7.1.

7.1 Experiment 1: Influence of the human's emotional state in the robot's decisions

Experiment 1 is composed of two scenarios, 1-1 and 1-2, and explores how the same action being performed by the human in the same context of the task, but carried out with a different emotional state, can trigger in the robot different decisions for the complementary action. Only the construction of the lower section of the task, attach the wheels and fixed them with nuts, was used.

The objects disposition for the current experiment is the following:

- Robot's workspace: 2x Nut;
- Human's workspace: 2x Wheel, Column 1, Column 2, Column 3, Column 4, Top Floor.

For Scenario 1-2, a Nut was added in the human's workspace but hidden from the robot's view. Video snapshots of the human-robot join action in Scenario 1-1 and Scenario 1-2 are shown in Figures 7.3 and 7.4 respectively.

In both scenarios the human starts with grasping a wheel (Figures 7.3a and 7.4a) and inserting it (Figures 7.3c and 7.4c). When the human grasps the wheel, the robot infers that he will insert it and decides to handover a nut to the human partner because it is the part he will need next.

The difference in the two scenarios happens here. While in Scenario 1-1 the human continues to display a neutral face (Figure 7.3f), the robot hands over the nut (Figure 7.3e), the human accepts and inserts it (Figure 7.3g). In Scenario 1-2, when the robot verbalizes its decision to handover a nut the human expresses anger (Figure 7.4f). This makes the robot understand that the human does not want the nut (Figure 7.4e), and as a consequence the robot changes its decision and asks the human to hand over a wheel (Figure 7.4g) to it, so that it can insert a wheel on its side of the construction.

In Scenario 1-1, the human working with the robot exhibits a neutral emotional state during all the interaction, and so, all the decisions made by the robot incorporate no positive nor negative emotions from its human partner. The field activity in the ESL codes the inferred

Figure 7.3: Video snapshots for scenario 1-1.

Online at: http://marl.dei.uminho.pt/public/videos/adb/Exp1-Scen1_1.html

(g) (h)

Figure 7.3: Video snapshots for scenario 1-1 (continued).
Online at: http://marl.dei.uminho.pt/public/videos/adb/Exp1-Scen1_1.html

human's emotional state. Figure 7.5a shows the field activation $u_{\mathrm{ESL}}(x, t)$ in this layer, which always has a bump of activity centred in in the same position ('Neutral') throughout the duration of the task. The change in the inferred emotional state of the human during interaction Scenario 1-2 is presented in Figure 7.5b. As can be seen, in the time interval T2-T3, a shift in the bump of activation from 'Neutral' to 'Anger' occurs.

The influence of the human emotional state in the robot's decisions regarding its complementary behaviour is clearly demonstrated by analyzing the DNF $u_{\mathrm{AEHA}}(x, t)$ in the Action Execution Layer (AEL) (Figure 7.6). This fields selects an adequate complementary goal-direct hand action. In Scenario 1-1, after the human grasped the wheel, the robot selected the action of handing over a nut (Figure 7.6a: see bump of activation coding 'Give nut'). In Scenario 1-2, the robot makes initially the same decision (Figure 7.6b: Field activation, times T1 to T2), but in response to the anger expressed by the human, the robot changes its decision to 'Request a wheel' (Figure 7.6b: Field activation, times T2 to T3). The preshaping present in Figures 7.6a and 7.6b, of the populations coding the actions 'Point to wheel' and 'Request wheel', means alternative actions the robot could in principle select.

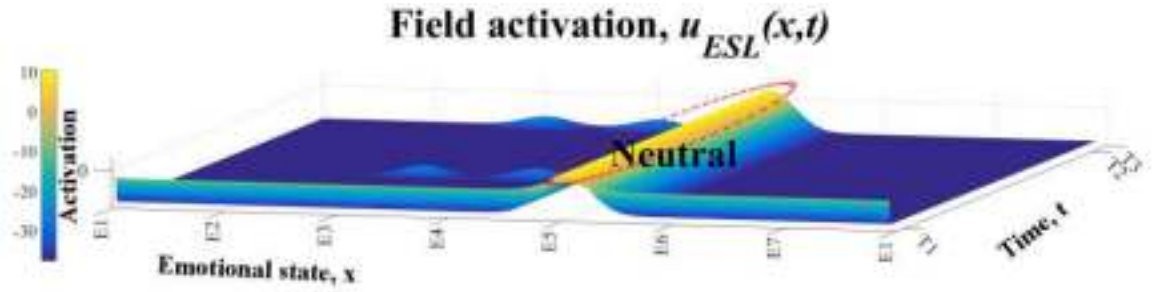
Figure 7.4: Video snapshots for scenario 1-2.

Online at: http://marl.dei.uminho.pt/public/videos/adb/Exp1-Scen1_2.html

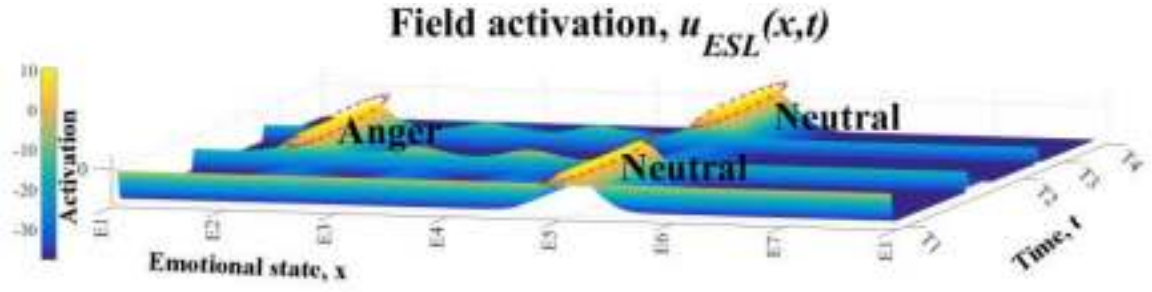
(g) (h)

Figure 7.4: Video snapshots for scenario 1-2 (continued).

Online at: `http://marl.dei.uminho.pt/public/videos/adb/Exp1-Scen1_2.html`

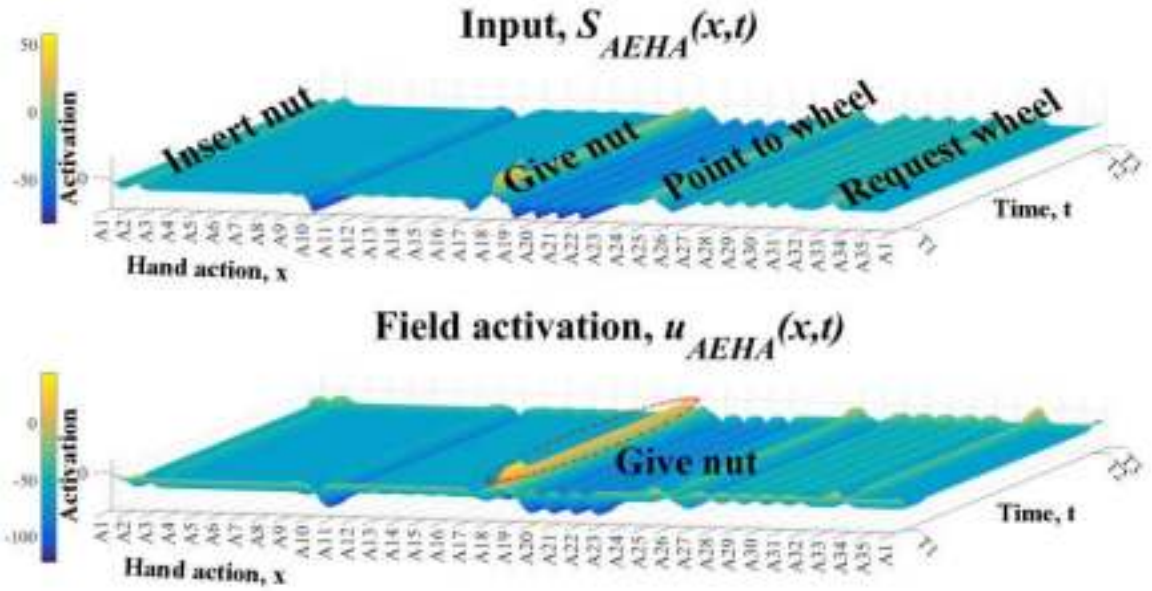


(a) Scenario 1-1: ESL.

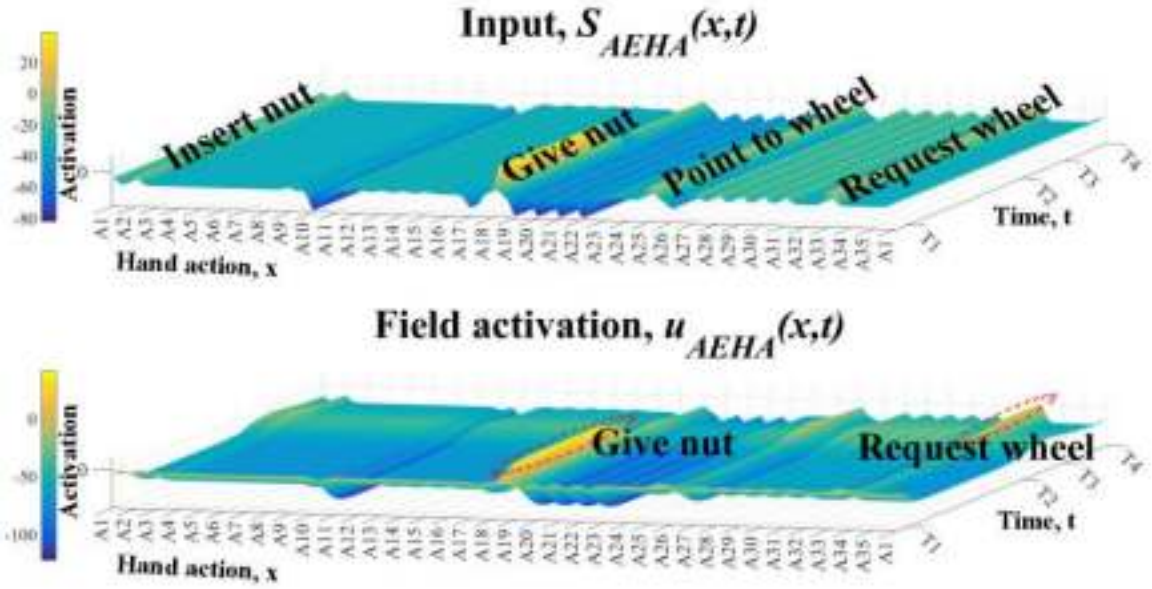


(b) Scenario 1-2: ESL.

Figure 7.5: Experiment 1: Emotional State Layer



(a) Scenario 1-1: AEHA.



(b) Scenario 1-2: AEHA.

Figure 7.6: Experiment 1: Action Execution Layer - Goal-directed hand actions.

7.2 Experiment 2: Influence of the human's emotional state in the robot's error detection and handling capabilities

Experiment 2 contains two scenarios, 2-1 and 2-2, and explores how the robot deals with errors in reaction to different inferred emotional states. While in Scenario 2-1 the human is displaying a happy expression (Figure 7.7b), in Scenario 2-2 the human has a fearful expression (Figure 7.8b). It is shown how the same error being committed during the construction task is detected in different ways, influenced by the human emotional state.

The two scenarios start with the lower section of the toy robot assembled, i.e. the Wheels and Nuts are already inserted in the Base. Thus, the next assembly steps consist of mounting the four columns. A specific serial order for plugging the columns was imposed: Column 1 $\longrightarrow$ Column 2 $\longrightarrow$ Column 3 $\longrightarrow$ Column 4. The different columns are identified by their color patterns. Given the reachable workspace of the two agents, it happens that Column 1 and Column 4 can only be mounted by the robot, while Column 2 and Column 3 can only be mounted by the human partner.

The objects disposition is:

- Robot's workspace: Wheel (inserted), Nut (inserted), Column 4;
- Human's workspace: Wheel (inserted), Nut (inserted), Column 1, Column 2, Column 3, Top Floor.

Both scenarios start in the same way, with the robot requesting the human to handover Column 1(See Figures 7.7a and 7.8a). However, the human ignores the robot's request and instead grasps Column 3 with the intention to insert it (Figures 7.7c and 7.8c). This is an error because Column 3 cannot yet be mounted.

When the human operator is in a positive emotional state the (expected) probability that he will commit errors is low because this signals that he is engaged in the joint task. In Scenario 2-1, the fact that the human is displaying since the beginning a happy facial expression, has made the robot to disable the processing of the DNFs in the Error Monitoring Layer (EML)

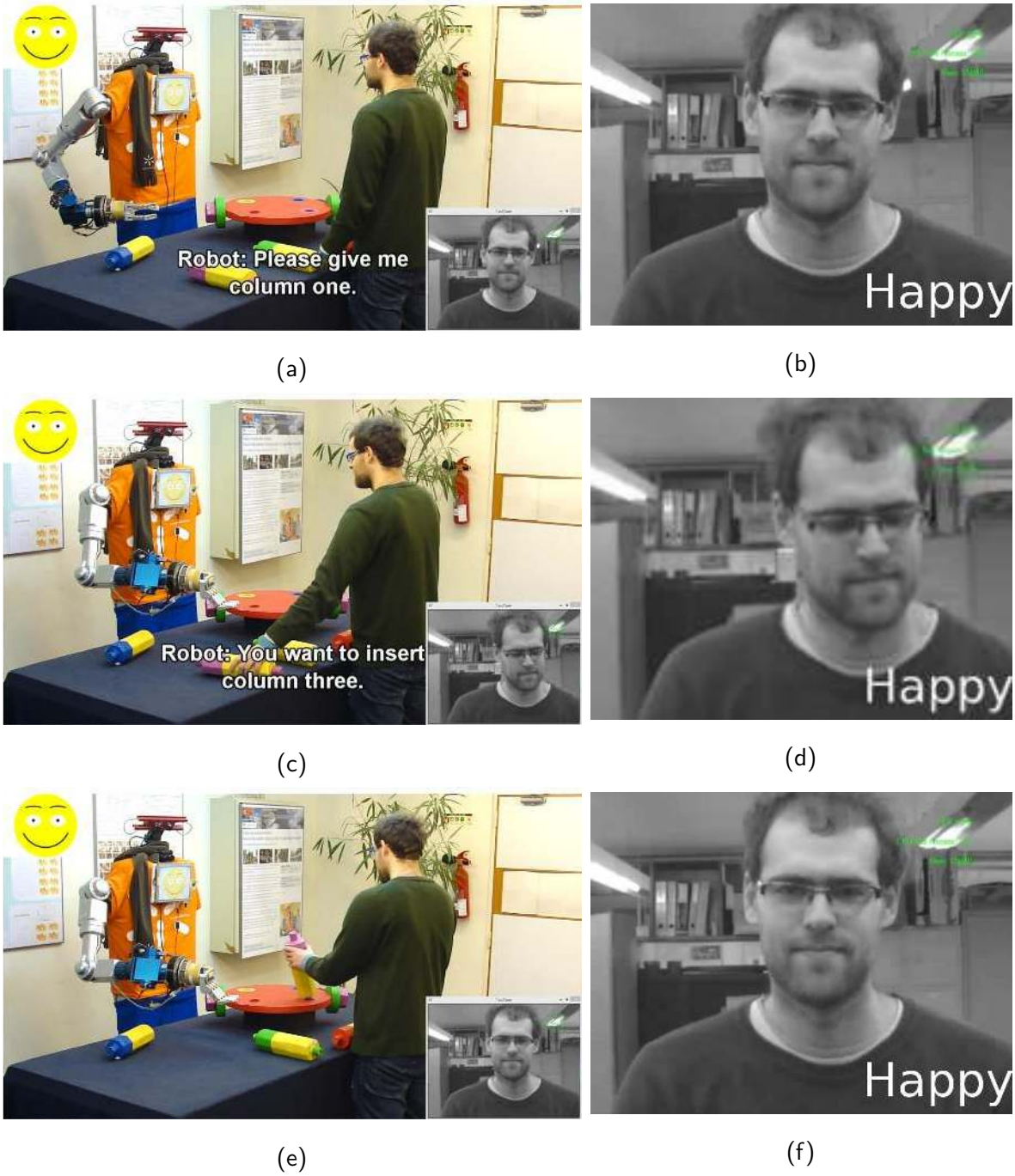


Figure 7.7: Video snapshots for scenario 2-1.

Online at: http://marl.dei.uminho.pt/public/videos/adb/Exp2-Scen2_1.html

(g) (h)

Figure 7.7: Video snapshots for scenario 2-1 (continued).
Online at: http://marl.dei.uminho.pt/public/videos/adb/Exp2-Scen2_1.html

responsible for detecting user's errors in intention and errors in the means. Thus, although the robot is able to infer, at the moment of grasping, that the intention of the human is to insert Column 1, it is not able to predict that the user's intention/goal is wrong. The human advances and inserts Column 3 (Figure 7.7e). The robot detects that this was error only after the column was plugged (error in execution) and orders the human to correct the error he has made (Figure 7.7g).

In scenario 2-2, the human is in a negative emotional state, this causes the robot to enable the processing of all the error detection components in EML enabled. As a consequence, as soon as the human grasps Column 3 to insert, the robot interprets this as an error in intention and prevents the error from occurring (Figure 7.8e).

The main difference in Scenarios 2-1 and 2-2 is due to the expressed emotional state by the human, whose inferred state by the robot is coded in activation of the DNF $u_{ESL}(x, t)$ in ESL. Figure 7.9a shows a bump of activation representing 'Happy' throughout the duration of Scenario 2-1, while Figure 7.9b shows a bump of activation in a different location representing 'Fear' during scenario 2-2.

The influence of the emotional state in the robot's error detection capabilities can be observed through in Error Monitoring Layer (EML) (Figure 7.10). While in Scenario 2-1 the

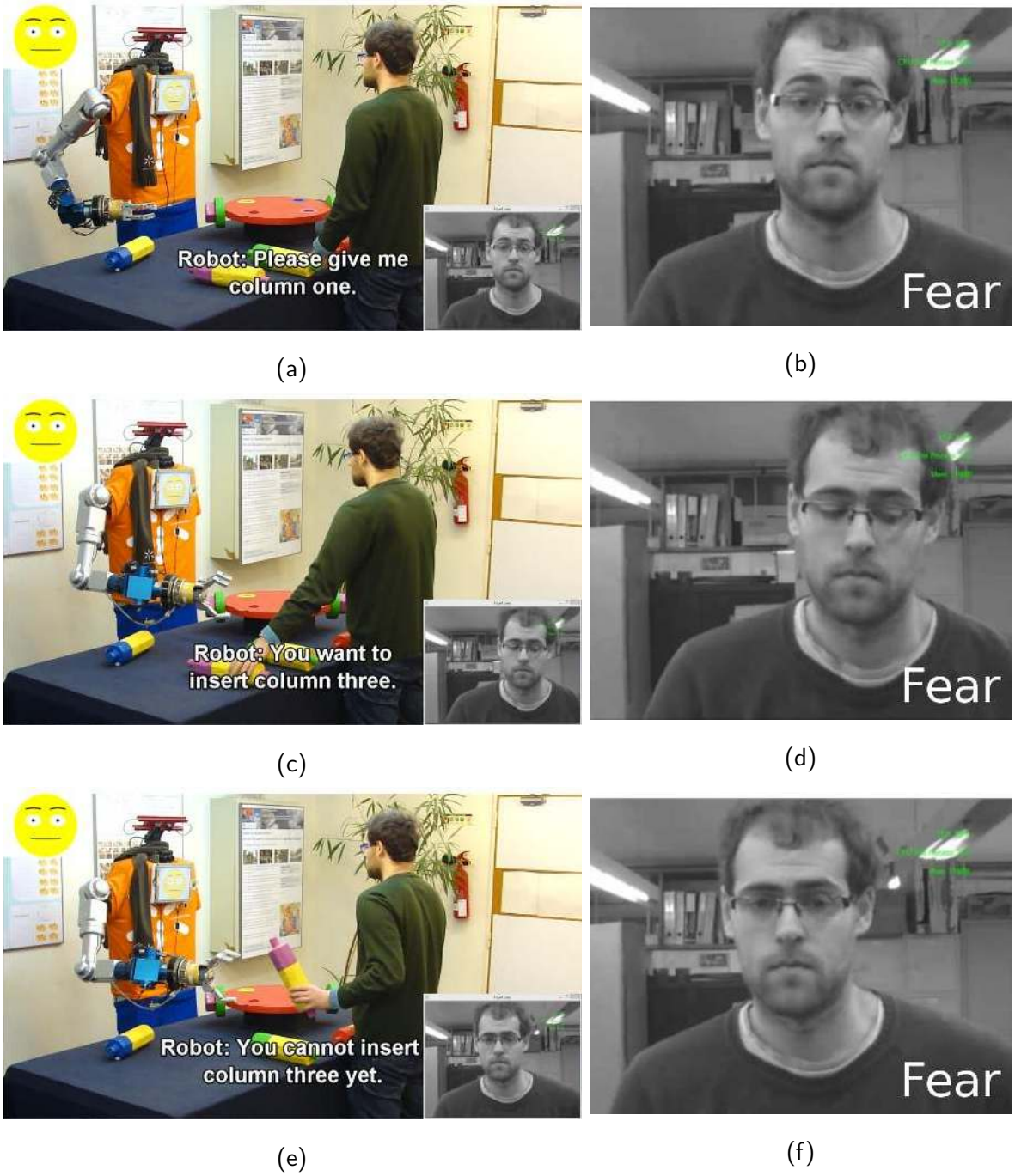


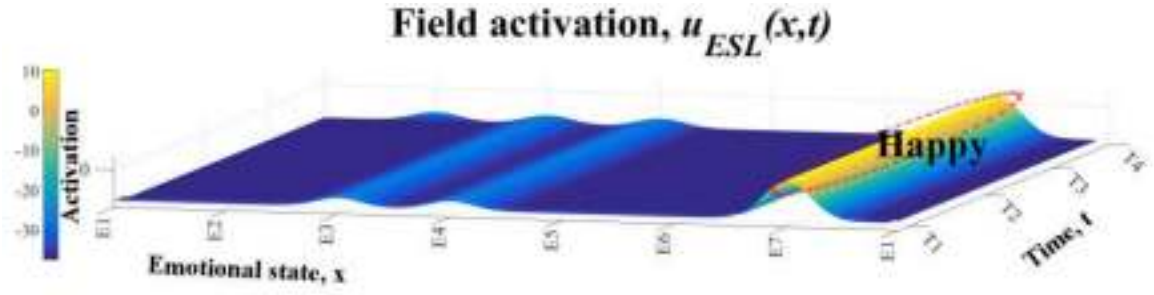
Figure 7.8: Video snapshots for scenario 2-2.

Online at: http://marl.dei.uminho.pt/public/videos/adb/Exp2-Scen2_2.html

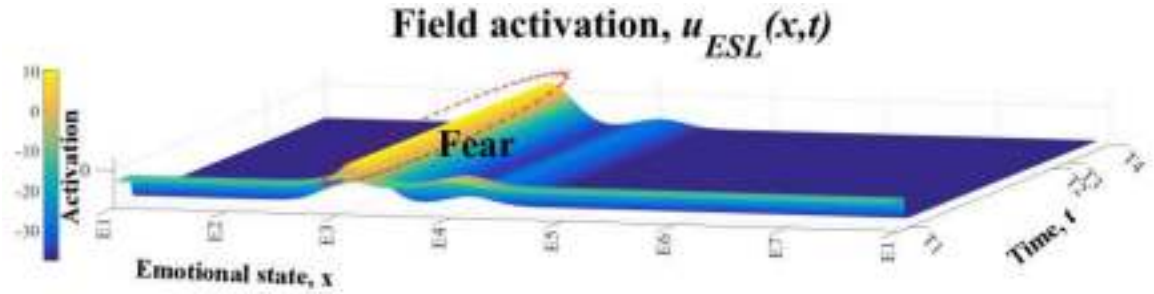
(g) (h)

Figure 7.8: Video snapshots for scenario 2-2 (continued).

Online at: http://marl.dei.uminho.pt/public/videos/adb/Exp2-Scen2_2.html

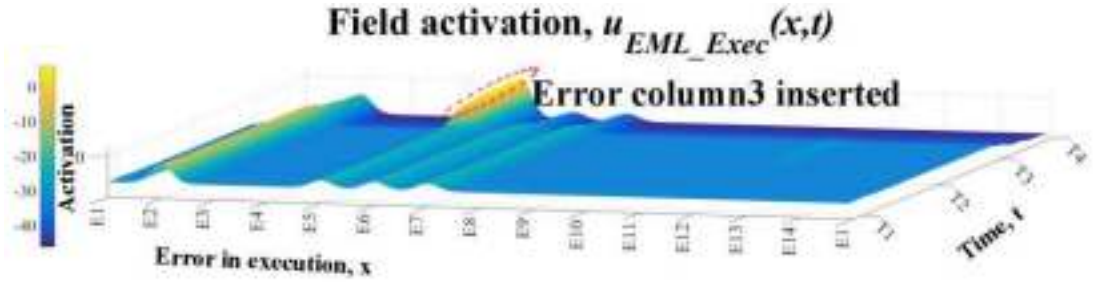


(a) Scenario 2-1: ESL.

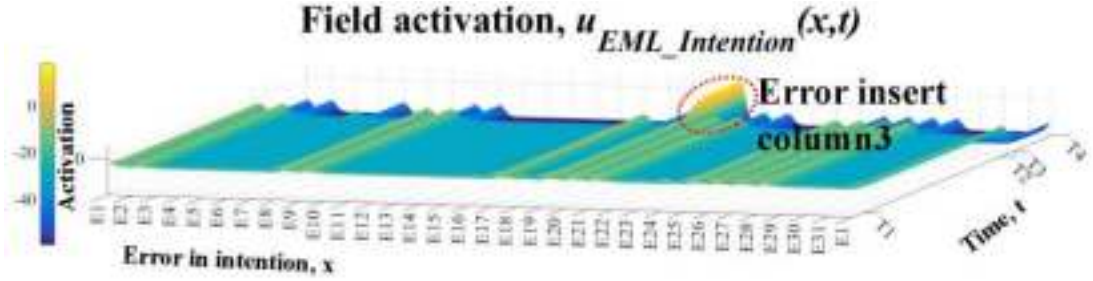


(b) Scenario 2-2: ESL.

Figure 7.9: Experiment 2: Emotional State Layer.



(a) Scenario 2-1: EML - Error in Execution.



(b) Scenario 2-2: EML - Error in Intention.

Figure 7.10: Experiment 2: Error Monitoring Layer.

robot detected the error "Insert Column 3" as an Execution Error (Figure 7.10a) in Scenario 2-2, the same error was anticipated and detected as an Error in Intention (Figure 7.10b).

The fact that the human was in a happy emotional state prevented the robot to anticipate the error. When the human displays a happy emotional state the robot assumes the construction is going well and disables the detection of errors in intention and errors in means, this way it can accelerate the processing and make decisions faster, with the downside of the robot being unable to anticipate errors the human can commit. However if an error is actually performed, the robot will be able to detect it and issue a warning or corrective order to this fact.

7.3 Experiment 3: Reaction of the robot to the human's persistence in error

In the interaction Scenarios 2-1 and 2-2 described in the previous section, the human partner has accepted the warnings and corrective orders issued by the robot. The robot has never displayed a negative emotional state toward the human partner.

Experiment 3 will explore how the robot, by producing expressive faces when required, can react to a stubborn human, and thus induce a change of his behaviour/attitude (see video snapshots in Figure 7.11).

The situation is the same as the previous Scenario 2-2, but this time the negative emotional state displayed by the human operator is 'Anger'. All DNFs in EML are therefore activated (their activation can be seen in Figure 7.12).

The robot starts of by requesting Column 1 to the human (Figure 7.11a). However, the human grasps Column 3 (Figure 7.11c) and the robot infers that the he will insert that column (see activation in $u_{\text{ASHA}}(x, t)$ in times T2-T3, Figure 7.13). As before, the robot detects that the human's goal to plug Column 4 is wrong (see activation in $u_{\text{EML\_Intention}}(x, t)$ in times T2-T3, Figure 7.12a), and warns that he will commit an error (Figure 7.11e).

Despite the warning, the human proceeds to insert Column 3 (Figure 7.11g), and as a consequence the robot now detects it as an execution error and issues a corrective action (see activation in $u_{\text{EML\_Exec}}(x, t)$ in times T3-T4, Figure 7.12b). Ignoring the robot, the human persist in the error. In response to this persistence and because the user is in an Angry state (see action in $u_{\text{ESL}}(x, t)$ times T1-T5, Figure 7.15), the robot takes a stand by expressing (also) an anger face (see activation u_{AEFA} in times T5-T6, Figure 7.14) and explaining again that an error was committed (Figures 7.11i and 7.11k).

Thus far, the robot had never displayed a negative emotion toward the human partner. Thus he gets surprised (Figure 7.11n) by the robot's anger. See activation in $u_{\text{ESL}}(x, t)$ at time T6 (Figure 7.15). The human finally accepts the robot's correction and removes the inserted column from the Base (Figure 7.11m). The robot then takes a neutral expression (Figure 7.14, times

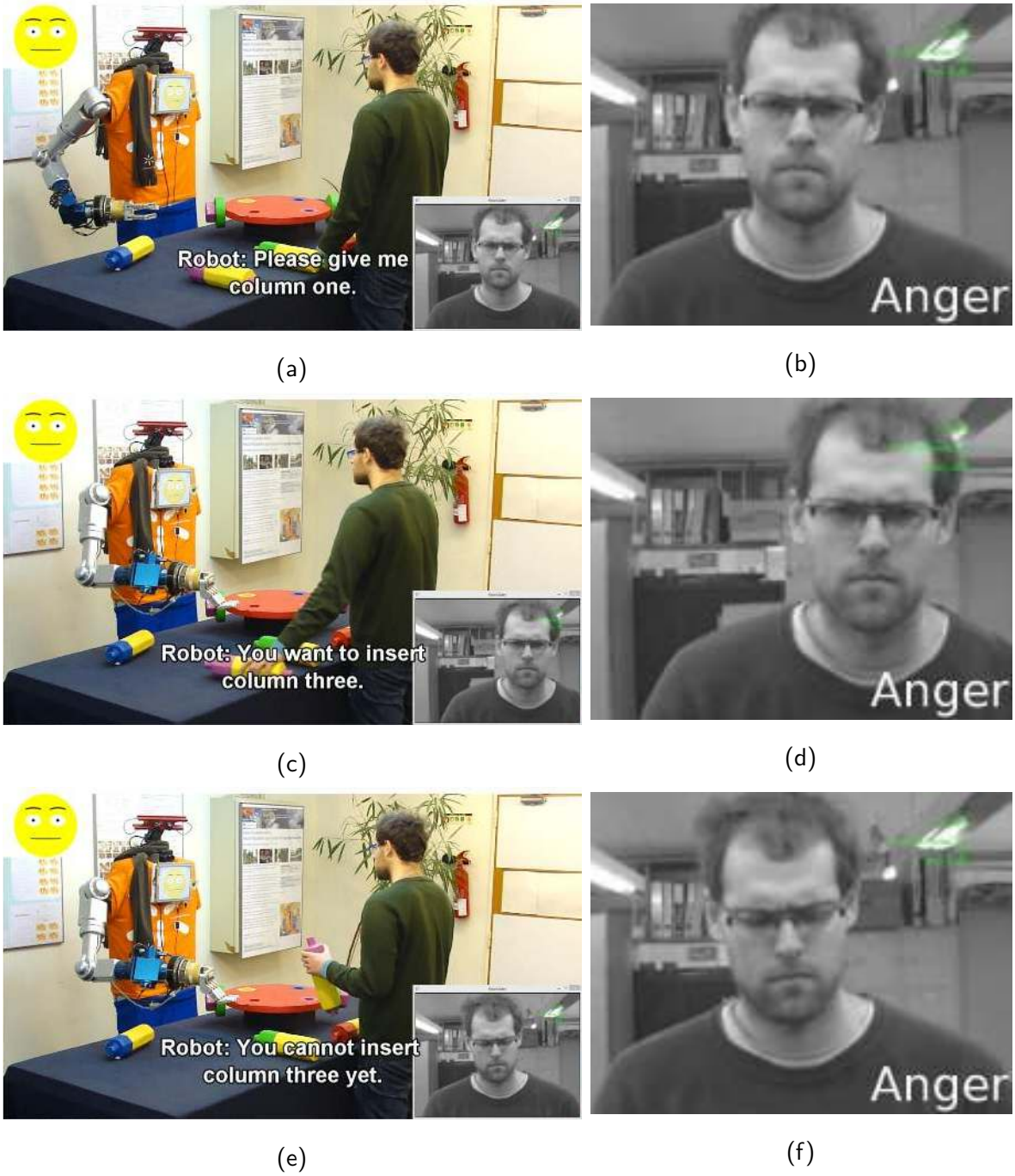


Figure 7.11: Video snapshots for experiment 3.

Online at: http://marl.dei.uminho.pt/public/videos/adb/Exp3.html

(g) (h)

(i) (j)

(k) (l)

Figure 7.11: Video snapshots for experiment 3 (continued).

Online at: `http://marl.dei.uminho.pt/public/videos/adb/Exp3.html`

(m) (n)

(o) (p)

(q) (r)

Figure 7.11: Video snapshots for experiment 3 (continued).

Online at: `http://marl.dei.uminho.pt/public/videos/adb/Exp3.html`

(s) (t)

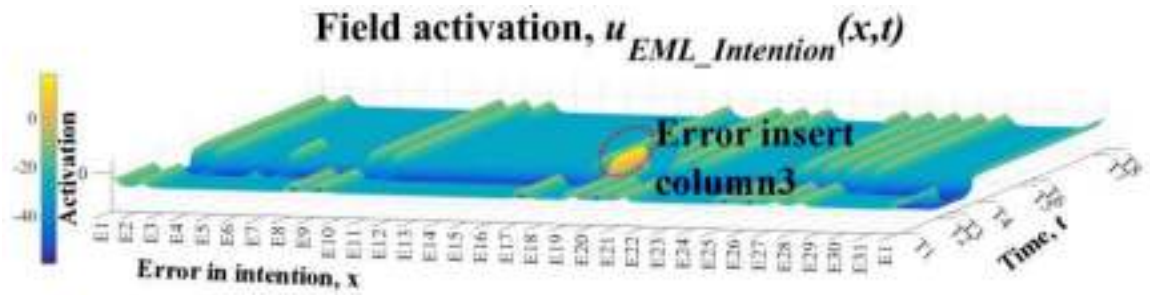
Figure 7.11: Video snapshots for experiment 3 (continued).
Online at: http://marl.dei.uminho.pt/public/videos/adb/Exp3.html

T7) and request again Column 1 to insert on its side (Figure 7.11o). But because the human expresses surprise in response to the robot's request, the decision of the robot changes from preparing to receive Column 1 to pointing toward to it (Figure7.11q). This gesture drives the attention of the human operator to the requested column. The human finally grasps and hands over Column 1 to the robot (Figure 7.11s), and the decision of the robot is to receive it. The temporal evolution of these changes in the selected goal-directed hand gestures of the robot can be seen in the activation of $u_{\mathrm{AEHA}}(x, t)$, times T6-T8, Figure 7.16.

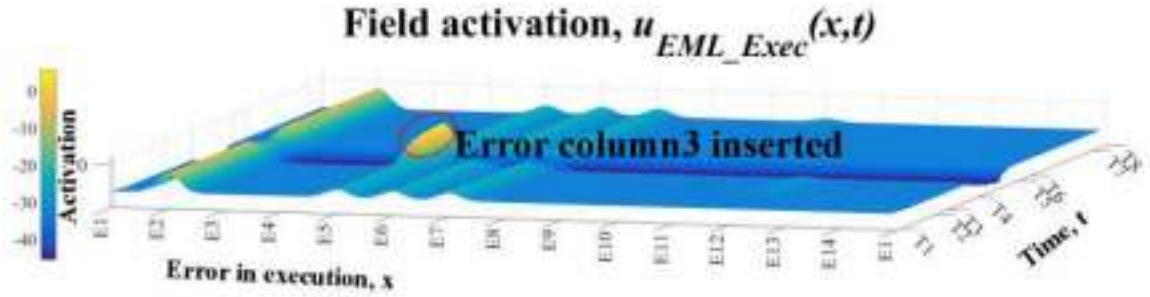
7.4 Experiment 4: Influence of the human's emotional state in task time

Experiment 4 explores how the human's emotional state might influence the time that it takes to complete the task. The construction of the lower section of the toy vehicle, was used as a test scenario.

Three scenarios were designed, in each scenario the human kept the expression of the same emotional state throughout the duration of the task. In the first scenario the human expressed a negative emotional state (Fear), in the second the human was in a neutral state, and in the



(a) Experiment 3: EML - Error in Intention.



(b) Experiment 3: EML - Error in Execution.

Figure 7.12: Experiment 3: Error Monitoring Layer.

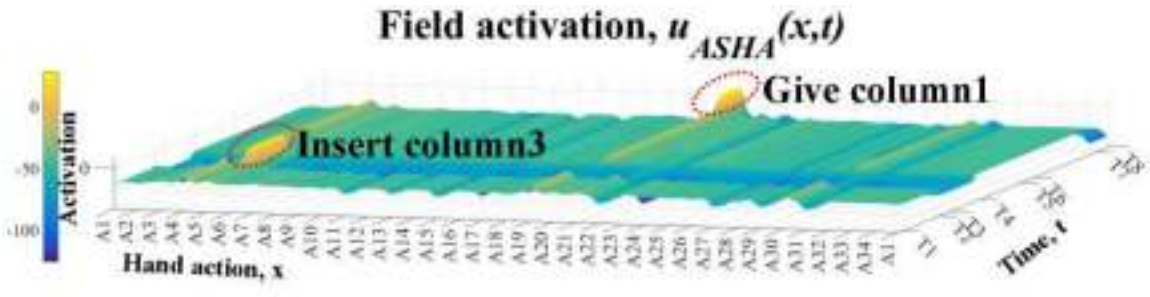


Figure 7.13: Experiment 3: Action Simulation Layer - Simulation of goal-directed hand actions.

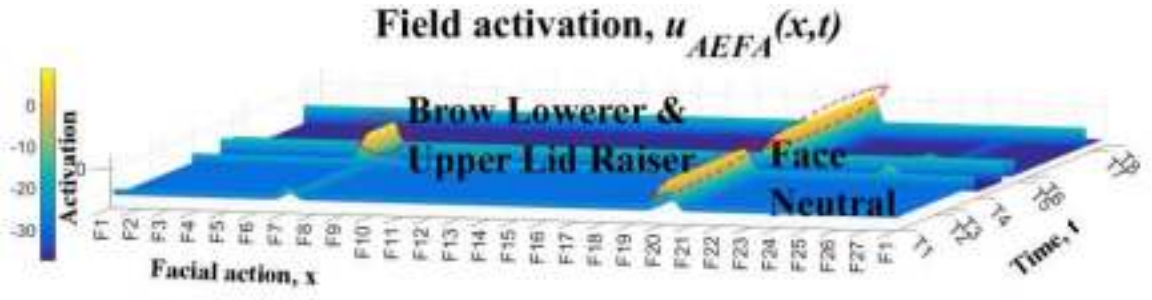


Figure 7.14: Experiment 3: Action Execution Layer - Facial actions sets execution.

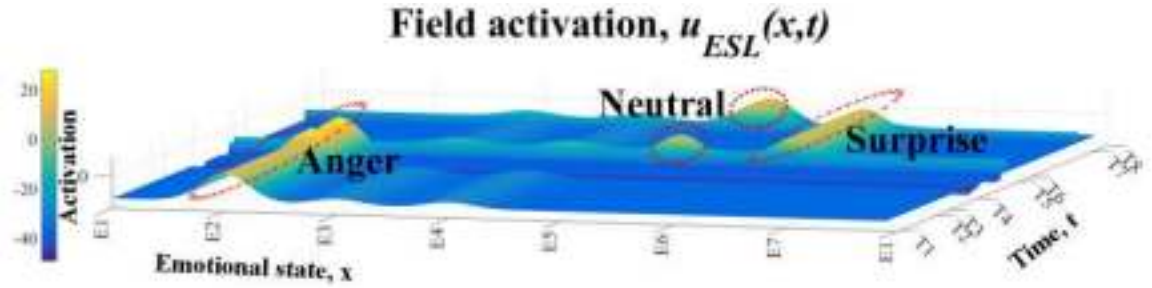


Figure 7.15: Experiment 3: Emotional State Layer.

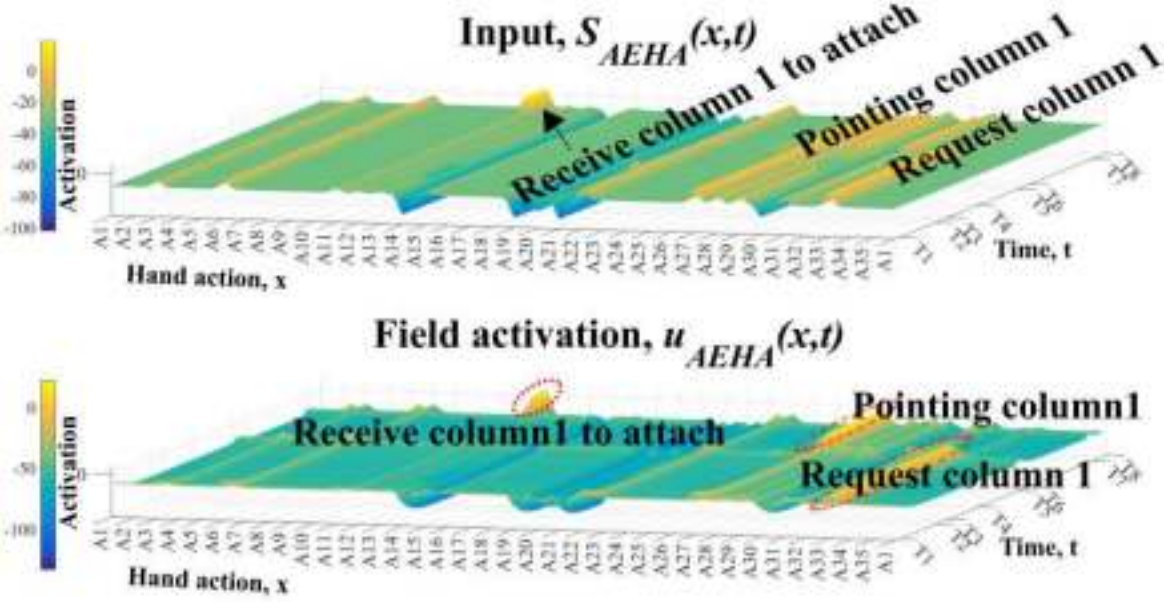


Figure 7.16: Experiment 3: Action Execution Layer - Goal-directed hand actions.

third the human displayed a positive emotional state (Happy). In all scenarios, the distribution of the objects in the robot's and human's workspace was the same.

Scenario	Emotional state	Time
4-1	Fear	2 min. 55 sec.
4-2	Neutral	2 min. 30 sec.
4-3	Happy	1 min. 50 sec.

Table 7.1: Experiment 4: Time to complete the task as a function of the human emotional state.

Videos online at:

4-1: http://marl.dei.uminho.pt/public/videos/adb/Exp4-Scen4_1.html

4-2: http://marl.dei.uminho.pt/public/videos/adb/Exp4-Scen4_2.html

4-3: http://marl.dei.uminho.pt/public/videos/adb/Exp4-Scen4_3.html

Table 7.1 shows the results of the three interaction scenarios. When the human is in a fearful state, the robot adjusts the arm movements to be slower and takes more time explaining its actions in order not to startle the human. In a neutral state, the robot uses a medium velocity for the arm movements. When the human displays a happy emotional state, the robot assumes the task is running smoothly, increases the velocity for the arm movements, disables the processing of DNFs responsible for the detection of some types of errors, decreasing the time it takes to make decisions.

What the results show in this particular experiment is, the negative expressions impact in the task time by increasing it when comparing to a neutral emotional state, 16% in this case. And when in a positive emotional state, the task time is reduced in 27% when comparing to neutral, but due to disabling the detection of some types of errors, its more prone for errors to occur, since the robot cannot anticipate them.

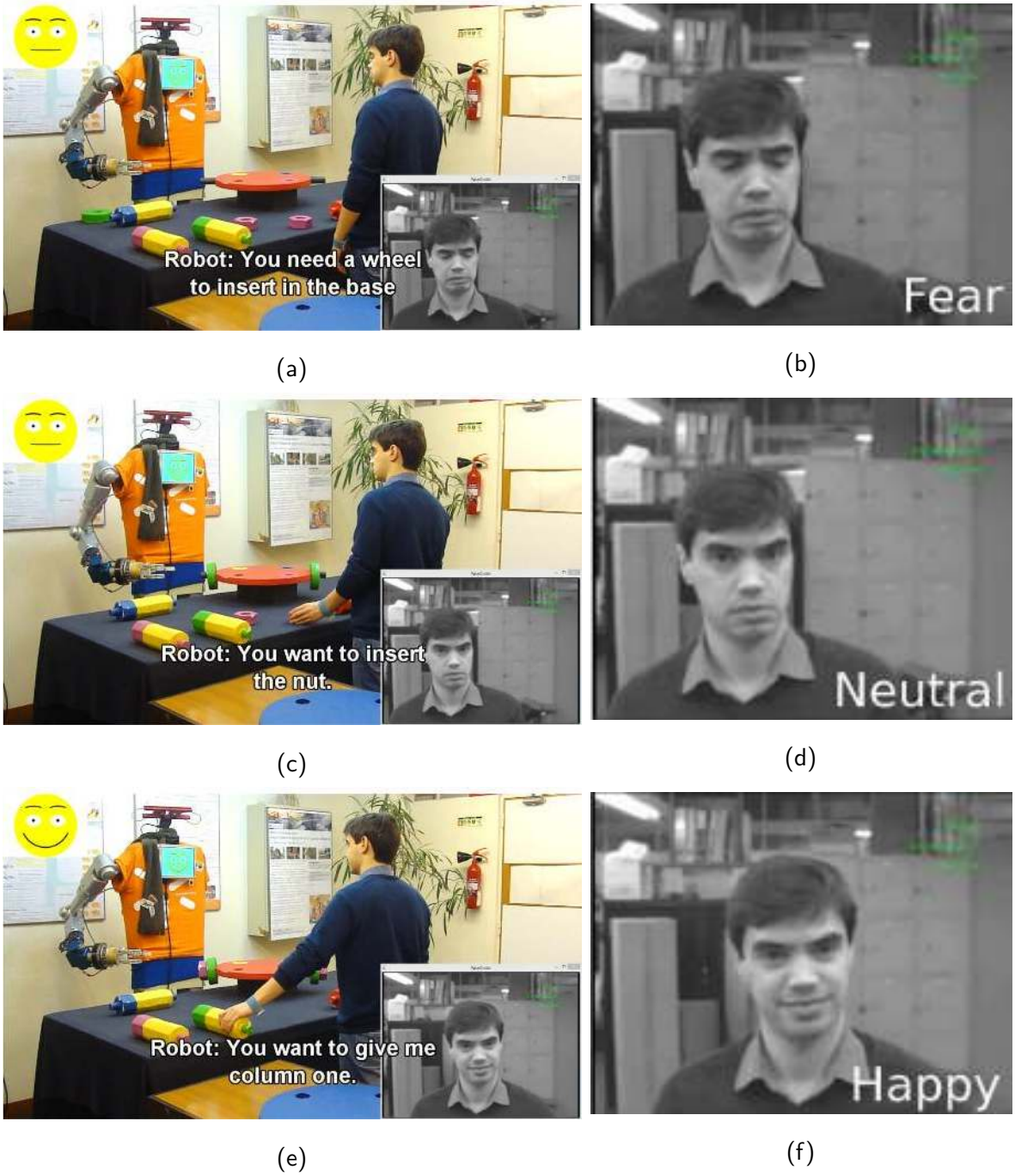
7.5 Experiment 5: A longer interaction scenario - dynamically adjusting behaviour to the expressed human emotional state.

As a final interaction scenario, the entire construction task, was performed, where the human cooperating with the robot shifts the expressed emotional state from negative (Fear) to neutral and then positive (Happy).

The task starts with the human presenting a fearful expression (see Figure 7.17b). The robot adjusts its arm movement velocity to be slower in order not to startle the human, also it takes more time explaining its actions (see Figure 7.17a).

After the wheels are inserted the human presents a neutral expression during the insertion of the nuts (see Figure 7.17d). The robot adjusts the movement velocity to medium and verbalizes less information.

When the middle section is assembled, the human is expressing happiness (see Figure 7.17f), so the robot also smiles and increases the movement velocity for the arm. Here one can see how the robot dynamically and in real time adjusts its behaviour – information verbalization and movement velocity – during the execution of the task.



(a) (b)

(c) (d)

(e) (f)

Figure 7.17: Video snapshots for experiment 5.

Online at: http://marl.dei.uminho.pt/public/videos/adb/Exp5.html

This page was intentionally left blank.

Chapter 8

Discussion, conclusion and future work

Decision making refers to the process of selecting a particular action from a set of alternatives. When acting alone, an individual may choose a motor behaviour that best serves a certain task based on the integration of sensory evidence and prior task knowledge. In a social context, this process is more complex since the outcome of one's decisions and emotions can be influenced by the decisions and emotions of others. A fundamental building block of social interaction is thus the capacity to predict and understand actions and emotional states of others. This allows an individual to select and prepare an appropriate motor behaviour in joint action tasks (Michael, 2011; Sebanz et al., 2006).

Here, a Dynamic Neural Field (DNF) architecture that combines the role of emotions in the decision making and movement execution of an autonomous and socially aware robot cooperating with human partners in real-world joint tasks, was presented. The proposed architecture is strongly inspired by converging evidence from cognitive and neurophysiological studies suggesting that mirror neurons encoding different levels of abstraction coexist and that there is an automatic but highly context-sensitive mapping from observed on to-be-executed actions as an underlying mechanism (Bekkering et al., 2009; Rizzolatti and Sinigaglia, 2008).

Dynamic neural fields model the emergence of persistent neural activation patterns that allow a cognitive agent to initiate and organize behaviour informed by past sensory experience, anticipated future environmental inputs and distal behavioural goals. The DNF-architecture

for joint action reflects the notion that cognitive representations, i.e. all items of memory and knowledge, consist of distributed, interactive, and overlapping networks of cortical populations ("cognit" from Fuster, 2006). Network neurons showing suprathreshold activity are participating in the selection of actions, emotional states and their associated consequences. Since the decision-making normally involves multiple, distributed representations of potential actions that compete for expression in overt performance, the robot's goal-directed behaviour is continuously updated for the current environmental and social context. Important for decision making in a collaborative setting, inferring others' goals and emotional states from their behaviour is realized by internal motor simulation based on the activation of the same joint representations of (hand and facial) actions and their environmental effects ("mirror mechanism" Rizzolatti and Sinigaglia, 2008); (for a recent review see Rizzolatti et al., 2014). Through this automatic motor resonance process, the observer becomes aligned with the co-actor in terms of actions, emotional states and goals. This alignment allows the robot to adapt dynamically its behaviour to that of the human co-actor, without explicit communication (for an integration of verbal communication in the DNF-architecture see Bicho et al., 2010).

The implementation of aspects of real-time social cognition in a robot based on continuously changing patterns of neuronal activity in a distributed, interactive network strongly contrasts with traditional views of human-like (social) intelligence. These realize the underlying cognitive processes as a manipulation (based on formal logic and formal linguistic systems) of discrete symbols that are qualitatively distinct and entirely separated from sensory and motor information. These approaches have provided many impressive examples of intelligent behaviour in artificial agents (for review see Vernon et al., 2007), in fact, the sequence of decisions shown in the presented robotics experiments could be also be implemented by symbolic planning. However, it is now widely recognized by the robotics and cognitive science communities that the symbolic framework based on has notorious problems to cope with real-time interactions in dynamic environments (Haazebroek et al., 2011; Kozma, 2008; Levesque and Lakemeyer, 2008). In human-robot joint tasks, the robot has to reason about a world that may change at any instance of time due to actions taken by the user. Even if the processing in the perceptual and decision modules would allow to continuously update the robot's plan in accordance with the user's

intention and emotional state, is considered, the extra processing step needed to embody the abstract action plan in the autonomous robot would challenge the fluent and seemingly effortless coordination of decisions and actions that characterizes human joint action in familiar tasks.

Bayesian models represent a quite popular alternative approach for modelling decision and integration processes in the face of uncertainty (Körding and Wolpert, 2006). It is important to note that the dynamic field framework is compatible with central aspects of probabilistic models. For instance, the pre-activation below threshold of several populations in the action execution layer due to prior task knowledge and contextual information may be interpreted in the sense of a probability density function for different complementary actions. This prior information has to be combined with evidence about the inferred goal and emotional state of the co-actor. In fact, it can be shown that in the input-driven regime the field dynamics may implement Bayes' rules (Cuijpers and Erlhagen, 2008). There are two major advantages of the dynamic neural field approach. First, stabilizing decision against noise, fluctuations and temporary absence of information in the input stream, is of particular importance. Second, as an example of the dynamical approach to cognition (Schöner, 2008), a DNF-based model allows us to address the important temporal dimension of coordination in joint action (Sebanz et al., 2006). The decision process linked to complementary actions unfolds over time under multiple influences which are themselves modelled as dynamic representations with proper time scales.

The DNF-architecture was tested in real-time human-robot joint action experiments in the context of a construction task.

In Experiment 1, it was demonstrated how the emotional state of the human partner can affect the decisions made by the robot. Specifically, it was shown that in the same context, a different emotional state displayed by the human can trigger a different complementary behaviour on the robot.

In Experiment 2, it was explored how the perceived emotions may play a role in the way the robot detects and handles different types of errors. When the human co-worker is in a positive emotional state, this is taken as a signal that the human is engaged in the task, and thus, it is not probably that he/she will commit errors. The load of the Error Monitoring processes can be decreased by deactivating the anticipation of errors in intention and errors in the action

means. The result is that the robot can make decisions faster. In case, the human co-worker makes an error this is detected *a posteriori* as an execution error. Reversely, when the human is in a negative emotional state (e.g. Anger) this is used as a signal that the human user is not committed to the task, and thus it is probably that he/she is more prone to make errors. All error monitoring processes are activated and this enables the robot to prevent the occurrence of errors by anticipating errors at the goal/intention level.

In Experiment 3 it was demonstrated how the robot can deal with a human operator persisting in making an error. It was shown that by expressing emotional states and verbalization of more information, the robot can induce the (stubborn) human to change his attitude and accept the robot's corrective suggestions.

The above summarized experiments have shown that perceived emotions play an important role in an early stage, during decision making and action preparation of a complementary action (Action Execution Layer (AEL)). In Experiment 4 it was shown that perceived emotions also play a role latter because they may affect the execution at the kinematics level (Motor control). In this experiment, three persons expressing different emotional states (Neutral, Fear, Happy) worked with the robot. When the human co-worker seemed to be in a fearful state, the robot adjusted the arm-hand movements to be slower and took more time verbalizing its reasoning in order not to startle the human. Reversely, when the human displayed a positive emotional state, the robot adjusted the arm-hand movements, and verbalization, to be faster. In a neutral state, the robot used a medium velocity for the arm-hand movements and verbalization. The over all result was that the time to complete the task decreases when the human partner is in a positive emotional state. However, to perform a more in depth study on this matter, a bigger study with more participants is required to make it possible to present statistically relevant results.

Finally, Experiment 5 has shown a longer interaction scenario – the complete construction of the toy vehicle – with the human shifting his emotional state and the way the robot adapted in real time its behaviour to these changes.

As it was shown, the adopted dynamic perspective offers in general a high degree of flexibility in joint task execution. However, in the present implementation of the DNF-architecture the neural representations and their connectivity were tailored by the designer. It is highly desirable

to endow the robot with a developmental program that would allow it to autonomously learn and represent new representations (Asada et al., 2009; Weng, 2004). Using correlation-based learning rules (Gerstner and Kistler, 2002) with a gating that signals the success of behaviour, it was shown for instance how goal-directed mappings between action observation and action execution that support an action understanding capacity may develop during learning and practice (Erlhagen et al., 2006a,b). Importantly, the developmental process, through Hebbian learning rules, may explain the emergence of new task-specific populations which have not been introduced to the architecture by the human designer (Erlhagen et al., 2007). Recently, it was demonstrated how the robot may autonomously develop – through tutor demonstration and feedback during joint performance – the connections between the populations in the two layers of the CSGL that code the possible serial orders and the longer term dependencies between subgoals. The work on learning and development in the DNF-architecture for joint action is consistent with the work of Keysers and Gazzola (2014) who have analysed how mirror neurons could develop and become a dynamic system that performs active inferences about the actions, sensations and emotions of others and allows joint actions despite sensory motor delays.

Various works have explored automatic facial expression recognition in human-computer interaction (see Pantic and Bartlett, 2007; Tian et al., 2005). However, a human-robot scenario presents additional challenges: lack of control over lighting conditions, relative poses, the inherent mobility of the robot and separation between robot and human. These are limitations imposed to the used robot that are also present in other works (e.g. Wimmer et al., 2008). The vision system limitations prevented us from performing experiments with a larger numbers of human subjects. Since it relies on the acquisition of a neutral face of the subject to perform the Action Units (AUs) coding, which might not be possible at all times. Also, the features extraction is not robust enough to detect subtle and micro expressions, which in more naturalistic scenarios would be the most common expressions. Tests conducted to the system by using the Cohn-Kanade face database (Kanade et al., 2000) reveal detection rates for some AUs above 70% (4, 12 15), others have detection rates just above 50% (1, 2, 5, 26) (Cunhal, 2014). This lead us to instruct the participants in our studies to perform posed expressions to improve the system detection rate.

Regardless of the sensory limitations, the DNF-architecture proved to be ready to cope with the demands of truly real world human-robot joint action scenarios. When dealing with multiple information sources, which in the real world might not be reliable or consistent, the DNF based cognitive architecture is able to cope with these situations, even when the information is not all available at the same time. Being able to synthesize in an embodied artificial agent the cognitive demands of real-time interactions with a human co-actor whose displayed emotional states modulate the robot's behaviour shows that the dynamic neural field theory provides a promising research program for bridging the gap that still exists in natural and (socially) intelligent human-robot joint action.

In the future, further user studies need to be conducted to assess how the robot can be more expressive, and also how the subject of face recognition can be explored to allow the robot to customize the interaction based on the person that is interacting with it.

Bibliographic references

J. Adelhardt, R. P. Shi, C. Frank, V. Zeißler, A. Batliner, E. Nöth, and H. Niemann, "Multimodal User State Recognition in a Modern Dialogue System," in *KI 2003: Advances in Artificial Intelligence*, ser. Lecture Notes in Computer Science, A. Günter, R. Kruse, and B. Neumann, Eds. Springer, 2003, vol. 2821, pp. 591–605. doi: 10.1007/978-3-540-39451-8_43

R. Adolphs, "How do we know the minds of others? Domain-specificity, simulation, and enactive social cognition," *Brain research*, vol. 1079, no. 1, pp. 25–35, mar 2006. doi: 10.1016/j.brainres.2005.12.127

S.-i. Amari, "Dynamics of pattern formation in lateral-inhibition type neural fields," *Biological Cybernetics*, vol. 27, no. 2, pp. 77–87, 1977. doi: 10.1007/BF00337259

N. Ambady and R. Rosenthal, "Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis." *Psychological Bulletin*, vol. 111, no. 2, pp. 256–274, 1992. doi: 10.1037/0033-2909.111.2.256

M. Asada, K. Hosoda, Y. Kuniyoshi, H. Ishiguro, T. Inui, Y. Yoshikawa, M. Ogino, and C. Yoshida, "Cognitive developmental robotics: a survey," *IEEE Transactions on Autonomous Mental Development*, vol. 1, no. 1, pp. 12–34, 2009.

S. Asteriadis, P. Tzouveli, K. C. Karpouzis, and S. D. Kollias, "Estimation of behavioral user state based on eye gaze and head pose-application in an e-learning environment," *Multimedia Tools and Applications*, vol. 41, no. 3, pp. 469–493, 2008. doi: 10.1007/s11042-008-0240-1

A. Austermann, N. Esau, L. Kleinjohann, and B. Kleinjohann, "Prosody based emotion recognition

for MEXI," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2005, pp. 1138–1144. doi: 10.1109/IROS.2005.1545341

J. A. Bargh, M. Chen, and L. Burrows, "Automaticity of social behavior: direct effects of trait construct and stereotype-activation on action," *Journal of personality and social psychology*, vol. 71, no. 2, pp. 230–44, aug 1996.

L. F. Barrett, "Discrete Emotions or Dimensions? The Role of Valence Focus and Arousal Focus," *Cognition and Emotion*, vol. 12, no. 4, pp. 579–599, 1998.

L. F. Barrett, B. Mesquita, K. N. Ochsner, and J. J. Gross, "The experience of emotion," *Annual review of psychology*, vol. 58, pp. 373–403, jan 2007. doi: 10.1146/annurev.psych.58.110405.085709

A. Beck, L. Cañamero, and K. A. Bard, "Towards an affect space for robots to display emotional body language," in *19th IEEE International Symposium on Robot and Human Interactive Communication*, Principe di Piemonte - Viareggio, Italy, 2010.

H. Bekkering, E. R. A. de Bruijn, R. H. Cuijpers, R. D. Newman-Norlund, H. T. van Schie, and R. G. J. Meulenbroek, "Joint Action: Neurocognitive Mechanisms Supporting Human Interaction," *Topics in Cognitive Science*, vol. 1, no. 2, pp. 340–352, apr 2009. doi: 10.1111/j.1756-8765.2009.01023.x

E. Bicho, *Dynamic approach to behavior-based robotics: design, specification, analysis, simulation and implementation*. Aachen: Shaker Verlag, 2000.

E. Bicho and G. Schöner, "The dynamic approach to autonomous robotics demonstrated on a low-level vehicle platform," *Robotics and Autonomous Systems*, vol. 21, no. 1, pp. 23–35, jul 1997. doi: 10.1016/S0921-8890(97)00004-3

E. Bicho, P. Mallet, and G. Schöner, "Using attractor dynamics to control autonomous vehicle motion," in *IECON '98. Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society (Cat. No.98CH36200)*, vol. 2, 1998, pp. 1176–1181.

——, "Target Representation on an Autonomous Vehicle with Low-Level Sensors," *The International Journal of Robotics Research*, vol. 19, no. 5, pp. 424–447, may 2000. doi: 10.1177/02783640022066950

E. Bicho, L. Louro, and W. Erlhagen, "Integrating verbal and nonverbal communication in a dynamic neural field architecture for human-robot interaction," *Frontiers in neurorobotics*, vol. 4, pp. 1–13, jan 2010. doi: 10.3389/fnbot.2010.00005

E. Bicho, W. Erlhagen, L. Louro, and E. Costa e Silva, "Neuro-cognitive mechanisms of decision making in joint action: a human-robot interaction study," *Human movement science*, vol. 30, no. 5, pp. 846–868, oct 2011. doi: 10.1016/j.humov.2010.08.012

E. Bicho, W. Erlhagen, L. Louro, E. Costa e Silva, R. Silva, and N. Hipólito, "A dynamic field approach to goal inference, error detection and anticipatory action selection in human-robot collaboration," in *New Frontiers in Human-Robot Interaction (Advances in Interaction Studies)*, 6th ed., K. Dautenhahn and J. Saunders, Eds. Amsterdam, The Netherlands: John Benjamins Publishing Company, 2011, pp. 135–164.

E. Bicho, W. Erlhagen, E. Sousa, L. Louro, N. Hipólito, E. Costa e Silva, R. Silva, F. Ferreira, T. Machado, M. Hulstijn, Y. Maas, E. R. A. de Bruijn, R. H. Cuijpers, R. D. Newman-Norlund, H. T. van Schie, R. G. J. Meulenbroek, and H. Bekkering, "The power of prediction: Robots that read intentions," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Vilamoura, Portugal: IEEE, oct 2012, pp. 5458–5459. doi: 10.1109/IROS.2012.6386297

R. J. R. Blair, "Facial expressions, their communicatory functions and neuro-cognitive substrates," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 358, no. 1431, p. 561, 2003.

S.-J. Blakemore and J. Decety, "From the perception of action to the understanding of intention," *Nature reviews. Neuroscience*, vol. 2, no. 8, pp. 561–567, aug 2001. doi: 10.1038/35086023

P. Branco, "Computer-based facial expression analysis for assessing user experience," Doctoral Thesis, University of Minho, 2006.

P. Branco, P. Firth, L. M. Encarnação, and P. Bonato, "Faces of emotion in human-computer interaction," in *CHI '05 extended abstracts on Human factors in computing systems - CHI '05*. New York, New York, USA: ACM Press, 2005, p. 1236. doi: 10.1145/1056808.1056885

P. Branco, L. M. Encarnação, and A. F. Marcos, "It's all in the face: studies on monitoring users' experience," in *Proceedings of SIACG - Third Ibero-American Symposium on Computer Graphics, Eurographics Proceedings*, P. Brunet, N. Correia, and G. Baranoski, Eds. Santiago de Compostela, Spain: Eurographics Association, 2006, pp. 45–51.

M. E. Bratman, "Shared cooperative activity," *The philosophical review*, vol. 101, no. 2, pp. 327–341, apr 1992.

——, "Shared Intention," *Ethics*, vol. 104, no. 1, pp. 97–113, oct 1993.

——, "I Intend that we J." in *Contemporary Action Theory volume 2: Social action*, ser. Synthese Library, Vol. 267, R. Tuomela and G. Holmström-Hintikka, Eds. Springer, 1997, pp. 49–63.

——, "Modest sociality and the distinctiveness of intention," *Philosophical Studies*, vol. 144, no. 1, pp. 149–165, mar 2009. doi: 10.1007/s11098-009-9375-9

C. Breazeal, "Early Experiments using Motivations to Regulate Human-Robot Interaction," in *In Proceedings of 1998 AAAI Fall Symposium: Emotional and Intelligent, The Tangled Knot of Cognition*, Orlando, FL., 1998, pp. 31–36.

——, "Regulating Human-Robot Interaction using "emotions", "drives" and facial expressions," in *Proceedings of Autonomous Agents*, vol. 98, 1998, pp. 14–21. doi: 10.1.1.56.6651

——, *Designing sociable robots (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2002.

——, "Toward sociable robots," *Robotics and Autonomous Systems*, vol. 42, no. 3-4, pp. 167–175, 2003.

——, "Emotion and sociable humanoid robots," *International Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 119–155, jul 2003. doi: 10.1016/S1071-5819(03)00018-1

——, "Emotive qualities in lip-synchronized robot speech," *Advanced Robotics*, vol. 17, no. 2, pp. 97–113, jan 2003. doi: 10.1163/156855303321165079

R. A. Brooks, "New approaches to robotics." *Science (New York, N.Y.)*, vol. 253, no. 5025, pp. 1227–32, sep 1991. doi: 10.1126/science.253.5025.1227

A. Bruce, I. R. Nourbakhsh, and R. Simmons, "The role of expressiveness and attention in human-robot interaction," in *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292)*, vol. 4. Washington D.C., USA: IEEE, 2002, pp. 4138–4142. doi: 10.1109/ROBOT.2002.1014396

J. T. Cacioppo, G. G. Berntson, J. T. Larsen, K. M. Poehlmann, and T. A. Ito, "The psychophysiology of emotion," in *Handbook of Emotions*, M. Lewis and J. M. Havil-Jones, Eds. New York: Guilford, 2000, ch. 11, pp. 173–191.

A. J. Calder and A. W. Young, "Understanding the recognition of facial identity and facial expression," *Nature Reviews Neuroscience*, vol. 6, no. 8, pp. 641–51, aug 2005. doi: 10.1038/nrn1724

L. Cañamero, "Emotions And Adaptation In Autonomous Agents: A Design Perspective," *Cybernetics and Systems: An International Journal*, vol. 32, no. 5, pp. 507–529, jul 2001. doi: 10.1080/01969720120250

——, "Emotion understanding from the perspective of autonomous robots research," *Neural networks : the official journal of the International Neural Network Society*, vol. 18, no. 4, pp. 445–55, may 2005. doi: 10.1016/j.neunet.2005.03.003

L. Cañamero and J. Fredslund, "How Does It Feel? Emotional Interaction with a Humanoid LEGO Robot," in *Socially Intelligent Agents: The Human in the Loop. Papers from the AAAI 2000 Fall Symposium*, K. Dautenhahn, Ed. Cape Cod, Massachusetts, USA: AAAI Press, 2000, pp. 23–28.

L. Carr, M. Iacoboni, M.-C. Dubeau, J. C. Mazziotta, and G. L. Lenzi, "Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 9, pp. 5497–502, apr 2003. doi: 10.1073/pnas.0935845100

G. Castellano, S. D. Villalba, and A. Camurri, "Recognising Human Emotions from Body Movement and Gesture Dynamics," in *Affective Computing and Intelligent Interaction*, ser. Lecture Notes in Computer Science, A. Paiva, R. Prada, and R. Picard, Eds. Springer Berlin / Heidelberg, 2007, vol. 4738, pp. 71–82. doi: 10.1007/978-3-540-74889-2_7

F. Chang, C.-J. Chen, and C.-J. Lu, "A linear-time component-labeling algorithm using contour tracing technique," *Computer Vision and Image Understanding*, vol. 93, no. 2, pp. 206–220, feb 2004. doi: 10.1016/j.cviu.2003.09.002

J. F. Cohn, "Foundations of human computing: facial expression and emotion," in *Proceedings of the 8th international conference on Multimodal interfaces - ICMI '06*. New York, New York, USA: ACM Press, 2006, p. 233. doi: 10.1145/1180995.1181043

J. D. Cole, *About Face*. Cambridge, Massachussetts: MIT Press, 1999.

——, "Empathy needs a face," *Journal of Consciousness Studies*, vol. 8, no. 5-7, pp. 51–68, 2001.

E. Costa e Silva, M. F. P. Costa, E. Bicho, and W. Erlhagen, "Nonlinear optimization for human-like movements of a high degree of freedom robotics arm-hand system," in *Computational Science and Its Applications-ICCSA 2011*. Springer Berlin Heidelberg, 2011, pp. 327–342.

E. Costa e Silva, M. F. P. Costa, J. P. F. Araújo, D. Machado, L. Louro, W. Erlhagen, and E. Bicho, "Towards human-like bimanual movements in anthropomorphic robots: a nonlinear

optimization approach," *Applied Mathematics & Information Sciences*, vol. 9, no. 2, pp. 619–629, 2015. doi: 10.12785/amis/090210

R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. D. Kollias, W. Fellenz, and J. G. Taylor, "Emotion recognition in human-computer interaction," *Signal Processing Magazine, IEEE*, vol. 18, no. 1, pp. 32–80, 2001. doi: 10.1109/79.911197

R. H. Cuijpers and W. Erlhagen, "Implementing Bayes' Rule with Neural Fields," in *Artificial Neural Networks - ICANN 2008*, ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 228–237. doi: 10.1007/978-3-540-87559-8_24

M. Cunhal, "Sistema de Visão para a Interação e Colaboração Humano-Robô: Reconhecimento de Objetos, Gestos e Expressões Faciais," MSc Dissertation, University of Minho, 2014.

C. Darwin, *The expression of emotion in man and animals*. New York: Oxford University Press, 1872.

M. Destephe, K. Hashimoto, and A. Takanishi, "Emotional gait generation method based on emotion mental model & Preliminary experiment with happiness and sadness," in *2013 10th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*. Jeju, Korea: IEEE, oct 2013, pp. 86–89. doi: 10.1109/URAI.2013.6677480

U. Dimberg and M. Thunberg, "Rapid facial reactions to emotional facial expressions," *Scandinavian Journal of Psychology*, vol. 39, no. 1, pp. 39–45, 1998. doi: 10.1111/1467-9450.00054

U. Dimberg, M. Thunberg, and K. Elmehed, "Unconscious facial reactions to emotional facial expressions," *Psychological science*, vol. 11, no. 1, pp. 86–9, jan 2000.

S. K. D'Mello and R. A. Calvo, "Beyond the Basic Emotions: What Should Affective Computing Compute?" in *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*. New York, New York, USA: ACM Press, 2013, pp. 2287–2294. doi: 10.1145/2468356.2468751

R. J. Douglas, C. Koch, M. Mahowald, K. A. Martin, and H. H. Suarez, "Recurrent excitation in neocortical circuits," *Science*, vol. 269, no. 5226, pp. 981–985, aug 1995. doi: 10.1126/science.7638624

R. Drillis and R. Contini, "Body Segment Parameters," New York University, School of Engineering, Research Division, New York, Tech. Rep. No. 1166-03, 1966.

P. Ekman, "Universals and cultural differences in facial expressions of emotion," in *Nebraska Symposium on Motivation*, J. Cole, Ed., vol. 19. Lincoln, Nebraska: Lincoln University of Nebraska Press, 1971, pp. 207–282.

——, "An argument for basic emotions," *Cognition & Emotion*, vol. 6, no. 3/4, pp. 169–200, may 1992. doi: 10.1080/02699939208411068

P. Ekman and D. Cordaro, "What is Meant by Calling Emotions Basic," *Emotion Review*, vol. 3, no. 4, pp. 364–370, sep 2011. doi: 10.1177/1754073911410740

P. Ekman and W. V. Friesen, "Facial action coding system: A technique for the measurement of facial movement," in *From appraisal to emotion: Differences among unpleasant feelings. Motivation and Emotion*, P. C. Ellsworth and C. A. Smith, Eds., vol. 12. Palo Alto: CA: Consulting Psychologists Press (1988), 1978, pp. 271–302.

P. Ekman and E. L. Rosenberg, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*, 2nd ed., ser. Series in Affective Science. Oxford University Press, 2005.

P. Ekman, W. V. Friesen, and J. C. Hager, *Facial Action Coding System*. Salt Lake City, USA: Research Nexus division of Network Information Research Corporation, 2002.

N. Endo and A. Takanishi, "Development of Whole-Body Emotional Expression Humanoid Robot for ADL-Assistive RT Services," *Journal of Robotics and Mechatronics*, vol. 23, no. 6, pp. 969–977, dec 2011. doi: 10.20965/jrm.2011.p0969

C. Engels and G. Schöner, "Dynamic fields endow behavior-based robots with representations," *Robotics and Autonomous Systems*, vol. 14, no. 1, pp. 55–77, feb 1995. doi: 10.1016/0921-8890(94)00020-3

W. Erlhagen and E. Bicho, "The dynamic neural field approach to cognitive robotics," *Journal of Neural Engineering*, vol. 3, pp. R36–R54, 2006.

——, "A Dynamic Neural Field Approach to Natural and Efficient Human-Robot Collaboration," in *Neural Fields*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 341–365. doi: 10.1007/978-3-642-54593-1_13

W. Erlhagen and G. Schöner, "Dynamic field theory of movement preparation." *Psychological review*, vol. 109, no. 3, pp. 545–72, jul 2002.

W. Erlhagen, A. Bastian, D. Jancke, A. Riehle, and G. Schöner, "The distribution of neuronal population activation (DPA) as a tool to study interaction and integration in cortical representations." *Journal of neuroscience methods*, vol. 94, no. 1, pp. 53–66, dec 1999.

W. Erlhagen, A. Mukovskiy, and E. Bicho, "A dynamic model for action understanding and goal-directed imitation," *Brain Research*, vol. 1083, no. 1, pp. 174–188, 2006.

W. Erlhagen, A. Mukovskiy, E. Bicho, G. Panin, C. Kiss, A. Knoll, H. T. van Schie, and H. Bekkering, "Goal-directed imitation for robots: A bio-inspired approach to action understanding and skill learning," *Robotics and Autonomous Systems*, vol. 54, no. 5, pp. 353–360, may 2006. doi: 10.1016/j.robot.2006.01.004

W. Erlhagen, A. Mukovskiy, F. Chersi, and E. Bicho, "On the development of intention understanding for joint action tasks," in *2007 IEEE 6th International Conference on Development and Learning*. London: Imperial College London, jul 2007, pp. 140–145. doi: 10.1109/DEVLRN.2007.4354022

N. Esau and L. Kleinjohann, "Emotional Robot Competence and Its Use in Robot Behavior Control," in *Emotional Engineering*, S. Fukuda, Ed. London: Springer London, 2011, pp. 119–142. doi: 10.1007/978-1-84996-423-4_7

N. Esau, L. Kleinjohann, and B. Kleinjohann, "Emotional Communication with the Robot Head MEXI," in *2006 9th International Conference on Control, Automation, Robotics and Vision*. Singapore: IEEE, 2006, pp. 1–7. doi: 10.1109/ICARCV.2006.345162

N. Esau, E. Wetzel, L. Kleinjohann, and B. Kleinjohann, "Real-Time Facial Expression Recognition Using a Fuzzy Emotion Model," in *2007 IEEE International Fuzzy Systems Conference*. London: IEEE, jun 2007, pp. 1–6. doi: 10.1109/FUZZY.2007.4295451

N. Esau, L. Kleinjohann, and B. Kleinjohann, "Integrating Emotional Competence into Man-Machine Collaboration," in *Biologically-Inspired Collaborative Computing*, ser. IFIP – The International Federation for Information Processing, M. Hinchey, A. Pagnoni, F. J. Rammig, and H. Schmeck, Eds. Boston, MA: Springer US, 2008, vol. 268, pp. 187–198. doi: 10.1007/978-0-387-09655-1_17

P. F. Ferrari, V. Gallese, G. Rizzolatti, and L. Fogassi, "Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex," *European Journal of Neuroscience*, vol. 17, no. 8, pp. 1703–1714, 2003. doi: 10.1046/j.1460-9568.2003.02601.x

F. Ferri, G. C. Campione, R. Dalla Volta, C. Gianelli, and M. Gentilucci, "To me or to you? When the self is advantaged." *Experimental brain research*, vol. 203, no. 4, pp. 637–46, jun 2010. doi: 10.1007/s00221-010-2271-x

F. Ferri, I. P. Stoianov, C. Gianelli, L. D'Amico, A. M. Borghi, and V. Gallese, "When action meets emotions: how facial displays of emotion influence goal-related behavior," *PLoS ONE*, vol. 5, no. 10, p. e13126, 2010. doi: 10.1371/journal.pone.0013126

T. Flash and N. Hogan, "The coordination of arm movements: an experimentally confirmed mathematical model." *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 5, no. 7, pp. 1688–703, jul 1985.

L. Fogassi, P. F. Ferrari, B. Gesierich, S. Rozzi, F. Chersi, and G. Rizzolatti, "Parietal lobe:

from action organization to intention understanding," *Science (New York, N.Y.)*, vol. 308, no. 5722, pp. 662–667, apr 2005. doi: 10.1126/science.1106138

T. Fong, I. R. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robotics and Autonomous Systems*, vol. 42, no. 3-4, pp. 143–166, mar 2003. doi: 10.1016/S0921-8890(02)00372-X

W. V. Friesen and P. Ekman, "EMFACS: Emotional Facial Action Coding System," *Unpublished manuscript, University of California, San Francisco*, 1982.

C. D. Frith and D. M. Wolpert, *The neuroscience of social interaction: Decoding, imitating, and influencing the actions of others*. New York, NY, USA: Oxford University Press, 2004.

M. Fujita, "On activating human communications with pet-type robot AIBO," *Proceedings of the IEEE*, vol. 92, no. 11, pp. 1804–1813, nov 2004. doi: 10.1109/JPROC.2004.835364

M. Fujita and H. Kitano, "Development of an Autonomous Quadruped Robot for Robot Entertainment," in *Autonomous Agents*. Boston, MA: Springer US, 1998, vol. 5, no. 1, pp. 7–18. doi: 10.1007/978-1-4615-5735-7_2

J. M. Fuster, "The cognit: a network model of cortical representation," *International Journal of Psychophysiology*, vol. 60, no. 2, pp. 125–132, 2006.

P. Gable and E. Harmon-Jones, "The motivational dimensional model of affect: Implications for breadth of attention, memory, and cognitive categorisation," *Cognition & Emotion*, vol. 24, no. 2, pp. 322–337, feb 2010. doi: 10.1080/02699930903378305

V. Gallese, L. Fadiga, L. Fogassi, and G. Rizzolatti, "Action recognition in the premotor cortex," *Brain*, vol. 119, no. 2, pp. 593–609, 1996. doi: 10.1093/brain/119.2.593

V. Gallese, C. Keysers, and G. Rizzolatti, "A unifying view of the basis of social cognition," *Trends in Cognitive Sciences*, vol. 8, no. 9, pp. 396–403, sep 2004. doi: 10.1016/j.tics.2004.07.002

A. Genovesio, P. J. Brasted, and S. P. Wise, "Representation of Future and Previous Spatial Goals by Separate Neural Populations in Prefrontal Cortex," *Journal of Neuroscience*, vol. 26, no. 27, pp. 7305–7316, jul 2006. doi: 10.1523/JNEUROSCI.0699-06.2006

W. Gerstner and W. M. Kistler, "Mathematical formulations of Hebbian learning," *Biological cybernetics*, vol. 87, no. 5-6, pp. 404–415, 2002.

M. Gilbert, "Walking Together: A Paradigmatic Social Phenomenon," *Midwest Studies In Philosophy*, vol. 15, no. 1, pp. 1–14, sep 1990. doi: 10.1111/j.1475-4975.1990.tb00202.x

A. Grecucci, R. P. Cooper, and R. I. Rumiati, "A computational model of action resonance and its modulation by emotional stimulation," *Cognitive Systems Research*, vol. 8, no. 3, pp. 143–160, 2007.

H. Gunes and M. Pantic, "Dimensional emotion prediction from spontaneous head gestures for interaction with sensitive artificial listeners," in *Proceedings of the 10th international conference on Intelligent virtual agents*, J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, and A. Safonova, Eds. Philadelphia, PA: Springer-Verlag Berlin Heidelberg, 2010, pp. 371–377.

——, "Automatic, Dimensional and Continuous Emotion Recognition," in *Creating Synthetic Emotions through Technological and Robotic Advancements*, J. Vallverdú, Ed. IGI Global, 2010, pp. 73–105. doi: 10.4018/978-1-4666-1595-3.ch005

P. Haazebroek, S. van Dantzig, and B. Hommel, "A computational model of perception and action for cognitive robotics," *Cognitive Processing*, vol. 12, no. 4, pp. 355–365, nov 2011. doi: 10.1007/s10339-011-0408-x

S. Hamann, "Mapping discrete and dimensional emotions onto the brain: controversies and consensus," *Trends in Cognitive Sciences*, vol. 16, no. 9, pp. 458–466, aug 2012. doi: 10.1016/j.tics.2012.07.006

T. Hashimoto, N. Kato, and H. Kobayashi, "Field trial of android-type remote class support system in elementary school and effect evaluation," in *2009 IEEE International Conference on*

Robotics and Biomimetics (ROBIO). IEEE, dec 2009, pp. 1135–1140. doi: 10.1109/ROBIO.2009.5420758

R. L. Hazlett, "Measuring emotional valence during interactive experiences," in *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06*, R. Grinter, T. Rodden, P. Aoki, E. Cutrell, R. Jeffries, and G. Olson, Eds. New York, New York, USA: ACM Press, apr 2006, pp. 1023–1026. doi: 10.1145/1124772.1124925

F. Hegel, T. P. Spexard, B. Wrede, G. Horstmann, and T. Vogt, "Playing a different imitation game: Interaction with an Empathic Android Robot," in *2006 6th IEEE-RAS International Conference on Humanoid Robots*, G. Sandini and A. Billard, Eds. Genova, Italy: IEEE, dec 2006, pp. 56–61. doi: 10.1109/ICHR.2006.321363

M.-K. Hu, "Visual pattern recognition by moment invariants," *IEEE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, feb 1962. doi: 10.1109/TIT.1962.1057692

E. Hudlicka, "Beyond Cognition: Modeling Emotion in Cognitive Architectures," in *Sixth International Conference on Cognitive Modeling*, M. Lovett, C. Lebiere, C. Schunn, and P. Munro, Eds., Mahwah, New Jersey, 2004, pp. 118–123.

M. Iacoboni, I. Molnar-Szakacs, V. Gallese, G. Buccino, J. C. Mazziotta, and G. Rizzolatti, "Grasping the intentions of others with one's own mirror neuron system," *PLoS Biology*, vol. 3, no. 3, p. e79, 2005. doi: 10.1371/journal.pbio.0030079

C. Jones and A. Deeming, "Affective Human-Robotic Interaction," in *Affect and Emotion in Human-Computer Interaction*, ser. Lecture Notes in Computer Science, C. Peter and R. Beale, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, vol. 4868, pp. 175–185. doi: 10.1007/978-3-540-85099-1_15

T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)*, F. M. Titsworth, Ed. Grenoble, France: IEEE Comput. Soc, 2000, pp. 46–53. doi: 10.1109/AFGR.2000.840611

J. Kędzierski and M. Janiak, "Construction of the social robot FLASH," *Scientific papers on electronics*, vol. 182, pp. 681–694, 2012.

J. Kędzierski, R. Muszyński, C. Zoll, A. Oleksy, and M. Frontkiewicz, "EMYS - Emotive Head of a Social Robot," *International Journal of Social Robotics*, vol. 5, no. 2, pp. 237–249, mar 2013. doi: 10.1007/s12369-013-0183-1

D. Keltner and P. Ekman, "Facial Expression of Emotion," in *Handbook of Emotions*, M. Lewis and J. M. Haviland-Jones, Eds. New York, NY, USA: Guilford Press, 2000, pp. 236–249.

M. L. Kesler/West, A. H. Andersen, C. D. Smith, M. J. Avison, C. E. Davis, R. J. Kryscio, and L. X. Blonder, "Neural substrates of facial emotion processing using fMRI," *Cognitive Brain Research*, vol. 11, no. 2, pp. 213–226, apr 2001. doi: 10.1016/S0926-6410(00)00073-2

C. Keysers and V. Gazzola, "Towards a unifying neural theory of social cognition." *Progress in brain research*, vol. 156, pp. 379–401, 2006. doi: 10.1016/S0079-6123(06)56021-2

——, "Hebbian learning and predictive mirror neurons for actions, sensations and emotions," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, no. 1644, pp. 1–11, apr 2014. doi: 10.1098/rstb.2013.0175

C. Keysers, B. Wicker, V. Gazzola, J.-L. Anton, L. Fogassi, and V. Gallese, "A Touching Sight: SII/PV Activation during the Observation and Experience of Touch," *Neuron*, vol. 42, no. 2, pp. 335–346, apr 2004. doi: 10.1016/S0896-6273(04)00156-4

R. Kirby, J. Forlizzi, and R. Simmons, "Affective social robots," *Robotics and Autonomous Systems*, vol. 58, no. 3, pp. 322–332, mar 2010. doi: 10.1016/j.robot.2009.09.015

S. Koelstra, M. Pantic, and I. Patras, "A Dynamic Texture-Based Approach to Recognition of Facial Actions and Their Temporal Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 1940–1954, nov 2010. doi: 10.1109/TPAMI.2010.50

K. P. Körding and D. M. Wolpert, "Bayesian decision theory in sensorimotor control." *Trends in cognitive sciences*, vol. 10, no. 7, pp. 319–26, jul 2006. doi: 10.1016/j.tics.2006.05.003

R. Kozma, "Intentional systems: Review of neurodynamics, modeling, and robotics implementation," *Physics of Life Reviews*, vol. 5, no. 1, pp. 1–21, mar 2008. doi: 10.1016/j.plrev.2007.10.002

I. Leite, A. Pereira, S. Mascarenhas, G. Castellano, C. Martinho, R. Prada, and A. Paiva, "Closing the loop: from Affect Recognition to Empathic Interaction," in *Proceedings of the 3rd international workshop on Affective interaction in natural environments - AFFINE '10*. Firenze, Italy: ACM Press, 2010, p. 43. doi: 10.1145/1877826.1877839

C. M. Leonard, E. T. Rolls, F. A. W. Wilson, and G. C. Baylis, "Neurons in the amygdala of the monkey with responses selective for faces," *Behavioural brain research*, vol. 15, no. 2, pp. 159–76, apr 1985.

K. R. Leslie, S. H. Johnson-Frey, and S. T. Grafton, "Functional imaging of face and hand imitation: towards a motor theory of empathy," *NeuroImage*, vol. 21, no. 2, pp. 601–7, feb 2004. doi: 10.1016/j.neuroimage.2003.09.038

H. Levesque and G. Lakemeyer, "Cognitive Robotics," in *Handbook of Knowledge Representation*. Elsevier, 2008.

N. Li, J. Bu, and C. Chen, "Real-time video object segmentation using HSV space," *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 2, pp. II85–II88, 2002. doi: 10.1109/ICIP.2002.1039893

K. D. Locke and L. M. Horowitz, "Satisfaction in interpersonal interactions as a function of similarity in level of dysphoria." *Journal of personality and social psychology*, vol. 58, no. 5, pp. 823–31, may 1990.

R. Lowe, C. Herrera, A. Morse, and T. Ziemke, "The Embodied Dynamics of Emotion, Appraisal and Attention," in *Attention in Cognitive Systems. Theories and Systems from an Interdisciplinary Viewpoint*, ser. Lecture Notes in Computer Science, L. Paletta and E. Rome, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, vol. 4840, pp. 1–20. doi: 10.1007/978-3-540-77343-6_1

P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*. San Francisco, CA: IEEE, jun 2010, pp. 94–101. doi: 10.1109/CVPRW.2010.5543262

T. Machado, T. Malheiro, S. Monteiro, E. Bicho, and W. Erlhagen, "Transportation of long objects in unknown cluttered environments by a team of robots: A dynamical systems approach," in *2013 IEEE International Symposium on Industrial Electronics*. Taipei, Taiwan: IEEE, may 2013, pp. 1–6. doi: 10.1109/ISIE.2013.6563794

S. Mann, "Wearable computing: a first step toward personal imaging," *Computer*, vol. 30, no. 2, pp. 25–32, 1997. doi: 10.1109/2.566147

H. K. M. Meeren, C. C. R. J. van Heijnsbergen, and B. de Gelder, "Rapid perceptual integration of facial expression and emotional body language," *Proceedings of the National Academy of Sciences*, vol. 102, no. 45, pp. 16 518–16 523, nov 2005. doi: 10.1073/pnas.0507650102

R. G. J. Meulenbroek, D. A. Rosenbaum, C. Jansen, J. Vaughan, and S. Vogt, "Multijoint grasping movements - Simulated and observed effects of object location, object size, and initial aperture," *Experimental Brain Research*, vol. 138, no. 2, pp. 219–234, may 2001. doi: 10.1007/s002210100690

J. Michael, "Shared Emotions and Joint Action," *Review of Philosophy and Psychology*, vol. 2, no. 2, pp. 355–373, may 2011. doi: 10.1007/s13164-011-0055-2

T. Minato, M. Shimada, H. Ishiguro, and S. Itakura, "Development of an Android Robot for Studying Human-Robot Interaction," in *Lecture Notes In Computer Science: Proceedings of the 17 th international conference on Innovations in applied artificial intelligence*, vol. 17, no. 20. Springer, 2004, pp. 424–434.

M. A. Miskam, S. Shamsuddin, M. R. A. Samat, H. Yussof, H. A. Ainudin, and A. R. Omar, "Humanoid robot NAO as a teaching tool of emotion recognition for children with autism

using the Android app," in *2014 International Symposium on Micro-NanoMechatronics and Human Science (MHS)*. IEEE, nov 2014, pp. 1–5. doi: 10.1109/MHS.2014.7006084

S. Monteiro and E. Bicho, "A dynamical systems approach to behavior-based formation control," in *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292)*, vol. 3. IEEE, 2002, pp. 2606–2611. doi: 10.1109/ROBOT.2002.1013624

——, "Robot formations: Robots allocation and leader-follower pairs," in *2008 IEEE International Conference on Robotics and Automation*. Pasadena, CA: IEEE, may 2008, pp. 3769–3775. doi: 10.1109/ROBOT.2008.4543789

S. Monteiro, M. Vaz, and E. Bicho, "Attractor dynamics generates robot formation: from theory to implementation," in *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004*, vol. 3. IEEE, 2004, pp. 2582–2586. doi: 10.1109/ROBOT.2004.1307450

A. Murata, L. Fadiga, L. Fogassi, V. Gallese, V. Raos, and G. Rizzolatti, "Object representation in the ventral premotor cortex (area F5) of the monkey," *Journal of neurophysiology*, vol. 78, no. 4, pp. 2226–30, oct 1997.

M. Neta and P. J. Whalen, "Individual differences in neural activity during a facial expression vs. identity working memory task," *NeuroImage*, vol. 56, no. 3, pp. 1685–92, jun 2011. doi: 10.1016/j.neuroimage.2011.02.051

R. D. Newman-Norlund, M. L. Noordzij, R. G. J. Meulenbroek, and H. Bekkering, "Exploring the brain basis of joint action: co-ordination of actions, goals and intentions," *Social neuroscience*, vol. 2, no. 1, pp. 48–65, jan 2007. doi: 10.1080/17470910701224623

R. D. Newman-Norlund, H. T. van Schie, A. M. J. van Zuijlen, and H. Bekkering, "The mirror neuron system is more active during complementary compared with imitative action," *Nature neuroscience*, vol. 10, no. 7, pp. 817–818, jul 2007. doi: 10.1038/nn1911

M. A. Nicolaou, H. Gunes, and M. Pantic, "Continuous Prediction of Spontaneous Affect from Multiple Cues and Modalities in Valence-Arousal Space," *IEEE Transactions on Affective Computing*, vol. 2, no. 2, pp. 92–105, apr 2011. doi: 10.1109/T-AFFC.2011.9

J. Novikova and L. Watts, "Towards Artificial Emotions to Assist Social Coordination in HRI," *International Journal of Social Robotics*, vol. 7, no. 1, pp. 77–88, feb 2015. doi: 10.1007/s12369-014-0254-y

K. Oatley and P. N. Johnson-Laird, "Cognitive approaches to emotions," *Trends in Cognitive Sciences*, vol. 18, no. 3, pp. 134–140, mar 2014. doi: 10.1016/j.tics.2013.12.004

——, "Towards a Cognitive Theory of Emotions," *Cognition & Emotion*, vol. 1, no. 1, pp. 29–50, mar 1987. doi: 10.1080/02699938708408362

K. N. Ochsner and J. J. Gross, "The cognitive control of emotion," *Trends in cognitive sciences*, vol. 9, no. 5, pp. 242–9, may 2005. doi: 10.1016/j.tics.2005.03.010

A. Ortony and T. J. Turner, "What's basic about basic emotions?" *Psychological Review*, vol. 97, no. 3, pp. 315–331, 1990. doi: 10.1037/0033-295X.97.3.315

H. Oster, "BabyFACS: Facial Action Coding System for infants and young children," *Unpublished manuscript, New York: New York University*, 2004.

J. Panksepp, "Neurologizing the Psychology of Affects: How Appraisal-Based Constructivism and Basic Emotion Theory Can Coexist," *Perspectives on Psychological Science*, vol. 2, no. 3, pp. 281–296, sep 2007. doi: 10.1111/j.1745-6916.2007.00045.x

M. Pantic and M. S. Bartlett, "Machine Analysis of Facial Expressions," in *Face Recognition*, ser. Advanced Robotics Systems, K. Kurihara, Ed. Vienna, Austria: I-Tech Education and Publishing, 2007, pp. 237–366.

M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: the state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424–1445, 2000. doi: 10.1109/34.895976

M. Pantic, A. Nijholt, A. P. Pentland, and T. S. Huang, "Human-Centred Intelligent Human Computer Interaction (HCI2): how far are we from attaining it?" *International Journal of Autonomous and Adaptive Communications Systems*, vol. 1, no. 2, p. 168, 2008. doi: 10.1504/IJAACS.2008.019799

A. Pereira, I. Leite, S. Mascarenhas, C. Martinho, and A. Paiva, "Using Empathy to Improve Human-Robot Relationships," in *Human-Robot Personal Relationships*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, M. H. Lamers and F. J. Verbeek, Eds. Springer Berlin Heidelberg, 2011, vol. 59, pp. 130–138. doi: 10.1007/978-3-642-19385-9_17

S. Petridis, A. Asghar, and M. Pantic, "Classifying laughter and speech using audio-visual feature prediction," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2010, pp. 5254–5257. doi: 10.1109/ICASSP.2010.5494992

M. Piccardi, "Background subtraction techniques: a review," in *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, vol. 4. IEEE, 2004, pp. 3099–3104. doi: 10.1109/ICSMC.2004.1400815

E. Poljac, H. T. van Schie, and H. Bekkering, "Understanding the flexibility of action-perception coupling," *Psychological research*, vol. 73, no. 4, pp. 578–86, jul 2009. doi: 10.1007/s00426-009-0238-y

F. E. Pollick, H. M. Paterson, A. Bruderlin, and A. J. Sanford, "Perceiving affect from arm movement," *Cognition*, vol. 82, no. 2, pp. B51–B61, dec 2001. doi: 10.1016/S0010-0277(01)00147-0

S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey," *Artificial Intelligence Review*, vol. 43, no. 1, pp. 1–54, jan 2015. doi: 10.1007/s10462-012-9356-9

G. Rizzolatti and L. Craighero, "The mirror-neuron system," *Annual review of neuroscience*, vol. 27, pp. 169–92, jan 2004. doi: 10.1146/annurev.neuro.27.070203.144230

G. Rizzolatti and C. Sinigaglia, *Mirrors in the Brain: How Our Minds Share Actions and Emotions*. New York: Oxford University Press, 2008.

G. Rizzolatti, R. Camarda, L. Fogassi, M. Gentilucci, G. Luppino, and M. Matelli, "Functional organization of inferior area 6 in the macaque monkey," *Experimental Brain Research*, vol. 71, no. 3, pp. 491–507, 1988. doi: 10.1007/BF00248742

G. Rizzolatti, L. Fadiga, V. Gallese, and L. Fogassi, "Premotor cortex and the recognition of motor actions," *Cognitive Brain Research*, vol. 3, no. 2, pp. 131–141, mar 1996. doi: 10.1016/0926-6410(95)00038-0

G. Rizzolatti, L. Fogassi, and V. Gallese, "Cortical mechanisms subserving object grasping and action recognition: a new view on the cortical motor functions," in *The New Cognitive Neurosciences, 2nd Edition*, M. S. Gazzaniga, Ed. A Bradford Book, The MIT Press, 2000, pp. 539–552.

——, "Neurophysiological mechanisms underlying the understanding and imitation of action," *Nature Reviews Neuroscience*, vol. 2, no. 9, pp. 661–70, sep 2001. doi: 10.1038/35090060

G. Rizzolatti, L. Cattaneo, M. Fabbri-Destro, and S. Rozzi, "Cortical Mechanisms Underlying the Organization of Goal-Directed Actions and Mirror Neuron-Based Action Understanding," *Physiological Reviews*, vol. 94, no. 2, pp. 655–706, apr 2014. doi: 10.1152/physrev.00009.2013

E. T. Rolls, "Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces," *Human neurobiology*, vol. 3, no. 4, pp. 209–22, jan 1984.

D. A. Rosenbaum, R. G. J. Meulenbroek, and J. Vaughan, "Planning reaching and grasping movements: theoretical premises and practical implications." *Motor control*, vol. 5, no. 2, pp. 99–115, apr 2001.

J. A. Russell, "Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies." *Psychological Bulletin*, vol. 115, no. 1, pp. 102–141, 1994. doi: 10.1037/0033-2909.115.1.102

——, "Emotion, core affect, and psychological construction," *Cognition & Emotion*, vol. 23, no. 7, pp. 1259–1283, nov 2009. doi: 10.1080/02699930902809375

J. A. Russell and L. F. Barrett, "Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant." *Journal of personality and social psychology*, vol. 76, no. 5, pp. 805–19, may 1999.

D. Sakamoto, T. Kanda, T. Ono, H. Ishiguro, and N. Hagita, "Android as a telecommunication medium with a human-like presence," in *Proceeding of the ACM/IEEE international conference on Human-robot interaction - HRI '07*. New York, New York, USA: ACM Press, 2007, p. 193. doi: 10.1145/1228716.1228743

A. Samal and P. A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: a survey," *Pattern Recognition*, vol. 25, no. 1, pp. 65–77, jan 1992. doi: 10.1016/0031-3203(92)90007-6

W. Sato, T. Kochiyama, S. Yoshikawa, E. Naito, and M. Matsumura, "Enhanced neural activity in response to dynamic facial expressions of emotion: an fMRI study," *Brain research. Cognitive brain research*, vol. 20, no. 1, pp. 81–91, jun 2004. doi: 10.1016/j.cogbrainres.2004.01.008

S. Schaal, "Is imitation learning the route to humanoid robots?" *Trends in Cognitive Sciences*, vol. 3, no. 6, pp. 233–242, jun 1999. doi: 10.1016/S1364-6613(99)01327-3

S. Scherer, M. Glodek, G. Layher, M. Schels, M. Schmidt, T. Brosch, S. Tschechne, F. Schwenker, H. Neumann, and G. Palm, "A generic framework for the inference of user states in human computer interaction: How patterns of low level behavioral cues support complex user states in HCI," *Journal on Multimodal User Interfaces*, vol. 6, no. 3-4, pp. 117–141, nov 2012. doi: 10.1007/s12193-012-0093-9

M. Scheutz, "The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots," in *Robot Ethics: The Ethical and Social Implications of Robotics*, P. Lin, K. Abney, and G. A. Bekey, Eds. Cambridge, Massachussetts: MIT Press, Cambridge MA, 2011, ch. 13, pp. 205–221.

M. Scheutz, P. Schermerhorn, and J. Kramer, "The utility of affect expression in natural language interactions in joint human-robot tasks," in *Proceeding of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction - HRI '06*, M. A. Goodrich, A. C. Schultz, and D. J. Bruemmer, Eds. New York, USA: ACM Press, 2006, p. 226. doi: 10.1145/1121241.1121281

G. Schöner, "Dynamical systems approaches to cognition," *Cambridge handbook of computational cognitive modeling*, pp. 101–126, 2008.

G. Schöner, M. Dose, and C. Engels, "Dynamics of behavior: Theory and applications for autonomous robot architectures," *Robotics and Autonomous Systems*, vol. 16, no. 2-4, pp. 213–245, dec 1995. doi: 10.1016/0921-8890(95)00049-6

M. Schulte-Rüther, H. J. Markowitsch, G. R. Fink, and M. Piefke, "Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: a functional magnetic resonance imaging approach to empathy," *Journal of cognitive neuroscience*, vol. 19, no. 8, pp. 1354–72, aug 2007. doi: 10.1162/jocn.2007.19.8.1354

P. G. Schyns, L. S. Petro, and M. L. Smith, "Transmission of facial expressions of emotion co-evolved with their efficient decoding in the brain: behavioral and brain evidence," *PLoS ONE*, vol. 4, no. 5, p. e5625, 2009. doi: 10.1371/journal.pone.0005625

N. Sebanz, H. Bekkering, and G. Knoblich, "Joint action: bodies and minds moving together," *Trends in cognitive sciences*, vol. 10, no. 2, pp. 70–6, feb 2006. doi: 10.1016/j.tics.2005.12.009

A. Senior, A. Hampapur, Y.-L. Tian, L. Brown, S. Pankanti, and R. Bolle, "Appearance models for occlusion handling," *Image and Vision Computing*, vol. 24, no. 11, pp. 1233–1243, nov 2006. doi: 10.1016/j.imavis.2005.06.007

R. P. Shi, J. Adelhardt, V. Zeißler, A. Batliner, C. Frank, E. Nöth, and H. Niemann, "Using Speech and Gesture to Explore User States in Multimodal Dialogue Systems," in *AVSP 2003 - International Conference on Audio-Visual Speech Processing*, St. Jorioz, France, 2003, pp. 151–156.

L. Sigal, S. Sclaroff, and V. Athitsos, "Skin color-based video segmentation under time-varying illumination," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 7, pp. 862–877, jul 2004. doi: 10.1109/TPAMI.2004.35

R. Silva, E. Bicho, and W. Erlhagen, "ARoS: An Anthropomorphic Robot For Human-Robot Interaction And Coordination Studies," in *Proceedings of the CONTROLO'2008 Conference - 8th Portuguese Conference on Automatic Control*, UTAD - Universidade de Trás-os-Montes e Alto Douro, APCA - Associação Portuguesa de Controlo Automático. UTAD - Vila Real, Portugal: UTAD, 2008, pp. 819–826.

R. Silva, L. Louro, T. Malheiro, W. Erlhagen, and E. Bicho, "Combining intention and emotional state inference in a dynamic neural field architecture for human-robot joint action," *Adaptive Behavior*, no. (in press), 2016. doi: 10.1177/1059712316665451

P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, "Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1948–1962, 2006. doi: 10.1109/JPROC.2006.884093

C. A. Smith and H. S. Scott, "A componential approach to the meaning of facial expressions," in *The Psychology of Facial Expression*, J. A. Russell and J. M. Fernández-Dols, Eds. Cambridge: Cambridge University Press, 1997, pp. 229–254.

R. Soares, E. Bicho, T. Machado, and W. Erlhagen, "Object transportation by multiple mobile robots controlled by attractor dynamics: theory and implementation," in *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*. San Diego, CA: IEEE, oct 2007, pp. 937–944. doi: 10.1109/IROS.2007.4399019

E. Sousa, W. Erlhagen, F. Ferreira, and E. Bicho, "Off-line simulation inspires insight: A neurodynamics approach to efficient robot task learning," *Neural Networks*, vol. 72, pp. 123–139, dec 2015. doi: 10.1016/j.neunet.2015.09.002

M. Sousa, S. Monteiro, T. Machado, W. Erlhagen, and E. Bicho, "Multi-robot cognitive

formations," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Vilamoura, Portugal: IEEE, oct 2012, pp. 545–550. doi: 10.1109/IROS.2012.6385833

J. P. Spencer and G. Schöner, "Bridging the representational gap in the dynamic systems approach to development," *Developmental Science*, vol. 6, no. 4, pp. 392–412, sep 2003. doi: 10.1111/1467-7687.00295

T. P. Spexard, M. Hanheide, and G. Sagerer, "Human-Oriented Interaction With an Anthropomorphic Robot," *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 852–862, oct 2007. doi: 10.1109/TRO.2007.904903

S. Steininger, F. Schiel, and A. Glesner, "User-state labeling procedures for the multimodal data collection of SmartKom," in *Proc. of the 3rd Int. conf. on Language Resources and Evaluation*, Las Palmas, Spain, 2002.

M. Talanov, J. Vallverdu, S. Distefano, M. Mazzara, and R. Delhibabu, "Neuromodulating Cognitive Architecture: Towards Biomimetic Emotional AI," in *2015 IEEE 29th International Conference on Advanced Information Networking and Applications*, L. Barolli, M. Takizawa, F. Xhafa, T. Enokido, and J. H. Park, Eds. Gwangiu: IEEE, mar 2015, pp. 587–592. doi: 10.1109/AINA.2015.240

F. Tanaka, K. Noda, T. Sawada, and M. Fujita, "Associated Emotion and Its Expression in an Entertainment Robot QRIO," in *Entertainment Computing – ICEC 2004*, ser. Lecture Notes in Computer Science, M. Rauterberg, Ed., 2004, vol. 3166, pp. 499–504. doi: 10.1007/978-3-540-28643-1_64

Y.-L. Tian, T. Kanade, and J. F. Cohn, "Facial Expression Analysis," in *Handbook of Face Recognition*. New York: Springer-Verlag, 2005, pp. 247–275. doi: 10.1007/0-387-27257-7_12

D. Tollefsen, "Let's Pretend!: Children and Joint Action," *Philosophy of the Social Sciences*, vol. 35, no. 1, pp. 75–97, mar 2005. doi: 10.1177/0048393104271925

G. Trovato, T. Kishi, N. Endo, K. Hashimoto, and A. Takanishi, "Development of facial expressions generator for emotion expressive humanoid robot," in *2012 12th IEEE-RAS*

International Conference on Humanoid Robots (Humanoids 2012). Osaka, Japan: IEEE, nov 2012, pp. 303–308. doi: 10.1109/HUMANOIDS.2012.6651536

R. Tuomela, "We-Intentions Revisited," *Philosophical Studies*, vol. 125, no. 3, pp. 327–369, sep 2005. doi: 10.1007/s11098-005-7781-1

C. van der Gaag, R. B. Minderaa, and C. Keysers, "Facial expressions: what the mirror neuron system can and cannot tell us," *Social neuroscience*, vol. 2, no. 3-4, pp. 179–222, 2007. doi: 10.1080/17470910701376878

H. T. van Schie, B. M. van Waterschoot, and H. Bekkering, "Understanding action beyond imitation: reversed compatibility effects of action observation in imitation and joint action," *Journal of experimental psychology. Human perception and performance*, vol. 34, no. 6, pp. 1493–500, dec 2008. doi: 10.1037/a0011750

J. Vaughan, D. A. Rosenbaum, and R. G. J. Meulenbroek, "Modeling Reaching and Manipulating in 2- and 3-D Workspaces: The Posture-Based Model," in *Proceedings of the ICDL 2006*. Bloomington: Proceedings of ICDL 2006 [CD-Rom], 2006.

D. Vernon, G. Metta, and G. Sandini, "A Survey of Artificial Cognitive Systems: Implications for the Autonomous Development of Mental Capabilities in Computational Agents," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 2, pp. 151–180, apr 2007. doi: 10.1109/TEVC.2006.890274

C. Vesper, S. Butterfill, G. Knoblich, and N. Sebanz, "A minimal architecture for joint action," *Neural networks : the official journal of the International Neural Network Society*, vol. 23, no. 8-9, pp. 998–1003, 2010. doi: 10.1016/j.neunet.2010.06.002

A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain," *Image and Vision Computing*, vol. 27, no. 12, pp. 1743–1759, nov 2009. doi: 10.1016/j.imavis.2008.11.007

A. Vinciarelli, M. Pantic, D. Heylen, C. Pelachaud, I. Poggi, F. D'Errico, and M. Schroeder, "Bridging the Gap between Social Animal and Unsocial Machine: A Survey of Social Signal

Processing," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 69–87, jan 2012. doi: 10.1109/T-AFFC.2011.27

K. Vytal and S. Hamann, "Neuroimaging support for discrete neural correlates of basic emotions: a voxel-based meta-analysis," *Journal of cognitive neuroscience*, vol. 22, no. 12, pp. 2864–85, dec 2010. doi: 10.1162/jocn.2009.21366

J. Weng, "Developmental robotics: Theory and experiments," *International Journal of Humanoid Robotics*, vol. 1, no. 02, pp. 199–236, 2004.

B. Wicker, C. Keysers, J. Plailly, J.-P. Royet, V. Gallese, and G. Rizzolatti, "Both of Us Disgusted in My Insula: The Common Neural Basis of Seeing and Feeling Disgust," *Neuron*, vol. 40, no. 3, pp. 655–664, oct 2003. doi: 10.1016/S0896-6273(03)00679-2

M. Wilson and G. Knoblich, "The case for motor involvement in perceiving conspecifics," *Psychological bulletin*, vol. 131, no. 3, pp. 460–473, may 2005. doi: 10.1037/0033-2909.131.3.460

M. Wimmer, B. A. MacDonald, D. Jayamuni, and A. Yadav, "Facial Expression Recognition for Human-Robot Interaction - A Prototype," in *Robot Vision*, ser. Lecture Notes in Computer Science, G. Sommer and R. Klette, Eds. Springer Berlin Heidelberg, 2008, vol. 4931, pp. 139–152. doi: 10.1007/978-3-540-78157-8_11

M. Zecca, N. Endo, S. Momoki, K. Itoh, and A. Takanishi, "Design of the humanoid robot KOBIAN - preliminary analysis of facial and whole body emotion expression capabilities-," in *Humanoids 2008 - 8th IEEE-RAS International Conference on Humanoid Robots*. Daejeon, South Korea: IEEE, dec 2008, pp. 487–492. doi: 10.1109/ICHR.2008.4755969

Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: audio, visual, and spontaneous expressions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 1, pp. 39–58, jan 2009. doi: 10.1109/TPAMI.2008.52

X. Zhao, E. Dellandrea, L. Chen, and I. A. Kakadiaris, "Accurate Landmarking of Three-Dimensional Facial Data in the Presence of Facial Expressions and Occlusions Using a

Three-Dimensional Statistical Facial Feature Model," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 41, no. 5, pp. 1417–1428, 2011. doi: 10.1109/TSMCB.2011.2148711

This page was intentionally left blank.

Appendices

Timings of results

Label	Time (s)
T1	5
T2	43
T3	46

Table 1: Timings for scenario 1.1.

Label	Time (s)
T1	5
T2	14
T3	21
T4	30

Table 2: Timings for scenario 1.2.

Label	Time (s)
T1	2
T2	8
T3	15
T4	23

Table 3: Timings for scenario 2.1.

Label	Time (s)
T1	3
T2	17
T3	20
T4	29

Table 4: Timings for scenario 2.2.

Label	Time (s)
T1	3
T2	13
T3	17
T4	23
T5	38
T6	42
T7	58
T8	64

Table 5: Timings for experiment 3.

Facial movements and emotions

Emotion	AUs and human movements
Disgust	4+12+25 / 4+25 / 9+15 / 9 / 4+9 / 4+5+9
Anger	4+5 / 4+7 / 4+Head Movement High
Fear	1+2+5+20+26 / 4+20 / 1+20
	5 + Hand Movement Low
Sadness	1+4+15 / 1+4 / 1+15 / 4+15
	1 + Hand Movement Low
	15 + Hand Movement Low
Neutral	Face detected
Surprise	1+2 / 1+2+5 / 1+2+26 / 1+2+5+26
Happiness	12 / 1+2+12 / 12+26

Table 6: Combinations of AUs and human movements capable of activating an emotional state detection (The detection of an Action Unit implies that the face is detected).

Emotion	**Movement velocity**	**Error detection**
Disgust	Medium	All
Anger	Medium	All
Fear	Low	All
Sadness	Medium	All
Neutral	Medium	All
Surprise	Medium	All
Happiness	High	Execution
Emotion	**Decisions**	**Verbalization**
Disgust		
Anger	Abort current action	
Fear		Provide more information
Sadness		
Neutral		
Surprise		
Happiness	Faster decisions	

Table 7: Influence of emotions in various aspects of the robot behavior.

Dynamic field neural populations

Four DNFs representing hand gestures	Sub-population 'label'	Meaning
Reaching/pointing towards an object:		
	R1	Wheel
	R2	Nut
	R3	Column 1
	R4	Column 2
	R5	Column 3
	R6	Column 4
	R7	Top Floor
	R8	Other
Reaching not towards an object:		
Hold Out gesture	H	Hold out empty hand
Grip type:		
	G1	Above grip
	G2	Side grip
	G3	Bottom grip

Table 8: Action Observation Layer: Hand gestures.

Three DNFs quantifying movements	Sub-population 'label'	Meaning
Body part:	A1	Idle (no movement)
Head	A2	Low (little movement)
Hand	A3	Medium
Body	A4	High (intense movement)

Table 9: Action Observation Layer: Quantity of Movement.

Four DNFs representing facial movements	Sub-population 'label'	Meaning
Face detection		
	F	Face is detected
Eyebrows		
	1	Action Unit 1
	2	Action Unit 2
	4	Action Unit 4
Eyes		
	5	Action Unit 5
	7	Action Unit 7
Mouth		
	12	Action Unit 12
	15	Action Unit 15
	18	Action Unit 18
	20	Action Unit 20
	25	Action Unit 25
	26	Action Unit 26

Table 10: Action Observation Layer: Facial movements.

Two DNF representing the distribution of relevant pieces in the two workspaces	**Sub-population 'label'**	**Meaning: in the workspace there exists a...**
Robot's workspace	O1	Wheel
Human's workspace	O2	Nut
	O3	Column 1
	O4	Column 2
	O5	Column 3
	O6	Column 4
	O7	Top Floor

Table 11: Object Memory Layer.

DNF representing	**Sub-population 'label'**	**Meaning**
Emotional state	E1	Disgust
	E2	Anger
	E3	Fear
	E4	Sadness
	E5	Neutral
	E6	Surprise
	E7	Happy

Table 12: Emotional State Layer.

DNF representing	Sub-population 'label'	Meaning: human wants to...
Intention	I1	Insert Wheel
	I2	Insert Nut
	I3	Insert Column 1
	I4	Insert Column 2
	I5	Insert Column 3
	I6	Insert Column 4
	I7	Insert Top Floor
	I8	Insert Other object
	I9	Hand over Wheel
	I10	Hand over Nut
	I11	Hand over Column 1
	I12	Hand over Column 2
	I13	Hand over Column 3
	I14	Hand over Column 4
	I15	Hand over Top Floor
	I16	Hand over Other object
	I17	Hold Base

Table 13: Intention Layer.

Two DNFs representing common sub goals	**Sub-population 'label'**	**Meaning**
Past	C1	Inserted Wheel 1
Present	C2	Inserted Wheel 2
	C3	Inserted Nut 1
	C4	Inserted Nut 2
	C5	Inserted Column 1
	C6	Inserted Column 2
	C7	Inserted Column 3
	C8	Inserted Column 4
	C9	Inserted Top Floor

Table 14: Common Sub Goals Layer

Table 15: Error Monitoring Layer: Intention.

One DNF representing errors in intention	**Sub-population 'label'**	**Meaning**
Human already inserted it	E1	Insert Wheel
	E2	Insert Nut
	E3	Insert Column 1
	E4	Insert Column 2
	E5	Insert Column 3
	E6	Insert Column 4
	E7	Insert Top Floor
Robot already inserted it	E8	Hand over Wheel
	E9	Hand over Nut
	E10	Hand over Column 1
	E11	Hand over Column 2
	E12	Hand over Column 3

Table 15: (continued).

One DNF representing errors in intention	**Sub-population 'label'**	**Meaning**
	E13	Hand over Column 4
	E14	Hand over Top Floor
Human cannot insert it yet	E15	Insert Wheel
	E16	Insert Nut
	E17	Insert Column 1
	E18	Insert Column 2
	E19	Insert Column 3
	E20	Insert Column 4
	E21	Insert Top Floor
Robot cannot insert it yet	E22	Hand over Wheel
	E23	Hand over Nut
	E24	Hand over Column 1
	E25	Hand over Column 2
	E26	Hand over Column 3
	E27	Hand over Column 4
	E28	Hand over Top Floor
Related with other objects	E29	Insert other object
	E30	Hand over other object
	E31	Hold Base

One DNF representing errors in execution	Sub-population 'label'	Meaning
Object inserted in the wrong order	E1	Inserted Wheel
	E2	Inserted Nut
	E3	Inserted Column 1
	E4	Inserted Column 2
	E5	Inserted Column 3
	E6	Inserted Column 4
	E7	Inserted Top Floor
Object dropped	E8	Dropped Wheel
	E9	Dropped Nut
	E10	Dropped Column 1
	E11	Dropped Column 2
	E12	Dropped Column 3
	E13	Dropped Column 4
	E14	Dropped Top Floor

Table 16: Error Monitoring Layer: Execution.

Table 17: Error Monitoring Layer: Means.

One DNF representing errors in means	**Sub-population 'label'**	**Meaning**
Errors in object insertion	E1	Wheel
	E2	Nut
	E3	Column 1
	E4	Column 2
	E5	Column 3
	E6	Column 4
	E7	Top Floor
Human grasping an object from the robot's hand	E8	Wheel
	E9	Nut
	E10	Column 1
	E11	Column 2
	E12	Column 3
	E13	Column 4
	E14	Top Floor
Hand over object	E15	Wheel
	E16	Nut
	E17	Column 1
	E18	Column 2
	E19	Column 3
	E20	Column 4
	E21	Top Floor
Point to an object that cannot be used	E22	Wheel
	E23	Nut
	E24	Column 1
	E25	Column 2

Table 17: (continued).

One DNF representing errors in means	**Sub-population 'label'**	**Meaning**
	E26	Column 3
	E27	Column 4
	E28	Top Floor
Other errors	E29	Hand over
	E30	Reach Base

Table 18: ASFA/AEFA: Action simulation/execution emotion directed facial actions.

One DNF representing chains of emotion directed facial actions	**Sub-population 'label'**	**Meaning**
Disgust	F1	Lower eyebrows, smile, open mouth slightly.
	F2	Lower eyebrows, open mouth slightly.
	F3	Nose wrinkles, pull lip corners down.
	F4	Nose wrinkles.
	F5	Lower eyebrows, nose wrinkles.
	F6	Lower eyebrows, eyes wide open, nose wrinkles.
Anger	F7	Lower eyebrows, eyes wide open.
	F8	Lower eyebrows, eyes half closed.
	F9	Lower eyebrows.
Fear	F10	Raise eyebrows + eyes wide open, stretch lips, mouth open.
	F11	Lower eyebrows, stretch lips.
	F12	Raise inner part of eyebrows, stretch lips.
	F13	Eyes wide open.
Sadness	F14	Raise inner part of eyebrows, lower eyebrows, lip corners down.
	F15	Raise inner part of eyebrows, Lower eyebrows.
	F16	Raise inner part of eyebrows, lip corners down.
	F17	Lower eyebrows, lip corners down.

Table 18: (continued).

One DNF representing chains of emotion directed facial actions	**Sub-population 'label'**	**Meaning**
	F18	Raise inner part of eyebrows.
	F19	Lip corners down.
Neutral	F20	Face detected with no AUs.
Surprise	F21	Raise eyebrows.
	F22	Raise eyebrows, eyes wide open.
	F23	Raise eyebrows, mouth open.
	F24	Raise eyebrows, eyes wide open, mouth open.
Happiness	F25	Smile.
	F26	Raise eyebrows, smile.
	F27	Smile, mouth open.

Table 19: ASHA/AEHA: Action simulation/execution of goal directed hand actions and communicative gestures.

One DNF representing chains of goal directed hand actions	**Sub-population 'label'**	**Meaning**
Reach to grasp object from the table and insert it	A1	Wheel
	A2	Nut
	A3	Column 1
	A4	Column 2
	A5	Column 3
	A6	Column 4
	A7	Top Floor
	A8	Other object
Reach to grasp object from the partner's hand and insert it	A9	Wheel
	A10	Nut
	A11	Column 1
	A12	Column 2
	A13	Column 3
	A14	Column 4
	A15	Top Floor
	A16	Other object
Reach to grasp object and hand over	A17	Wheel
	A18	Nut
	A19	Column 1
	A20	Column 2
	A21	Column 3
	A22	Column 4
	A23	Top Floor
	A24	Other object

Table 19: (continued).

One DNF representing chains of goal directed hand actions	**Sub-population 'label'**	**Meaning**
Pointing to object	A25	Wheel
	A26	Nut
	A27	Column 1
	A28	Column 2
	A29	Column 3
	A30	Column 4
	A31	Top Floor
	A32	Other object
Other	A33	Hold Out empty hand
	A34	Reach Base
	A35	Communicate error (AEHA only)

This page was intentionally left blank.

Numerical values for the dynamic field parameters

Parameters	Values
Number of pools of neurons, i.e. subpopulations: N_{pools}	8
Centers of the subpopulations: x_{m} (in Eq. 3.7)	$x_{\text{m}} = (m-1)\times 10$, $m = 1, ..., N_{\text{pools}}$
Sampling distance along x : dx' (Eq. 3.1)	1
τ_{i} (Eq. 3.1)	$5.0 \times dt$ (Where dt is the computation cycle)
h_{i} (Eq. 3.1)	$-0.1 \times max\ W$
Intrafield interaction: $w(x) = w_{\text{exc}}(x) - w_{\text{ini}}(x)$, where w_{exc} and w_{ini} are given by Eq. 3.5	
w_{exc}:	
A_{i} (Eq. 3.5)	10
σ_{i} (Eq. 3.5)	1.5
$w_{\text{inhib,i}}$ (Eq. 3.5)	0
w_{ini}:	
A_{i} (Eq. 3.5)	8
σ_{i} (Eq. 3.5)	2
$w_{\text{inhib,i}}$ (Eq. 3.5)	0

Table 20: Layer AOL: Reaching & Pointing.

Parameters	**Values**
Number of pools of neurons, i.e. subpopulations: N_{pools}	1
Centers of the subpopulations: x_{m} (in Eq. 3.7)	$x_{\text{m}} = (m-1)\times 10$, $m = 1, ..., N_{\text{pools}}$
Sampling distance along x : dx' (Eq. 3.1)	1
τ_{i} (Eq. 3.1)	$5.0 \times dt$ (Where dt is the computation cycle)
h_{i} (Eq. 3.1)	$-0.1 \times max\ W$
Intrafield interaction: $w(x) = w_{\text{exc}}(x) - w_{\text{ini}}(x)$, where w_{exc} and w_{ini} are given by Eq. 3.5	
w_{exc}:	
A_{i} (Eq. 3.5)	10
σ_{i} (Eq. 3.5)	1.5
$w_{\text{inhib,i}}$ (Eq. 3.5)	0
w_{ini}:	
A_{i} (Eq. 3.5)	8
σ_{i} (Eq. 3.5)	2
$w_{\text{inhib,i}}$ (Eq. 3.5)	0

Table 21: Layer AOL: Hold Out & Face Detect.

Parameters	**Values**
Number of pools of neurons, i.e. subpopulations: N_{pools}	3
Centers of the subpopulations: x_{m} (in Eq. 3.7)	$x_{\text{m}} = (m-1)\times 10$, $m = 1, ..., N_{\text{pools}}$
Sampling distance along x : dx' (Eq. 3.1)	1
τ_{i} (Eq. 3.1)	$5.0 \times dt$ (Where dt is the computation cycle)
h_{i} (Eq. 3.1)	$-0.1 \times max\ W$
Intrafield interaction: $w(x) = w_{\text{exc}}(x) - w_{\text{ini}}(x)$, where w_{exc} and w_{ini} are given by Eq. 3.5	
w_{exc}:	
A_{i} (Eq. 3.5)	10
σ_{i} (Eq. 3.5)	1.5
$w_{\text{inhib,i}}$ (Eq. 3.5)	0
w_{ini}:	
A_{i} (Eq. 3.5)	8
σ_{i} (Eq. 3.5)	2
$w_{\text{inhib,i}}$ (Eq. 3.5)	0

Table 22: Layer AOL: Grip Type.

Parameters	**Values**
Number of pools of neurons, i.e. subpopulations: N_{pools}	3
Centers of the subpopulations: x_{m} (in Eq. 3.7)	$x_{\text{m}} = (m-1)\times 10$, $m = 1, ..., N_{\text{pools}}$
Sampling distance along x : dx' (Eq. 3.1)	1
τ_{i} (Eq. 3.1)	$5.0 \times dt$ (Where dt is the computation cycle)
h_{i} (Eq. 3.1)	$-0.1 \times max\ W$
Intrafield interaction: $w(x) = w_{\text{exc}}(x) - w_{\text{ini}}(x)$, where w_{exc} and w_{ini} are given by Eq. 3.5	
w_{exc}:	
A_{i} (Eq. 3.5)	10
σ_{i} (Eq. 3.5)	1.5
$w_{\text{inhib,i}}$ (Eq. 3.5)	0
w_{ini}:	
A_{i} (Eq. 3.5)	8
σ_{i} (Eq. 3.5)	2
$w_{\text{inhib,i}}$ (Eq. 3.5)	0

Table 23: Layer AOL: Eyebrows.

Parameters	**Values**
Number of pools of neurons, i.e. subpopulations: N_{pools}	2
Centers of the subpopulations: x_{m} (in Eq. 3.7)	$x_{\text{m}} = (m-1)\times 10$, $m = 1, ..., N_{\text{pools}}$
Sampling distance along x : dx' (Eq. 3.1)	1
τ_{i} (Eq. 3.1)	$5.0 \times dt$ (Where dt is the computation cycle)
h_{i} (Eq. 3.1)	$-0.1 \times max\ W$
Intrafield interaction: $w(x) = w_{\text{exc}}(x) - w_{\text{ini}}(x)$, where w_{exc} and w_{ini} are given by Eq. 3.5	
w_{exc}:	
A_{i} (Eq. 3.5)	10
σ_{i} (Eq. 3.5)	1.5
$w_{\text{inhib,i}}$ (Eq. 3.5)	0
w_{ini}:	
A_{i} (Eq. 3.5)	8
σ_{i} (Eq. 3.5)	2
$w_{\text{inhib,i}}$ (Eq. 3.5)	0

Table 24: Layer AOL: Eyes.

Parameters	**Values**
Number of pools of neurons, i.e. subpopulations: N_{pools}	6
Centers of the subpopulations: x_{m} (in Eq. 3.7)	$x_{\text{m}} = (m-1)\times 10$, $m = 1, ..., N_{\text{pools}}$
Sampling distance along x : dx' (Eq. 3.1)	1
τ_{i} (Eq. 3.1)	$5.0 \times dt$ (Where dt is the computation cycle)
h_{i} (Eq. 3.1)	$-0.1 \times max\ W$
Intrafield interaction: $w(x) = w_{\text{exc}}(x) - w_{\text{ini}}(x)$, where w_{exc} and w_{ini} are given by Eq. 3.5	
w_{exc}:	
A_{i} (Eq. 3.5)	10
σ_{i} (Eq. 3.5)	1.5
$w_{\text{inhib,i}}$ (Eq. 3.5)	0
w_{ini}:	
A_{i} (Eq. 3.5)	8
σ_{i} (Eq. 3.5)	2
$w_{\text{inhib,i}}$ (Eq. 3.5)	0

Table 25: Layer AOL: Mouth.

Parameters	**Values**
Number of pools of neurons, i.e. subpopulations: N_{pools}	4
Centers of the subpopulations: x_{m} (in Eq. 3.7)	$x_{\text{m}} = (m-1)\times 10$, $m = 1, ..., N_{\text{pools}}$
Sampling distance along x : dx' (Eq. 3.1)	1
τ_{i} (Eq. 3.1)	$5.0 \times dt$ (Where dt is the computation cycle)
h_{i} (Eq. 3.1)	$-0.1 \times max\ W$
Intrafield interaction: $w(x) = w_{\text{exc}}(x) - w_{\text{ini}}(x)$, where w_{exc} and w_{ini} are given by Eq. 3.5	
w_{exc}:	
A_{i} (Eq. 3.5)	10
σ_{i} (Eq. 3.5)	1.5
$w_{\text{inhib,i}}$ (Eq. 3.5)	0
w_{ini}:	
A_{i} (Eq. 3.5)	8
σ_{i} (Eq. 3.5)	2
$w_{\text{inhib,i}}$ (Eq. 3.5)	0

Table 26: Layer AOL: QoMHand & QoMBody & QoMHead.

Parameters	**Values**
Number of pools of neurons, i.e. subpopulations: N_{pools}	7
Centers of the subpopulations: x_{m} (in Eq. 3.7)	$x_{\text{m}} = (m-1)\times 10$, $m = 1, ..., N_{\text{pools}}$
Sampling distance along x : dx' (Eq. 3.1)	1
τ_{i} (Eq. 3.1)	$5.0 \times dt$ (Where dt is the computation cycle)
h_{i} (Eq. 3.1)	$-0.1 \times max\ W$
Intrafield interaction: $w(x) = w_{\text{exc}}(x) - w_{\text{ini}}(x)$, where w_{exc} and w_{ini} are given by Eq. 3.5	
w_{exc}:	
A_{i} (Eq. 3.5)	10
σ_{i} (Eq. 3.5)	1.5
$w_{\text{inhib,i}}$ (Eq. 3.5)	0
w_{ini}:	
A_{i} (Eq. 3.5)	8
σ_{i} (Eq. 3.5)	2
$w_{\text{inhib,i}}$ (Eq. 3.5)	0

Table 27: Layer OML.

Parameters	**Values**
Number of pools of neurons, i.e. subpopulations: N_{pools}	9
Centers of the subpopulations: x_{m} (in Eq. 3.7)	$x_{\text{m}} = (m-1)\times 10$, $m = 1, ..., N_{\text{pools}}$
Sampling distance along x : dx' (Eq. 3.1)	1
τ_{i} (Eq. 3.1)	$5.0 \times dt$ (Where dt is the computation cycle)
h_{i} (Eq. 3.1)	$-0.1 \times max\ W$
Intrafield interaction: $w(x) = w_{\text{exc}}(x) - w_{\text{ini}}(x)$, where w_{exc} and w_{ini} are given by Eq. 3.5	
w_{exc}:	
A_{i} (Eq. 3.5)	10
σ_{i} (Eq. 3.5)	1.5
$w_{\text{inhib,i}}$ (Eq. 3.5)	0
w_{ini}:	
A_{i} (Eq. 3.5)	8
σ_{i} (Eq. 3.5)	2
$w_{\text{inhib,i}}$ (Eq. 3.5)	0

Table 28: Layer CSGL.

Parameters	**Values**
Number of pools of neurons, i.e. subpopulations: N_{pools}	34
Centers of the subpopulations: x_{m} (in Eq. 3.7)	$x_{\text{m}} = (m-1)\times 10$, $m = 1, ..., N_{\text{pools}}$
Sampling distance along x : dx' (Eq. 3.1)	1
τ_{i} (Eq. 3.1)	$5.0 \times dt$ (Where dt is the computation cycle)
h_{i} (Eq. 3.1)	$-0.7 \times max\ W$
A_{i} (Eq. 3.5)	12.5
σ_{i} (Eq. 3.5)	2.0
$w_{\text{inhib,i}}$ (Eq. 3.5)	9.5

Table 29: Layer ASL: ASHA.

Parameters	**Values**
Number of pools of neurons, i.e. subpopulations: N_{pools}	27
Centers of the subpopulations: x_{m} (in Eq. 3.7)	$x_{\text{m}} = (m-1)\times 10$, $m = 1, ..., N_{\text{pools}}$
Sampling distance along x : dx' (Eq. 3.1)	1
τ_{i} (Eq. 3.1)	$5.0 \times dt$ (Where dt is the computation cycle)
h_{i} (Eq. 3.1)	$-0.7 \times max\ W$
A_{i} (Eq. 3.5)	12.5
σ_{i} (Eq. 3.5)	2.0
$w_{\text{inhib,i}}$ (Eq. 3.5)	9.5

Table 30: Layer ASL: ASFA.

Parameters	**Values**
Number of pools of neurons, i.e. subpopulations: N_{pools}	17
Centers of the subpopulations: x_{m} (in Eq. 3.7)	$x_{\text{m}} = (m-1)\times 10$, $m = 1, ..., N_{\text{pools}}$
Sampling distance along x : dx' (Eq. 3.1)	1
τ_{i} (Eq. 3.1)	$5.0 \times dt$ (Where dt is the computation cycle)
h_{i} (Eq. 3.1)	$-1.1 \times max\ W$
A_{i} (Eq. 3.5)	6.0
σ_{i} (Eq. 3.5)	2.0
$w_{\text{inhib,i}}$ (Eq. 3.5)	4.5

Table 31: Layer IL.

Parameters	**Values**
Number of pools of neurons, i.e. subpopulations: N_{pools}	7
Centers of the subpopulations: x_{m} (in Eq. 3.7)	$x_{\text{m}} = (m-1)\times 10$, $m = 1, ..., N_{\text{pools}}$
Sampling distance along x : dx' (Eq. 3.1)	1
τ_{i} (Eq. 3.1)	$5.0 \times dt$ (Where dt is the computation cycle)
h_{i} (Eq. 3.1)	$-1.1 \times max\ W$
A_{i} (Eq. 3.5)	6.0
σ_{i} (Eq. 3.5)	2.0
$w_{\text{inhib,i}}$ (Eq. 3.5)	4.5

Table 32: Layer ESL.

Parameters	**Values**
Number of pools of neurons, i.e. subpopulations: N_{pools}	35
Centers of the subpopulations: x_{m} (in Eq. 3.7)	$x_{\text{m}} = (m-1)\times 10$, $m = 1, ..., N_{\text{pools}}$
Sampling distance along x : dx' (Eq. 3.1)	1
τ_{i} (Eq. 3.1)	$10.0 \times dt$ (Where dt is the computation cycle)
h_{i} (Eq. 3.1)	$-0.5 \times max\ W$
A_{i} (Eq. 3.5)	12.5
σ_{i} (Eq. 3.5)	2.0
$w_{\text{inhib,i}}$ (Eq. 3.5)	9.5

Table 33: Layer AEL: AEHA.

Parameters	**Values**
Number of pools of neurons, i.e. subpopulations: N_{pools}	27
Centers of the subpopulations: x_{m} (in Eq. 3.7)	$x_{\text{m}} = (m-1)\times 10$, $m = 1, ..., N_{\text{pools}}$
Sampling distance along x : dx' (Eq. 3.1)	1
τ_{i} (Eq. 3.1)	$10.0 \times dt$ (Where dt is the computation cycle)
h_{i} (Eq. 3.1)	$-0.5 \times max\ W$
A_{i} (Eq. 3.5)	10
σ_{i} (Eq. 3.5)	2.0
$w_{\text{inhib,i}}$ (Eq. 3.5)	8.5

Table 34: Layer AEL: AEFA.

Parameters	**Values**
Number of pools of neurons, i.e. subpopulations: N_{pools}	31
Centers of the subpopulations: x_{m} (in Eq. 3.7)	$x_{\text{m}} = (m-1)\times 10$, $m = 1, ..., N_{\text{pools}}$
Sampling distance along x : dx' (Eq. 3.1)	1
τ_{i} (Eq. 3.1)	$5.0 \times dt$ (Where dt is the computation cycle)
h_{i} (Eq. 3.1)	$-2.0 \times max\ W$
A_{i} (Eq. 3.5)	6.0
σ_{i} (Eq. 3.5)	2.0
$w_{\text{inhib,i}}$ (Eq. 3.5)	4.5

Table 35: Layer EML: Error in Intention.

Parameters	**Values**
Number of pools of neurons, i.e. subpopulations: N_{pools}	30
Centers of the subpopulations: x_{m} (in Eq. 3.7)	$x_{\text{m}} = (m-1)\times 10$, $m = 1, ..., N_{\text{pools}}$
Sampling distance along x : dx' (Eq. 3.1)	1
τ_{i} (Eq. 3.1)	$5.0 \times dt$ (Where dt is the computation cycle)
h_{i} (Eq. 3.1)	$-2.0 \times max\ W$
A_{i} (Eq. 3.5)	6.0
σ_{i} (Eq. 3.5)	2.0
$w_{\text{inhib,i}}$ (Eq. 3.5)	4.5

Table 36: Layer EML: Error in Means.

Parameters	**Values**
Number of pools of neurons, i.e. subpopulations: N_{pools}	14
Centers of the subpopulations: x_{m} (in Eq. 3.7)	$x_{\text{m}} = (m-1)\times 10$, $m = 1, ..., N_{\text{pools}}$
Sampling distance along x : dx' (Eq. 3.1)	1
τ_{i} (Eq. 3.1)	$5.0 \times dt$ (Where dt is the computation cycle)
h_{i} (Eq. 3.1)	$-2.0 \times max\ W$
A_{i} (Eq. 3.5)	6.0
σ_{i} (Eq. 3.5)	2.0
$w_{\text{inhib,i}}$ (Eq. 3.5)	4.5

Table 37: Layer EML: Error in Execution.

Numerical values for the inter-field synaptic weights

Table 38: Synaptic weights from OML to ASL.

DNFs in OML→ASL	**Synaptic link $a_{\text{ASL,OML}}$**	**Weight**
OML Robot→ASHA	$a_{\text{A17,O1}}$	-1.2
	$a_{\text{A25,O1}}$	0.8
	$a_{\text{A18,O2}}$	-1.2
	$a_{\text{A26,O2}}$	0.8
	$a_{\text{A19,O3}}$	-1.2
	$a_{\text{A27,O3}}$	0.8
	$a_{\text{A20,O4}}$	-1.2
	$a_{\text{A28,O4}}$	0.8
	$a_{\text{A21,O5}}$	-1.2
	$a_{\text{A29,O5}}$	0.8
	$a_{\text{A22,O6}}$	-1.2
	$a_{\text{A30,O6}}$	0.8
OML Human→ASHA	$a_{\text{A1,O1}}$	0.5
	$a_{\text{A9,O1}}$	0.5
	$a_{\text{A17,O1}}$	-1.2
	$a_{\text{A25,O1}}$	-1.2

Table 38: (continued).

DNFs in OML→ASL	**Synaptic link $a_{\mathbf{ASL,OML}}$**	**Weight**
	$a_{A2,O2}$	0.5
	$a_{A10,O2}$	0.5
	$a_{A18,O2}$	-1.2
	$a_{A26,O2}$	-1.2
	$a_{A3,O3}$	0.5
	$a_{A11,O3}$	0.5
	$a_{A19,O3}$	-1.2
	$a_{A27,O3}$	-1.2
	$a_{A4,O4}$	0.5
	$a_{A12,O4}$	0.5
	$a_{A20,O4}$	-1.2
	$a_{A28,O4}$	-1.2
	$a_{A5,O5}$	0.5
	$a_{A13,O5}$	0.5
	$a_{A21,O5}$	-1.2
	$a_{A29,O5}$	-1.2
	$a_{A6,O6}$	0.5
	$a_{A14,O6}$	0.5
	$a_{A22,O6}$	-1.2
	$a_{A30,O6}$	-1.2
	$a_{A7,O7}$	0.5
	$a_{A23,O7}$	-1.2
	$a_{A31,O7}$	-1.2

Table 39: Synaptic weights from AOL to ASL.

DNFs in AOL→ASL	**Synaptic link $a_{\mathbf{ASL,AOL}}$**	**Weight**
Reaching→ASHA	$a_{A1,R1}$	1.5
	$a_{A17,R1}$	1.5
	$a_{A2,R2}$	1.5
	$a_{A18,R2}$	1.5
	$a_{A3,R3}$	1.5
	$a_{A19,R3}$	1.5
	$a_{A4,R4}$	1.5
	$a_{A20,R4}$	1.5
	$a_{A5,R5}$	1.5
	$a_{A21,R5}$	1.5
	$a_{A6,R6}$	1.5
	$a_{A22,R6}$	1.5
	$a_{A7,R7}$	1.5
	$a_{A8,R8}$	2.0
Hold Out→ASHA	$a_{A33,H}$	2.5
Grip Type→ASHA	$a_{A1,G1}$	0.5
	$a_{A17,G1}$	-1.2
	$a_{A2,G1}$	-1.2
	$a_{A18,G1}$	0.5
	$a_{A3,G1}$	0.5
	$a_{A19,G1}$	-1.2
	$a_{A4,G1}$	0.5
	$a_{A20,G1}$	-1.2
	$a_{A5,G1}$	0.5
	$a_{A21,G1}$	-1.2
	$a_{A6,G1}$	0.5

Table 39: (continued).

DNFs in AOL→ASL	**Synaptic link $a_{\mathbf{ASL,AOL}}$**	**Weight**
	$a_{\text{A22,G1}}$	-1.2
	$a_{\text{A7,G1}}$	-1.2
	$a_{\text{A23,G1}}$	-1.2
	$a_{\text{A1,G2}}$	-1.2
	$a_{\text{A17,G2}}$	0.5
	$a_{\text{A2,G2}}$	0.5
	$a_{\text{A18,G2}}$	-1.2
	$a_{\text{A3,G2}}$	-1.2
	$a_{\text{A19,G2}}$	-1.2
	$a_{\text{A4,G2}}$	-1.2
	$a_{\text{A20,G2}}$	-1.2
	$a_{\text{A5,G2}}$	-1.2
	$a_{\text{A21,G2}}$	-1.2
	$a_{\text{A6,G2}}$	-1.2
	$a_{\text{A22,G2}}$	-1.2
	$a_{\text{A7,G2}}$	0.5
	$a_{\text{A1,G3}}$	-1.2
	$a_{\text{A17,G3}}$	-1.2
	$a_{\text{A2,G3}}$	-1.2
	$a_{\text{A18,G3}}$	-1.2
	$a_{\text{A3,G3}}$	-1.2
	$a_{\text{A19,G3}}$	0.5
	$a_{\text{A4,G3}}$	-1.2
	$a_{\text{A20,G3}}$	0.5
	$a_{\text{A5,G3}}$	-1.2
	$a_{\text{A21,G3}}$	0.5

Table 39: (continued).

DNFs in AOL→ASL	**Synaptic link $a_{\mathbf{ASL,AOL}}$**	**Weight**
	$a_{A6,G3}$	-1.2
	$a_{A22,G3}$	0.5
	$a_{A7,G3}$	-1.2
Pointing→ASHA	$a_{A25,P1}$	2.0
	$a_{A26,P2}$	2.0
	$a_{A27,P3}$	2.0
	$a_{A28,P4}$	2.0
	$a_{A29,P5}$	2.0
	$a_{A30,P6}$	2.0
	$a_{A31,P7}$	2.0
	$a_{A32,P8}$	2.0

Table 41: Synaptic weights from AOL to ASL.

DNFs in AOL→ASL	**Synaptic link $a_{\mathbf{ASL,AOL}}$**	**Weight**
Face Detect→ASFA	$a_{F20,F}$	1.1
Eyebrows→ASFA	$a_{F10,1}$	0.3
	$a_{F12,1}$	0.55
	$a_{F14,1}$	0.4
	$a_{F15,1}$	0.55
	$a_{F16,1}$	0.55
	$a_{F18,1}$	0.95
	$a_{F20,1}$	-0.2
	$a_{F21,1}$	0.55
	$a_{F22,1}$	0.4
	$a_{F23,1}$	0.4
	$a_{F24,1}$	0.35

Table 41: (continued).

DNFs in AOL→ASL	**Synaptic link $a_{\mathbf{ASL,AOL}}$**	**Weight**
	$a_{F26,1}$	0.4
	$a_{F10,2}$	0.3
	$a_{F20,2}$	-0.2
	$a_{F21,2}$	0.55
	$a_{F22,2}$	0.4
	$a_{F23,2}$	0.4
	$a_{F24,2}$	0.35
	$a_{F26,2}$	0.4
	$a_{F1,4}$	0.4
	$a_{F2,4}$	0.55
	$a_{F5,4}$	0.55
	$a_{F6,4}$	0.4
	$a_{F7,4}$	0.65
	$a_{F8,4}$	0.55
	$a_{F9,4}$	0.95
	$a_{F11,4}$	0.55
	$a_{F13,4}$	0.35
	$a_{F14,4}$	0.4
	$a_{F15,4}$	0.55
	$a_{F17,4}$	0.55
	$a_{F20,4}$	-0.2
Eyes→ASFA	$a_{F6,5}$	0.4
	$a_{F7,5}$	0.75
	$a_{F10,5}$	0.35
	$a_{F13,5}$	0.95
	$a_{F20,5}$	-0.2

Table 41: (continued).

DNFs in AOL→ASL	**Synaptic link $a_{\mathbf{ASL,AOL}}$**	**Weight**
	$a_{F22,5}$	0.4
	$a_{F24,5}$	0.35
	$a_{F8,7}$	0.55
	$a_{F20,7}$	-0.2
Mouth→ASFA	$a_{F1,12}$	0.4
	$a_{F20,12}$	-0.2
	$a_{F25,12}$	0.95
	$a_{F26,12}$	0.4
	$a_{F27,12}$	0.55
	$a_{F14,15}$	0.4
	$a_{F16,15}$	0.55
	$a_{F17,15}$	0.55
	$a_{F19,15}$	0.95
	$a_{F20,15}$	-0.2
	$a_{F10,20}$	0.3
	$a_{F11,20}$	0.55
	$a_{F12,20}$	0.55
	$a_{F20,20}$	-0.2
	$a_{F1,25}$	0.4
	$a_{F2,25}$	0.55
	$a_{F20,25}$	-0.2
	$a_{F10,26}$	0.3
	$a_{F20,26}$	-0.2
	$a_{F23,26}$	0.4
	$a_{F24,26}$	0.35
	$a_{F27,26}$	0.55

DNFs in CSGL→ASL	**Synaptic link $a_{\text{ASL,CSGL}}$**	**Weight**
CSGL Present→ASHA	$a_{\text{A17,C1}}$	0.8
	$a_{\text{A1,C2}}$	0.85
	$a_{\text{A9,C2}}$	0.8
	$a_{\text{A25,C2}}$	0.8
	$a_{\text{A18,C3}}$	0.8
	$a_{\text{A2,C4}}$	0.85
	$a_{\text{A10,C4}}$	0.8
	$a_{\text{A26,C4}}$	0.8
	$a_{\text{A19,C5}}$	0.8
	$a_{\text{A4,C6}}$	0.85
	$a_{\text{A12,C6}}$	0.8
	$a_{\text{A28,C6}}$	0.8
	$a_{\text{A5,C7}}$	0.85
	$a_{\text{A13,C7}}$	0.8
	$a_{\text{A29,C7}}$	0.8
	$a_{\text{A6,C8}}$	0.85
	$a_{\text{A14,C8}}$	0.8
	$a_{\text{A29,C8}}$	0.8
	$a_{\text{A7,C9}}$	0.85

Table 40: Synaptic weights from CSGL to ASL.

Table 42: Synaptic weights from ASL to IL.

DNFs in ASL→IL	**Synaptic link $a_{IL,ASL}$**	**Weight**
ASHA→IL	$a_{I1,A1}$	1.0
	$a_{I2,A2}$	1.0
	$a_{I3,A3}$	1.0
	$a_{I4,A4}$	1.0
	$a_{I5,A5}$	1.0
	$a_{I6,A6}$	1.0
	$a_{I7,A7}$	1.0
	$a_{I8,A8}$	1.0
	$a_{I1,A9}$	1.0
	$a_{I2,A10}$	1.0
	$a_{I3,A11}$	1.0
	$a_{I4,A12}$	1.0
	$a_{I5,A13}$	1.0
	$a_{I6,A14}$	1.0
	$a_{I7,A15}$	1.0
	$a_{I8,A16}$	1.0
	$a_{I9,A17}$	1.0
	$a_{I10,A18}$	1.0
	$a_{I11,A19}$	1.0
	$a_{I12,A20}$	1.0
	$a_{I13,A21}$	1.0
	$a_{I14,A22}$	1.0
	$a_{I15,A23}$	1.0
	$a_{I16,A24}$	1.0
	$a_{I1,A25}$	1.0
	$a_{I2,A26}$	1.0

Table 42: (continued).

DNFs in ASL→IL	**Synaptic link $a_{IL,ASL}$**	**Weight**
	$a_{I3,A27}$	1.0
	$a_{I4,A28}$	1.0
	$a_{I5,A29}$	1.0
	$a_{I6,A30}$	1.0
	$a_{I7,A31}$	1.0
	$a_{I8,A32}$	1.0
	$a_{I1,A33}$	0.5
	$a_{I2,A33}$	0.5
	$a_{I3,A33}$	0.5
	$a_{I4,A33}$	0.5
	$a_{I5,A33}$	0.5
	$a_{I6,A33}$	0.5
	$a_{I7,A33}$	0.5

DNFs in CSGL→IL	**Synaptic link $a_{IL,CSGL}$**	**Weight**
CSGL Present→IL	$a_{I1,C2}$	0.25
	$a_{I2,C4}$	0.25
	$a_{I4,C6}$	0.25
	$a_{I5,C7}$	0.25
	$a_{I6,C8}$	0.25
	$a_{I7,C9}$	0.25

Table 43: Synaptic weights from CSGL to IL.

Table 44: Synaptic weights from ASL to ESL.

DNFs in ASL→ESL	**Synaptic link $a_{\mathbf{ESL,ASL}}$**	**Weight**
ASFA→ESL	$a_{E1,F1}$	1.0
	$a_{E1,F2}$	1.0
	$a_{E1,F3}$	1.0
	$a_{E1,F4}$	1.0
	$a_{E1,F5}$	1.0
	$a_{E1,F6}$	1.0
	$a_{E2,F7}$	1.0
	$a_{E2,F8}$	1.0
	$a_{E2,F9}$	0.5
	$a_{E5,F9}$	0.5
	$a_{E3,F10}$	1.0
	$a_{E3,F11}$	1.0
	$a_{E3,F12}$	1.0
	$a_{E3,F13}$	1.0
	$a_{E4,F14}$	1.0
	$a_{E4,F15}$	1.0
	$a_{E4,F16}$	1.0
	$a_{E4,F17}$	1.0
	$a_{E4,F18}$	0.5
	$a_{E5,F18}$	0.6
	$a_{E4,F19}$	0.5
	$a_{E5,F19}$	0.6
	$a_{E5,F20}$	1.0
	$a_{E6,F21}$	1.0
	$a_{E6,F22}$	1.0
	$a_{E6,F23}$	1.0

Table 44: (continued).

DNFs in ASL→ESL	**Synaptic link $a_{ESL,ASL}$**	**Weight**
	$a_{E6,F24}$	1.0
	$a_{E7,F25}$	1.0
	$a_{E7,F26}$	1.0
	$a_{E7,F27}$	1.0

DNFs in AOL→ESL	**Synaptic link $a_{ESL,AOL}$**	**Weight**
QoM_Hand→ESL	$a_{E3,L}$	1.0
	$a_{E4,L}$	1.0
	$a_{E2,H}$	1.0
QoM_Body→ESL	$a_{E2,H}$	1.0
QoM_Head→ESL	$a_{E2,H}$	1.0

Table 45: Synaptic weights from AOL to ESL.

Table 46: Synaptic weights from OML to AEL.

DNFs in OML→AEL	**Synaptic link $a_{AEL,OML}$**	**Weight**
OML Robot→AEHA	$a_{A1,O1}$	1.2
	$a_{A9,O1}$	-2.5
	$a_{A17,O1}$	1.3
	$a_{A25,O1}$	-1.0
	$a_{A2,O2}$	1.2
	$a_{A10,O2}$	-2.5
	$a_{A18,O2}$	1.3
	$a_{A26,O2}$	-1.0
	$a_{A3,O3}$	1.2
	$a_{A11,O3}$	-2.5

Table 46: (continued).

DNFs in OML→AEL	**Synaptic link $a_{\mathbf{AEL,OML}}$**	**Weight**
	$a_{A19,O3}$	1.3
	$a_{A27,O3}$	-1.0
	$a_{A4,O4}$	1.2
	$a_{A12,O4}$	-2.5
	$a_{A20,O4}$	1.3
	$a_{A28,O4}$	-1.0
	$a_{A5,O5}$	1.2
	$a_{A13,O5}$	-2.5
	$a_{A21,O5}$	1.3
	$a_{A29,O5}$	-1.0
	$a_{A6,O6}$	1.2
	$a_{A14,O6}$	-2.5
	$a_{A22,O6}$	1.3
	$a_{A30,O6}$	-1.0
OML Human→AEHA	$a_{A9,O1}$	0.25
	$a_{A17,O1}$	-2.5
	$a_{A25,O1}$	1.0
	$a_{A10,O2}$	0.25
	$a_{A18,O2}$	-2.5
	$a_{A26,O2}$	1.0
	$a_{A11,O3}$	0.25
	$a_{A19,O3}$	-2.5
	$a_{A27,O3}$	1.0
	$a_{A12,O4}$	0.25
	$a_{A20,O4}$	-2.5
	$a_{A28,O4}$	1.0

Table 46: (continued).

DNFs in OML→AEL	**Synaptic link $a_{\mathbf{AEL,OML}}$**	**Weight**
	$a_{\mathrm{A13,O5}}$	0.25
	$a_{\mathrm{A21,O5}}$	-2.5
	$a_{\mathrm{A29,O5}}$	1.0
	$a_{\mathrm{A14,O6}}$	0.25
	$a_{\mathrm{A22,O6}}$	-2.5
	$a_{\mathrm{A30,O6}}$	1.0

Table 48: Synaptic weights from IL to AEL.

DNFs in IL→AEL	**Synaptic link $a_{\mathbf{AEL,IL}}$**	**Weight**
IL→AEHA	$a_{\mathrm{A9,I1}}$	-1.0
	$a_{\mathrm{A17,I1}}$	1.0
	$a_{\mathrm{A10,I2}}$	-1.0
	$a_{\mathrm{A18,I2}}$	1.0
	$a_{\mathrm{A11,I3}}$	-1.0
	$a_{\mathrm{A19,I3}}$	1.0
	$a_{\mathrm{A12,I4}}$	-1.0
	$a_{\mathrm{A20,I4}}$	1.0
	$a_{\mathrm{A13,I5}}$	-1.0
	$a_{\mathrm{A21,I5}}$	1.0
	$a_{\mathrm{A14,I6}}$	-1.0
	$a_{\mathrm{A22,I6}}$	1.0
	$a_{\mathrm{A9,I9}}$	2.0
	$a_{\mathrm{A17,I9}}$	-1.0
	$a_{\mathrm{A25,I9}}$	-1.5
	$a_{\mathrm{A33,I9}}$	-1.5
	$a_{\mathrm{A10,I10}}$	2.0

Table 48: (continued).

DNFs in IL→AEL	**Synaptic link $a_{\mathbf{AEL,IL}}$**	**Weight**
	$a_{A18,I10}$	-1.0
	$a_{A26,I10}$	-1.5
	$a_{A33,I10}$	-1.5
	$a_{A11,I11}$	2.0
	$a_{A19,I11}$	-1.0
	$a_{A27,I11}$	-1.5
	$a_{A33,I11}$	-1.5
	$a_{A12,I12}$	2.0
	$a_{A20,I12}$	-1.0
	$a_{A28,I12}$	-1.5
	$a_{A33,I12}$	-1.5
	$a_{A13,I13}$	2.0
	$a_{A21,I13}$	-1.0
	$a_{A29,I13}$	-1.5
	$a_{A33,I13}$	-1.5
	$a_{A14,I14}$	2.0
	$a_{A22,I14}$	-1.0
	$a_{A30,I14}$	-1.5
	$a_{A33,I14}$	-1.5

Table 53: Synaptic weights from IL to EML.

DNFs in IL→EML	**Synaptic link $a_{\mathbf{EML,IL}}$**	**Weight**
IL→Error in Intention	$a_{E1,I1}$	1.0
	$a_{E15,I1}$	1.0
	$a_{E2,I2}$	1.0
	$a_{E16,I2}$	1.0

Table 53: (continued).

DNFs in IL→EML	**Synaptic link $a_{\mathbf{EML,IL}}$**	**Weight**
	$a_{E3,I3}$	1.0
	$a_{E17,I3}$	1.0
	$a_{E4,I4}$	1.0
	$a_{E18,I4}$	1.0
	$a_{E5,I5}$	1.0
	$a_{E19,I5}$	1.0
	$a_{E6,I6}$	1.0
	$a_{E20,I6}$	1.0
	$a_{E7,I7}$	1.0
	$a_{E21,I7}$	1.0
	$a_{E29,I8}$	2.0
	$a_{E8,I9}$	1.0
	$a_{E22,I9}$	1.0
	$a_{E9,I10}$	1.0
	$a_{E23,I10}$	1.0
	$a_{E10,I11}$	1.0
	$a_{E24,I11}$	1.0
	$a_{E11,I12}$	1.0
	$a_{E25,I12}$	1.0
	$a_{E12,I13}$	1.0
	$a_{E26,I13}$	1.0
	$a_{E13,I14}$	1.0
	$a_{E27,I14}$	1.0
	$a_{E14,I15}$	1.0
	$a_{E28,I15}$	1.0
	$a_{E30,I16}$	2.0

DNFs in CSGL→AEL	**Synaptic link** $a_{\mathbf{AEL,CSGL}}$	**Weight**
CSGL Present→AEHA	$a_{A1,C1}$	0.9
	$a_{A9,C1}$	0.25
	$a_{A25,C1}$	0.75
	$a_{A33,C1}$	1.85
	$a_{A17,C2}$	0.9
	$a_{A2,C3}$	0.9
	$a_{A10,C3}$	0.25
	$a_{A26,C3}$	0.75
	$a_{A33,C3}$	1.85
	$a_{A18,C4}$	0.9
	$a_{A3,C5}$	0.9
	$a_{A11,C5}$	0.25
	$a_{A27,C5}$	0.75
	$a_{A33,C5}$	1.85
	$a_{A20,C6}$	0.9
	$a_{A21,C7}$	0.9
	$a_{A22,C8}$	0.9

Table 47: Synaptic weights from CSGL to AEL.

DNFs in ESL→AEL	**Synaptic link $a_{\mathbf{AEL,ESL}}$**	**Weight**
ESL→AEHA	$a_{\text{A25,E6}}$	0.5
	$a_{\text{A26,E6}}$	0.5
	$a_{\text{A27,E6}}$	0.5
	$a_{\text{A28,E6}}$	0.5
	$a_{\text{A29,E6}}$	0.5
	$a_{\text{A30,E6}}$	0.5

Table 49: Synaptic weights from ESL to AEL

DNFs in EML→AEL	**Synaptic link $a_{\mathbf{AEL,EML}}$**	**Weight**
Error in Execution→AEHA	$a_{\text{A17,E8}}$	1.0
	$a_{\text{A18,E9}}$	1.0
	$a_{\text{A19,E10}}$	1.0
	$a_{\text{A20,E11}}$	1.0
	$a_{\text{A21,E12}}$	1.0
	$a_{\text{A22,E13}}$	1.0

Table 50: Synaptic weights from EML to AEL.

DNFs in ESL→AEL	**Synaptic link $a_{\mathbf{AEL,ESL}}$**	**Weight**
ESL→AEFA	$a_{\text{F20,E1}}$	0.5
	$a_{\text{F7,E2}}$	0.4
	$a_{\text{F20,E2}}$	0.45
	$a_{\text{F20,E3}}$	0.5
	$a_{\text{F20,E4}}$	0.5
	$a_{\text{F20,E5}}$	0.5
	$a_{\text{F20,E6}}$	0.5
	$a_{\text{F25,E7}}$	0.65

Table 51: Synaptic weights from ESL to AEL.

DNFs in EML→AEL	**Synaptic link $a_{\mathbf{AEL,EML}}$**	**Weight**
Error in Execution→AEFA	$a_{F7,E1}$	0.5
	$a_{F7,E2}$	0.5
	$a_{F7,E3}$	0.5
	$a_{F7,E4}$	0.5
	$a_{F7,E5}$	0.5
	$a_{F7,E6}$	0.5
	$a_{F7,E7}$	0.5

Table 52: Synaptic weights from EML to AEL.

Table 54: Synaptic weights from CSGL to EML.

DNFs in CSGL→EML	**Synaptic link $a_{\mathbf{EML,CSGL}}$**	**Weight**
CSGL Past→Error in Intention	$a_{E1,C2}$	1.0
	$a_{E2,C4}$	1.0
	$a_{E4,C6}$	1.0
	$a_{E5,C7}$	1.0
	$a_{E6,C8}$	1.0
	$a_{E7,C9}$	1.0
	$a_{E8,C1}$	1.0
	$a_{E9,C3}$	1.0
	$a_{E10,C5}$	1.0
CSGL Present→Error in Intention	$a_{E23,C1}$	1.0
	$a_{E24,C1}$	0.5
	$a_{E18,C1}$	0.5
	$a_{E19,C1}$	0.5
	$a_{E20,C1}$	0.5
	$a_{E21,C1}$	0.5

Table 54: (continued).

DNFs in CSGL→EML	Synaptic link $a_{\mathbf{EML,CSGL}}$	Weight
	$a_{\mathrm{E15,C2}}$	1.0
	$a_{\mathrm{E24,C2}}$	0.5
	$a_{\mathrm{E18,C2}}$	0.5
	$a_{\mathrm{E19,C2}}$	0.5
	$a_{\mathrm{E20,C2}}$	0.5
	$a_{\mathrm{E21,C2}}$	0.5
	$a_{\mathrm{E24,C3}}$	0.5
	$a_{\mathrm{E18,C3}}$	0.5
	$a_{\mathrm{E19,C3}}$	0.5
	$a_{\mathrm{E20,C3}}$	0.5
	$a_{\mathrm{E21,C3}}$	0.5
	$a_{\mathrm{E24,C4}}$	0.5
	$a_{\mathrm{E18,C4}}$	0.5
	$a_{\mathrm{E19,C4}}$	0.5
	$a_{\mathrm{E20,C4}}$	0.5
	$a_{\mathrm{E21,C4}}$	0.5
	$a_{\mathrm{E21,C5}}$	0.5
	$a_{\mathrm{E19,C6}}$	0.5
	$a_{\mathrm{E20,C6}}$	0.5
	$a_{\mathrm{E21,C6}}$	0.5
	$a_{\mathrm{E20,C7}}$	0.5
	$a_{\mathrm{E21,C7}}$	0.5
	$a_{\mathrm{E21,C8}}$	0.5

DNFs in ASL→EML	**Synaptic link $a_{\mathbf{EML,ASL}}$**	**Weight**
ASHA→Error in Means	$a_{E15,A17}$	1.0
	$a_{E16,A18}$	1.0
	$a_{E17,A19}$	1.0
	$a_{E18,A20}$	1.0
	$a_{E19,A21}$	1.0
	$a_{E20,A22}$	1.0
	$a_{E21,A23}$	1.0
	$a_{E22,A25}$	1.0
	$a_{E23,A26}$	1.0
	$a_{E24,A27}$	1.0
	$a_{E25,A28}$	1.0
	$a_{E26,A29}$	1.0
	$a_{E27,A30}$	1.0
	$a_{E28,A31}$	1.0

Table 55: Synaptic weights from ASL to EML.

DNFs in OML→EML	**Synaptic link $a_{\mathbf{EML,OML}}$**	**Weight**
OML Robot→Error in Means	$a_{E15,O1}$	1.0
	$a_{E16,O2}$	1.0
	$a_{E17,O3}$	1.0
	$a_{E18,O4}$	1.0
	$a_{E19,O5}$	1.0
	$a_{E20,O6}$	1.0
	$a_{E21,O7}$	1.0
OML Human→Error in Means	$a_{E22,O1}$	1.0
	$a_{E23,O2}$	1.0
	$a_{E24,O3}$	1.0
	$a_{E25,O4}$	1.0
	$a_{E26,O5}$	1.0
	$a_{E27,O6}$	1.0
	$a_{E28,O7}$	1.0

Table 56: Synaptic weights from OML to EML.

Table 57: Synaptic weights from CSGL to EML.

DNFs in CSGL→EML	**Synaptic link $a_{\mathbf{EML,CSGL}}$**	**Weight**
CSGL Past→Error in Execution	$a_{E2,C4}$	1.25
	$a_{E4,C6}$	1.25
	$a_{E5,C7}$	1.25
	$a_{E6,C8}$	1.25
	$a_{E7,C9}$	1.25
CSGL Present→Error in Execution	$a_{E4,C1}$	0.5
	$a_{E5,C1}$	0.5
	$a_{E6,C1}$	0.5
	$a_{E7,C1}$	0.5
	$a_{E2,C2}$	0.5
	$a_{E4,C2}$	0.5
	$a_{E5,C2}$	0.5
	$a_{E6,C2}$	0.5
	$a_{E7,C2}$	0.5
	$a_{E4,C3}$	0.5
	$a_{E5,C3}$	0.5
	$a_{E6,C3}$	0.5
	$a_{E7,C3}$	0.5
	$a_{E4,C4}$	0.5
	$a_{E5,C4}$	0.5
	$a_{E6,C4}$	0.5
	$a_{E7,C4}$	0.5
	$a_{E7,C5}$	0.5
	$a_{E5,C6}$	0.5
	$a_{E6,C6}$	0.5
	$a_{E7,C6}$	0.5

Table 57: (continued).

DNFs in CSGL→EML	**Synaptic link $a_{\mathbf{EML,CSGL}}$**	**Weight**
	$a_{\mathrm{E6,C7}}$	0.5
	$a_{\mathrm{E7,C7}}$	0.5
	$a_{\mathrm{E7,C8}}$	0.5

DNFs in ASL→EML	**Synaptic link $a_{\mathbf{EML,ASL}}$**	**Weight**
ASHA→Error in Execution	$a_{E8,A1}$	0.2
	$a_{E8,A9}$	0.2
	$a_{E8,A17}$	0.2
	$a_{E9,A2}$	0.2
	$a_{E9,A10}$	0.2
	$a_{E9,A18}$	0.2
	$a_{E10,A3}$	0.2
	$a_{E10,A11}$	0.2
	$a_{E10,A19}$	0.2
	$a_{E11,A4}$	0.2
	$a_{E11,A12}$	0.2
	$a_{E11,A20}$	0.2
	$a_{E12,A5}$	0.2
	$a_{E12,A13}$	0.2
	$a_{E12,A21}$	0.2
	$a_{E13,A6}$	0.2
	$a_{E13,A14}$	0.2
	$a_{E13,A22}$	0.2
	$a_{E14,A7}$	0.2
	$a_{E14,A15}$	0.2
	$a_{E14,A23}$	0.2

Table 58: Synaptic weights from ASL to EML.