Philosophiae doctor thesis

Exploring Different Means of Human-Robot Interactions by Using Environmental Data to Adapt the Robot's Behaviours

From 1st October 2019 to 31st March 2023

Siméon Capy
Tokyo University of Agriculture and Technology

Faculty of Engineering

Department of Mechanical System Engineering
GVLab
Student ID: 19833014
Institutional e-mail: s192106v@st.go.tuat.ac.jp
Personal e-mail: simeoncapy@gmail.com

Tokyo University of Agriculture and Technology
東京農工大学
Naka-chō 2-24-16
186-0012 Koganei-shi
Tōkyō-to, JAPAN

Supervisors

Prof. Gentiane Venture
Director of GVLab
Tel.: +81 42-367-7092
E-Mail: venture@cc.tuat.ac.jp

Prof. Yasutaka Tagawa
Director of Tagawa's lab

E-Mail: tagawa@cc.tuat.ac.jp

Abstract

English

With the development of robots for the general public and their usage outside of factories, they would have to interact with humans in a closer way, further from the controlled environment of the laboratories or factories. To improve the interaction between humans and robots, to make them closer to a partner than just automate nay a tool, they have to take advantage of the environmental data.

In this thesis, we will explore with two robotic projects how we can create scenarios where the robots change their behaviour according to, either, the user's emotions or the weather data. The first project uses an industrial robot, that has no initial purpose for social interactions. The scenario was created to have no prior training from the user, and to achieve a very specific task during a fair in Japan with the general public. The answer from about 200 participants was gathered with four questionnaires to gather their opinion about the scenario. Results suggest that participants considered our proposed scenario as enjoyable, safe, and interesting. For the second project, we created from scratch an abstract presence robot for the home. With the aim to serve as a link between a couple in their home, for a long-term interaction. This robot, called Yokobo, uses non-verbal expressive means of communication (movement and light), and weather data to change its behaviours. Two two-week experiments were done in our laboratories to evaluate the robustness and usability of Yōkobo and the perception and reception of the participants. The results show that Yōkobo can sustain long-term interaction and serve as a welcoming partner. In order to be integrated inside the home, a smartphone application has been developed and tested to configure the Raspberry Pi (the central unit of Yōkobo) without using a screen, mouse or keyboard. The solution was tested with 11 users that managed to configure the robot by themselves. Finally, for the purpose of long-term interaction, we propose a solution to adapt the behaviours of Yōkobo from the world data thanks to a causal representation, and a reinforcement learning algorithm to create the action.

Keywords: human-robot interaction, social robotics

日本語

題名：環境データを用いたロボットの行動適応による、人とロボットの多様なインタラクションの手段に関する研究

一般向けのロボットが開発され、工場の外で使われるようになると、実験室や工場の管理された環境から離れ、より身近なところで人間と接する必要が出てきます。人間とロボットのインタラクションを向上させ、ロボットを単なる自動化、道具としてではなく、パートナーに近づけるためには、環境データを活用することが必要です。

本論文では、2つのプロジェクトを通じて、ユーザーの感情や気象データに応じてロボットが行動を変化させるシナリオを作成する方法を検討しました。最初のプロジェクトは、社会的な相互作用を目的としない産業用ロボットを使用します。このシナリオは、ユーザーからの事前トレーニングを必要とせず、日本で開催された一般向けの展示会で非常に具体的なタスクを達成するために作成されました。このシナリオについて、約200名の参加者から、4つのアンケートで回答を得ることができました。その結果、参加者は私たちの提案したシナリオを「楽しい」「安全」「面白い」と考えていることが分かりました。2つ目のプロジェクトでは、家庭用の抽象的なプレゼンスロボットをゼロから作成しました。これは、家庭内でカップルが長期的に交流するためのつなぎ役となることを目的としています。このロボットは「Yokobo」と呼ばれ、非言語的な表現手段（動きや光）と気象データを使って行動を変化させます。研究室では2週間の実験を2回行い、「ヨコボ」の堅牢性や使い勝手、参加者の知覚や受け止め方を評価しました。その結果、ヨコボは長期的なインタラクションを維持し、歓迎されるパートナーとして機能することがわかりました。家庭内に組み込むために、画面やマウス、キーボードを使わずにRaspberry Pi（Yokoboの中央装置）を設定するスマートフォンアプリを開発し、テストを行いました。このソリューションは、11人のユーザーが自分でロボットを設定することができることを確認しました。最後に、長期的なインタラクションを目的として、カジュアル表現により世界データからヨコボの行動を適応させ、強化学習アルゴリズムにより行動を作成するソリューションを提案します。

※ 備考、abstractのみ日本語、論文本文は英語。

キーワード： 人間とロボットのインタラクション、ソーシャルロボティクス

Français

Titre : Étude de différents moyens d'interaction entre les humains et les robots en utilisant les données de l'environnement pour en changer le comportement

Avec le développement des robots pour le grand public et leur utilisation en dehors des usines, ceux-ci devront interagir avec les humains d'une manière plus proche, loin de l'environnement contrôlé des laboratoires ou des usines. Pour améliorer l'interaction entre les humains et les robots, et pour les rendre plus proches d'un partenaire que d'un simple outil automatisé, ils devront tirer parti des données de leur environnement.

Dans cette thèse, nous allons explorer, avec deux projets robotiques, comment nous pouvons créer des scénarios où les robots modifient leur comportement en fonction, soit des émotions de l'utilisateur, soit de données météorologiques. Le premier projet utilise un robot industriel, qui n'a pas été conçu initialement pour avoir des interactions sociales. Le scénario a été créé pour que les utilisateurs n'aient pas besoin de formation initiale, et pour accomplir une tâche très spécifique lors d'une exposition au Japon ouvert au grand public. Les réponses d'environ 200 participants ont été recueillies à l'aide de quatre questionnaires destinés à recueillir leur opinion sur le scénario. Les résultats suggèrent que les participants ont considéré le scénario proposé comme agréable, sûr et intéressant. Pour le deuxième projet, nous avons créé de zéro un robot de présence abstrait à destination de la maison. L'objectif est, ici, de servir de lien entre les membres d'un couple à son domicile, avec une interaction de long terme. Ce robot, appelé Yōkobo, utilise des moyens de communication expressifs non verbaux (mouvement et lumière), et des données météorologiques pour modifier son comportement. Deux expériences de deux semaines ont été réalisées dans nos laboratoires pour évaluer la robustesse et l'ergonomie de Yōkobo, ainsi que la perception et la réception des participants. Les résultats montrent que Yokobo peut soutenir une interaction à long terme et servir de partenaire accueillant. Afin d'être intégré à l'intérieur de la maison, une application smartphone a également été développée et testée pour configurer la Raspberry Pi (l'unité centrale de Yōkobo) sans utiliser d'écran, de souris ou de clavier. La solution a été testée avec 11 utilisateurs qui ont réussi à configurer le robot par eux-mêmes. Enfin, dans le but d'une interaction à long terme, nous proposons une solution pour adapter les comportements de Yokobo à partir des *données du monde* (environnement) grâce à une représentation causale, et un algorithme d'apprentissage par renforcement pour mouvoir Yōkobo ainsi que changer la couleur de sa lumière.

NB : seul le résumé (abstract) *est traduit en français, le reste de la thèse est en anglais.*

Mots-clefs : interactions humain-robot, robotique sociale

Acknowledgement

I would like to thank warmly Professor **Gentiane Venture** who gave me the opportunity to join the GVLab to achieve my PhD and gave me all the resources and support I needed.

I would also thank Professor **Yasutaka Tagawa** and the jury members who review my work: Professors **Kondo**, **Maeda**, **Mizuuchi** and **Raksincharoensak**.

I also warmly thank Professor **Liz Rincon** who supervised me at the beginning of my thesis and gave me precious advice and help.

I also thank Dr **Enrique Coronado** who was my senior in the lab and acted like a *senpai* to me by giving me precious advice.

I acknowledge also Dr **Dominique Deuff**, who gave me the chance to be involved in the Yōkobo project which became a central part of my PhD. In addition, I also thank **Corentin Aznar** and **Dora Garcin** who created the design of Yōkobo.

I would also like to acknowledge Professor **Pongsathorn Raksincharoensak** who accepted to supervise my special research project.

Then, I would also like to thank the members of the GVLab, past and present, who warmly welcomed me and helped me to settle in Japan. I especially thank **Shohei Hagane** and **Shōyo Ishida** who helped me a lot with the administrative paper works and to understand the cultural differences between France and Japan. I also acknowledge **Pablo Osorio** who became a really close friend during this PhD. It was really enjoyable and enriching to work with them.

Then, I thank warmly the members of the International Club of TUAT and especially **Jingzu Yee**, they shared many moments of my life in Japan.

Next, I thank Professor **Tomoko Hongo** and the members of the EAGLe office, but also Mmes **Matsumoto** and **Matsuoka**, two volunteers of the city of Koganei. All of them helped me a lot to learn the Japanese language.

Penultimately, I thank **Sylvain Riondet**, without him, I would not have obtained the contact to be able to do this PhD.

Finally, I thank all the members of my family who have always supported me in my projects, even on the other side of the world.

Contents

Abstract **ii**

Acknowledgement **v**

List of Figures **x**

List of Tables **xi**

List of Publications **xii**

1 Introduction **1**
- 1.1 Social Robots . . . 1
  - 1.1.1 History . . . 1
  - 1.1.2 Definition . . . 3
- 1.2 Interaction With Robots . . . 4
- 1.3 Application of Social Robots . . . 5
- 1.4 Motivation . . . 6
- 1.5 Outline of the Thesis . . . 7

2 Design of the Experiments **8**
- 2.1 Research questions & Relevant concepts . . . 8
- 2.2 Relevant Scales & Tools . . . 9
  - 2.2.1 Likert Scale & Semantic Difference Scale . . . 9
  - 2.2.2 Negative Attitude towards Robots Scale . . . 10
  - 2.2.3 Godspeed questionnaire . . . 10
  - 2.2.4 CH-33 questionnaire . . . 11
  - 2.2.5 Inclusion of Others in Self Scale . . . 11
  - 2.2.6 Self-Assessment Manikin scale . . . 11
  - 2.2.7 Big Five personality test . . . 12
  - 2.2.8 System Usability Scale . . . 13
  - 2.2.9 Demographic Data . . . 13
- 2.3 Adapting the scales and design of the questionnaire . . . 13
- Summary . . . 13

3 Preliminary Work **14**
- 3.1 Introduction . . . 14
- 3.2 Related work . . . 15
- 3.3 Objectives and contributions . . . 16
- 3.4 Design and Implementation . . . 17
  - 3.4.1 Hardware . . . 17
  - 3.4.2 Software . . . 17
- 3.5 Validation *in the wild* . . . 22
  - 3.5.1 Subjective validation of the proposed system . . . 22
  - 3.5.2 Participants . . . 23
- 3.6 Results . . . 25
  - 3.6.1 User perceptions and emotional reactions . . . 25

3.6.2 Negative Attitudes Towards Robots . . . 25
3.6.3 User needs and desires . . . 25
3.6.4 Hypothesis . . . 25
3.7 Discussion . . . 27
3.7.1 Regarding Q1 and Q2 . . . 27
3.7.2 Regarding Q3 . . . 28
3.7.3 Limitations . . . 29
3.8 Conclusion and Future Work . . . 29
Summary . . . 30

4 Design of a Non-Verbal Behaviour Robot: Yōkobo 31
4.1 Conception . . . 31
4.2 Related Works . . . 32
4.2.1 Creating Encounters and Links between Persons . . . 32
4.2.2 Form Factor . . . 33
4.2.3 Perception through Reduced Robotic Movement . . . 33
4.3 Contribution . . . 34
4.4 Initial Step of the Design . . . 34
4.4.1 Yōkobo's Shape Design . . . 34
4.4.2 Services and Associated Functions . . . 35
4.4.3 Behaviour Design . . . 35
4.5 First implementation . . . 36
4.5.1 Selection of the devices . . . 36
4.5.2 Printed Circuit Board . . . 37
4.5.3 Encountered problems . . . 38
4.6 Final Hardware . . . 39
4.7 Software . . . 39
4.7.1 Architecture . . . 42
Summary . . . 43

5 Technical Experiment with Yōkobo 44
5.1 Experiment Settings . . . 44
5.1.1 Participants and schedule . . . 45
5.1.2 Tools . . . 45
5.2 Results . . . 47
5.2.1 Robustness, Stability, and Usability . . . 47
5.2.2 Perception and Reception . . . 49
Summary . . . 52

6 Integration of Yōkobo into Smart Homes 53
6.1 Related Works . . . 53
6.1.1 Robots in Smart Home . . . 53
6.1.2 Configuration of IoRT . . . 56
6.1.3 Usage of Single-board Computers . . . 56
6.2 Contribution . . . 56
6.3 Design of the Configuration Process . . . 57
6.3.1 On the Raspberry Pi . . . 57
6.3.2 On the Smartphone . . . 59
6.4 Design of the Application . . . 59
6.5 Experimental Tests, Results and Discussion . . . 61
Summary . . . 63

7 How Can We Have Long Term Interactions With Robots? 64
7.1 Literature Review . . . 64
7.1.1 Creating a Representation of the World . . . 64
7.1.2 Creating Adaptable Actions . . . 65
7.1.3 Intention estimation . . . 65

7.2 BeWoDa's Architecture . . . 69
7.2.1 Inputs . . . 70
7.2.2 Outputs . . . 70
7.2.3 User's Intention Estimation . . . 70
7.2.4 Reinforcement learning algorithms . . . 74
7.2.5 Robot state . . . 77
7.3 Training of the algorithm . . . 81
7.3.1 Foreword . . . 81
7.3.2 Data Generation . . . 81
7.3.3 Results and Discussions . . . 82
7.4 Future work and Improvement . . . 83
Summary . . . 84

8 Conclusion **85**
8.1 Limitations . . . 86
8.2 Open Challenges and Future Works . . . 86

List of Acronyms **87**

Bibliography **89**

A Questionnaires Scale **106**
A.1 The Big Five questionnaire . . . 106
A.2 The System Usability Scale (SUS) scale . . . 106

B NEXTAGE OPEN Characteristics **108**
B.1 Dimensions . . . 108
B.2 Specifications . . . 109

C Yōkobo's Hardware **110**
C.1 Bill of Material . . . 110
C.2 Final PCB Circuit Diagram . . . 112
C.3 Final PCB Routing Diagram . . . 113
C.3.1 Front . . . 113
C.3.2 Bottom . . . 113

D Questionnaires and Data of the Technical Experiment **114**
D.1 Questionnaires . . . 114
D.2 Graffiti wall . . . 118

E Behaviours from World Data (BeWoDa) **120**
E.1 Algorithm . . . 120
E.2 Results . . . 120

List of Figures

1.1 Timeline of some steps in the robotic history. Concerning the recent time, we focus on successful commercial robots. 2
1.2 Example of a long-term experiment in school with children and the robot Pepper . 4
1.3 Two LOVOTs 5
1.4 Number of studies in *Xplore* regarding Nao robot the last 5 years 6
1.5 Example of popular robots from Science fiction 7

2.1 Standardised process for using questionnaires in HRI research 8
2.2 Example of a 5-point Likert and Semantic scales. The $\overline{Adjective}$ stands for the opposite of *Adjective* 10
2.3 The NARS questionnaire 11
2.4 The Inclusion of Others in Self (IOS) scale 12
2.5 The PAD and SAM scale 12

3.1 Robot on the booth 17
3.2 Main elements of the designed Graphical User Interface (GUI) 18
3.3 General architecture of the project. The GUI program is using client-server architecture, the server is in Python and the client in JavaScript 18
3.4 Example of markers used by the staff 19
3.5 Real scene with the objects the robot has to detect and grasp 20
3.6 Examples of elements shown in the GUI according to the status of the interaction 21
3.7 Flowchart of the command of the robot 21
3.8 The content of each questionnaire P1–P4 23
3.9 The robot was in front of an alley with a lot of traffic, this was challenging for the algorithms that were developed in a controlled environment. 24

4.1 Kinematic diagram of Yōkobo (H = 33; L = 36; W = 24; ø = 15) 32
4.2 Representation of the proximity levels around Yōkobo (the black dot). 35
4.3 The anthropomorphic inspiration for the design of Yōkobo's behaviours according to the weather data 36
4.4 The Raspberry Pi 4B, the model used for Yōkobo has 8 GB of RAM 37
4.5 Sensors used for Yōkobo. The PIR will be finally abandoned to be replaced by the HCSR04. 38
4.6 The v2 of the PCB 38
4.7 Hardware architecture 39
4.8 Axis of the Denavit–Hartenberg parameters of Yōkobo 40
4.9 Finite state machines designed to command Yōkobo's services and behaviours. . . 40
4.10 Architecture of the different algorithms of Yōkobo 42

5.1 The mainlines of the experimental setup 45
5.2 Example of clock-in/out message in the Slack channel 45
5.3 The composition of the questionnaires to evaluate technical viability of Yōkobo . . 47
5.4 Participants cumulative interaction time for both experiments 50
5.5 Interaction time throughout both experiments 51

6.1 Structure of the message sent *via* Bluetooth 57
6.2 Process of the first configuration 58

6.3 Screens of the application 60
6.4 Schematic diagram of the Model-View-Controller (MVC) architecture 60
6.5 Experimental setup 62

7.1 Process of the search strategy 66
7.2 General architecture of the proposed solution. 69
7.3 Example of a simple Bayesian Network 72
7.4 The Dynamic Bayesian Network used to map the causal relationship from the world's data 73
7.6 The parameters of the filter 73
7.5 Representation of the difference between the two periods of the algorithm. The hat ($\hat{t}_1$) symbolises the estimation the algorithm is calculating on the future slice. . . . 74
7.7 Schematic diagram of the Neural Network of the Deep Q-Learning algorithm . . . 76
7.8 Reminder of the Denavit–Hartenberg parameters of Yōkobo 79
7.9 The green band symbolises the position where the training algorithm will generate the user's position 82
7.10 Screenshot of the animation generated by Blender. The 3D model of Yōkobo is using the second version. 83

D.1 Picture of Yōkobo used in the questionnaire to know the vocabulary used by participants to describe the robot. 116

E.1 The angular position (in radian) of each motor over steps, for some episodes 121
E.2 The PAD of Yōkobo over steps, for some episodes 122
E.3 Posture of Yōkobo with the colour of the light for four consecutive steps at episode 0 123
E.4 Posture of Yōkobo with the colour of the light for four consecutive steps at episode 498 124

List of Tables

3.1 Research articles proposing interactive robotic systems with industrial robots for grasping perceptions towards robots after performing direct Human-Robot Interaction 16
3.2 Demographic data of each questionnaire 24
3.3 Semantic analysis results (K-1) 26
3.4 Semantic analysis results (K-2) 26
3.5 Average (μ) and standard deviation (σ) values for each NARS type 27
3.6 Average (μ) and standard deviation (σ) of the NARS questionnaire according to each different group 27

4.1 Denavit–Hartenberg parameters of Yōkobo 40

5.1 Semantic analysis results (second experiment). Lowest score—positive adjective; highest score—negative adjective. 48

6.1 Comparison of some domotic robots (NM: not mentioned) 55
6.2 The ASCII character set 59
6.3 SUS results with 3 custom questions. Scores from *Strongly Disagree* (1) to *Strongly Agree* (5). The positive questions are in bold. 62
6.4 Result of the behavioural analysis 62

7.1 Comparison of the studies to estimate pedestrians/cyclists' intention 67
7.3 The state space of the environment 70
7.4 The palette of colours of the LED 71
7.5 The actions space 71
7.6 The states and actions for the reinforcement learning algorithm that create the trajectories of Yōkobo 75
7.7 The reward of the Deep Q-Learning (DQL) algorithm that commands the motors. If the interaction time is below 10 s, we assume that the user was just passing in front of Yōkobo. That is why we do not penalise it. A continuous reward is added at each step, and a terminal one is only added at the terminal step. 77
7.8 The mapping table between emotion, PAD and colour 78
7.9 The rewards of the DQL algorithm for the light 78
7.10 Denavit–Hartenberg parameters of Yōkobo 79
7.11 The parameters of the two DQL algorithms. 81

B.1 Robot specifications 109

D.1 Questions asked in QS, QW1 and QW2 115
D.2 Questions in QW2 for the selected participants 116
D.3 SUS questions in QW2 for the selected participants, using a Likert scale 117

List of Publications

Main Publications

1. Siméon Capy, Liz Rincon, Enrique Coronado, Shohei Hagane, Seiji Yamaguchi, Victor Leve, Yuichiro Kawasumi, Yasutoshi Kudou, and Gentiane Venture. Expanding the frontiers of industrial robots beyond factories: Design and in the wild validation. *Machines*, 10(12), 2022. ISSN 2075-1702. doi: 10.3390/machines10121179. URL `https://www.mdpi.com/2075-1702/10/12/1179` [1]

2. Siméon Capy, Pablo Osorio, Shohei Hagane, Corentin Aznar, Dora Garcin, Enrique Coronado, Dominique Deuff, Ioana Ocnarescu, Isabelle Milleville, and Gentiane Venture. Yōkobo: A robot to strengthen links amongst users with non-verbal behaviours. *Machines*, 10(8), 2022. ISSN 2075-1702. doi: 10.3390/machines10080708. URL `https://www.mdpi.com/2075-1702/10/8/708` [2]

3. Siméon Capy, Enrique Coronado, Pablo Osorio, Shohei Hagane, Dominique Deuff, and Gentiane Venture. Integration of a presence robot in a smart home. *International Conference on Computer, Control and Robotics (ICCCR) 2023, Shanghai, China*, (accepted) [3]

4. Siméon Capy, Gentiane Venture, and Pongsathorn Raksincharoensak. Pedestrians and cyclists' intention estimation for the purpose of autonomous driving. *International Journal of Automotive Engineering*, 14(1):10–19, 2023. doi: 10.20485/jsaeijae.14.1_10 [4]

Related Publications

1. Dominique Deuff, Isabelle Milleville-Pennel, Ioana Ocnarescu, Dora Garcin, Corentin Aznar, Siméon Capy, Shohei Hagane, Pablo Felipe Osorio Marin, Enrique Coronado Zuniga, Liz Rincon Ardila, and Gentiane Venture. Together alone, yōkobo, a sensible presence robject for the home of newly retired couples. In *Designing Interactive Systems Conference*, DIS '22, page 1773–1787, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450393584. doi: 10.1145/3532106.3533485 [5]

Chapter 1

Introduction

During my master's degree project, I studied the impact of unusual sensors, namely electroencephalogram, for the Human–Robot Interaction (HRI), to go beyond the social robotics [6]. With the same aim of exploring the HRI further, I started this PhD thesis on their improvement. With the objective to have a robot that *understands* the user's actions, behaviour or emotion, and then interacts in the best way possible.

This thesis is led by three projects to achieve the aforementioned goal. In a preliminary work (chapter 3), work has been done with an industrial robot to learn the reaction of the users with a robot that has no first aim to do social interactions. In that scenario, the robot has to read users' emotions to adapt its own behaviours. The second project (see chapter 4) is the creation of a *presence robot* for the home entrance, with the aim to strengthen the link between two people in their household. Using a robot at home (or in the lab for the experiment, see chapter 5), allows us to see how people react when they have to interact with a robot on daily basis. Moreover, the built robot is only using non-verbal behaviours, in order to explore another part of HRI. Finally, using this platform, an algorithm has been developed to create an adaptive reaction by the robot (see chapter 7), according to the user's action and environment data (such as weather). Before detailing those three projects, this introduction chapter will sum up the current state of the art in social robotics; moreover, the chapter of each project will also do so in their specific area.

1.1 Social Robots

1.1.1 History

Already during Ancient history, like in Greece, humanity was creating complex mechanisms, called **automatons**, to execute tasks. Such as the Antikythera mechanism which is the first known analogue computer, used to predict astronomical positions, but also the automata of Hero of Alexandria. He created several mechanisms, presented in his mechanical treatise called *Automata* [7], such as the theatres. With small characters, they could recreate life-like scenes for entertainment [8].

But we had to wait until the 15th century, with the rediscovery of the past knowledge in the Western World, and with the development of clock-making, to see increasingly sophisticated automata appear. One can mention Strasbourg's astronomical clock[1], that, besides displaying the time and other calendar-related data, has some animated characters turning around at specific moments of the day. Then we can mention the Jaquet-Droz automata in the 18th century, mimicking humans playing music, writing or drawing [9]. The inventors had the wish to create mechanisms that could reproduce the complexity of the human being. This desire led some to create hoaxes, such as the famous Mechanical Turk [10], contemporaries thought it was an automaton with the capability to beat human players at chess. Instead, it was just a puppet manipulated by a person hidden in the machine.

Nonetheless, those mechanisms had the limitation of being not programmable, the action done were always the same. At the beginning of the 19th century, Joseph Marie Jacquard created the

[1]The second version is from the 16th century, but a previous one was made during the 14th century.

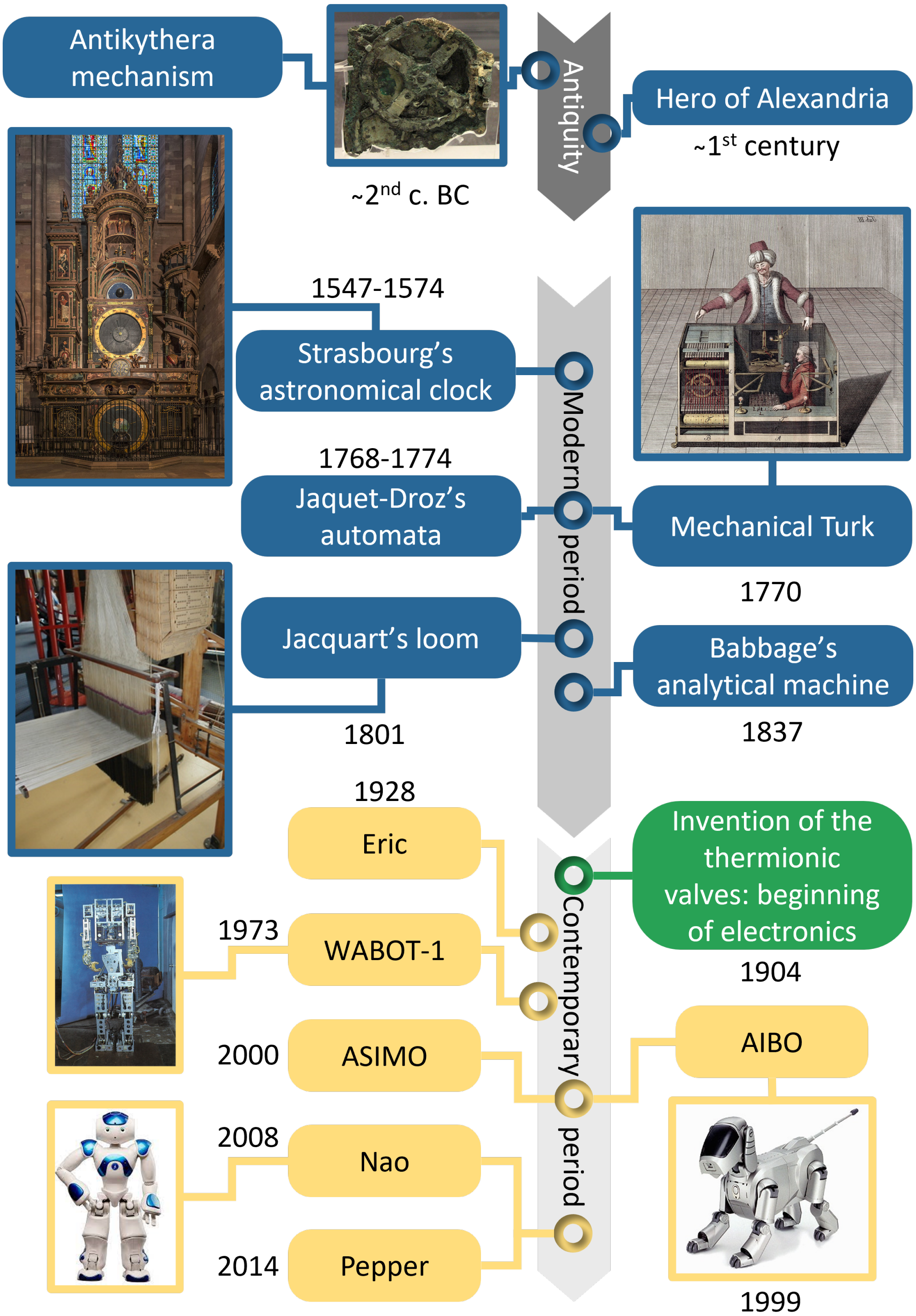


Figure 1.1: Timeline of some steps in the robotic history. Concerning the recent time, we focus on successful commercial robots.

eponymous loom [11]. Using punched cards, it opened the way for adaptable automata, that could change their behaviour according to the needs of the user. During the same century, Charles Babbage conceptualised the analytical engine, a mechanical computer using Jacquard's punched cards as inputs; opening the way for modern computers [12].

Finally, with the development of electronics at the beginning of the 20th century, the increased and firsts electronic automata appeared, which will now be called robots, from the Czech *robota* meaning *forced labour*. Nowadays, the main difference with the automata is the programming, the robot can adapt their tasks according to inputs and be reprogrammed. During the 19th century, the robot industry is grown exponentially, with different kinds adapted to many domains, for the industry, home, emergency... Still, with the wish to recreate a human, androids are as old as the robot word itself, such as *Eric* made by William Richards and Alan Reffell. Then comes robots that have to deeply interact with humans, in a similar way as another person would do, to have *social interactions* and then called **social robots.**

1.1.2 Definition

The definition of a social robot can vary amongst roboticists, as mentioned by [13]. We will present here the definition we will use in this thesis. As a first base, we can use the definition from the *International Journal of Social Robotics*:

> Social Robotics is the study of robots that interact and communicate among themselves, with humans, and with the environment, within the social and cultural structure attached to their roles.
>
> — International Journal of Social Robotics [14]

Social Robot (SR) are designed to achieve many purposes, such as care [15], entertainment [16], education, or personal assistance [17]. Their modes of interaction with humans are diverse, like voice, screen, or gestures. Similar to human beings, robots can also transmit social cues [18] and express emotions [19] through movement, and not only by voice or face. In recent years, more and more robots have been made with the main purpose of interacting with people, and not doing simple tasks like assembling, packing (for industrial robots), hoovering or moving (for chore robots). Even if all those robots somehow interact with humans, they are limited to receiving instructions.

One can mention several recent commercial robots made to have social interactions (see Figure 1.1), with one of the pioneers: AIBO from Sony [16]. This dog robot was created to be used as its living counterpart, and to have interactions like them. The main function of AIBO is entertainment, and they "*communicate among themselves, with humans, and with the environment, within the social and cultural structure attached to their roles* [being a dog]". Even if AIBO was created for entertainment, people used it in other ways, such as football [20] or care [21]. Social robots, that have been made for interaction, are more prone to be used in other ways than their original goal; contrary to hoover robots, or industrial ones (even if we will see in the next chapter that is still feasible).

We can continue with robots that were made with one main function where social interactions are used, such as Paro from AIST [22, 23] and Pepper from SoftBank Robotics [24]. The first one is a therapeutical robot used to comfort and help old people by interacting with it as they would do with a pet, without the constraints of a real animal (hygiene, food, need to care...). Paro adapts its behaviours according to the patient's humour, like would do a real animal and *a fortiori* human. The robot is no more a simple object but can be personified by the user and be considered a member of the family. Thereby, users attached so much importance to their AIBO companion that they offered him a funeral ceremony when it *dies* [25]. Moreover, the personification can lead people to interpret some behaviours in a manner is was not programmed, see chapter 5 or to justify some reaction of the robot; e.g. mentioned that children explained to the researcher that AIBO was tired to justify its no reaction [26].

Pepper, on the other hand, was made to welcome people in stores or exhibitions as a receptionist. Contrary to AIBO and Paro, it is also using voice to communicate making its interaction deeper, in addition to its chest-screen. Even if it was mainly used as a receptionist at the beginning, developers and researchers tackle it and create diverse usages such as teaching [27, 28], or care [29]. In those fields, sociality is important to achieve their task and help to see the robot as more than just a tool. Another example is the robot Nao from Aldebaran Robotics [30], mostly used as a development platform for teaching robotics or for research, it is used in many projects. In addition to football like AIBO [20], it has been used in a TV show in France to act as a commentator and interact with the public and the (human) presenter [31]. It has also been used to help therapists to interact with children with autistic disorders [32, 33], using the ability of the robot to be similar to a human in his way of interacting.

Finally, we can mention avatar robots, such as Orihime [34] or Double [35]. Even if they are not autonomous and rely solely on the user, they are used to achieve social interactions when it is not possible otherwise, because of illness, for Orihime, or distance for Double.

Therefore, by using all of those examples, we can highlight what is a **social robot** and it will be used as a definition for the rest of this manuscript:

1. SR should interact with people besides their main task, deeper than just receiving instructions or giving information.
2. SR can be seen as more than just a tool and be personified, nay as a companion.
3. SR should move, at least. Verbal behaviours are not mandatory.

We can notice that personification is a necessary but not sufficient condition. Indeed, anything can be personified in some cases, even a computer, especially when it is not working "*why are* you *crashing?*"

1.2 Interaction With Robots

The way the robots are interacting is also important. Thus like humans, that not only use voice to interact but also gestures, gaze, and facial expressions..., a robot also has to take advantage of the movement. As mentioned before, we can also transmit some emotions in the movement, even unconsciously. To be considered as a robot and not just an object, or an enhanced AI (such as a Vocal Assistant (VA)), it should be able to use movement to express itself. Hence, in the case of a receptionist Pepper, if it was not moving its arms/head, it could also not complies with the first point of the definition. Indeed, the voice or screen would be here only to execute the main task: giving information. The hand gestures add another level of interaction. Moreover, as it will be more exhaustively explained in chapter 4, it is not required to have a humanoid/zoomorphic shape to create a social robot. Since the movement is enough to communicate emotion, an arm is, for instance, enough to transmit social clues [36]. However, social interactions can also be added to the speech with *unneeded* sentences to accomplish the main task, such as "*How are you?*", "*You look good today*", "*I remember you like coffee, so please find the nearby coffee shops*". Those little attentions make the robot more *human*, a social being and not just a machine.

Figure 1.2: Example of a long-term experiment in school with children and the robot Pepper [27]

The detection of human emotions or expressions is also important to manage good HRI, hence, Hayashi et al. also studied the impact of the transmission of emotion to a robotic arm by touch [37], the participant got a better impression of the robot with the touching capabilities, and the researchers could detect the transmitted emotions. Spezialetti et al. identified three main axis into the *robot-and-emotion* field [38]:

1. Formalisation of robot's emotional state.
2. Expression of this state.
3. Inferring human's emotional state.

In order to "*make the interaction more intuitive, genuine, and natural*". The transmission of emotions, along with the understanding of people's emotions is important to create a *two-way interaction*. Henschel et al. mentioned that is one of the key components of SR: "*When a robot failed to* [respond to a human in a social manner]*, people were disappointed and experienced a sense of dissonance*", in addition, the robot should be able to "*display thoughts and feelings*" [13].

Finally, another dimension that is important with SR is **time**. Indeed, in most cases, we will interact with those robots for a long time, and repetitively, especially if we want to bring them into our homes. That is why in this thesis the experiments last for a long time (during several days or two weeks) and also *in the wild* to have a glimpse of how people react with SR in unscripted scenarios. Already in 2013, Leite et al. realised a survey on long-term interaction with robots, even if some of them, like Roomba, are not social robots [39]. For the robots used for care and education, the duration of the interactions depends on the medical session or class duration as well as the repetition of use.

Thence, one of the two objectives of this thesis is to explore the HRI with non-verbal behaviours on a long-time basis and in a *wild* environment, and how to improve them.

1.3 Application of Social Robots

Figure 1.3: Two LOVOTs [40]

Leite et al [39] identified four environments where social robots were used: health care, education, work environment and public space, and home. And we can find a similar trend if we analyse the use, for example, of the Nao robot in the recent literature, most of the applications concern the health and education domains. If we do a small review on the *XPlore* search engine from IEEE with the keyword *'Nao robot'* in the last five years, 202 results emerged and most of them (130) concern technical solutions to improve, for example, the walk, or object detection. The rest (72) is divided into 5 domains presented in Figure 1.4, and a third of the studies concern the health and education fields. The robot is often used as a middleman or a companion in health therapy, or as a catalyser for education. Indeed those domains require a lot of social abilities, and then a *normal* robot would be less convenient[2]. However, no research uses the capabilities of Nao at home on a daily basis (besides for therapy). To highlight the importance that having the health and education fields for SR, we can mention the case of the robot Jibo [41], which was originally made as a personal assistant for the home, but after the purchase of the project by the Nippon Telegraph and Telephone company (NTT) has been changed to be used for health or education [13].

Chapter 6 will sum up the recent (social) robots used in a home context. As a first example of an experiment at home, there is Fribo [42], a social networking *robot* used to share with friends the activities of each one while living in their apartment. Even if it does not comply with our earlier definition, because it is not moving, we can still analyse the concept. The robot is used here to be an enhanced social network, without therapy/education aim. It is used as a catalyser in the relationship between the group of friends and helped them to bring some topic to discuss when they meet up face to face. Most of the

[2] Even if one can find logistic robots in hospitals or pharmacies.

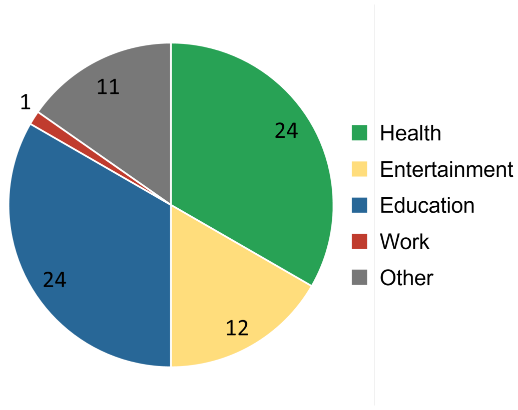


Figure 1.4: Number of studies in *IEEE Xplore* regarding Nao robot the last 5 years. The health category concern mostly therapist robots for autistic children, or to be a companion for old people in retired houses. The entertainment robot comports studies to improve the capabilities of Nao in the RoboCup. The "work" study concerns a case where Nao pretends to work in a supermarket to see the effect of explicit emotional adaptation.

current social robots for the home are either for entrainment[3] (AIBO) or assistant robots like the first version of Jibo, Elliq [17] or Nabaztag/Karotz [43]; being close to the current vocal assistant, such as Amazon Alexa or Google Assistant. Another more disruptive robot is LOVOT [40, 44], build to be a companion in the home and interact without verbal behaviours (it can use some sound reactions, such as tweeting). With more advanced AI than the previous robots to create more *natural* behaviours than AIBO for instance, its goal is to really replace a pet. It can be used in pairs, and in that case, both of them will have a different personality and interact differently, having also, in addition to HRI, robot-robot interactions. Since this robot is new (2018) no studies on its impact on long-term interaction, and the expectation of the users exists; but we can already find experiment for therapy with dementia patient [45], showing again the importance of health for the social robots.

Therefore, the second objective of this thesis is to explore the use of SR in domains that are less common, i.e. not for therapy or education, but, for example, in the home or work environments.

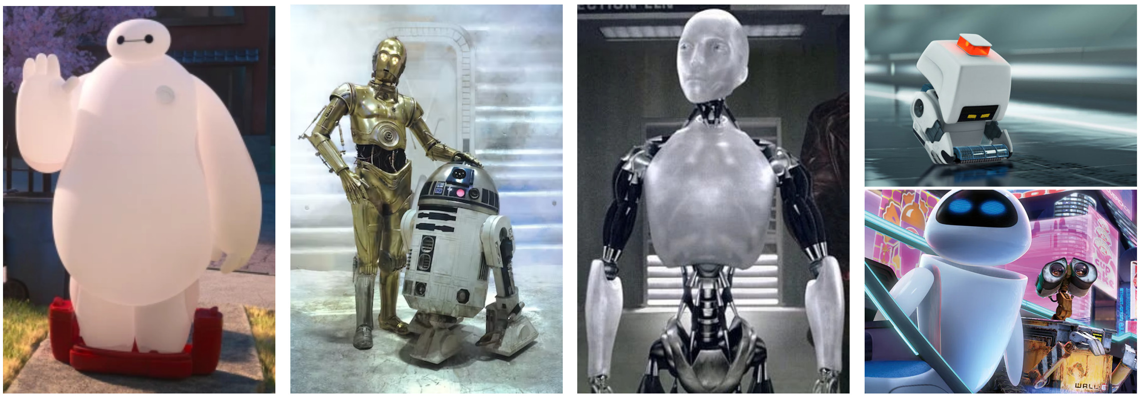
1.4 Motivation

Nowadays, the expectation people have about robotics is mostly shaped by Science fiction, where they interact freely with humans, see some examples in Figure 1.5. Their social interaction is similar to the one between humans, even if the shape is not humanoid. The imaginary portrayed in those works are often a world where humans and robots live in a close relationship, where robots are no more than just a tool. Whether those interactions are positive (e.g. Baymax, Fig. 1.5a or WALL-E, Fig. 1.5d) or dystopic (e.g. *I, Robot*, Fig. 1.5c or *Terminator*).

My goal, and the motivation of this thesis, is to lay one stone on the bridge that will close this gap between the perception and wishes that people have about robots in their way to interact with them, and the current reality. Nowadays, HRI is similar to the one with VA: a scripted communication. If the user does not follow the preset pattern, the agent either does not understand (*"Sorry, I didn't understand"*) or do something else. Thereby, we will try to explore how to make the interactions smoother, and novel kinds of interactions do not focus on voice only. But also to see how the general public perceives robots, indeed, even if they are getting more and more common, they are still rare in one's home; similar to the TVs during the 1950s.

That can lead to the following General Research Question (GRQ):

[3] Here we can include *being a companion* in this group.



(a) Baymax from *Big Hero 6*, Disney ©

(b) C-3PO and R2-D2 from *Star Wars*, Lucasfilm Ltd. ©

(c) Sonny from *I, Robot*, 20th Century Fox ©

(d) M-O, EVE and WALL-E from *WALL-E*, Pixar ©

Figure 1.5: Example of popular robots from Science fiction

GRQ How can we develop robotic systems that can communicate with humans in a more versatile manner, using less automatism by using their environmental data to change their behaviours?

We will reply by looking at different angles of the problem in the two main projects described in this thesis, each of them with its own sub-research questions. Using the environmental data (people's emotions, current temperature...) is a key in human-human interactions. If a robot keeps the same behaviours all the time, it becomes predictable for the user that have difficulties to see a social partner, but more an automaton.

1.5 Outline of the Thesis

The thesis manuscript is organised as follow: in chapter 2 we will describe how to create a questionnaire for experiments involving robot interacting with people, and the different tools available. Chapter 3 will present the experimentation of an industrial robotic system used outside factories, with the development of cognitive capabilities. Chapter 4 will describe the design and creation of a presence robot for the home, and in chapter 5 will be presented the technical evaluation of the robot. Chapter 6 details the creation of a smartphone application in order to integrate the robot into the smart home of the user, in an easy way. Chapter 7 will present an algorithm made for having long-term interactions with the robot developed in the previous chapters. Finally, chapter 8 will conclude this manuscript.

Chapter 2

Design of the Experiments

In order to prove the efficiency of their methods, researchers need metrics and create a specific protocol for their experiments. Their goal is to be able to compare their new methods with state-of-the-art or to prove the efficiency of their work. Then, before going to each project, we will see how to evaluate them and what kind of different tool exists for evaluating social robotics.

In the domain of HRI, the questionnaires are a common tool, since we evaluate the interactions a person had with a robot. Even if each study is different, it exists similarities between their questionnaires. Indeed, Rueben et al. [46] do not recommend creating one's own questionnaire from scratch, but rather reusing (and adapting) questionnaires that answer the selected concepts in previous studies. Then, some standardised questionnaires exist. But like a Lego game, researchers generally combine different blocks together to create the form that will fulfil their needs, as we will do in this thesis.

We follow the process described by Rueben et al. [46] to create a questionnaire suitable for the experiment, see Figure 2.1.

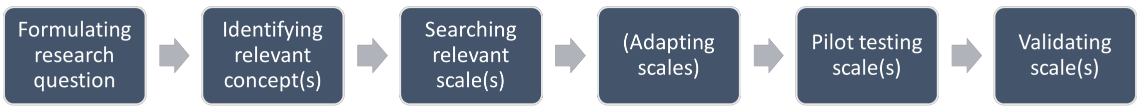


Figure 2.1: Standardised process for using questionnaires in HRI research, adapted from [46]

Even if the questionnaire seems to be a good tool to evaluate the HRI, and grasp the point of view of the participants, some other way exists. For example, physiological sensors can be used, such as electrocardiogram, electroencephalogram, or eye tracking. In the case of this thesis, we will see in chapter 5 that the logs of the robot are also used. The questionnaires are one metric, but not the only one, they have for them their simplicity of usage for the participants, but they require to be well constructed to be relevant.

2.1 Research questions & Relevant concepts

The first step of the process described in Figure 2.1, is to find the research questions and the relevant linked concepts. Since it depends on each experiment, it will be done on each relevant chapter (3, 6 and 5). Since each of them belongs to the same topic of social robotics and is testing close concepts, we can do here a literature review of the relevant scale for our field, and then use some of them in the experiments. We will see that each experiment shares some tools even if the research questions are different.

We can mention some concepts that will be studied in this thesis:

Naturalness The definition of the *naturalness* of a movement could seem as obvious at first sight, but it is in fact more complex. Like the definition of *life*. Everyone can tell what is alive or not, but when we try to create a precise definition, it is more complex. The studies that evaluate the naturalness of movements often concern anthropomorphic ones.

For example, Knopp et al. [47] compared the generated motion of their primitive model with actor movements. In order to evaluate the naturalness of the moment, they display to participants, at the same time, the same movement done by the two methods on an avatar. Then, they simply asked participants: "**On which side did you perceive the more natural movement?**". similarly, Aransih and Edison [48] evaluate the naturalness of movements done by autistic persons by comparing them with *neurotypical motions*. In addition to asking if the participants felt the movements were natural, they add a second question: "**Are you certain of your perception?**". This second question adds more precision to the binarity of the first one.

Legibility A robot action is defined as *legible* if its intentions are **understandable** by humans [49]. Indeed, when humans observe an action, they matched it with the internal representation they made of it. If this action is different from what they expected, they would perceive it as *unnatural* [48].

Predictability Strongly linked with the previous concept. Considering people create a mental representation of the action, if this representation matched the actual action, the person would have *predicted* the action. Actually, most of the studies measuring legibility are measuring predictability [49].

Perception of the robot/behaviours It is defined as the identification and interpretation of the robot and its behaviour. Because the body motions can also transmit emotions [50], it also implies the perception of the latter. Those mechanisms are done implicitly by humans. Concerning this concept, it can be related to robots *in general* or the robot that is evaluated.

Usability Is not only linked to robots, but to any system and refers to how a system is easy or hard to use. It can be linked to the similar concept of *user-friendliness* for software or websites. For a new product, it permits to the designers to observe if the users have difficulties with using it or not; and then doing the suitable corrections.

2.2 Relevant Scales & Tools

The next step of the process is to look for the relevant scales and tools. Hence, a literature review of studies in HRI evaluation has been done and several scales emerged. Based also on the previous works done in the lab, where, for example, the emotion of a generated motion has been evaluated with a questionnaire [51, 52]; since body movements transmit also emotions [50]. Additional parameters that might influence the perception of the movement are the personality of the person and their familiarity with robots. Based on the literature, different metrics can be used to evaluate the perception *per se*, or concept that influences it, the legibility/predictability and usability.

2.2.1 Likert Scale & Semantic Difference Scale

Both scales are similar in their construction, using *point* on an axis, from 1 to N, see Figure 2.2. The number of points N vary depending on the study, generally, it is an odd number in order to allow a *neutral* answer, but some studies can use an even number to force the participant to make a choice, called asymmetric scale [53]. The common values for N are 5 or 7 nay 9; in this thesis, the 5-point scale will be used.

The Likert scale [54] made in 1932 by Rensis Likert evaluates the answer to a question from *Totally Disagree* (1) to *Totally Agree* (N), and became one of the most popular ways to ask questions in forms, and not only for robotics. On the other hand, the Semantic Difference (SD) scale[1] uses the axis to evaluate the participant's opinion between two opposite adjectives, such as *small–big* or *slow–fast*. In order to select the items of this questionnaire, we follow recommendations presented in the Kansei Engineering ergonomics discipline [56**?**]. Kansei Engineering is a suitable human-centred and ergonomic approach often used to grasp and analyze the emotional, need, and

[1]Also called Godspeed scale [55]

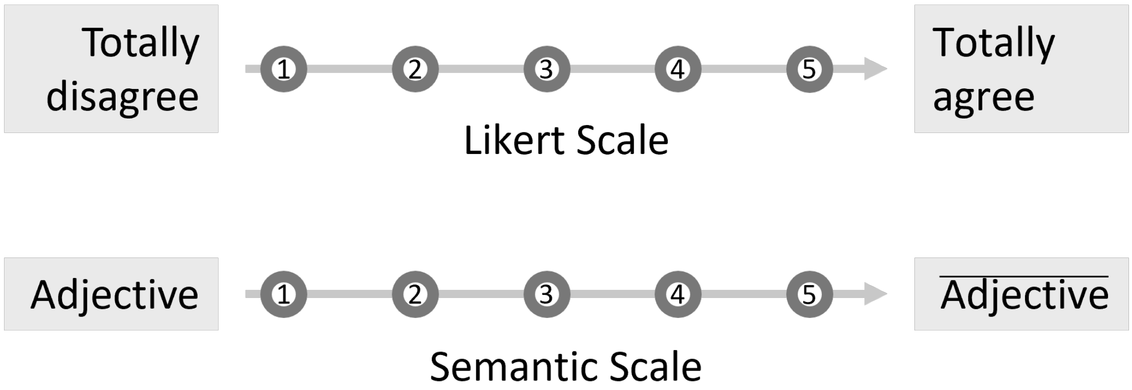


Figure 2.2: Example of a 5-point Likert and Semantic scales. The $\overline{Adjective}$ stands for the opposite of *Adjective*

social values of people towards products, interfaces, and services [57]. The first step in developing a self-reporting questionnaire is to collect a certain number of emotional words and adjectives that are relevant to the design of some specific product or application. Therefore, these adjectives can vary according to the application domain [56]. Additionally, these adjectives can be obtained after consulting experts in the domain or reviewing state-of-art articles.

For example, Dragan et al. used a 7-point Likert scale to evaluate the predictability of a robot motion, if it matches the expectation of the participants [58]. Eyssel et al. also used a 7-point Likert scale as an evaluation tool. Finally, Lichtenthäler et al. reviewed the literature about the legibility of robots, and the Likert is a mentioned tool with a 5-point and 9-point scale.

2.2.2 Negative Attitude towards Robots Scale

The NARS [59, 60] is a questionnaire made of 14 questions (see Figure 2.3) using 5-point Likert scale, in order to evaluate the feeling of people regarding robots. Thanks to this scale, we can observe if there is any bias in the relationship people have regarding robots. We can hypothesise that people that are not familiar with robots or even afraid will have more difficulties interacting with them. According to [61], NARS is the most popular questionnaire used in robotics to measure attitudes towards robots. These questions are classified into three sections: NARS-S1, NARS-S2, NARS-S3. Generally, the average value of these three sections is calculated and reported separately. These average values range between 1 to 5. The NARS-S1 section is composed of six questions and is designed to obtain attitudes toward situations of interactions with robots. The NARS-S2 section is composed of 5 questions and is designed to obtain attitudes toward the social influence of robots. Finally, the NARS-S3 section is composed of 3 questions and is designed to obtain attitudes toward emotions in interaction with robots [62]. Then, NARS-S1, NARS-S2, and NARS-S3 values are combined (average) to obtain the general NARS score[2], which value is also between 1 to 5.

2.2.3 Godspeed questionnaire

The Godspeed questionnaire [63] is a questionnaire used to gather the impression of the participants about the robot using a semantic 5-point difference scale where the adjectives have already been chosen (to be reusable). The questionnaire is divided into five sections, about anthropomorphism, animacy, likeability, perception of intelligence and perception of safety. It contains, in total, 24 questions, and has been translated into several languages (about a dozen).

[2]Because the NARS-S3 is using positive questions, contrary to others, the five-complement has to be used $(6 - score)$ before calculating the global score.

1. (S2) I would feel uneasy if robots really had emotions.
2. (S2) Something bad might happen if robots developed into living beings.
3. (S3) I would feel relaxed talking with robots.
4. (S1) I would feel uneasy if I was given a job where I had to use robots.
5. (S3) If robots had emotions, I would be able to make friends with them.
6. (S3) I feel comforted being with robots that have emotions.
7. (S1) The word "robot" means nothing to me.
8. (S1) I would feel nervous operating a robot in front of other people.
9. (S1) I would hate the idea that robots or artificial intelligence were making judgments about things.
10. (S1) I would feel very nervous just standing in front of a robot.
11. (S2) I feel that if I depend on robots too much, something bad might happen.
12. (S1) I would feel paranoid talking with a robot.
13. (S2) I am concerned that robots would be a bad influence on children.
14. (S2) I feel that in the future society will be dominated by robots.

Figure 2.3: The NARS questionnaire

2.2.4 CH-33 questionnaire

The CH-33 questionnaire [64] is also used to grasp the impression of the participants regarding the robots. It uses 33 Likert-scale questions, to evaluate six categories: performance, acceptance, toughness, harmlessness, humanness and agency. This questionnaire is originally done in Japanese, but an English version is available [65].

For example, this tool, in addition to the Godspeed one, has been used for example by Hu et al. to evaluate the perception of the participant towards the robot [65].

2.2.5 Inclusion of Others in Self Scale

The IOS scale evaluates the relationship of the participant with *other*, in our case a robot, as mentioned in [55]. It represents the relationship thanks to overlapping circles, with Venn-like diagrams. This representation is less ambiguous than words that have a more fuzzy meaning.

It can be useful for researchers to evaluate the impact of a robot on a person, especially for home robots. Indeed, if the person will have to stand alongside the robot every day, it is better for a good relationship. On another hand, it can be also a good metric to see the evolution of the relationship; like between humans, we are not close with our friends *at first sight* but get closer when we start to better know each other.

2.2.6 Self-Assessment Manikin scale

Based on the experimental evaluation of Ishida S. [51], we can use the following tools to evaluate the user's perception of the robot's action. It is linked to the Pleasure-Arousal-Dominance (PAD) model (see Figure 2.5a) that represents the emotion thanks to a three-dimensional model:

Pleasure (P) represents the positive or negative emotions caused by a situation;

Arousal (A) represents the degree of activity, psychological and physical arousal;

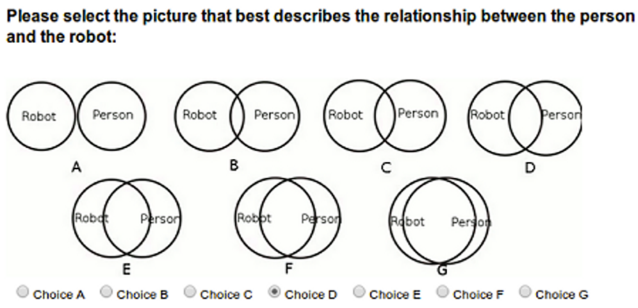
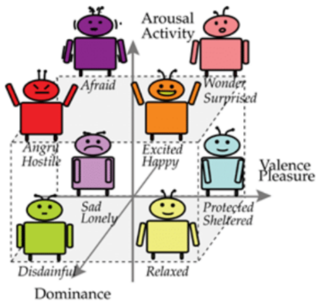
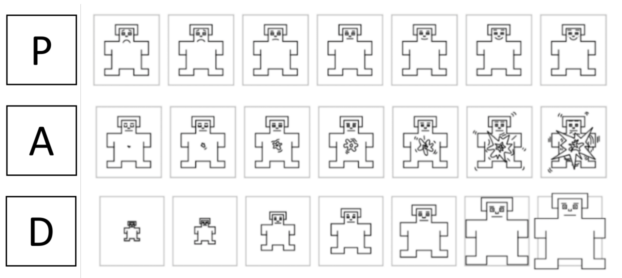


Figure 2.4: The IOS scale from [55]



(a) PAD [66]



(b) SAM [67]

Figure 2.5: The PAD and SAM scale

Dominance (D) represents how dominant you feel about a situation.

The Self-Assessment Manikin (SAM) scale [67] evaluates the PAD thanks to images, see Figure 2.5b. The drawing helps to avoid some misunderstandings or linguistic issues that words could bring. Moreover, since the drawings are more intuitive, answering the questions is easier for the participants, and do not need to be translated.

2.2.7 Big Five personality test

This test [68] uses 5 indicators to evaluate the human personality, it is widely used in psychology and cognitive science. This 15-question test is used to get the personality of the person, and then to see if there is a relationship with the emotion recognition found with SAM. The different questions are indicated in Figure A.1, and the different personality traits below:

Extraversion Extraverts get their energy from interacting with others, while introverts get their energy from within themselves. Extraversion includes the traits of being energetic, talkative, and assertive. Questions 1-3 relate to this.

Neuroticism Neuroticism is also sometimes called Emotional Stability. This dimension relates to one's emotional stability and degree of negative emotions. People that score high on neuroticism often experience emotional instability and negative emotions. Traits include being moody and tense. Questions 4-6 relate to this.

Openness People who like to learn new things and enjoy new experiences usually score high in openness. Openness includes traits like being insightful and imaginative and having a wide variety of interests. Questions 7-9 relate to this.

Conscientiousness People that have a high degree of conscientiousness are reliable and prompt. Traits include being organized, methodic, and thorough. Questions 10-12 relate to this.

Agreeableness These individuals are friendly, cooperative, and compassionate. People with low agreeableness may be more distant. Traits include being kind, affectionate, and sympathetic. Questions 13-15 relate to this.

2.2.8 System Usability Scale

One popular way to measure the usability of a system is the SUS developed by Brooke J. [69, 70]. With a 0–100 grade, it evaluated the system's success, while also providing data about the trends regarding the interaction flow and primary users' takeaways, using ten 5-value Likert scale questions, see Figure A.2. Brooke described its own scale as "quick and dirty", giving a first glance of the usability while being simple to use and implement. Even if the authors of [71] mention limitations of the SUS scale, such as the lack of precision (what is exactly the usability problem), it is still a pertinent tool to have a first quick glance about the product's usability.

2.2.9 Demographic Data

Finally, demographic questions are also a common tool in questionnaires. Those personal items allow the researcher to better *know* each participant, to categorise them. Indeed, even if every person is unique, we often react in the same way that person belongs to the same group; and it can also be useful to discard some bias, hence in chapter 3 we will that all demography-based hypotheses could not be validated. Common criteria are *age*, *gender*, *nationality*, *profession* or any relevant personal question, such as *lnowledge about robots*.

2.3 Adapting the scales and design of the questionnaire

After grasping all the tools, the questionnaire can be created by assembling them together and then answering the research question. However, researchers should be careful about the length of that questionnaire. Thus, Martinez et al. describe in [72] that questionnaires for visitors in museums and expositions must be breve and simple, otherwise "it is unlikely that visitors will fill out the questionnaires if they are too long to complete or too difficult to understand, and responses will not reflect the real experience". Moreover, many practitioners, textbooks, and research articles suggest that long questionnaires should be avoided [73, 74, 75] to prevent careless responses and respondent fatigue as well as to motivate visitors to participate in the survey.

Summary

— It is recommended to not create a questionnaire from scratch, but to reuse already made questionnaires.

— The first step is to find the research question(s) and the related concept(s).

— Seven tools were described: (i) Likert scale, (ii) Semantic Difference scale, (iii) Negative Attitude towards Robots Scale (NARS), (iv) IOS scale, (v) SAM scale, (vi) Big Five test and (vii) SUS. We will choose amongst them to create the questionnaire for each experiment.

— Researchers should pay attention to the length of the questionnaire.

Chapter 3

Preliminary Work: Expanding the Frontiers of Industrial Robots Beyond Factories

With this first work, we will evaluate how people could perceive an industrial robot that was not made to have social interactions. As a preliminary work to experiment more deeply on the social interactions at home/work, in a familiar environment for the user, here the robot will be working outside the controlled environment of a laboratory. In addition, participants will use it without prior training, in order to have spontaneous interactions and reactions.

3.1 Introduction

Robots with advanced interactive capabilities are promising alternatives toward more efficient and flexible industrial systems [76]. However, employees' lack of acceptance and positive attitudes towards robots have limited the spread of collaborative robots in factories [77, 78]. Moreover, it is expected that the current academic and commercial interest in the use of these advanced robots expands its frontiers beyond factories to more ecological and everyday-life scenarios, such as restaurants, shops, and homes [79]. Unlike traditional efforts in industrial robotics generally focused on increasing task performance, the next generation of robotics applications enabling interaction with humans require the consideration of hedonics (i.e., emotions and desires) as well as ergonomics factors [80, 81, 82]. These novel research activities will enable manufacturers and designers to build alternatives or counter-measures to increase the acceptability, the social impact, and the desirability of interactive robots [83, 77]. Many research articles, such as [84, 85, 86] suggest that a low acceptance and negative attitudes toward robots are in part produced by the mismatch between people's expectations (in part due to social media, movies, or inexperience with robots) and the real capabilities of robots. Robots must be able to work outside laboratories to evaluate more realistic expectations. However, due to the complexity required to create robust applications with robots, Human-Robot Interaction (HRI) applications are rarely tested in uncontrolled and public settings. Moreover, they are often presented and evaluated with convenient samples (e.g., laboratory and faculty members) [87, 88]. This approach makes data collection more manageable and avoids many technical issues often presented when robots are required to perform in open, uncertain, and crowded environments. However, there is a need in the HRI community to move towards approaches that enable the acquisition of more valuable theoretical and technical insight through the use and validation of robotic systems in natural, open, everyday environments. This approach is called in literature as HRI *in the wild* [87, 88]. Common types of robots used in this emergent paradigm are social and service robots. Some examples are presented in [27, 28, 89]. In many cases, these robots are either pre-programmed or remotely controlled rather than fully autonomous. Some recent examples using this methodology are [90, 91, 92]. In this article, we affronted the challenge of bringing an industrial robot *to the wild* through the development of a multi-modal and distributed system architecture. This architecture integrates advanced sensors, effective deep learning methods, and a human-machine graphical interface to enable fully autonomous HRI.

With the aim to first evaluate the interactions of humans with robots, a project has been developed with an industrial robot made by Kawada Robotics and called NEXTAGE Open [93], see the Appendix B for the detailed characteristics of the robot. The goal was here to see how people can interact with a robot not initially made for social interactions.

3.2 Related work

Exposure to robots is basically done in three ways: *no HRI* (i.e., participants are not asked to imagine, view, or interact with a robot), *indirect HRI* (i.e., participants are asked to imagine an HRI task or observe an interaction with the robot using videos or images) and *direct HRI* (i.e., the robot is physically present and interact with participants) [61]. Several cross-cultural studies about acceptance and attitudes toward robots in different domains have been presented in previous works, such as [94, 95, 96]. However, in these types of studies, no HRI or indirect HRI is used. While the study of these non-functional and human-centred aspects through direct interaction is a popular topic in some social robotics areas, such as Robot-Assisted Therapy (RAT) [97], education [98] and elderly-care [99], the attention of these aspects when performing advanced Human-Robot Interaction (HRI) activities using industrial and collaborative robots is still limited [100]. A recent review of research articles evaluating attitudes, anxiety, acceptance, and trust in social robotics domains is presented in [61]. They discovered that "studies providing direct HRI may report different attitudes to studies where participants do not directly interact with a robot". Moreover, attitudes can change depending on the application domain and the design of the robot (e.g., humanoid-like or anthropomorphic). In the context of industrial robotics, studies collecting attitudes, expectations or perceptions towards robots after a direct HRI interaction are still rare. For example, Aaltonen et al. [101] grasped possible expectations of factory workers from the industrial and academic points of view. However, data collection was performed with an online questionnaire with no direct HRI. Moreover, experimental research with industrial and collaborative robots has been mainly limited to lab experiments with convenience samples (e.g. students or people working in a research lab) [102]. Laboratory studies are still mainstream in HRI. However, validations of robotics systems *in the wild* will increase more and more in relevance as increases the need to have advanced robotic systems able to be used outside factories and laboratories.

Museums and exhibitions are suitable *in the wild* scenarios for exposing novel technological achievements to people with different backgrounds and interests [103]. An example is proposed in [104], where an ethnographic study is performed in a museum with a humanoid robot. Other examples of HRI demonstrations *in the wild* scenarios using FROG (a tour guide robot) are presented in [105, 106]. Other suitable locations for *in the wild* experiment are shopping malls, such as the work done by Kand et al. [107]. They used a robot (Robovie-IIF) to help customers to find their way in the mall. The robot was partially teleoperated, to overcome speech recognition of algorithm difficulties, and to keep the interaction smooth. We can also mention some of our previous work with robot interaction in school [27, 28]. In those studies, the Pepper robot was used to entertain children in their school with some activities, in uncontrolled scenarios. In the current article, we affront the technical challenge of designing and executing an HRI application with a dual-arm industrial robot in a public and crowded scenario, specifically a robotic exhibition.

Table 3.1 summarises recent and similar works reporting the development of robotic systems with industrial robots able to share the same space with people and that collect the users' perceptions towards robots after performing direct HRI. This table shows that one of this article's main novelties/differences against previous works is the experimental setting, which is not performed in a closed room or laboratory. Instead, validation of our proposed system is performed in a public and noisy scenario with no control of environmental conditions (e.g., illuminations or the number of people in the field of view of the robot). Similar to [100], our proposed system is integrated into a dual-arm industrial robot. Unlike [100], where the robot is remotely controlled, the robotic system proposed in this article is fully autonomous. Moreover, many systems developed in previous and similar works were developed to follow an industrial task, which in most cases requires some previous training. An exception is [108] where a straightforward but suitable task is proposed for

Article	Robotic platform	Setting	Task	Autonomy	Training required	Participants
Müller-Abdelrazeq et al. (2019) [77]	Universal Robot 5 robot arm	Laboratory	Assembly task	Fully autonomous	Yes	90 subjects mainly students from a technical university
Rossato et al. (2021) [109]	Universal Robot 10e robot arm	Laboratory	Collaborative task	Fully autonomous	Yes	20 industrial senior and younger workers
Drolshagen et al. (2021) [108]	KUKA LBR iiwa 7 R800 (robot arm)	Closed room	The robot pick up wooden sticks to hand them over to the worker	Fully autonomous	No	10 participants with mental or physical disabilities.
Elprama et al. (2017) [100]	Baxter dual-arm robot	Closed room	Participants instructs the robot to put blocks inside boxes	Remote controlled	Yes	11 car factory employees
This work	NEXTAGE Open dual-arm robot	Public space	The robot gives gifts to visitors according to their instructions and facial expression	Fully autonomous	No	Hundreds, but only 207 answered some questionnaires.

Table 3.1: Research articles proposing interactive robotic systems with industrial robots for grasping perceptions towards robots after performing direct Human-Robot Interaction

enabling people with mental and physical disabilities to interact with an industrial robot. These complex industrial tasks, such as assembly, are inappropriate for a robotic exhibition where people interact with the robot voluntarily and have no time for complex explanations. Therefore, we designed an intuitive socio-emotional scenario where the robot can guide and adapt to human actions and emotions.

3.3 Objectives and contributions

This article proposes a novel application where a dual-arm industrial robot, originally designed to be used by industrial workers, can interact with people from different backgrounds and outside laboratories and factories. The main objective of the project presented in this article is to develop an intuitive, social, and engaging interactive scenario for the International Robot EXhibition (IREX) using the NEXTAGE Open robot by *KAWADA ROBOTICS CORPORATION* [93]. Rather than proposing a typical scenario where robots require to be isolated from visitors or a cooperative industrial task (which often requires previous training), we developed an application where humans intuitively interact with an industrial robot through emotions and body motions. In this way, visitors can experience what is to interact with an advanced robot with advanced grasping, perceptual, and cognitive capabilities in an everyday-like situation. Therefore, the main contributions of this work are (i) the creation of an intuitive and engaging HRI scenario that integrates the NEXTAGE Open dual-arm industrial robot; (ii) the execution and validation of the proposed HRI scenario in an unconstrained, crowded, and dynamic environment; and (iii) validation of the proposed scenario using self-reports that grasp the emotional experience towards the robot platform and the proposed HRI scenario of participants as well as their potential needs and attitudes towards industrial robots with social capabilities.

Moreover, this system is validated with more types of users, rather than only factory workers or faculty members (e.g., students).

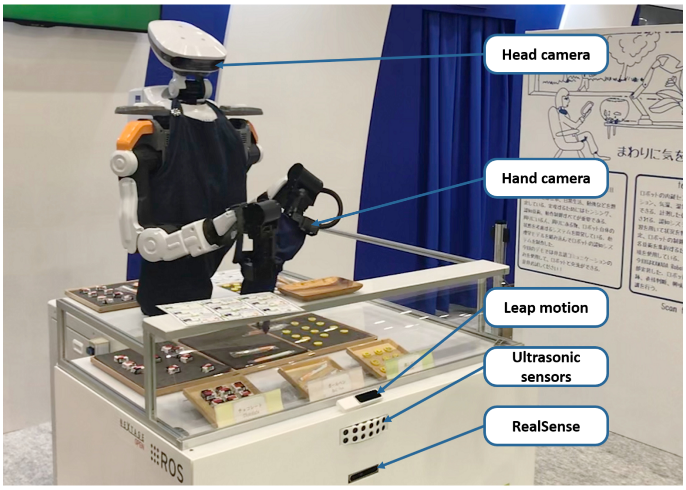


Figure 3.1: Robot on the booth

3.4 Design and Implementation

3.4.1 Hardware

The experimental setup is composed of four parts. The main one is the *NEXTAGE Open* robot from Kawada Robotics [93]. It is an upper-body anthropomorphic robot with 15 DoFs, 6 for each arm, 2 in the neck and 1 for the waist. The robot also has 4 cameras, 2 in the head and 1 in each hand. The hand are composed of a three-fingered pneumatic gripper, and the payload for each is 1.5 kg. The robot is controlled by an Intel *NUC* PC on Ubuntu 16.

The second part is composed of three sensors used to detect the user's interactions with the robots. The **people detection** is achieved with two devices (redundancy for robustness): a *RealSense* depth camera (RS), and a set of ultrasonic sensors (US). The latter is a custom made array of six *HC-SR04* sensors, see Figure 3.1. Both are used to estimate the proximity of the user to start and sustain the interaction and are placed below the robot on the front panel. Only the sensitive part is visible, thanks to two slits. The **user's choice detection** is done with a *Leap motion* placed on the booth's edge, in front of the object's window. It detects the choice between three positions: left (chocolate), centre (pen) and right (eraser). A mark was placed on the floor to indicate the user's ideal position which maximizes detection.

The cameras are used for two purposes. The embedded **eye camera** (EC) detects the person's facial expression and the *engagement* of the user. According to those, the robot will change its behaviour, by knowing, for example, if the user is engaged in the interaction. The **hand cameras** (HC) are used to detect the position of the objects on the table, and to grasp them.

The third part is placed behind the robot: a monitor was installed on a wall to display the Graphic User Interface (GUI). It indicates to the user *what to do*, see Figure 3.2. The instructions are written in English and Japanese; visual feedback from the cameras (user's face with his engagement and the recognition of the objects from the HCs) is also displayed.

The last part consists of the computers. Besides the computer used to control the robot, two computers composed the apparatus. They run the algorithms that are divided on both to maximise robustness, see section 3.4.2.

3.4.2 Software

The software architecture is described in Figure 3.3. The first PC (Desktop: RAM 32 GB, CPU Ryzen 1900X 8 cores 3.8 GHz, GPU Nvidia GeForce RTX 2070) is dedicated to vision, with three resource-consuming algorithms. The second PC (laptop Dell *Inspiron* 14 5000: RAM 8 GB, CPU i7-7500 2.70 GHz) runs the sensors scripts. Data are gathered on two blackboard scripts (i.e., shared working memories) and sent to the control PC and the GUI using NEP [110], on a WLAN.

Figure 3.2: Main elements of the designed Graphical User Interface (GUI)

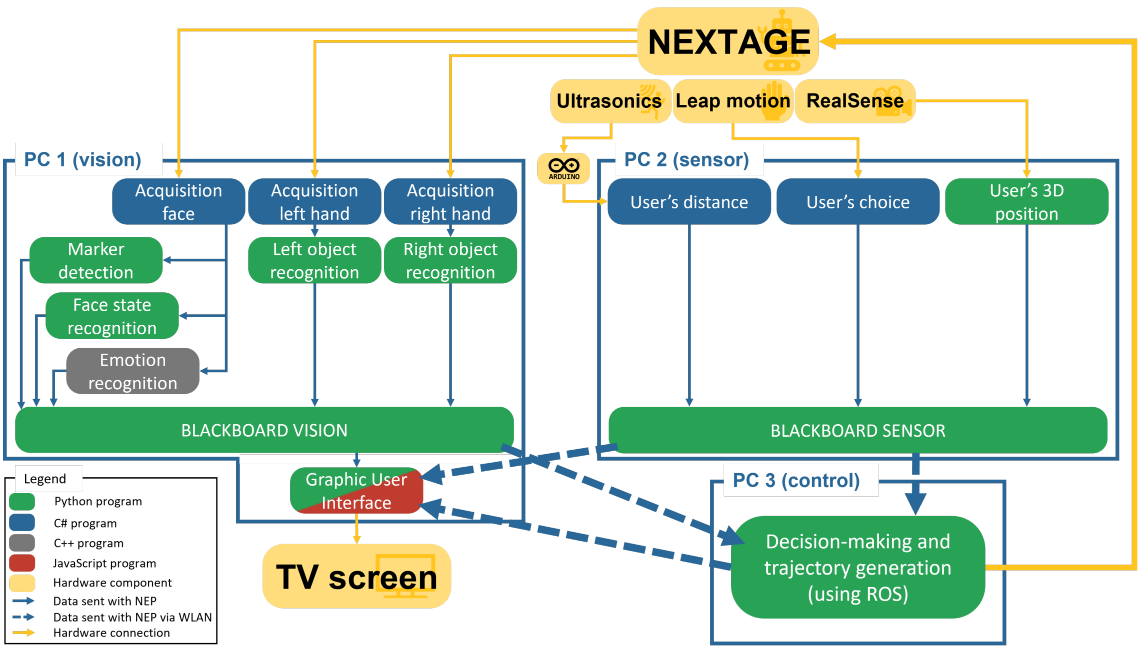


Figure 3.3: General architecture of the project. The GUI program is using client-server architecture, the server is in Python and the client in JavaScript

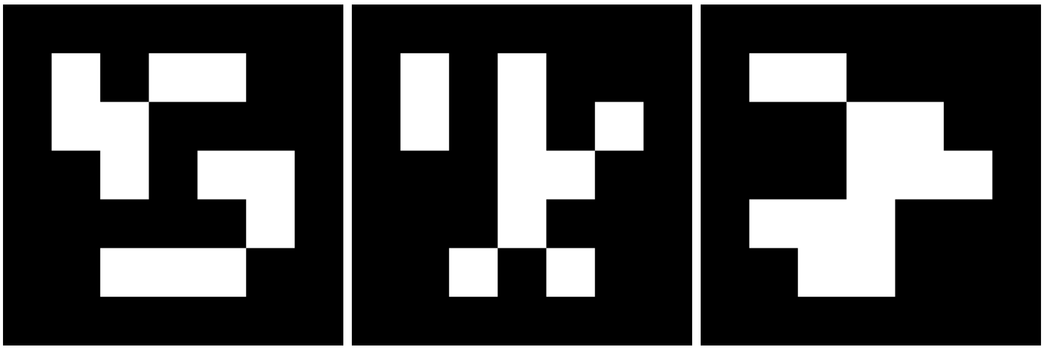


Figure 3.4: Example of markers used by the staff

a) Eye camera algorithms

The first algorithm **detects the gaze direction and head orientation** (and then determines the engagement) of the user regarding the robot. If s/he is not looking it in the eye, it will continue its task until the user is *engaged* in the interaction. We are using an algorithm we developed, described in [111].

The second one **recognises five different emotions** of the user, if s/he is *happy*, *sad*, *surprised*, *angry* or *neutral*. The emotions are used to determine if the robot will give or not the gift. If the emotion of the user is negative (e.g. *angry*) the robot will take the gift back and let the user choose another one. We assume the user changed their mind and would prefer another gift.

The last algorithm using the video stream from the EC is to **detect the markers**. The marker is a QR-code-like image (see Figure 3.4) printed on the badge of the staff members to activate particular functionalities of the robot. For instance, the robot will give the object directly in the user's hand by tracking it. It was not available to the general public for safety reasons, and the gift was dropped off on a tray in front of the user.

b) Hand cameras algorithms

Two identical algorithms run simultaneously to process the video stream of each hand. This algorithm performs object recognition by feature extraction and classification. Different customized models were created corresponding to the objects the robot has to manipulate. The Faster RCNN was applied, with these models taking advantage of its high capacity to detect different objects with enough precision.

The algorithm was trained to detect the three objects the robot has to grasp: a *pen*, a *chocolate* and an *eraser* (with a smiley shape), Figure 3.5. The model was trained with about 300 images; with one or several objects, fully displayed or partially covered or with different luminosity and background.

The algorithm's output is the object's name, the bounding box (BB), and the confidence score (the probability of being a true positive). Only the objects detected with a confidence score over 75% are used, and the bounding box coordinates are sent to the control algorithm. Moreover, for the chocolate and the eraser, their BB is expected to be square, so the shape of the BB is checked. If it is a rectangle and not a square, it means the object is partially in the field of vision. Hence, the centre of the BB is not the centre of the object, and to avoid grasping problems, the object is discarded until it becomes fully visible.

c) Sensors algorithms

The algorithm is separated into two parts. The first one is on an Arduino where the six ultrasonic sensors (US) are connected. The script loops on them to get the current distance data and send it *via* serial communication to the sensor PC. A second algorithm grasps the data from the six sensors and sends them by NEP to the blackboard sensor script that gathers all the sensor data before sending them to the control algorithm.

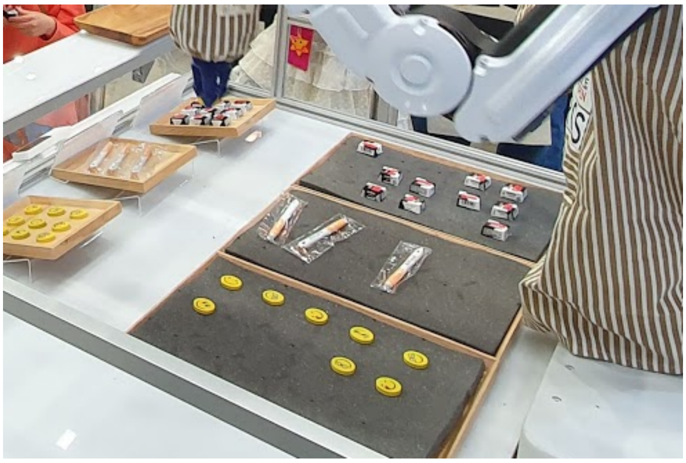


Figure 3.5: Real scene with the objects the robot has to detect and grasp

d) Control algorithm

The algorithm is detailed in Figure 3.7. Four threads are running in the background to gather the data sent via NEP. Moreover, the algorithm is using a virtual grid to track the objects on the plates. The quantity and position of each object on that grid are recorded. The cameras are used to know the precise position of the different objects and facilitate the grasping task.

The **Working** task is used when no one interacts with the robot. In this case, the robot is moving objects on the grid randomly. When a user is detected, the robot stops this task and turns its head to the user. If it has an object in its hand, it proposes it to the user and gives it to him/her, if the facial expression and engagement are positive.

The robot then **Presents the objects** with a hand gesture (and on the TV screen as well) and waits for the user to point at the object of their choice, the pointing is detected by the Leap Motion sensor. The robot picks the selected object and offers it to the user. While doing that, the user's facial expression and engagement are checked, and according to them, the robot will give the object or pick a new one (see section a)).

The **Refill task** is done when the robot has nothing to do with the user (i.e. during the **Working** task) if the plates are almost empty. If they are empty, the task is done even if a user is present. To refill, the robot turns 90° to access a big plate with different objects. This plate is easily accessible by the staff even during the running of the robot, with safety. Because the objects are placed randomly on that plate, the robot relies only on the hand cameras with the object detection algorithm to pick the correct object.

For simplification, two details are not in the graph. The first one is the detection of the **QR Marker**, which is done when the user approaches. Thus, the **Give object** action will be different, as described in subsection a). The second one is the **Invite next user** action, which is done after two interactions of the same user. Indeed, we limited users to 2 consecutive interactions to avoid monopolisation.

e) Human-Robot Graphical User Interface

We designed a Graphical User Interface (GUI) able to: a) provide feedback to people interacting with the robot about the decisions and actions taken by the robot (e.g., the current facial expression of the user that has been detected, and selected gift), b) give instructions to users about how to proceed in the interaction, and c) enable developers to monitor the status of sensors and the results from deep learning algorithms. This GUI was developed using modern web technologies and JavaScript libraries, such as Node.js, HTML, CSS, and Vue.js. The interface subscribes to the blackboard sensor and blackboard vision modules and the decision-making modules using NEP. The information obtained from these external modules is used to change dynamically the elements displayed to the users. Figure 3.2 presents the basic sections of this interface. Figure 3.6 shows examples of the different elements changed according to the robot and sensors' status. When there

a) b) c)

Figure 3.6: Examples of elements shown in the GUI according to the status of the interaction: a) No human interaction is performed, b) the interface shows which object is selected by the human, c) the robot provides information about what is doing and how it feels

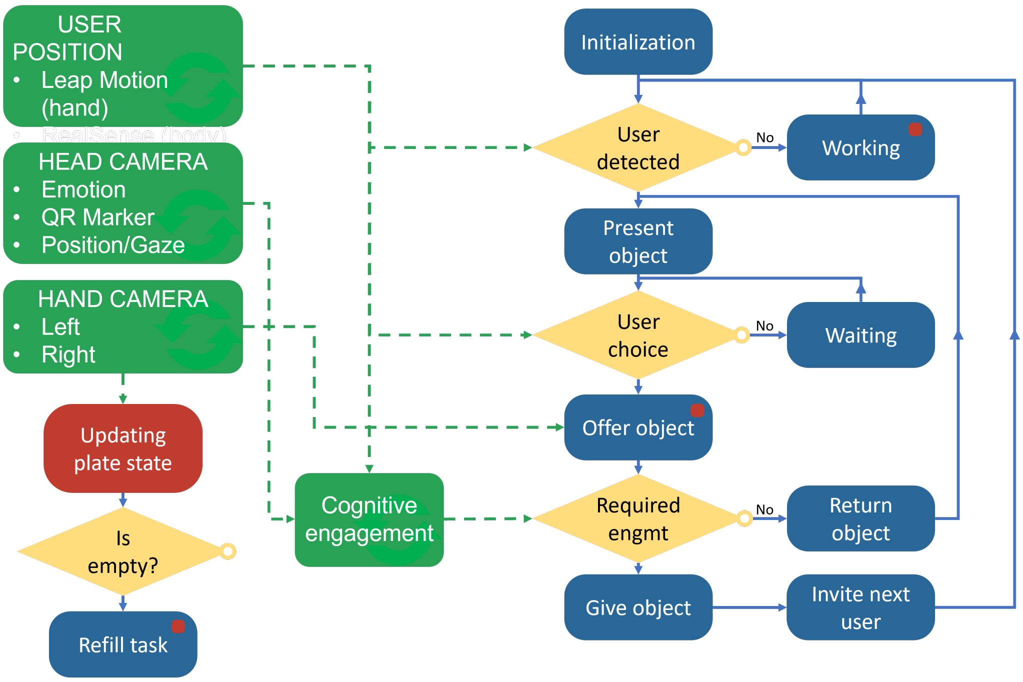


Figure 3.7: Flowchart of the command of the robot. The green boxes are threads running in the background and updating data. The actions with a red mark are calling the *Updating plate state* one. The *Cognitive engagement* thread estimates the engagement of the user towards the robot thanks to the user's position, emotion or gaze.

is no interaction between the robot and a user, the background colour of the interface is green; otherwise, it is blue.

3.5 Validation *in the wild*

The experimentation took place at the International Robotics Exhibition (IREX) 2019, in Tokyo, Japan. It is a public event for four days at the end of December, before the COVID-19 pandemic. As presented in section 3.3, the main objective of this work was to build a system architecture that integrates state-of-the-art perceptual tools to provide engaging experiences for visitors of IREX 2019. In order to know if this objective was met and to identify possible improvements for posterior iterations, we formulated the research question Q1 (see below). Additionally to the main goal of this article, we also grasped the attitudes and expectations of visitors of IREX after a direct interaction with an industrial robot with affective and cognitive skills. Therefore, we formulated the next two additional research questions Q2 and Q3.

Q1: What are the emotional reactions and perceptions of visitors towards the proposed interactive scenario?

Q2: What is the attitude of visitors towards robots after direct interactions with an industrial robot with affective and cognitive skills?

Q3: What are the potential expectations of visitors towards robots in their working and everyday environment?

3.5.1 Subjective validation of the proposed system

We use the semantic difference (SD) (q.v. section 2.2.1) self-reporting questionnaires to grasp participants' emotional reactions to the robot and the proposed HRI. The results from these questionnaires are used to answer the research question Q1. We collected 30 possible pairs from state-of-art works on industrial and social robotics. Finally, 18 pairs were selected to be applied in different 5-point Kansei (K) questionnaires. We classified these pairs of words into two sections: K-1 and K-2, they are presented in Tables 3.3 and 3.4 respectively. While K-1 was designed to grasp the feeling of visitors about the HRI scenario (using 6 pairs), K-2 was designed to grasp the impressions about robot design, usefulness, and skills (using 13 pairs). Additionally, two open questions, identified as OP-1 and OP-2, were applied to answer question Q3. These two questions were defined as follows:

OP-1: *If you had to work together with a robot, what would be the main characteristics you think the robot should have?*

OP-2: *If you had to live with a robot, what would be the main characteristics you think the robot should have?*

We use the NARS scale (refer to section 2.2.2) to grasp the attitudes of participants after a direct interaction with the robot and in this way answer question Q2. We use the original version of NARS defined in [60], which is in the Japanese language and its version in English. In total NARS-S1, NARS-S2, and NARS-S3, and the proposed questions of K-1, K-2, OP-1, and OP-2 are composed of 35 questions.

It is relevant to highlight that the research methodologies and objectives of this article are different from most classical HRI research activities performed in laboratories or structured environments, where participants are hired or are members of the same laboratory or school. This type of participant has time and motivation to effectively answer large questionnaires built with five or more items for the same psychological or usability construct. This traditional approach is often applied to prove the validity and reliability of the results using tools such as Cronbach's alpha [112]. In the application presented in this article, qualitative data is obtained from visitors

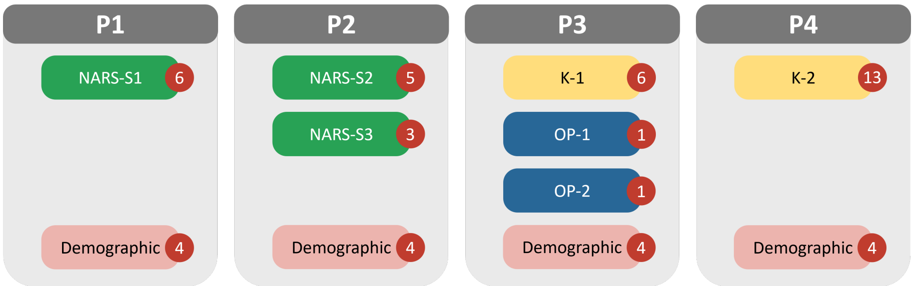


Figure 3.8: The content of each questionnaire P1–P4. The number in the red circle indicates the number of questions.

of a robotic exhibition that participated voluntarily. Therefore, many typical considerations for increased validity, sensitivity, and reliability done in structured, descriptive, or explanatory research performed in laboratories are not suitable and are out of this project's scope. In this context, a common topic of discussion in the literature is when to use single or multi-items for the same construct in self-reporting questionnaires. For example, [113] discusses how the use of multiple-item measures is costly, aggravates respondent behaviour, and increases response errors. Moreover, even "a second or third item of the same construct contributes little to the information obtained from the first item". Diamantopoulos et al. in [114] suggest that single-item approaches are viable options in exploratory research. This type of research is usually performed at a preliminary stage, in unstructured settings, such as in this work. Recent works discussing the suitability of single-item or multi-item constructs are [115, 116]. In practice, single-item approaches can be considered suitable for well-understood constructs, such as those measuring satisfaction. Moreover, multi-item approaches are more suitable for complex constructs, such as trust and attitudes [117, 114]. In this work, we use a single-item approach for measuring satisfaction-related constructs and a multi-item approach (NARS) to measure attitudes. Moreover, as described in section 2.3 about the length of the questionnaires, we divide the proposed 35 questions among four questionnaires (P1, P2, P3, and P4) that were applied on different days in the IREX exposition. NARS-S1 questions are asked in P1. NARS-S2 and NARS-S3 questions are asked in P2. K-1 (6 items), OP-1, and OP-2 questions are asked in P3. Finally, K-2 (13 items) questions are asked in P4. The first part of the questionnaire P4 (composed of 6 items) was used to grasp impressions of perceived intelligence and animacy towards the proposed robotics system. The second part of the questionnaire P4 (composed of 7 items) was used to grasp design-related aspects of the robot platform. Four demographic questions (age range, gender, country, robotics experience) are included in P1, P2, P3, and P4. We provided two versions of each questionnaire, one in Japanese and the other in English, participants were free to choose the language that suited them the most. Figure 3.8 sums up the content of each questionnaire.

3.5.2 Participants

Participants of this study are visitors of IREX 2019 that voluntarily interacted with the proposed HRI system. After they interacted with the robot, we asked them if they could fill out one of the questionnaires described before and by doing so give consent to use their answers for research purposes. No personal data were collected. The total number of participants answering is 207. From the questionnaires P1 and P2, participants (5 and 2 respectively) were discarded because they did not answer all of the questions. We discarded participants only in P1 and P2 because the NARS score is built from the answer to every question. Contrastingly, the answers to the questions of P3 and P4 can be taken individually, then, empty answers are not discarded. Table 3.2 sums up the demographic data of each questionnaire. Because IREX is held in Tokyo, Japan, most of the participants are Japanese (between 70% and 86% depending on the questionnaires). We also divided the participants into groups according to their knowledge of robotics: *novice* (1/5 and 2/5 on the Likert scale), and people that are knowledgeable about robots, called *expert* (3/5, 4/5 and 5/5).

Table 3.2: Demographic data of each questionnaire. Some participant did not mention their gender or country. For the latter, we assume they were Japanese or not depending on the version of the questionnaire they used (Japanese or English).

	P1		P2		P3		P4	
Considered answers	78	100%	67	100%	47	100%	44	100%
Japanese	57	73%	47	70%	40	85%	38	86%
Non Japanese	22	28%	20	30%	7	15%	6	14%
Male	62	79%	55	82%	26	55%	27	61%
Female	15	19%	12	18%	21	45%	17	39%
Novice	21	27%	22	33%	25	53%	23	52%
Expert	57	73%	45	67%	22	47%	20	46%
OP-1					18	38%		
OP-2					21	45%		

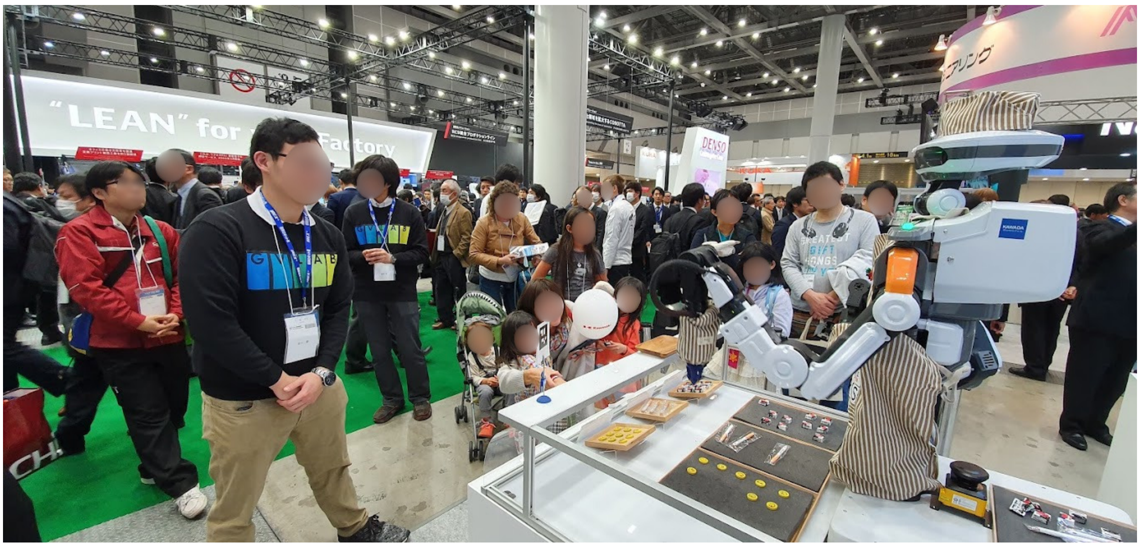


Figure 3.9: The robot was in front of an alley with a lot of traffic, this was challenging for the algorithms that were developed in a controlled environment.

3.6 Results

3.6.1 User perceptions and emotional reactions

We use the results from questionnaires P3 and P4 to evaluate the answer to Q1 (*What are the emotional reactions and perceptions of visitors towards the proposed interactive scenario?*). Results from emotional reactions (P3) are shown in Table 3.3. Table 3.4 contains the result from the P4 questionnaire, regarding the feeling of the participants regarding the robot. We can gather some dimensions into similar *concepts* ($\mathcal{C}$), regarding the **satisfaction** (α) and **comfort** (β) of the participants after interaction with the robot (K-1); and **behaviour** (γ), **interactions** (δ) and **appearance** (ε) for their impressions (K-2).

3.6.2 Negative Attitudes Towards Robots

We use the results from questionnaires P1 and P2 to answer research question Q2 (*What is the attitude of visitors towards robots after direct interactions with an industrial robot with affective and cognitive skills?*). The mean value and standard deviation (SD) for NARS-S1, NARS-S2 and NARS-S3 are summarised in Table 3.5. The mean values of each NARS section can be interpreted as positive (values close to 1), neutral (values close to 3), and negative (values close to 5). We divided the participants into *male–female*, *novice–expert* and *Japanese–non-Japanese* groups. Table 3.6 shows the mean and standard deviation values for each of these groups.

3.6.3 User needs and desires

We use the open OP-1, and OP-2 to grasp the potential needs and design desires of users towards robots. Relevant words used by participants to describe needs and desired features of robots in working environments are: *kindness*, *safety*, *intuitive*, *responsive*, *cuteness*, *convenient*, *fast*, *accurate* and *efficient*. On the other hand, participants consider that robot in their every-life environment must be *fun*, *kind*, *safe* and *cute*. Moreover, they consider that robots must be able to have effective and interesting communication skills, understand feelings and emotions, use clothes, and have convenient and interesting functionalities such as being able to cook and dance.

3.6.4 Hypothesis

With the collected data and the different groups created, we can draw some hypotheses. They will lead, in addition to the research question the analysis of the results.

H1: The cultural background (country of origin) has an influence on their perception of the robot.

H2: The gender of the participants has an influence on their perception of the robot.

H3: The knowledge about the robots has an influence on their perception of the robot.

Regarding the cultural background (H1), Jiang and Cheng pointed out better acceptance of robots by Chinese people [118]; and Bröhl et al. mentioned that Japanese people are more used to seeing robots in everyday life than Chinese or US people [119].

Concerning the knowledge about robots (H3), we can hypothesise that *expert* people have a more rational preconception about robots, on their real capability and not based on a fictional image of robots [120].

Table 3.3: Semantic analysis results (K-1), about the feelings the participant had regarding the robot. The *concepts* ($\mathcal{C}$) are regarding to the **satisfaction** (α) and **comfort** (β).

$\mathcal{C}$	Dimension		Semantic Evaluation (μ^σ)									
	Positive (1)	Negative (5)	Japanese	Non-J.	p	Male	Female	p	Novice	Expert	p	**Total**
α	Happy	Unhappy	$1.43^{0.68}$	$1.14^{0.38}$	0.136	$1.46^{0.71}$	$1.29^{0.56}$	0.346	$1.24^{0.60}$	$1.55^{0.67}$	0.109	$1.38^{0.64}$
α	Interested	Boring	$1.28^{0.55}$	$1.14^{0.38}$	0.447	$1.27^{0.53}$	$1.24^{0.54}$	0.844	$1.24^{0.52}$	$1.27^{0.55}$	0.836	$1.26^{0.52}$
α	Disappointed	Amused	$4.60^{0.63}$	$5.00^{0.00}$	**0.000**	$4.54^{0.65}$	$4.81^{0.51}$	0.116	$0.72^{0.61}$	$4.59^{0.59}$	0.467	$4.66^{0.59}$
β	Relaxed	Anxious	$2.10^{1.22}$	$1.86^{1.46}$	0.690	$1.88^{0.99}$	$2.29^{1.49}$	0.297	$2.28^{1.43}$	$1.82^{0.96}$	0.196	$2.06^{1.23}$
β	Safe	Danger	$1.25^{0.54}$	$1.71^{1.50}$	0.447	$1.23^{0.51}$	$1.43^{0.98}$	0.409	$1.28^{0.89}$	$1.36^{0.58}$	0.702	$1.32^{0.75}$
β	Confused	Clear	$4.20^{1.14}$	$3.71^{1.70}$	0.491	$4.23^{1.18}$	$4.00^{1.30}$	0.532	$4.20^{1.29}$	$4.05^{1.17}$	0.669	$4.13^{1.21}$

Table 3.4: Semantic analysis results (K-2). The *concepts* ($\mathcal{C}$) are regarding to the **behaviour** (γ), **interactions** (δ) and **appearance** (ε).

$\mathcal{C}$	Dimension		Semantic Evaluation (μ^σ)									
	Positive (1)	Negative (5)	Japanese	Non-J.	p	Male	Female	p	Novice	Expert	p	**Total**
γ	Smart	Stupid	$1.61^{0.72}$	$1.33^{0.52}$	0.290	$1.44^{0.58}$	$1.76^{0.83}$	0.176	$1.57^{0.66}$	$1.57^{0.75}$	0.977	$1.57^{0.69}$
γ	Simple	Complicated	$3.58^{1.00}$	$3.17^{1.72}$	0.590	$3.67^{1.11}$	$3.29^{1.10}$	0.284	$3.61^{1.12}$	$3.43^{1.12}$	0.597	$3.52^{1.10}$
γ	Dynamic	Static	$3.53^{1.20}$	$2.20^{1.10}$	**0.050**	$3.26^{1.10}$	$3.56^{1.50}$	0.488	$3.36^{1.33}$	$3.38^{1.20}$	0.964	$3.37^{1.24}$
γ	Responsive	Slow	$3.00^{1.23}$	$2.50^{0.55}$	0.16	$2.96^{1.26}$	$2.88^{1.05}$	0.820	$3.04^{1.26}$	$2.81^{1.08}$	0.511	$2.93^{1.16}$
δ	Lifelike	Artificial	$2.89^{1.29}$	$2.17^{1.17}$	0.205	$2.48^{1.19}$	$3.29^{1.31}$	**0.046**	$2.57^{1.27}$	$3.05^{1.28}$	0.218	$2.80^{1.27}$
δ	Emotional	Emotionless	$3.05^{1.11}$	$2.83^{1.33}$	0.714	$2.93^{1.11}$	$3.18^{1.19}$	0.489	$3.04^{1.15}$	$3.00^{1.14}$	0.900	$3.02^{1.12}$
δ	Useful	Useless	$1.89^{1.01}$	$1.50^{0.84}$	0.330	$1.81^{0.96}$	$1.88^{1.05}$	0.832	$1.65^{0.88}$	$2.05^{1.07}$	0.192	$1.84^{0.98}$
δ	Familiar	Unknown	$3.39^{1.35}$	$2.50^{0.84}$	**0.053**	$3.11^{1.45}$	$3.53^{1.07}$	0.278	$3.26^{1.39}$	$3.29^{1.27}$	0.951	$3.27^{1.30}$
ε	Desirable	Undesirable	$1.84^{0.89}$	$1.20^{0.45}$	**0.029**	$1.63^{0.84}$	$2.00^{0.89}$	0.189	$1.78^{1.04}$	$1.75^{0.64}$	0.901	$1.77^{0.86}$
ε	Cute	Ugly	$1.95^{1.09}$	$1.33^{0.52}$	**0.043**	$1.81^{0.96}$	$1.94^{1.20}$	0.716	$1.87^{0.97}$	$1.86^{1.15}$	0.969	$1.86^{1.04}$
ε	Modern	Old	$1.57^{0.83}$	$1.33^{0.52}$	0.374	$1.35^{0.69}$	$1.82^{0.88}$	0.070	$1.61^{0.78}$	$1.45^{0.83}$	0.523	$1.53^{0.79}$
ε	Attractive	Unattractive	$1.79^{1.04}$	$1.33^{0.52}$	0.116	$1.67^{1.00}$	$1.82^{1.01}$	0.619	$1.74^{1.01}$	$1.71^{1.01}$	0.935	$1.73^{0.99}$
ε	Like	Dislike	$1.68^{0.87}$	$1.00^{0.00}$	**0.000**	$1.44^{0.75}$	$1.82^{0.95}$	0.175	$1.65^{0.78}$	$1.52^{0.93}$	0.623	$1.59^{0.83}$

Table 3.5: Average (μ) and standard deviation (σ) values for each NARS type. Values close to one represent positive attitudes and values close to five represent negative attitudes. Because the NARS-S3 is using positive questions, it is the five's complement of the mean that is shown.

Type	μ	σ
Interaction (S1)	1.90	0.72
Social (S2)	2.60	0.96
Emotion (S3)	2.62	0.97

Table 3.6: Average (μ) and standard deviation (σ) of the NARS questionnaire according to each different group. The column p represents the p-value of Welch's t-test. Because the NARS-S3 is using positive questions, it is the five's complement of the mean that is shown.

	Interaction (S1)			Social (S2)			Emotion (S3)		
Groups	μ	σ	p	μ	σ	p	$\bar{\mu}$	σ	p
Japanese	1.90	0.64	0.890	2.47	0.97	0.084	2.73	1.01	0.130
Non Japanese	1.87	0.92		2.90	0.88		2.37	0.82	
Male	1.91	0.71	0.890	2.61	1.02	0.809	2.62	0.99	0.963
Female	1.88	0.78		2.55	0.66		2.61	0.86	
Novice	2.07	0.65	0.168	2.71	0.85	0.484	2.59	0.92	0.852
Expert	1.83	0.74		2.54	1.01		2.64	1.00	

3.7 Discussion

3.7.1 Regarding Q1 and Q2

In Table 3.3, the visitors expressed to have a highly positive experience by feeling **happy** ($\mu = 1.38; \sigma = 0.64$), **relaxed** ($\mu = 2.06; \sigma = 1.23$), and **interested** ($\mu = 1.26; \sigma = 052$) in the proposed HRI scenario. Moreover, rather than considering interaction with the NEXTAGE Open industrial robot a dangerous task, visitors mostly expressed feeling **safe** ($\mu = 1.32; \sigma = 0.75$). They also considered the proposed HRI scenario as **clear** ($\bar{\mu} = 0.87; \sigma = 1.21$[1]) or intuitive enough, with a value really close to 1, which was the main objectives of this project. We can sum up the feeling of the participants and reply to Q1 by analysing the different concepts, thus they were **satisfied[α]** (scores around 1.3) and felt **comfortable[β]** (scores between 1.3 and 2) after their interaction with the robot.

Results shown in Table 3.4 suggest that visitors perceived the robot as a **smart** ($\mu = 1.67; \sigma = 0.67$) and **complex** machine ($\mu = 3.52; \sigma = 1.10$), which is between a **lifelike** and an **artificial** entity ($\mu = 2.80; \sigma = 1.27$). Even though the robot was able to recognise emotions as well as make decisions regarding them, or the engagement, visitors had a neutral impression between **emotional - emotionless** ($\mu = 3.02; \sigma = 1.12$). One factor influencing this result may be the lack of whole-body expressive movements presented in the proposed application, which could be one possible improvement in the system to have a better response from the participants [19]. This lack of emotion could also be an explanation for the perception of the robot's movement by the participants. They considered them neither responsive nor slow as well as mostly static rather than dynamic, with a score of around 3 for both of them. It is confirmed by the analysis of their impressions regarding the **behaviour[γ]** and **interactions[δ]** of the robot; participants have difficulties interpreting them, and grade them neutrally with scores of around 3, aside from the intelligence and usefulness that are judged positively.

From Table 3.3, we can see that the perception of the robot is positive, the visitors see it as **cute** ($\mu = 1.86; \sigma = 1.04$), **desirable** ($\mu = 1.77; \sigma = 0.86$) or **attractive** ($\mu = 1.73; \sigma = 0.99$). During the experiment, the robot was dressed in a hat and an apron, which could also explain the positive reaction of the public. However, they were quite neutral regarding their **familiarity** ($\mu = 3.27; \sigma = 1.30$) with NEXTAGE, even if it is not famous to the general public, its design is still

[1] In order to keep coherence, the values in the text are presented to have the value close to 1 for the described adjective. When an adjective is close to the *5-side*, then the *five's complement* ($6 - \mu$) is used, and symbolised by a bar: $\bar{\mu}$.

close to what a person could expect to be a robot, and the participants judged its **apperance[ε]** positively (scores around 1.7).

Nevertheless, the analysis of the different subgroups can not allow us to answer the three hypotheses (H1-3). Indeed almost all the p-values of Welch's t-test do not permit to discard of the null hypothesis. This is clearer with the *novice-expert* group, where most of the p-values are close to 1. We can still point out some significant differences, with $p < 0.05$. We can observe some differences in the perception of the robot, especially regarding the country of origin. Thus, the Japanese participants described the presented robot as less **dynamic** ($\mu = 3.53; \sigma = 1.20$) than the foreign participants ($\mu = 2.20; \sigma = 1.10$). One explanation could be the habituation of robots in everyday life in Japan. Those robots, such as Pepper, try to be dynamic to catch the audience's attention. Yet, NEXTAGE is originally an industrial robot and then has more *rigid* movements. We can see with our results that foreigners ($\mu = 2.50; \sigma = 0.84$) seem more **familiar** with that kind of robot that Japanese participants ($\mu = 3.39; \sigma = 1.35$). However, both groups found the robot **cute** and **liked** it in a similar proportion. On the other hand, men tend to see more **life** ($\mu = 2.48; \sigma = 1.19$) in the robot than women ($\mu = 3.29; \sigma = 1.31$).

To answer the second research question (Q2), we can also use the NARS analysis of P1 and P2. As shown in Table 3.5, participants have a positive attitude towards interacting with industrial robots, a neutral and slightly positive attitude towards the social influence of robots, as well as a neutral attitude toward emotions in interaction with robots. Results from the Welch's t-test shown in Table 3.6 suggest that there is not a statistically significant effect between genders (p-value higher than 0.05), dismissing again H2. This result differs from those reported in [77], which suggested that women have more negative attitudes towards robots than men. However, our results agree with those recently reported in [78], which performed a systematic mapping of research articles exploring attitudes, trust, and acceptance in different contexts of social robotics. They found out that "the gender of the participants [is] not associated with their affective attitudes toward social robots". Similarly, the t-test applied to *novice* and *expert* groups indicates there are no statistically significant effects regarding robotic-related experiences (H3). A hypothesis could be that experts in robotics can be people working in marketing for robotic companies, or workers in factories where robots and humans work isolated. Therefore, it may also be the first time they actually share the same space with an industrial robot in an interactive scenario for most of them. This can explain why results from people with more experience in robotics presented similar values to those identified as novices (in many cases families and tourists). On the other hand, the novice group, even though they know less about robots, can still be interested in them, and that is why they visited IREX. The analysis of the countries is also non-significant and does not allow us to verify H1. Even if [119] pointed out some differences in the perceptions of robots by countries. Even if some analysis using NARS already pointed out that the country of origin has an influence regarding robot [121].

As explained in section 3.2, studies reporting results of attitudes, acceptance and trust towards robots change depending on the application domain, the type of exposure, and the design of the robot. Romeo and Lado found positive perceptions about robots during the pandemic in Spain for Generation Z (1997-2012) [122]. However, those results could be just a confirmation of the general tendency, regarding the acceptance of robots in Spain or by the young generation. Therefore, robotics systems such as those that we presented in this article shall become more relevant in the near future. This study mainly showed positive attitudes and impressions of visitors towards the use of robots. This suggests that the main objective of the project presented in this work was met, proving that people with no experience could interact with a robot without prior teaching, in an enjoyable scenario. However, these impressions can be biased by the fact that humans received a gift. How this factor and other complex factors (such as the cultural background) may or not result in positive attitudes, acceptance or trust and how they affect the robot's perception of different attributes such as its behaviour or appearance is out of the scope of this article and can be discussed in future iterations/studies of the proposed cognitive system.

3.7.2 Regarding Q3

Concerning the third research question (Q3, *What are the potential expectations of visitors towards robots in their working and everyday environment?*), the words and desired features used for replying to the OP-1 and OP-2 suggest that providing robots with pleasing design/aesthetics, intuitive communication skills, and enjoyable HRI activities can be relevant aspects to improving the user experience and desirability of robots in both working and every-life settings. This contrasts with the traditional utilitarian objectives of robotics, which focus on performance (i.e, efficiency and effectiveness) and, this agrees with recent studies in HCI and HRI showing the importance of hedonic factors (e.g., aesthetics, pleasure, and emotions) for the successful integration of novel technological [123, 124, 125, 56]. The identification of the needs and desires of possible users of robots is a relevant task for developing robots and applications with greater acceptance, social impact, and market penetration [126, 127].

The participants also focus a lot on safety, especially in work settings, 39% of answers to OP-1 mention it. For instance, one participant mentioned that "*a human should be able to turn it off*". This comment can be influenced by science fiction books or films. Liang and Lee pointed out that people with exposure to that kind of media have higher fear of robots and AI [120]. They also identified that only 30% of the US population has no fear of robots and AIs, and actually most of them make no difference between fear of robots and AIs. This fear of robots/AIs can explain the sake of safety, it was one of the conditions of their acceptance in factories from the 60s [128].

Even if we mention before that participants are not solely looking for efficiency, it is still a recurrent comment, especially in the case of working with robots. The robots are essential for the so-called *Industry 5.0* to improve the efficiency of the process [129]. The people are expecting them to do their task "*fastly*" and "*accurately*".

3.7.3 Limitations

Because of the nature of the experiment, without any selection of participants, we have some bias in the population. Then, besides the questionnaire P4, all of the others have an unbalanced distribution, with around 80% of males for P1 and P2, and around 60% for P3. Even if the 'I' of IREX stands for *international*, the number of non-Japanese is below 30%. Due to the lack of foreign people, we gather all the non-Japanese together, even if some disparities exist, even between close countries, such as France and Germany [121]. Being a heterogeneous group could explain the difficulty to interpret the results.

The difference in the size of the different subgroups can be an explanation for the non-statistical difference in the results, and why we cannot clearly answer the hypothesis. Since it was not the main goal of our experiment, it is not a problem, but can be taken into account for future works.

3.8 Conclusion and Future Work

Industrial robots are advanced machines able to generate engaging HRI scenarios that are difficult to reach with most social robots, such as those tasks requiring advanced manipulation of objects. However, most of them are evaluated and used in factories or research laboratories. In this work, we proposed the initial iteration of an advanced industrial robotic system able to deal with a noisy, crowded, and public environment (an international robotic exhibition). We successfully affronted the challenge of bringing this robot to the wild and made a step ahead toward developing systems helping in the understanding of those social and technical factors influencing the adoption of robots in everyday environments. The proposed system could perform during the whole event (four days) without any technical issues and could manage the interactions with hundreds of visitors. The system can recognise human faces, states, actions, and emotions from the user. Then, in an unscripted scenario, the robot could adapt its behaviours according to the user's ones. We demonstrated by using an industrial robot in this explorative study that they could be used with the general public, in the wild, without prior specific training. With the addition of the results from applied questionnaires that suggest that visitors considered our proposed scenario enjoyable, safe, and interesting, it suggests that the proposed work's main objective was met. Results collecting expectations of this preliminary study also coincide with previous studies showing the importance of non-functional elements, such as aesthetics, in HRI and HCI. Future studies will be focused on exploring how non-functional features, emotional expression, and different robot personalities, as well as the different errors presented in the interaction, can influence or result in an improvement in the attitudes, positive experiences, trust, and acceptance toward robots in everyday and public scenarios. In addition to checking with a wider population about the impact of gender, knowledge about robots and country of origin. Concerning the latter, it could be better to have a group for each nationality.

Summary

— We experimented with the industrial robot NEXTAGE OPEN in an uncontrolled environment (*in the wild*).

— The participants could interact with it to obtain the gift of their choice.

— The robot was using social cues (gaze or facial expression) to estimate the engagement of the user in the interaction.

— The participants enjoyed our scenario and felt comfortable and satisfied.

— **Achievement**: Siméon Capy, Liz Rincon, Enrique Coronado, Shohei Hagane, Seiji Yamaguchi, Victor Leve, Yuichiro Kawasumi, Yasutoshi Kudou, and Gentiane Venture. Expanding the frontiers of industrial robots beyond factories: Design and in the wild validation. *Machines*, 10(12), 2022. ISSN 2075-1702. doi: 10.3390/machines10121179. URL `https://www.mdpi.com/2075-1702/10/12/1179` [1].

Chapter 4

Design of a Non-Verbal Behaviour Robot: Yōkobo

We saw in the previous chapter the effect of an industrial robot with social behaviours on the general public. However, this robot has an anthropomorphic shape that implies expectations from the users, notably in the interaction; they can also feel uncomfortable. In addition, even if the robot ran for a long time, the interaction with each participant was short. In this second project, we will aim to push further the study of HRI by seeing the effect of a long time of interaction with an abstract-shaped robot; by keeping non-verbal behaviours only. For this study, we will build a robot from scratch, which is also an important component in the HRI field, in robotics in general. Indeed, creating its own platform allows the researchers to have a robot that fulfils all their needs; contrary to using a ready-to-use robot (like NEXTAGE OPEN), which will have some limitations and require some compromises. In our case, we aimed to answer specific questions that would not have been possible with a commercial robot.

This project is a collaboration with designers, regarding the design of the object itself, and the design of the interaction; with researchers/students from France, Japan and Costa Rica. The pooling of knowledge from different fields or cultural backgrounds was essential in this project and it is also important *to me*. Since the *social* in social robotics is a truly subjective topic, especially regarding culture. It was important to share information to design social interactions.

In this chapter, we will describe the conception of the robot created to address the objectives described in the introductory chapter: how to achieve social interactions without voice, and, here, we few degrees of freedom and a non-anthropomorphic shape. In addition, the robot is designed to be used for a long time, at home.

Source code
`https://github.com/GVLaboratory/yokobo`

The documentation of Yōkobo regarding hardware and software.

4.1 Conception

We propose a presence robot with a new HRI approach called Human–Robot–Human Interaction (HRHI), in which the robot plays the role of *middleman* between two persons. With this approach, our robot (Figure 4.1), was designed to be included in a smart home with the main goal of reinforcing the link between people (e.g., a couple) [2, 5]. With that in mind, the robot will place at the home entrance, and its objective is to welcome visitors or family members when they are back. For reinforcing the link between users, it will transmit *motion messages* from one user to another. Those *messages* are a mime of the person's movement (limited by the robot's movement capabilities). Moreover, besides this, the robot has also a non-robotic function: being a 'key bowl'. It has been named Yōkobo, a portmanteau word from the Japanese word *yōkoso* (welcome) and

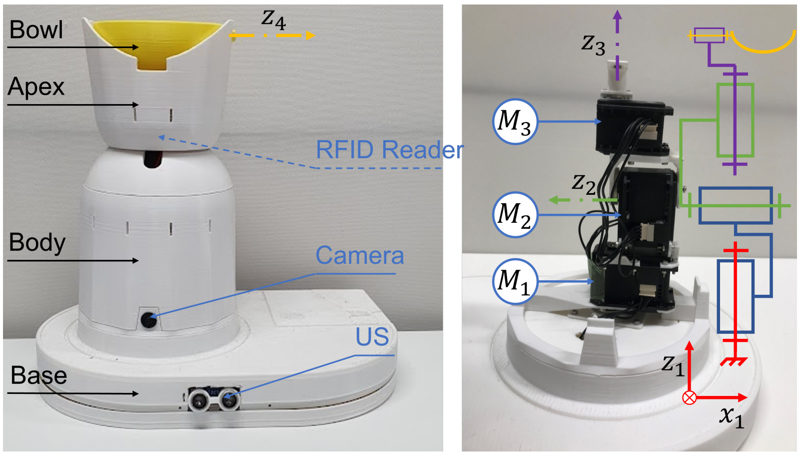


Figure 4.1: Parts and kinematic diagram of Yōkobo. Dimensions of the robot in centimetres: H = 33; L = 36; W = 24; ⌀ = 15.

the French pronunciation of the word *robot* (the *t* is silent). The purpose of Yōkobo is to be included in a smart home, then it can use the data from the available sensors like a hygrometer or thermometer. Thus, it can use those data to adapt its behaviours. Concerning the design, unlike most robots with social interactions for homes sold on the market, Yōkobo was designed with an abstract shape. Moreover, its only way of communication is through its movements and lights. Yōkobo's design approach follows the principles of *slow technology* [130]. Time is a key point in this approach, it plays a role in the adoption of the object and encourages the user to reflect on technology. By using this design method, Yōkobo is created in a way that allows its user to slow down, be surprised, and be reminded of the environment and their partner. It was proven by Odom et al. [130] that this approach can "create feelings of anticipation, open spaces for questioning the role of technology and even help to change routines".

some Robot Assistant (RA) also appeared. They can accomplish several tasks, such as checking emails and calendars or controlling smart home devices. The difference between VA and RA is the physical presence of the latter, thanks to their movement abilities. One can mention, Elliq which uses LEDs and body language to facilitate communication with users, with *expressive motions* [19] or facial expressions. Thus, Haru [131] can express its mood by moving its eyes or changing its shapes. Luria et al. [132] showed that RAs is more enjoyable than VAs and they are often preferred because of their social embodiment [133]. Moreover, RAs provide direct interaction between the user and the robot, and no other person is included in this interaction, even if the whole family can use it; contrary to HRHI robots. Each person has a private *canal* with the RA.

With HRHI, the robot is **designed** to be in the centre of a relationship between two or more people, being a catalyst in their relationship. Those robots are planned to create encounters between the users, and not just to serve as communication tools, such as social networking robots [42]. Furthermore, they can be distinguished from telecommunication robots, even if they are also using an HRHI relationship, but here the robot is more an extension of the human and serves as an enhanced phone.

Albeit Yōkobo uses motions (like an RA), it has, in addition, the specificity of being a *robject*: "embedding useful robotic technologies within everyday life objects" [134]. In our case, Yōkobo has been built around a key bowl. Thereby, even if it is not working, it can still be used, besides its robotic function.

4.2 Related Works

4.2.1 Creating Encounters and Links between Persons

Recent efforts have been made to enable the formation of social encounters, for example, the Abstract Machine by Anderson-Bashan et al. [135], which uses a greeting process usually only implemented for humanoids [136, 137] to new forms. One can also mention the *weak* robots [138], which, by displaying signs of weakness, encourage humans to interact with them. Similarly, the Ranger robot [139] catches the attention of children and helps them to tidy up their rooms. These examples have proven ways for enhancing HRI or helping to accomplish tasks. However, one important notion to consider when developing encounters is to deal with any possible feeling of ostracism that the user may have because of the perceived rejection when a robot is fixated on a task or with another user [140].

The telepresence robots are one example of the currently applied HRHI method, with, for example, Haru [141]. Here, the robotic body serves as a user's extension in real-time communication. Nonetheless, the robot does not act as a bridge for communication, rather it serves as a tool to provide an almost physical video chat experience.

Another case of robots implementing HRHI solutions are robots made to improve social interactions in autistic children [142]. They help to make the relationships between the children and the therapists better. Even if we can consider those robots using HRHI, it is more of a one-way relationship. The goal of these robots is to help children to communicate, and they normally do not help people to communicate with children. It is more of a H→R→H relationship than an HRHI; moreover, they are created for a specific target.

Rifinski et al. [143] conducted research about extended relationships with a robot. They investigated the effect of a robot considering a human–human–robot interaction (HHRI), where the robot is not central in the discussion, instead, it acts as a third-party observer. They experimented on robotic influence and its associated movement when two people talked.

The robot Fribo created by Jeong et al. [42] went deeper, being what they called a social networking robot (SN-Robot), with the aim to decrease the feeling of loneliness. The robot, using HRHI, transmits some information about the person's actions in their homes to their friends. It acts as a *middleman* to inform, with voice, the group of friends (three persons equipped with their robot) when one, for instance, opened the fridge; everything is shared anonymously. The participants became attached to the robot during the experiment, especially because of the robot's voice. Moreover, the robot helped to catalyse conversation in the group of friends. Nevertheless, Fribo is static and used only voice and screen to communicate, no movements are involved.

4.2.2 Form Factor

According to Campa [144], SRs are mostly humanoid or animal-like. Still, we could extend this classification, based on their form factors, into four categories [145], i.e., humanoids, zoomorphic, semi-abstract, and abstract ones.

The **humanoid** form factor facilitates social interactions since the user can rely on their own experience to interact with the robot. However, the human figure can create distress in the person [146]. That is why the **zoomorphic** robots can be more comfortable to people [147]. They have diverse form factors, ranging from a dog, such as AIBO, to a sea urchin [148]. Hence, people are driven to interact with the robot as they would with a pet [149]. Nonetheless, those two categories have psychological consequences, like creating strong attachments [150]. The next category is the **semi-abstract** robots, they do not look similar to any living creatures, but the users can imagine some human or animal shape and behaviour by pareidolia or anthropomorphism. They can also be inspired by either a real or imaginary animal. One can cite Lovot [40], which might share characteristics close to a penguin but with a face closer to an anime character. Finally, the **abstract** robots resemble nothing biological, even if while interacting with them, it is possible to deduce their behaviours, as shown in [135]. With (semi-)abstract form factors, the designers are less constrained, and the users have fewer expectations about the robot's behaviours [151]. However, their interpretations could differ because of the user's backgrounds [152].

4.2.3 Perception through Reduced Robotic Movement

Studies have been done by Duarte et al. [153] about the concept of Non-Verbal Behaviour (NVB) in order to transmit intentions with a humanoid robot. They demonstrated the possibility of understanding intention only with motion. The participants of the experiment could understand the robot's intent thanks to the head/arm movements or the gaze. This finding can be extended to other robotic embodiments and not only human-like body parts. For example, Broers et al. [154] showed it with a trash-can-like robot and Bevins et al. [155] with drones. We can also add the study of Lehmann et al. [156], which showed that non-anthropomorphic robot movements are essential to help users perceive engagement. Thus, with a 2-DoF robot, Anderson-Bashan et al. [135], showed that people could see social cues with simple motions.

Furthermore, Hoffman et al. [123] also demonstrated the importance of NVB for a robot, highlighting the value of the design process. They showed that a robot with positive emotions and expressive motions is more inclined to be accepted than a complex anthropomorphic robot with "unaffectionate motion". The majority of studied robots are non-anthropomorphic, for instance, Travis [157], a robotic speaker dock, and a listening companion. With only 1 DoF, this robot was designed with music-aided movement to create a relaxing environment. Similarly, Luria et al. [132] designed Vyo, a smart home assistant robot with five DoFs, and the motions were designed to be respectful or reassuring. By using such design notions while in the design phase, it makes it possible to build an agreeable atmosphere around the robot, accentuating the movement.

4.3 Contribution

We propose a novel approach in social robotics, where the robot is used as a bond between two persons inside their homes. The robot is using greeting motions and is also interacting with them at the home entrance. Yōkobo has a semi-abstract shape to make it self-effacing in the HRH relation, in order to help it to achieve its goal of being just a medium, and not creating a strong attachment at the partner's expense.

We also developed our specific protocol to test the previous aspects during two two-week experiments organised *in the wild*. The participants could interact freely with Yōkobo without any specific scripted scenario. We propose a set of tools to analyse the interactions, the robustness of the robot, and the users' perceptions.

4.4 Initial Step of the Design

Yōkobo was created with a three-stage design process. The first stage enables the greeting concept identification, with the brainstorming by the designers to find the concept of our behavioural object. The second focuses on Yōkobo's modelling, with shape and behaviour designs, by including the engineers in the loop as well. The last one is aimed at realising the functioning prototype, applying the Agile method [158].

4.4.1 Yōkobo's Shape Design

Yōkobo's shape was imagined alongside its movements. It is centred around two imaginaries: Japanese ceramic and robot-like depictions, organic and mechanic ones, mixing circular shapes (on the main pieces) and angular (on the edges). The ceramic imaginary provides Yōkobo with more precious, personal, and unique attributes. While the robotic one evokes the popular trend known as *mecha*, via its edges and vents.

Based on previous fieldwork [159], it was essential to design a discreet, yet useful, object. Yōkobo's physiognomy is an object that blends effortlessly in a home, specifically in the entrance. Its shape is composed of four parts: a **base**, a **body**, an **apex**, and a **bowl** (Figure 4.1). Its dimensions were chosen considering its ideal placement. The shape of Yōkobo, abstract, is enough to provoke social cues, and reduce the gap between expectations of people about the robot, and its real capacities [5].

The location of Yōkobo in the house has also been chosen for its symbolism. For example, in Japan, the *genkan* (玄関) represents the separation between outside and inside, where guest removes their shoes. It is the place where visitors are welcomed in the house even in Western cultures and can be linked to the butler symbolism. The placement had an impact on Yōkobo's shape, like the size, or physical placement in the room: hanging, on the wall, on the floor... For the entrance, it has been decided to have Yōkobo placed on a piece of furniture. The technologies

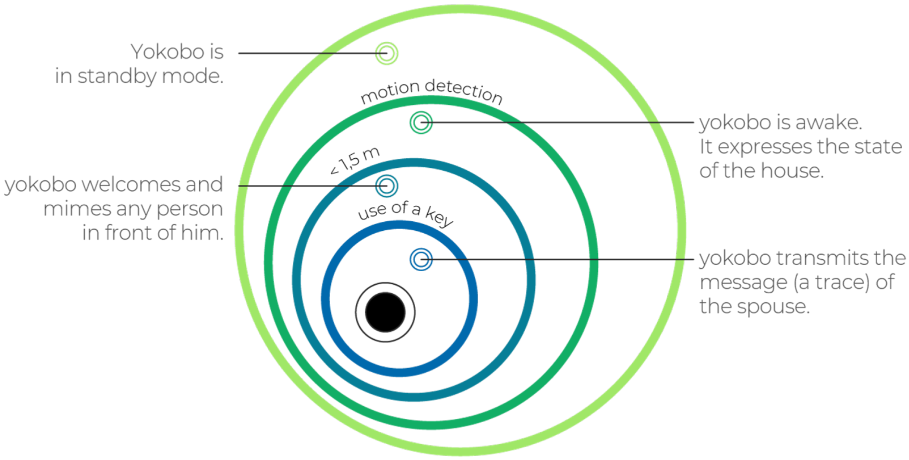


Figure 4.2: Representation of the proximity levels around Yōkobo (the black dot).

chosen (see sections 4.5 and 4.6) also had an impact on Yōkobo's design, for instance, the size of the **base** has been enlarged to host all the electronics.

4.4.2 Services and Associated Functions

Yōkobo's concept, as a *robject*, allows it to function as a standard key bowl, for objects such as keys or coins, or as a medium for house members to interact with each other. This non-robotic function is linked to its location, for example in France, it is common to have a bowl or plate called *vide-poches* (literally: empty one's pockets) when one enters one's home. Yōkobo primary communication means are movement and light. We drive the user to focus on Yōkobo's interaction by dismissing the vocal notions, instead of just listening. The main question we had to answer to build its functionality was, *how to condense the richness of human greeting into a robject?*.

Heenan et al. [137] showed that the greeting process was formed by a series of physical states, such as handshakes or nods, coupled via a transition sequence using proxemics. Guided by the above notions, the proposed solution focuses on proximity levels that serve as communication guidelines for how Yōkobo responds. It results in four levels of interaction that serve as the bases for designing the behaviours of the robot: **standby**, **state of the house**, **mimic**, and **record**, see Figure 4.2.

Standby is the robot status when nobody is in the entrance hall; Yōkobo is *active* and has periodical movements, guided by the Atmospheric Pressure (AP) and air quality (CO_2) data. **State of the house** starts once a person is near the robot, it continues the periodical movement and expresses the home state through movements. **Mimic** is triggered when someone is close to the robot. The robot then follows and reproduces human gestures. **Record** allows the user to leave a gesture message by saving the data from the mimicking process to later show to the other person. The lights, placed in the robot's **apex** change when a message is played or recorded.

As a summary, Yōkobo has four main functions:

1. Delivering messages between the members of the couple;
2. Being a key bowl;
3. Expressing the state of the house;
4. Greeting the visitors.

4.4.3 Behaviour Design

Previous research found that minimal robotic movement allows humans to perceive social cues and can cause feelings of surprise or engagement, according to the robot's attributes [135, 152]. Inspired by these results, the designers chose to shape Yōkobo's animations based on human reactions to the ambient conditions [5], see Figure 4.3. Therefore, the reactions are:

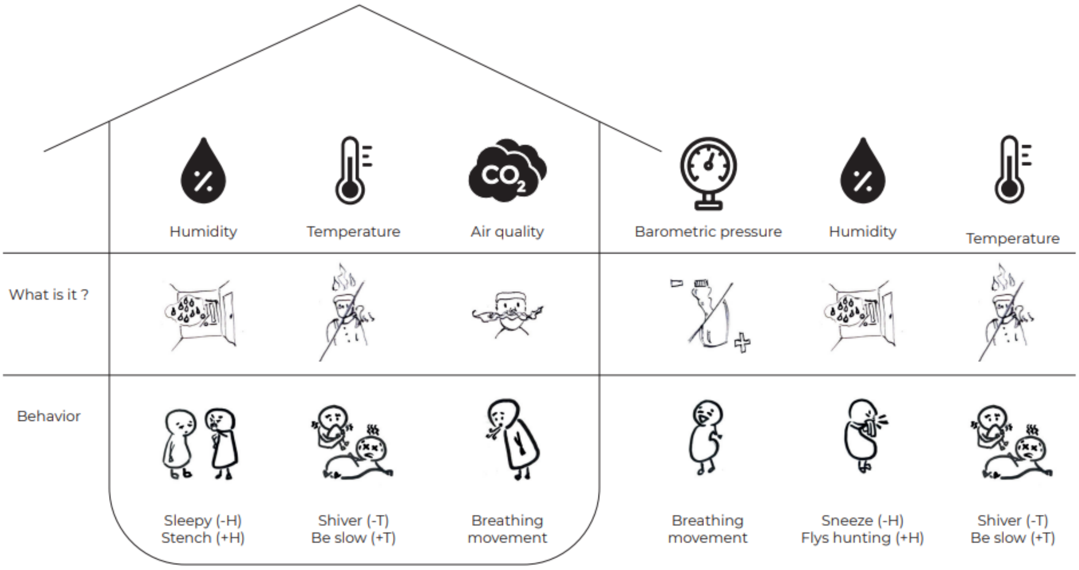


Figure 4.3: The anthropomorphic inspiration for the design of Yōkobo's behaviours according to the weather data, from [5]

Humidity inside: when the value is high, Yōkobo displays characteristics of human sleepiness, hinting at slowness and human body stretching.

Humidity outside: it displays a movement, combining the **body** and **apex**, to suggest a human sneeze.

Temperature: depending on the temperature variation, it shakes the **body** and **apex**, or has slow movements.

CO_2 and AP: these values are used through Yōkobo's periodical movements to increase or decrease the motor speed. They are used to imply the human breathing alteration by a higher concentration of CO_2 or AP increments.

Time: in addition to the weather data, the current time has an impact on Yōkobo. The light's colour changes depending on the hour with a shading that is orangish during the nigh and white around noon.

By using all those parameters together, and even if they follow defined rules, it adds some unpredictableness to Yōkobo's reaction. It makes then Yōkobo more *natural*, indeed, if the robot has predictable actions, it get closer to an automaton. Contrary to a human or animal that can have unpredictable reactions.

4.5 First implementation

After the designers wrote the functional specifications they need to achieve, we create the first architecture to fulfil their needs. We decide to use a Raspberry Pi (RPi, see Fig. 4.4) as the main unit of Yōkobo, because it is compact, affordable and with a big community that already developed many solutions for it. Moreover, we needed something to interact with low-level sensors. The RPi provides 40 connection pins, called GPIO (General Purpose Input/Output) to interact with electronic components, in addition to common IT interfaces: USB, HDMI or RJ45.

4.5.1 Selection of the devices

For each need a specific sensor or actuator has been chosen:

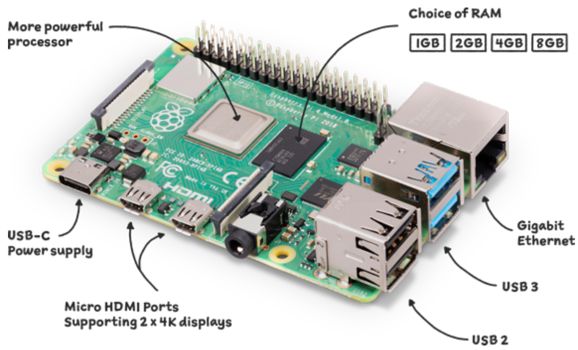


Figure 4.4: The Raspberry Pi 4B, the model used for Yōkobo has 8 GB of RAM

Movement
The main way of expressivity of Yōkobo is the movement, thus 3 motors have been selected to move it, see section 4.6 for details.

Light
In addition to the movement, Yōkobo can express itself with light. To have a wider range of expressivity, 2 RGB LEDs were chosen to add colours. They are placed on Yōkobo's apex, under the bowl, one for each side.

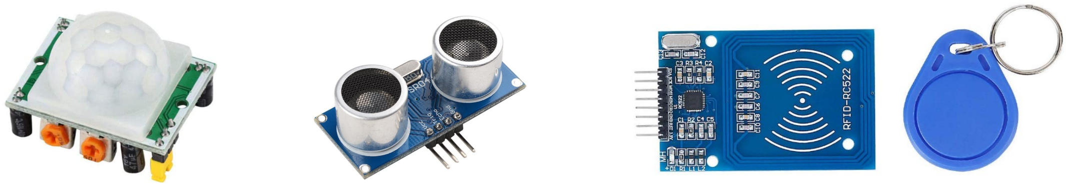
User detection
When a user is approaching Yōkobo, it should be detected in order to start the interaction process. To that aim, passive infrared sensors (PIRs) were chosen, see Fig. 4.5a. Those sensors detect the infrared emitted or reflected by persons, animals or objects. It is used as a motion detector.

Body movement tracking
For mimicking the user, the body posture and gesture have to be caught. The Raspberry Pi has a specific connector for embedded cameras and it is natively supported by the OS. Those cameras are small and connected with a ribbon cable, which facilitates the integration inside Yōkobo. Concerning the detection algorithm, it already exists several of them to detect body posture, such as OpenPose [160] or MediaPipe [161]. Because we are using an RPi, some limitation appears: first, the framework has to be available on ARM architecture and second, it has to be light enough to run on the RPi. We decided to use the solution provided by Intel called OpenVINO [162], it is based on OpenPose, and has been optimised to work on an RPi, using the Intel Neural Compute Stick 2 [163]. It is a CPU specialised to run AI algorithms, it is used to overcome the limitations of the RPi's processor.

Identification of user
Another function of Yōkobo is to act differently if it is a visitor or a member of the family. Hence, the user has to be detected precisely. To that aim, two technology has been selected: **Bluetooth** and **RFID**. The Bluetooth will be used to detect when the user approaches, from a further distance and wakes up Yōkobo; using the Bluetooth Low Energy (BLE) technology, with a key tag, it can be detected in a 10 m range. On the other hand, the Radio Frequency IDentification (RFID), is used when the user puts their keys inside the bowl, and the RFID works with a short distance, within a few centimetres. The RPi has already an embedded Bluetooth device, however for the RFID, it requires the use of an additional sensor: MFRC-522.

4.5.2 Printed Circuit Board

In order to have a robot compact, clean and easy to assemble and maintain, it has been decided to create a Printed Circuit Board (PCB), and not use breadboards. It will have a double function: power and connection.



(a) PIR **(b)** HCSR04 **(c)** MFRC-522

Figure 4.5: Sensors used for Yōkobo. The PIR will be finally abandoned to be replaced by the HCSR04.

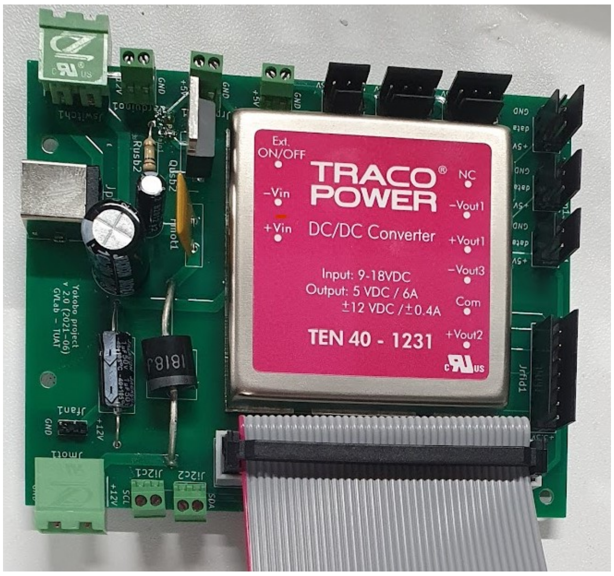


Figure 4.6: The v2 of the PCB with the TRACO to step down voltage from 12 to 5 V. The big ribbon cable on the bottom goes to the RPi. The other connectors are used for each sensor. The big connector on the top left is for the power switch, the one on the bottom for the motor, and, finally, the metallic connector in the middle is the power input.

First, it has to power everything according to the need of each component. The motors are powered in 12 V whereas RPi and other sensors[1] with 5 V. Then, the first task of the PCB is to downgrade the voltage from 12 to 5. The first design was using a buck converter module (LM2576-5.0WT), but the 5 V output was not working well so it was decided to switch to a TRACO, which is more reliable, see Fig. 4.6. Another point that has to be taken into account is power consumption. The RPi can consume at maximum 3 A (then 15 W), for the motor it depends on the load; then we decided to provide at maximum 5 A (then 60 W). Those have an impact on the choice of the TRACO and the power supply (that has to supply at minimum 75 W).

Secondly, to manage the connection, all the sensors are plugged into the RPi with ribbon cables and specific connectors. Thanks to them, there is no need to check the GPIO pins on the RPi, and it avoid also false contact.

The circuit and routing diagrams of the PCB are available in the Appendix C.2 and C.3. To undergo the aforementioned current, the PCB's copper layer weight is 4 oz/ft^2 (11.3 g·m^{-2}), hence the thickest track is 80 th (2 mm)[2].

4.5.3 Encountered problems

In addition to the problem of the buck converter, three others occurred. The most serious one was about the RPi. Indeed, during the first tests, we realised it was struggling with all the algorithms running. The Human Position Estimation Algorithm (HPEA) already uses 100% of one of the 4

[1] The RFID reader is powered by 3.3 V that are provided by the RPi.
[2] The PCB-making industry traditionally uses the imperial units

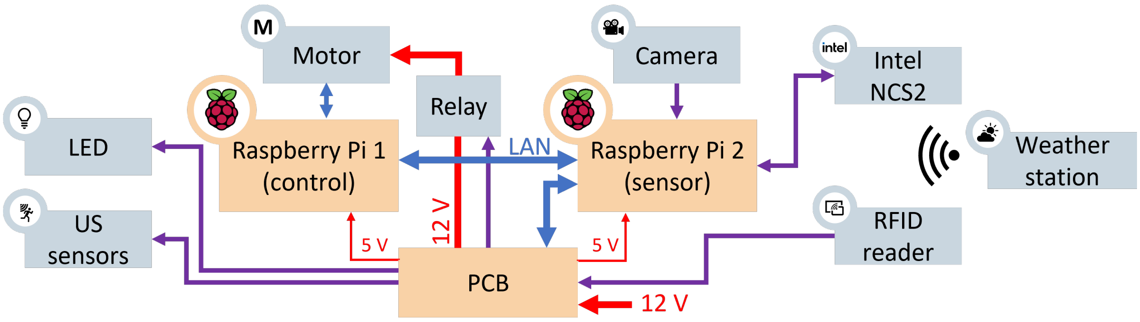


Figure 4.7: Hardware architecture. The red arrows symbolise the power link, the blue ones the data, and the purple ones both.

cores of the RPi's CPU. The main loop script (see section 4.7) uses another one. Then the RPi crashed often and the temperature of the CPU reached the limit of 80 °C. Then it was decided to use 2 RPis instead of one, to split the load, one for the sensors and the other one for the main algorithm and motors command.

The second problem concerns the LEDs. After testing, the rendering of the colours was really bad, too pinkish. Moreover, the luminosity was quite low. To avoid a long calibration process to find a good way to render the colours, it was decided to use an RGB LED strip instead. It has the advantage to have a better fidelity for the colour and a stronger luminosity. Finally, it had another level of expressivity because, now, Yōkobo has more than 2 LEDS, they are running all around the apex.

The third problem is related to user detection. The performance of the PIR was not good enough, especially because the designer asked to remove the plastic Fresnel lens around the sensor. It was not compatible with the design they wanted. However, removing it decreased a lot the performance of the sensor. We decided then to change the technology to Ultrasonic Sensor (US), it avoids having a protuberance on Yōkobo's front.

4.6 Final Hardware

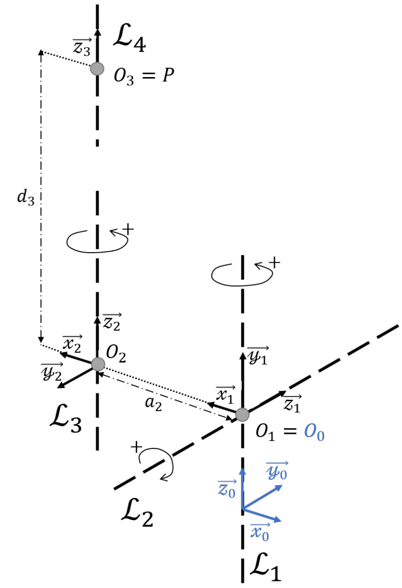
The final version of Yōkobo has four DoFs, driven by three Dynamixel motors (Figure 4.1). M_1 yaws the body around axis z_1, M_2 bows the apex, and M_3 yaws the apex around axis z_3. The three motors are located inside the body, and the bowl freely rolls by gravity. The Denavit-Hartenberg parameters of Yōkobo are indicated in Figure 4.8.

Figure 4.7 describes the hardware architecture. The central units are two Raspberry Pis (RPi) 4B, one for the control, and the other for the sensors. The latter is upgraded with a USB Intel Neural Compute Stick 2 (NCS) to run a human pose estimation algorithm (HPEA) on the OpenVINO toolkit [160]. The LED strip and sensor (RFID reader and two ultrasonic sensors (USs)) connection is made via the RPi2 GPIO pins through a two-layer self-made printed circuit board (PCB). The motors are connected directly to the RPi1 with USB and powered via the PCB. A 12 V-8 A power supply is used. The PCB dispatches that power to the motors (12 V-5 A) and the RPis (5 V-6 A) with a Traco. All the electronics are hidden inside Yōkobo; only the power cable is visible.

To detect when someone is nearing the robot, we use two USs around the base. To identify a user personally. We use the house keys with an RFID tag to detect whether the key is in the bowl. A wide-angle camera embedded in the body uses the RPis *camera port*. It is used for the mimicking procedure.

4.7 Software

By improving the work in [137], two finite state machines (FSMs), one named the main decision loop (MDL) and the other the motor control (MC), were constructed. The central software ar-



	a_i	α_i	d_i	θ_i
1	0	$\pi/2$	0	θ_1
2	a_2	$-\pi/2$	0	θ_2
3	0	0	d_3	θ_3

Figure 4.8 & Table 4.1: Denavit–Hartenberg parameters of Yōkobo. With $a_2 = 10$ mm that represents the eccentric between M_2 and M_3; and $d_3 = 100$ mm that represents the height of the bowl.

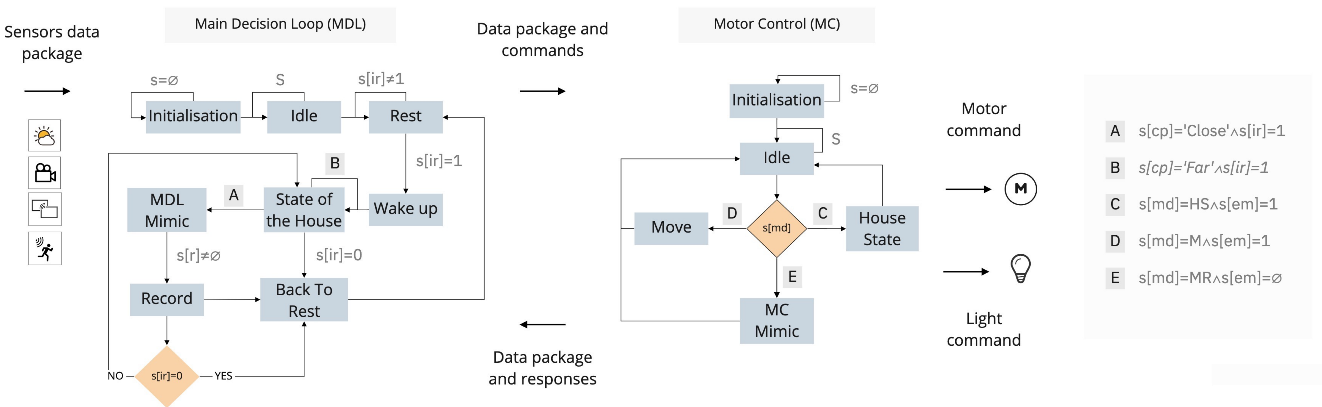


Figure 4.9: Finite state machines designed to command Yōkobo's services and behaviours.

chitecture can be seen in Figure 4.9. It encapsulates the data acquisition process, along with the central automata. Based on current sensor values, the first (MDL) decides how to switch between behaviours. The MC receives the commands from the MDL to send to the motors. All data transfers are done via the NEP framework we developed [110]. Since our approach follows a sequential loop of actions, there was no need for parallelisation in the main state machines, the only subsystem running in parallel to MDL and MC is the data acquisition, which constantly reads the sensors and publishes the data through NEP topics. These data are read whenever the FSM calls for it. The parallel process is in charge of acquiring the data and writing it to two specific buffers where the most up-to-date sensor values will reside. One buffer is uniquely given and accessed per FSM. Inside each FSM, to avoid reading the buffer while the parallel program is writing, a semaphore method is implemented as a precaution. As a proof of concept, all programs are coded in Python, except for the HPEA, which runs natively in C++.

Yōkobo's movement is either conducted by predefined motions or by a periodical movement guided through a sinusoidal wave linked to its **apex**. This generator, in the MC, has the formula $hm = \alpha \cdot sin(\omega \cdot t)$, where α is a value between 5 and 35 in proportion to the AP, ω is chosen from a range, 0.4–1.4, based on the CO_2 readings, t is the current time value.

The MDL state (Figure 4.9) switch is guided by variable s; it belongs to the set of weather data (wd), USs (ir), RFID sensor (r), and the person's distance to the robot (cp).

To summarise the MDL states:

1. **Initialisation**: it ensures the correct system boot-up and coordination among FSMs.
2. **Idle**: pause state while the first sensor's data package is gathered.
3. **Rest**: the robot **apex** follows the movement defined by hm.
4. **Wake-up**: Yōkobo displays an animation selected based on the current house temperature.
5. **State of the house**: Yōkobo's motion is guided by hm. It can also do pre-defined animations based on the two humidity sensor readings.
6. **Go back to rest**: an analogous process to the *Wake-Up* state, the difference being that the motion is selected based on the outdoor temperature.
7. **MDL mimic**: The MDL commands the MC to enact its mimic behaviour or play a recording of the previous trace. If r has no value, the robot mimics the user's motion. Otherwise, it moves to play a trace provided that the current RFID tag is not the same as the one that left the message. The system then plays the current trace before recording. The light colour (blue) signals the user the trace is playing.
8. **Record**: still mimicking, but now the person's movements are saved for 10 s and the light is set to green.

For the MC (Figure 4.9), s is outlined by the weather conditions (wd), HPEA body points (bp), motor commands (em), and the MDL command (md), with values MR for mimicking or record, M for playing an animation, or HS for house state.

The MC's first state opens the motor ports, pre-loads all animation files, and sets the initial positions for all joints. After the preparation finishes, the *Idle* state starts. Inside, the MC waits for the MDL commands to either go to:

1. The *Move* state, where the animation data points are sent to the motors.
2. The *House State*, the node in charge of using the humidity-guided planned trajectory, applying the hm generator.
3. Continue *Idle*.
4. Move to the *MC Mimic* subprocess.

Furthermore, if the automaton keeps the same state, the periodical movement is enabled. The motors are controlled through the Python Dynamixel SDK. The commands are given as a motor's position, then the Dynamixel motors reach the commanded position using their inner PID controllers.

Lastly, the **MC Mimic** subprocess recognises and reproduces the following human movements: sidesteps, bowing and twisting. These movements are also the ones used to record the trace message. The algorithm reproduces the person's motion by using the bp data and, in response, outputs a valid motor command for the robot. The method is outlined below:

Sidesteps: the system obtains the human waist centre XY coordinates from the image data and rotates the base motor so that the human is always seen in the field of (centre) view.

Bowing: the application checks the vertical motion of the user's shoulders, neck, and hips. If one of the first two is lowered below a given threshold, and the hip position has not changed, the second motor lowers the apex.

Twist: the system captures the human shoulder width (Ls) and the torso length (Lt), viewed from the front. It also continually calculates the shoulder width to torso ratio (Ls/Lt). This ratio decreases when the user turns to the side because Ls becomes smaller. When this situation is detected, the program determines there is a twist. It then rotates the top motor 90° and reproduces the gestures.

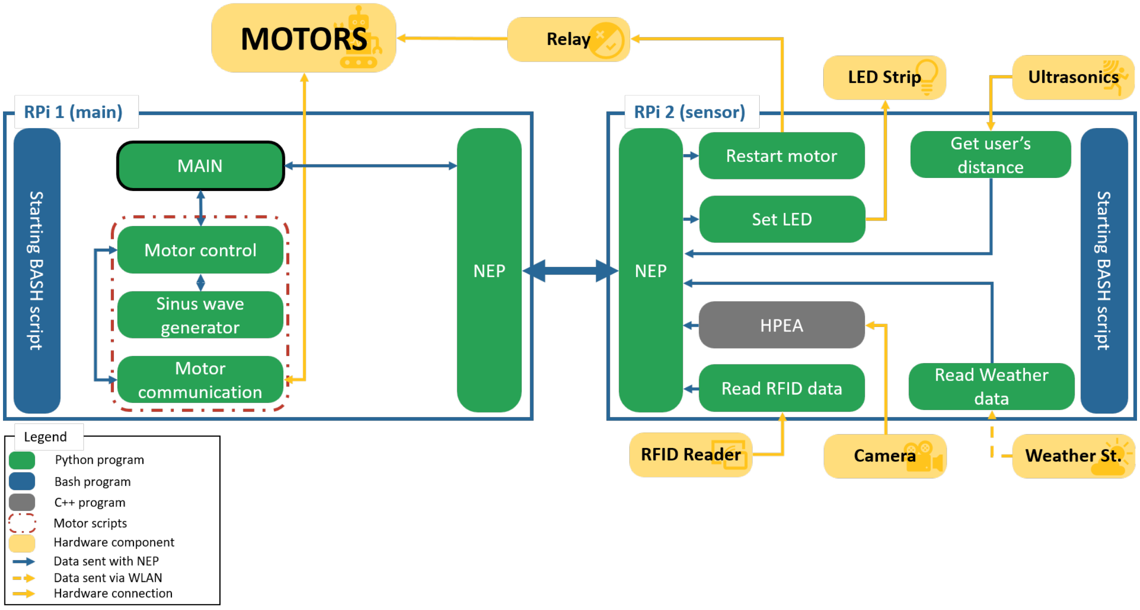


Figure 4.10: Architecture of the different algorithms of Yōkobo. The program `MAIN` corresponds to the MDL, and the program `Motor control` correspond to MC. The motor scripts (dashed in red) were split into 3 different programs for technical reasons (clarity of the code). For a better understanding of data flow, only the `NEP` programs that share the data between the 2 RPis are shown; but inside the RPi 1 there is another script that exchanges the data between each program.

4.7.1 Architecture

As mentioned in the hardware sections, the different software was split into two RPis, to divide the processing load. The two FSMs are running on the first RPi and command the motor according to the data share through NEP from the other RPi. On the second RPi, an algorithm for each sensor is running in the background and sends/receives constantly the data via NEP. We decided to split each block to have easier maintenance of the whole code. Moreover, the program for reading the weather data works only every 10 min due to the update frequency of the weather data on the server.

Finally, two bash scripts were made, for each RPi, to run all those programs when the RPis start. The bash scripts called the scripts in the right order, and allow Yōkobo to be ready-to-use when it is switched on, without requiring someone to run each script. In chapter 6 we will see that another program is called to allow the Bluetooth configuration. Figure 4.10 sums up the software architecture.

Summary

- Yōkobo is a presence robot for the home entrance, with the aim to create a link with HRHI between two members of a couple.
- It uses movement (with 3 motors) and light (RGB LED strip) to express itself.
- It mimics the human gesture thanks to a Human Position Estimation Algorithm, and transfer them as *kinematic messages*.
- The software uses two Finite state machines to adapt the behaviour of Yōkobo according to sensors' data.
- **Achievement:** Siméon Capy, Pablo Osorio, Shohei Hagane, Corentin Aznar, Dora Garcin, Enrique Coronado, Dominique Deuff, Ioana Ocnarescu, Isabelle Milleville, and Gentiane Venture. Yōkobo: A robot to strengthen links amongst users with non-verbal behaviours. *Machines*, 10(8), 2022. ISSN 2075-1702. doi: 10.3390/machines10080708. URL `https://www.mdpi.com/2075-1702/10/8/708` [2] (Editor's choice paper).
- Yōkobo also got the "Kawaii Design Award 2021" from the Kansei Engineering Society.

Chapter 5

Technical Experiment with Yōkobo

THE purpose of the experiment is to validate Yōkobo's technical stability. Although Yōkobo was designed to be used by young retired people, we can still validate the technical stability regarding the human greeting, and the smooth operation of the robot. The experiment was split in two, the first one as a pilot study to set all the parameters of Yōkobo and test the reaction of the user in order to adjust everything before doing the second one to deeply test the HRI. The two experiments are the same, except for the content of the questionnaires, see the next section 5.1. The experiments are called Experiment 1 (E1) and Experiment 2 (E2). As mentioned, E1 was designed to serve as a preliminary test, in order to gather initial stability characteristics, reception metrics, usability score, and first perceptual accounts. During this experiment, Yōkobo ran non-stop for the first week. However, several issues happened when the mimic/record state was enabled. The principal difficulty happened with the motors' commands from the Motor Control (MC) because of a faulty middleware that converted the motor data package to commands. Regardless of these errors, the situation was rapidly handled, and it was possible to continue the experiment for three days (12–16 h/d) and two reduced days (6–10 h/d).

This first experiment pointed out the necessary technical changes that Yōkobo needed to improve its stability and user reception. From the experimental tools, it was possible to observe that the problematic aspects involved the message-handling procedure and the sudden motor stoppage. These hampered the users from fully interacting with the robot.

To handle the aforementioned problems, we decided to improve the hardware architecture. This change provided more computational resources for the HPEA and the motor control. Moreover, the software part was upgraded to avoid false sensor triggers and improve the transition functions logic. These changes permitted a faster interaction among motors and data acquisition, which translated into optimised behaviours.

The experiment E2 started after we integrated the updates into Yōkobo. In contrast to E1, Yōkobo could run non-stop during the two weeks of E2; besides a few resets during the weeknights needed to prevent motor failure in case of wrong positioning. We can mention another problematic situation that happened on the last day of the experiment: a false switching between the states of the FSM occurred while an animation ran. It reduced the experimental day by two hours, but hopefully, this problem was resolved before most of the participants arrived at the office, and then it did not impact the results.

The next sections detail the experimental settings and the results. A comparative analysis is also presented, which shows the upgrade's impact on Yōkobo's overall qualities and capabilities.

5.1 Experiment Settings

The experiments were designed to be done in our laboratory. Figure 5.1 sums ups the mainlines of the experiment. The research questions that drove the experiment were:

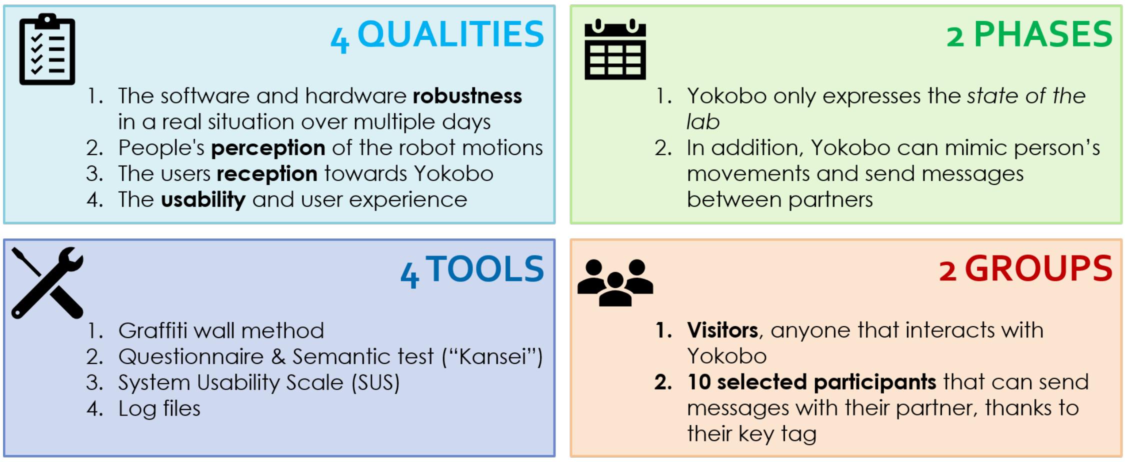


Figure 5.1: The mainlines of the experimental setup

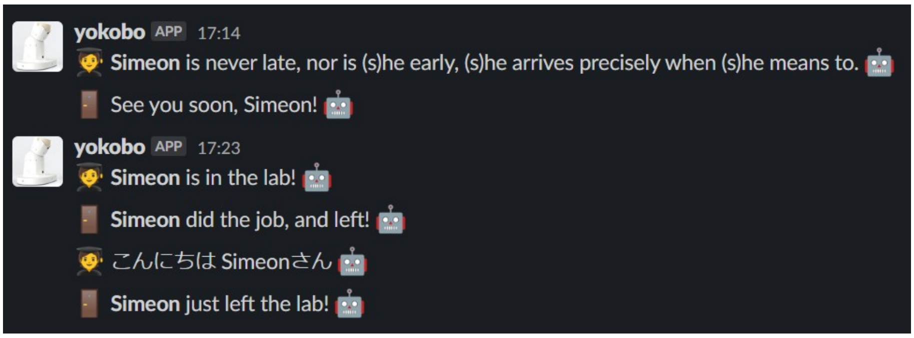


Figure 5.2: Example of clock-in/out message in the Slack channel. The messages are randomly chosen from a list. The name of the user is read from the RFID key tag.

Q1 How long is Yōkobo capable of running continuously?

Q2 Are the users capable of perceiving motion qualities from their partner by using non-verbal gestures and the proposed mimic algorithm?

5.1.1 Participants and schedule

Each two-week experiment was split into two phases of one week each. During the first week, the robot is accessible to anyone in the lab, and the participants can interact freely with it. The *mimic* functionality is available to the participants, and they can explore the interaction with Yōkobo and look at how it behaves according to its environment.

During the second week, some Selected Participant (SP), can interact deeper with Yōkobo, thus we made the messaging functionality available. They can exchange a specific *kinematic message* with their partner. Each SP are paired, and has a specific RFID key tag. Then Yōkobo knows who is approaching and to whom to send the message. To make Yōkobo a part of the laboratory, when a participant clock-in/out, a message is sent to the lab's Slack channel (see Fig. 5.2).

5.1.2 Tools

The main goal is to gather the participants feeling about the robot during the long-term experiment, to see if there is some evolution. But also to check if Yōkobo can work without problems during the two weeks. Then, several tools are used with a specific purpose.

a) Graffiti wall

The first tool is the *graffiti wall* method [164]. A whiteboard is located close to Yōkobo and different questions are asked to let the participants express their impressions freely while they are using the robot. Thus, they can give a spontaneous reaction. We replaced the questions every two days with a total of four questions, to gather information regarding the *feelings*, *experiences*, and *perceptions* the participants have about Yōkobo. These are asked in English, and the participants can answer by drawing or writing in any language. This tool was available to all participants, selected or not, and visitors of the lab as well. A sample of the participants' answers is available in Appendix D.2.

b) Questionnaires

The next tools, two and three, are part of four questionnaires, send at different moments. Concerning E1, a unique questionnaire (Q) is sent at the end of the experiment. For E2, one (QS) is sent just before the start of the experiment. The second one (QW1) at the end of the first week, and QW2 at the end of the experiment. The questionnaires Q and QW2 are sent only to the SP. The difference in the number of questionnaires between E1 and E2, relied solely on additional semantic analysis and the evolutive metrics we wanted to observe; since the first experiment was essentially a pilot study.

Figure 5.3 sums up the content of each questionnaire described in the next subsections. The *Likert* part for Q and QW2 has also a question about the attitude of Yōkobo after check-in. This question uses a free-speech box. There is also a free-speech box at the end of QW2 for any additional remarks.

α. **Semantic test** The second tool is a semantic test based on Kansei Engineering (q.v. section 2.2.1), combined with custom questionnaires (see β.). With bipolar adjectives, the semantic differential questionnaire allowed us to obtain how the participant will characterise the robot, given a number of selected concepts [165]. The 18 adjectives used for this semantic test were chosen by following the recommendations presented in [56, 124]. The word selection is decisive to guarantee that they are relevant to the design of the robot. These word sets were collected after consulting experts in the domain or reviewing state-of-the-art articles. The proposed semantic assessment uses the semantic scale to measure the concepts of interest, *Behaviour*, *Interaction*, and *Appearance*. The semantic scale uses two bipolar adjectives with a scale between 1 and 5; 1 means the answer is close to the *positive* adjective, and 5 is closer to the *negative* one.

β. **Other questions** The rest of the questionnaire is composed of several parts. First **demographic questions** about the *age*, *gender*, *nationality*, *knowledge about robots* and *knowledge about Yōkobo* (the latter only for QS). For E2, we also ask the students to enter their *student ID* in order to match each questionnaire (QS, QW1 and QW2) without asking for their names. It guarantees anonymity since the researchers do not know the link between IDs and student identity.

The second bunch of questions is using **Likert scale** to collect information about *behaviours*, *motions* and *experience*, and to evaluate the *Curiosity*, *Happiness*, *Fearfulness*, *Enthusiasm*, *Confusion* and *Friendliness* sentiments experienced by the users throughout the two weeks of experimental sessions. For QW2 additional questions are asked about the messages sent in Slack, if they found them *Useful*, *Funny*, *Annoying* or *Intrusive*.

Then comes from different blocks, one about the **kinematic messages** exchanged through Yōkobo. We ask the number of messages they send and received and if that is easy or hard. We also question the feeling of their partner by those messages, and if they can feel a connection. Those questions are asked with Likert scale questions and free-speech boxes.

Then we ask questions about the **design** of Yōkobo. Since it has an abstract shape, we are curious about the word the participants will spontaneously choose to describe it. We use for that a picture with markers (see Figure D.1) and let the participants use their own words.

We also ask participants, for QW1 and QW2, how many times they **interacted** with Yōkobo.

Finally, for Q, we have three questions about the **partner**. Because it is a pilot study, we want to test an additional point: *can the participants identify their partner?* Then they are paired by

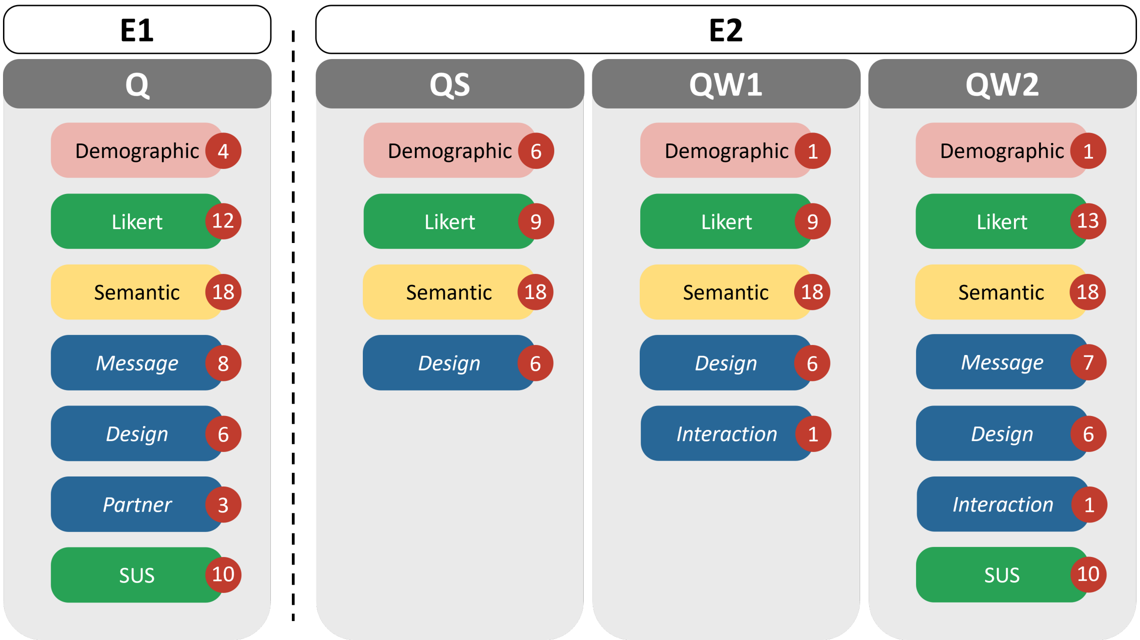


Figure 5.3: The composition of the questionnaires, for the first experiment (E1) and the second one (E2), with the questionnaires sent at the beginning (QS), midtime (QW1) and end of the experiments (Q and QW2). The red circle indicates the number of questions.

the researchers without knowing their "soulmate". Hence, the questionnaire of the first experiment is not anonymous.

The questionnaires are available in Tables D.1 and D.2 of the Appendix D.1, the different scales used for each question are indicated.

γ. **SUS** The third tool measures Yōkobo's usability with the SUS (refer to section 2.2.8). The questions and results are available in Table D.3 of the Appendix D.1.

c) Logs

The last tool is the interaction logs collected by the robot. These are used to verify the hardware and software robustness and provide metrics about the interaction time, sensor variations, and motor usage. The data are gathered each day and transferred to a cloud drive during the night, in order to be accessible. A second log file, not used for the experiment, is also sent with the temperature of the RPis' CPU, to monitor any failure.

5.2 Results

5.2.1 Robustness, Stability, and Usability

During E1, 6/10 participants pointed out that the recording process was complicated and not intuitive (from the graffiti wall, and discussions with the participants at the end of the experiment). They had difficulties interpreting the light signal and then understanding which state the robot was in. Moreover, there was a time delay between the person's motion and Yōkobo's replication, it was too long and did not seem to follow the user's movement concurrently. The log files pointed out that movements and messages were being recorded correctly; nevertheless, the motors' unexpected failure impacted the transmission of the message. The questionnaire (Q) for E1 pointed out that three participants could not send or receive any messages, and two couples managed to do both. It is confirmed by the log files since 7/10 participants tried to send at least one message, (Figure 5.4). On the other hand, regarding E2, all participants interacted with Yōkobo. The

Table 5.1: Semantic analysis results (second experiment). Lowest score—positive adjective; highest score—negative adjective.

Cpt	Dimension (Score)		Semantic Evaluation $\mu^{(\sigma)}$		
	Positive (1)	**Negative (5)**	**QS**	**QW1**	**QW2**
Behaviour	Dynamic	Static	$2.7^{(1.0)}$	$2.2^{(1.0)}$	$2.6^{(1.0)}$
	Smart	Stupid	$2.2^{(0.8)}$	$2.7^{(0.7)}$	$2.4^{(0.5)}$
	Simple	Complicated	$2.5^{(0.9)}$	$2.3^{(0.9)}$	$3.0^{(0.9)}$
	Responsive	Slow	$3.1^{(0.9)}$	$2.9^{(1.2)}$	$3.2^{(0.8)}$
Interaction	Lifelike	Artificial	$3.2^{(0.8)}$	$2.7^{(0.9)}$	$3.4^{(1.1)}$
	Emotional	Emotionless	$2.7^{(1.2)}$	$2.6^{(1.0)}$	$2.9^{(0.9)}$
	Familiar	Unknown	$2.1^{(0.8)}$	$2.6^{(1.1)}$	$2.1^{(0.9)}$
	Useful	Useless	$2.4^{(0.5)}$	$2.7^{(0.7)}$	$2.0^{(0.7)}$
Appearance	Desirable	Undesirable	$2.9^{(0.8)}$	$3.0^{(1.3)}$	$2.6^{(1.1)}$
	Cute	Ugly	$1.5^{(0.5)}$	$1.9^{(0.3)}$	$1.6^{(0.5)}$
	Modern	Old	$1.5^{(0.5)}$	$1.9^{(0.8)}$	$1.5^{(0.5)}$
	Attractive	Unattractive	$1.9^{(0.5)}$	$2.1^{(0.3)}$	$2.1^{(0.7)}$
	Like	Dislike	$1.8^{(0.6)}$	$2.1^{(0.9)}$	$1.7^{(0.7)}$

majority of participants (9/10) managed to send and receive the same amount of messages with a mean of two messages sent with a standard deviation of ±1, meaning one partner sent one, and the other member received it and sent it. The participant that suggested that s/he did not receive messages (from the questionnaire answer) may have confused the motion message by the usual robot movements, explaining why it was the only case that presented this situation. The logs also confirm this idea because both users' data are present (sending the messages), and the motion played was related to the actual body point data observed in their movements. We can also mention that 6/10 participants also indicated the movement as coming from Yōkobo rather than their partners, which can be explained by the animations that may have been triggered as part of the mimic and recording sub-process. Nonetheless, during E2, 40% of participants agreed that Yōkobo helped them feel their partners and 30% were neutral (from the questionnaire). Yōkobo was acting like a bridge between them, validating the possibility of transmitting the presence of someone through a robot. Although some of the results may, at first sight, hint toward a negative result, it was the opposite. If we compare the difference between the results of E1 and E2, 6/10 participants agreed that the recording process was more intuitive than in E1, validating the improvement done on the hardware/software performance and on the better light indication about Yōkobo's current state. The time lag between the robot's and user's movement was reduced, which translated to further exploration of the robot's capabilities and an improvement in the average interaction time: with an increase of 38.85 s (indicated with the red line in Figure 5.5). The latter can be observed as a partition by participants and daily cumulative interaction time in Figure 5.4, where it can be seen that there is an increment of the interaction time by participants in E2.

Yōkobo's SUS score after E1 was **66.11**. Concerning E2, Yōkobo obtained an average grade of **63.25**. Nevertheless, after splitting the result between the participants who were part of the first and second experiments and, the participant of only the second one, we obtain a score of **61.43**, and **67.5** respectively. The old participants have a lower rating on average by 2.5 points, with two edge cases with decrements of 10. Even if the grade seems low, the value is a passing grade. Actually, on average, a system rating is around 68; an example of a grade 61 product is the Apple Watch [166]. The difference between the two groups can be explained by the difference between the previous experience that the participants already had with the robot and the expectation of what the current version might have had as an added feature. Both grades pointed out that the interaction still needed improvement to be more reflective of the pairs. Overall, half of the participants (5/10) agreed the robot was easy to use, intuitive enough to learn by themselves, would like to continue using it, and was well integrated but needed further improvement in the recording function; 2/10 were neutral; 3/10 rated that the system needed more improvement. With this feedback, we could increase Yōkobo's *usability* by improving the recording behaviour and filtering which points are saved on the message (disregarding the animation).

5.2.2 Perception and Reception

The graffiti wall showed a positive perception of the greeting motions. Participants liked being welcomed by Yōkobo. Some even interpreted them (i.e., the greeting motions) as good manners, even if it was not programmed so. For instance, two participants thought the bowl moved down to interact with it more easily.

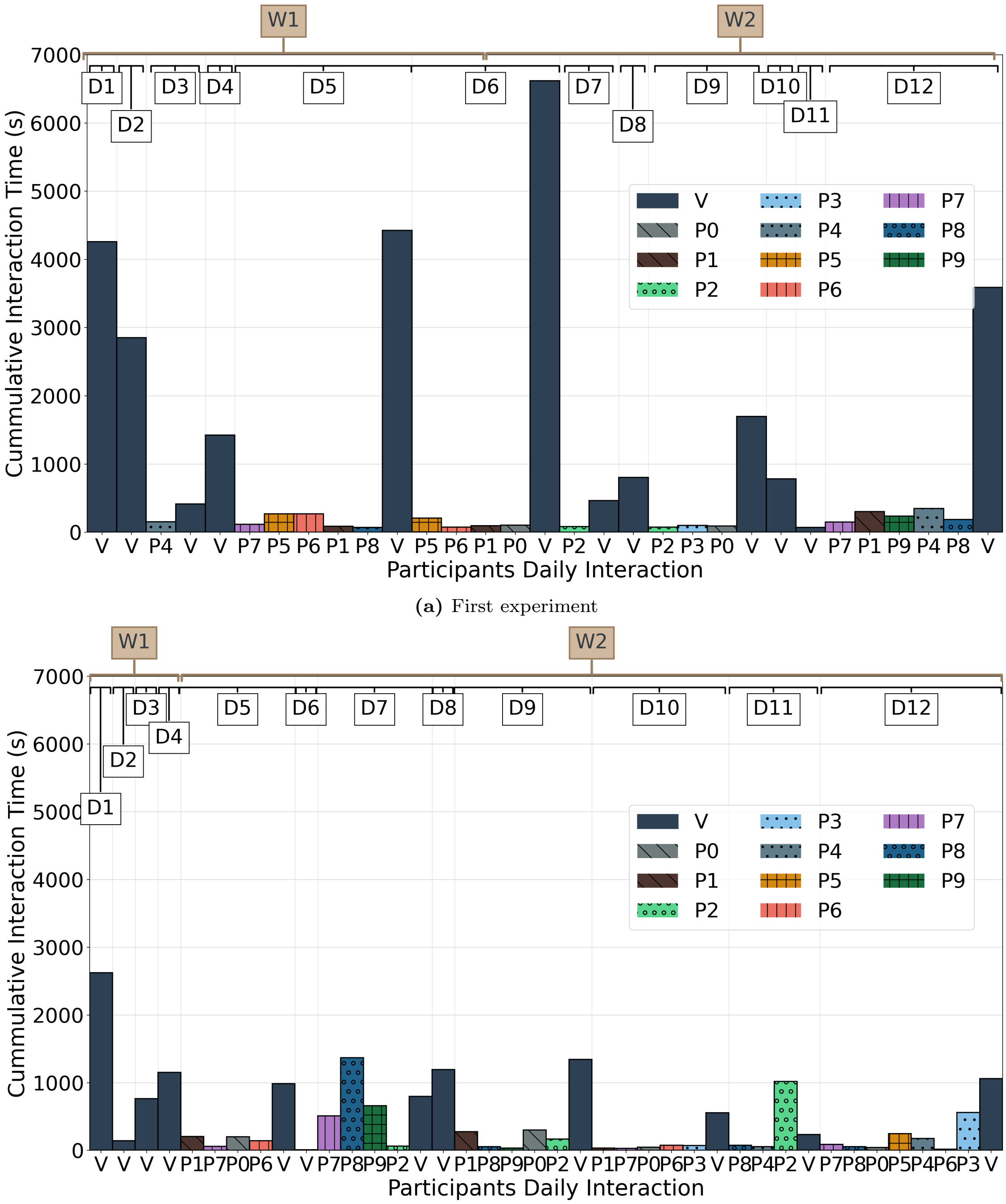
From the log files, the time users spent interacting with Yōkobo, from the *Wake-Up* to the start of the *Go Back to Rest* state (Figure 5.5) was analysed. The maximum interaction time spent for E1 was 826 s (13 min 46 s), the minimum was 16 s, with an average, over the two weeks, of 76 s. We can also mention that this time for the second week showed an average increase of 21.9 s; this corresponds to more *visitors* interacting with Yōkobo, as can be observed in Figure 5.4a. Concerning E2, the average was 113.24 s, with a maximum of 1307 s (21 min 47 s) and a minimum of 13.39 s. The minimum time corresponded to a person detected but not a complete interaction in both cases. The main difference that could be seen in Figure 5.5a between experiments is an increment in the average exchange and significantly fewer faulty sensor triggers. It shows how the upgrades manage to reduce this false interaction, which also explains the reduction in the sample quantity. The trend also demonstrates that extensive exchanges appeared during the week and were not condensed to single days. In contrast to E1, the average visitor interaction time for the second week did reduce; this can be partly attributed to the experimental days being in the middle of the holiday sessions and the end of the school semester. Still, there was a significant increment in the SP interaction when comparing Figures 5.4. However, the maximum time can be associated with the mimic procedure and the participants' curiosity; measured with the questionnaire, by the end of week two of both experiments, were high, with an average rating of 3.7/5.

Participants had longer interactions with Yōkobo during E2 with an average increment of 37.24 s, not only the new participants but also the ones that participated in E1. For example, participant 2 (P2) presented more extended and frequent interactions than before. In general, this trend and the increment of the average interaction time plus the improvement of the semantic analysis evaluations (*Useful* and *Familiar*) showed that the user wanted to use the robot continuously. However, in both experiments during the second week, the visitors, especially for lengthy interactions, may have had a perceptual bias due to their *desire* [167] and perceived *scarcity* [168] to use the robot to its full capabilities as the SP could. Nonetheless, both experiments provided enough arguments to show that participants wanted to further interact with the robot, providing a clear guide towards surpassing the novelty effect. However, more data are needed to make a conclusive argument. Due to the slow technology principles used for Yōkobo, the more time the users spend interacting with the robot, the better it is. We are looking for them to discover the functioning of the robot by themselves, we are not looking for efficiency (as it can be with usual robots or products).

Another interesting fact from the logs of both experiments is that several participants split into subgroups of two or three users. This formation introduced an increase in interaction time. Yōkobo may have been a conversation starter and a tool for users to find common ground. By creating these social scenarios, the participants further understood Yōkobo's capabilities on their own or in their subgroups, making the interaction with the robot clearer. This may explain the average 'decreasing' rating for the fearfulness sentiment.

A related observation is that the extended interactions happened when clocking out with the RFID tag, meaning between 16:00 and 18:00. We hypothesise that one possible explanation is because of the daily schedule situations. When users first tag in, they might be rushing to start their daily activities; however, when tagging out, they may have extra time to spend with Yōkobo and their colleagues.

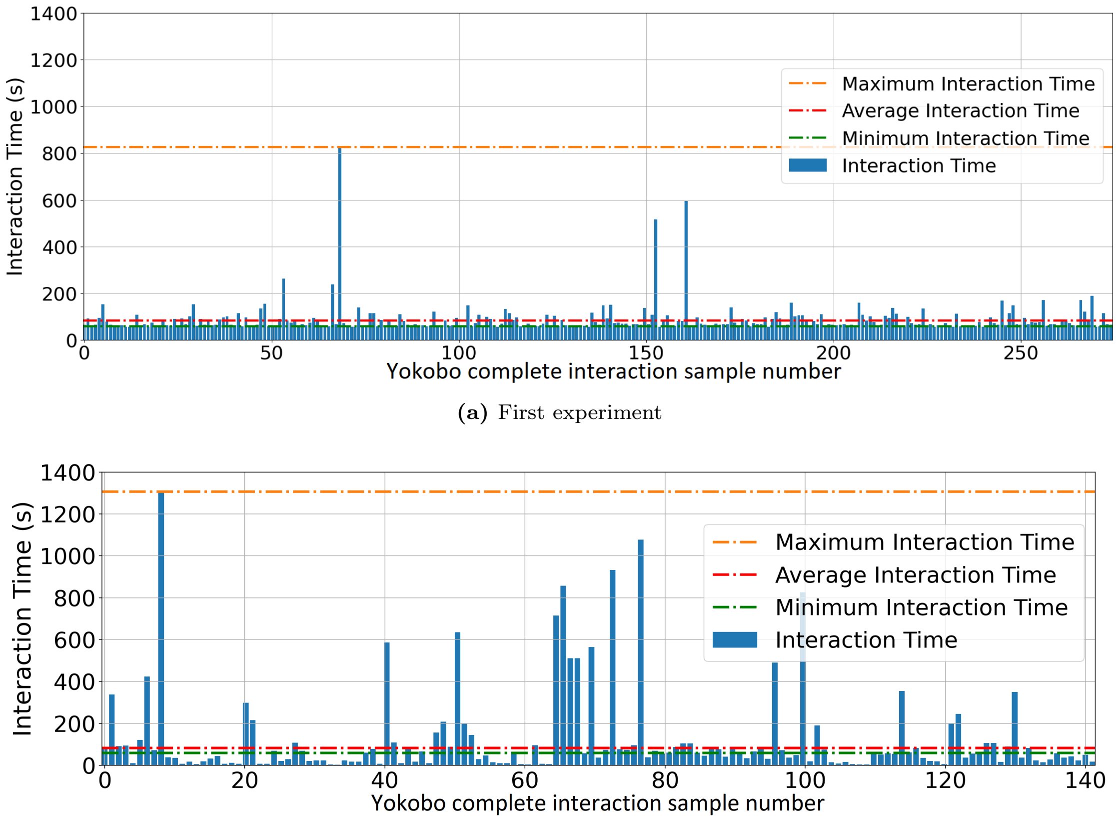
The collected data from the semantic evaluation are shown in Table 5.1. The *Concepts* (Cpt) that were analysed are presented, along with their respective *Dimensions*. Dimensions tend to the positive side (lower than 3) with a standard deviation around one. When comparing the different concepts, we found that the reception towards *Behaviour* was positive. A total of 6/10 participants agreed that the system was dynamic, smart, and straightforward, in line with the proposed design notions; 2/10 were neutral, and the other two participants thought it was static and unintelligent.



(a) First experiment

(b) Second experiment

Figure 5.4: Participants cumulative interaction time for both experiments (from T-L), **D** for days (from 1 to 12), **W1** and **W2** for weeks; each bar corresponds to a user. New participants for E2 are denoted as **P2** plus their respective tag numbers. The pattern for each participant represents the couple they belong to. Each bar corresponds to the cumulative interaction time that each SP and the V group had with Yōkobo per day. The V group is also considered since their interaction is valuable to differentiate between groups and how each one decides to interact with the robot. The graphs also show that, on average, each SP left at least one message per experiment. This result, in combination with the data acquired through the questionnaire, allows us to discern patterns associated with the robot's interactions, robustness, or personal perceptions, which (later on) is beneficial to discover the pain points associated with Yōkobo's interactions and qualifies the perceptual impacts it had on the user.



(a) First experiment

(b) Second experiment

Figure 5.5: Interaction time throughout both experiments. Each sample represents an interaction with Yōkobo that completes the states loop, i.e., *Wake-Up* trigger to the start of the *Go Back to Rest* state. This interaction time is measured in seconds and composes every interaction without differentiating between SPs or V. The values associated with the minimum, maximum, and average times for both experiments are also shown. The maximum interaction time spent for E1 is 826 s, the minimum is 16 s, and an average of 76 s. For E2 the average is 113.24 s, the maximum is 1307 s, and the minimum is 13.39 s.

Responsiveness is the dimension with the highest grade, indicating that the robot's system flow was not yet appropriate, and users expected a more fluid interaction. The latter was associated with the recording process difficulties described above and the transitional animations between the states, which may be perceived as too *artificial*. Concerning the dimensions of *Interaction*, even though on the positive side, they presented a negative tendency with the *lifelike* and *emotional* dimensions having a mean value close to 3. As seen in the semantic analysis (from the questionnaire) and the logs, the messages recorded and the behaviours were not fully understood. Participants were neutral regarding the emotional content of the animations, with an average value of 2.9 and a small standard deviation. On the other hand, *Appearance* followed the same positive trend as *Behaviour*, with desirability being the least positive dimension. The first two concepts (*Behaviour* and *Interaction*) could be improved by making the robot more intuitive (legible trajectories or special gestures) and adding other perceptual signals for the user, such as a sound or a clearer light display.

Interestingly, 7/10 users found Yōkobo more complicated over time, especially for E2. The *Confusion* sentiment by the end of E2 increased by 0.3. It would be necessary to improve the message cues and sequences for the user to understand it better and differentiate between Yōkobo states, movements, and animations. Similarly, the *Curiosity* increased once the recording functionality was enabled before decreasing at the end of E2. This may have appeared to hinder the overall results; however, it was, in fact, the opposite. Yōkobo is designed with the notion of slow technology, such that the users cannot fully understand the robot in a single interaction. Instead, it is a medium or a reminder of what is around them [130]. The users may be surprised or confused by the robot at different moments of their interactions and have continuous shifts in their perceptions.

Moreover, participants were asked about the word they would use for each part of Yōkobo. A total of 7/10 participants used anatomic words, such as head, body, and torso. Using these words to describe the robot, they identified it as a living creature, utilising pareidolia. The qualities participants used to describe Yōkobo were *smart*, *modern*, and *cute*.

According to the graffiti wall, participants seemed interested in Yōkobo, and the shape was trustworthy. The concept of *robject* was also evaluated. We regularly observed in the answers some mentions about the bowl. Some users wrote about what was inside: words such as "pick up" or "offering" were used.

Finally, regarding Yōkobo's office integration, 8/10 participants described its interaction as welcoming and 2/10 were neutral. After setting their RFID tags on the reader, they felt it displayed a *greeting attitude*, *energetic*, and *whimsical*. A total of 8/10 found the clock-in–out function useful, and preferred to do it this way; the other two participants were neutral about it.

Summary

— We performed two different experiments in our office to test Yōkobo's robustness, perception, reception, and usability.

— It proved its ability to work for extended periods of time.

— Participants felt welcomed and greeted by it. They perceived it as intelligent and a "welcoming partner".

— During our two-week experiment, 40% of the participants could feel his/her partner, proving the capability of this concept.

— **Achievement:** Siméon Capy, Pablo Osorio, Shohei Hagane, Corentin Aznar, Dora Garcin, Enrique Coronado, Dominique Deuff, Ioana Ocnarescu, Isabelle Milleville, and Gentiane Venture. Yōkobo: A robot to strengthen links amongst users with non-verbal behaviours. *Machines*, 10(8), 2022. ISSN 2075-1702. doi: 10.3390/machines10080708. URL `https://www.mdpi.com/2075-1702/10/8/708` [2].

Chapter 6

Integration of Yōkobo into Smart Homes

Since the goal of Yōkobo is to be a part of the smarthome to explore the use of a social robot in this environment, it is required to configure it. As a proof of concept, Yōkobo is using data from a unique IoT device: a weather station; but additional devices could be used. The interactions with and the perception of the robot by the users are studied in chapter 5 for the technical experiment. This chapter focuses on the integration of Yōkobo into the smart home, with the configuration process. Because Yōkobo will have to be settled in people's homes, the incorporation should be easy and non-intrusive.

With the democratisation of the Internet of Things (IoT) devices, the number of smart homes is increasing. Those homes use automation systems to achieve some function remotely, such as lighting, surveillance or heating. Thanks to a user interface, the users can monitor or control their home, even from outside. Those interfaces are diverse, like wall-mounted panels, vocal assistants or smartphones [169, 170]. A plethora of devices can be connected, such as light bulbs, weather stations (WS), thermostats, cameras... The connection between the devices can be made with wires or wireless, using ZigBee, Bluetooth or Wi-Fi protocol. The latter tends to become the most popular way to connect the devices in one's home, thanks to its ease of use, and nowadays, most homes are equipped with a Wi-Fi router [170]. Finally, the democratisation of vocal assistants, for instance, Amazon Alexa or Google Assistant, creates ready-to-use ecosystems for the users. Indeed, the assistants' manufacturers provide an API that allows third-party firms to incorporate their own devices into the assistant ecosystem. Then, the end-user has just one system for managing everything.

Source code
`https://github.com/GVLaboratory/yokobo-app-android`

The source code of the Android application

6.1 Related Works

6.1.1 Robots in Smart Home

Romeo et al. [171] described the merging of robotics and IoT as the combination of Communication, Perception, Control and Computation components, taking advantage of cloud computing. The robots should communicate with their environment to share data, generally by Bluetooth or Wi-Fi. Thanks to the latter connection, they can also externalise some computation tasks on the cloud. They have also to perceive their environment with various sensors, and they can also affect it with actuators, using control intelligence to make decisions based on all that data. The authors called it the Internet of Robotic Things (IoRT). While they mainly focused on (service) industrial applications, some can be extended to smart homes, such as surveillance, education, and healthcare. Several commercial products have been proposed for those purposes that we can call simply *domotic robots*.

One of the earliest IoRT for home is Nabaztag/Karotz [43] with the purpose, for example, to read users' emails or fetch weather forecast data on the Internet. However, it does not take advantage of the home's data. It only searches for information on the Internet. The second version adds one functionality with the RFID tags to read books. Even if the company, Violet, stopped the service, some community-driven projects have been created, such as Open-Karotz. The users can develop additional functionalities than the default ones and then incorporate the robot into their smart home.

Other everyday robots are for chores such as hoovering [172], mowing or cleaning, but they are not taking advantage of the smart home besides using voice control or allowing to start/stop the routines by an application.

Concerning the three domains cited by [171], one can mention Amazon Astro. A 2021-realised surveillance robot for one's home [173]. It can be controlled remotely or warn the user. It can also serve as a telecommunication robot thanks to its screen. Taking advantage of the Amazon's assistant, Alexa, the robot can achieve familiar tasks such as controlling the lights or the shutters.

Education is also a common purpose for a home's robot, for instance, to teach programming or robotics to children, like Clicbot [174] or Roybi [175]. Jibo [41] is a robot that can be used for education and care. Nonetheless, their integration is limited to being connected to the internet. They do not take advantage of the smart home.

Nevertheless, one of the most common applications of IoRT is for healthcare, for example, Elliq [17]. This robot was created to assist older adults' daily life with tasks such as news reading, weather forecasts, music, video calls, and more. It was built to be a robotic counterpart to the vocal assistant, yet it is not a *member* of the smart home, not being able to interact with other devices.

Do et al. went further into the IoRT with RiSH, a robot-integrated smart home [176]. Here it is more the robot that creates the smart home by adding many sensors/actuators for elderly care; e.g. it can be used for fall detection thanks to microphones. The user's (thanks to wearable devices), the home's or the robot's data is gathered in a *home gateway* and can be sent to a remote server for further processing. Then the robot can have access to more information rather than only relying on its own.

One can finally cite Michal Luria et al. work, in which they created a robot called Vyo to be the central interface of the smart home, and they compared it with a voice assistant, wall panel and a smartphone [177, 178]. They designed an arm-shaped robot using non-verbal behaviours and tangible icons to control the home devices and interact with the user. Although here the robot has the same purpose as existing devices such as a wall panel or a vocal assistant, the robot is *only* a robotic remote controller.

When we compare those robots, see Table 7.1, especially the commercial ones, we can see a typical pattern for communication. Firstly, as described in the introduction, Wi-Fi is a significant characteristic of domestic robots and smart devices. Secondly, we can extend this with Bluetooth; only the oldest robot on the table, Karotz, is not compatible.

We can even add vocal assistants like Amazon Echo, Google Home, or Apple Homepod that use both norms and communicate only remotely. Even some other devices like surveillance cameras, doorbells or thermostats that do not frequently move, like vocal assistants, do not have a wired connection; due to the inconvenience to set cables everywhere in the home. The previous highlights that users expect smart devices (and robots) to be connected wirelessly and managed remotely.

Finally, we can study the home's IoT data usage by the domotic robots. When the robot accesses them, it is often for sending a command or displaying some data like Vyo or Astro (like vocal assistants or most home automation systems [170]); they do not adapt their behaviours. When the robot uses them, it is primarily for care purposes. In the case of Jibo, it can connect to medical devices, but only to warn the user/doctor when something abnormal happens. RiSH [176] goes further by adding plenty of sensors in the home, such as microphones, cameras or infrared sensors. However, all of this ecosystem is created for the sole purpose of the robot.

Table 6.1: Comparison of some domotic robots (NM: not mentioned)

	Yōkobo	Karotz [43]	Vyo [177]	Astro [173]	ElliQ [17]	Jibo [41]	Roomba [172]
Year	2021	2006	2017	2021	2017	2015	2002
Commercial	No	Yes	No	Yes	Yes	Yes	Yes
Wi-Fi	Yes	Yes	NM	Yes	Yes	Yes	Yes
Bluetooth	Yes	No	NM	Yes	NM	Yes	Yes
Mobile	No	No	No	Yes	No	No	Yes
DoF (except moving)	3 + 1	2	5	3	2	3	
Camera/ human recognition	Yes/Yes	Yes (v2)/No	Yes/Yes	Yes/Yes	Yes (tablet)/ NM	Yes/Yes	Yes/No
Speaker/ voice recognition	No/No	Yes/Yes	Yes/No	Yes/Yes	Yes/Yes	Yes/Yes	Yes/No
Screen/ Expressive light	No/Yes	No/Yes	Yes/No	Yes/No	Yes (tablet)/ Yes	Yes/No	No/No
Application	Yes	No	No	Yes	Yes (tablet)	Yes	Yes
Shape	Abstract	Rabbit	Arm	Rover	Abstract	Abstract	Round
Use IoT data	Yes	No (possible with API)	Yes	Yes (with Alexa)	No	Medical devices	No
Purpose	Care Communication	Entertainment	Control	Surveillance Communication	Care	Care Education	Hoovering
Configuration	App	PC program	NM	Screen	Screen	QR-code	App

6.1.2 Configuration of IoRT

Comparing the different robots again shows that all recent robots are provided with an application, especially devices without a screen. Nowadays, people using smart devices are used to having an application to manage them. Indeed, nowadays smartphone is a standard tool that most people possess. For example, in France, as of 2021, the possession rate of a smartphone is at least 82.8% for people under 59 years old [179]. Bröhl et al. studied the device type usage preference amongst PCs, tablets and smartphones for daily activities by user generation group. Even if they did not look for configuration activities, their analysis shows the adaption of smartphones [180]. Most users from the young generation (19-36 y. o.) prefer to use a smartphone or a tablet. Even the next generation (37-52 y. o.) selects the smartphone, even if it is close to the PC (32.4% and 31%, respectively). One can see that a smartphone is now a standard tool for daily activities for people under 60.

Concerning the configuration, when the robots come with a touchscreen, it is generally used as an input device; otherwise, an application is used. Jibo uses an original way by generating a QR code, from the application, with the network credential info. The user just has to present it to the robot's camera. Roomba, the vocal assistants and the weather station Netatmo (see chapter 4) use the smartphone by establishing a direct connection with the robot/device (either by Bluetooth or Wi-Fi) to send the Wi-Fi information. This method is simple to use thanks to already known interfaces and reduces the number of operations. Moreover, it has the advantage of not requiring to physically access the device by connecting a screen or a cable, which is less user-friendly. Due to the latter, we decided to use the smartphone method to install Yōkobo into the home. According to Anderson et al., *familiarity* is one of the key elements of an engaging User Experience (UX) [181].

6.1.3 Usage of Single-board Computers

As described in section 4.5, the Raspberry Pi (RPi) is a credit-card-size single-board computer (SBC), available at a low price (less than 100 US$), with diverse types of wired and wireless connections.

Because the RPis are a popular tool for prototyping systems involving electronics, several projects for smart homes using them exist in the literature. It is an ordinary, cheap, easy-to-use, and fully customisable tool for IoT projects. Projects span from for controlling the temperature in a room [182], watering plants [183] to managing a whole home [184]. One can also find projects to manage the parking lots of a smart campus [185]. Nonetheless, all those projects share the same way to configure the RPi, beforehand with a Screen, Mouse, and Keyboard (SMK); this means that setting up the system in a new environment, requires some technical skills.

Indeed, it is not possible to configure a raw RPi in a user-friendly way, as described in section 6.1.2. To connect it to a network, one has to either connect an SMK, or a LAN cable or modify the configuration file directly on the micro-SD card [186]. Those methods are not possible due to the nature of Yōkobo since they are not user-friendly and contradict the current standard methods of configuring IoT devices.

We can also mention some other SBCs, such as Orange Pi, Beagle Board, or Nvidia Jetson. Those use similar methods as RPi for the configuration, especially the connection to the Internet. For example, Nvidia's board, besides using an SMK, can also be configured by connecting a PC by USB. But it requires also moving some jumpers on the board and setting up the SSH connection. Again, in the perspective to configure the device in several places, with people with basic IT skills, those methods are not recommended.

To our knowledge, no work to configure a system using an RPi *via* a smartphone application has ever been presented.

6.2 Contribution

We created a robot to serve as a link for people in their smart homes. Yōkobo was designed with compact size (H×L×W: 33×36×24 cm), therefore, we used an SBC for the central architecture. For our prototype, we selected the RPi because it is easy to use and customize, is versatile, and has a low cost. We described the creation process of Yōkobo in chapter 4. In order to incorporate

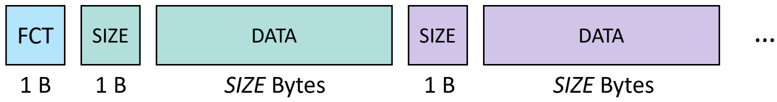


Figure 6.1: Structure of the message sent *via* Bluetooth

the robot into the users' smart home, we need a simple configuration. Indeed, the purpose of Yōkobo is to be used at home and not in a lab setting, and during the experiments, Yōkobo will be moved to several houses, and then configured several times. For those reasons, an easy way of configuration was required.

Hence, we propose a solution to configure the RPi thanks to a smartphone application connected by Bluetooth with a specific protocol. The Raspberry Pi configuration with a smartphone on Bluetooth is a first to the best of our knowledge, and because it is a popular tool for prototyping, we hope our work can be useful for the community. Moreover, even if the proposed method is done on an RPi, it is not strongly linked to it. Indeed, our method could be used with SBCs (or even *normal* computers) with a Linux architecture and Bluetooth capability. Finally, we propose also a user study with the System Usability Scale (SUS) [70] and behavioural analysis to validate our solution.

6.3 Design of the Configuration Process

The configuration process, which consists of connecting the RPi to the home network and linking the weather station, is composed of 2 parts : (1) an application for smartphones; (2) some scripts on the RPi. The configuration is done following the chart presented in Fig. 6.2. It is composed of several steps done alternately on each device, and this is done by exchanging messages *via* Bluetooth. Those messages are using our original protocol described in Fig. 6.1. First, a byte corresponding to the required function (`0` = close the connection, `1` = Wi-Fi configuration and `2` = weather station configuration), then comes the data. The first byte indicates the size (in byte) of the following data (hence, the max size of the data is 255 chars), and then comes the data per se. Next comes another bloc of data, with the size first and after the data, etc.

6.3.1 On the Raspberry Pi

The RPi configuration program is split into two parts. First, a master script in Python calls the other scripts and lights Yōkobo during the configuration (see Fig. 6.2). It also waits for the user to push the Bluetooth button to run the next part. This script is launched on the boot of the RPi thanks to *systemd*. Second, a script in Node.js that achieves the Bluetooth connection using the BLE protocol (Bluetooth Low Energy). That script's task is to receive the application's messages and replies; according to the function received, the script does a different task.

When the function is `1 Wi-Fi`, the script uses the credentials sent by the application and sets up the connection with the Network Manager [187]; since the default method using the WPA supplicant has difficulties setting up the connection through the command line without a reboot of the system. The local IP address of Yōkobo is sent back to the application as an acknowledgement.

When the function is `2 Weather Station`, we use the *pyatmo* library [188] (in Python) to deal with the *OAuth 2.0* protocol used by the weather station [189]. Thanks to this protocol, the *Resource Owner* (the user) can let the *Client* (third-party API, here, our app) access the data/services of the *Provider* (here, Netatmo). Hence, the configuration occurs in two phases:

1. First, an URL is generated to approve the connection. The user has to log in to give their consent.

2. Secondly, a redirect URL is created. It is used to generate the token. This token grants access to the weather station data without keeping the user's credentials.

As an acknowledgement of the linkage, the current temperature is sent back to the app, see Fig. 6.3d.

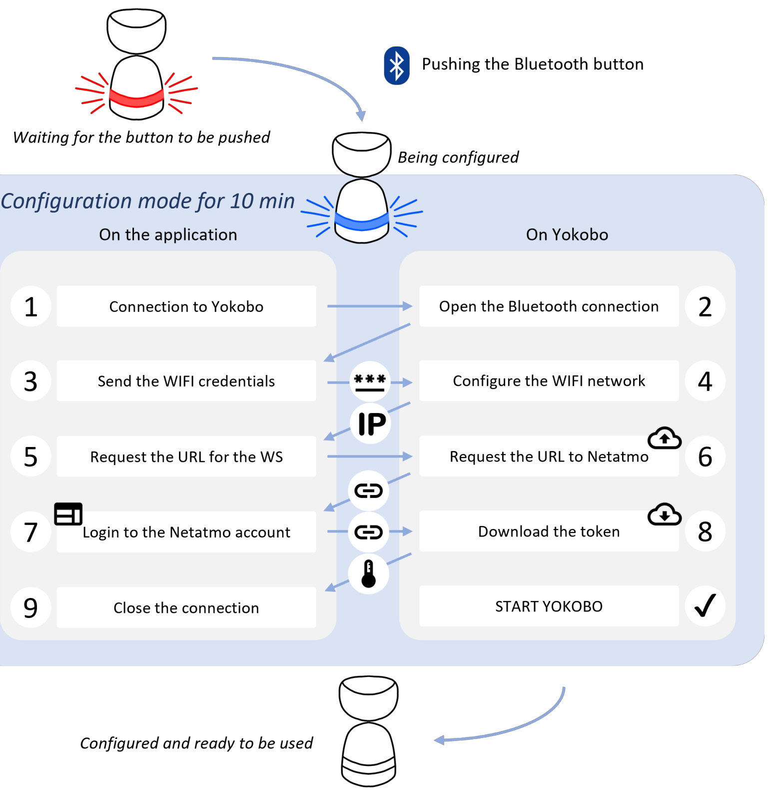


Figure 6.2: Process of the first configuration

Let us see an example of a message that the program on the RPi can receive:

```
1  PYTHON: Waiting for msg
2  value == □□SSID□PASSWORD□WPA-PSK□1□JP
3  Client connected to: tcp://127.0.0.1:10000
4  [
5      'updateWifi.py',
6      'SSID',
7      'PASSWORD',
8      'WPA-PSK',
9      '1',
10     'JP'
11 ]
12 EchoCharacteristic - onReadRequest: value = 192.168.11.31/1
```

The second line corresponds to the command received by the RPi, and the □ represents the replacement character because it is not displayable. Indeed, because of our protocol, the first character and the one before each block of data is just a number and then is not necessarily a visible character. For example, the character before `SSID` should be `4` (the length of the word "SSID"), and if we look at the ASCII[1] character set (Table 6.2) we can see it correspond to `EOT` (End of Transmission), that cannot be displayed. But it has no impact on the protocol since those bytes are not read as *characters* but *numbers*.

Then, in green is printed the reply of the program on the RPi. One can see it has correctly understood the function `1 Wi-Fi` by calling the appropriate file and then sending each parameter. The last line corresponds to the reply that will be sent to the application, with here, the IP address of Yōkobo and `1` which means it is connected to the Internet.

[1]American Standard Code for Information Interchange

Table 6.2: The ASCII character set, the dashed boxes correspond to unprintable characters

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0x	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
1x	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
2x	SP	!	"	#	$	%	&	'	(	)	*	+	,	-	.	/
3x	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4x	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5x	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6x	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7x	p	q	r	s	t	u	v	w	x	y	z	{	|	}	~	DEL

6.3.2 On the Smartphone

We decided to have the minimum step to make it simple for the user, taking as inspiration the *three-click rule*. This rule was first made for websites and stipulates that the user will get bored if the information is more than three clicks away from the home page. Even if Jiménez et al. showed that it is less true in the case of smartphones, they added that "*the longer it took to browse a website, the harder the experience* [is] *for the users*." [190]. Hence, we did the process in 3 steps with a few operations:

1. The **connection to Yōkobo with Bluetooth**. The user does it by pushing the Bluetooth button on the robot and selecting Yōkobo on the list of devices discovered by the app, see Fig. 6.3a.

2. The **Wi-Fi credential input**. The user selects their Wi-Fi name and enters the password, see Fig. 6.3b.

3. The **weather station linkage**. The user has to log in to the Netatmo website in an embedded web browser, see Fig. 6.3c.

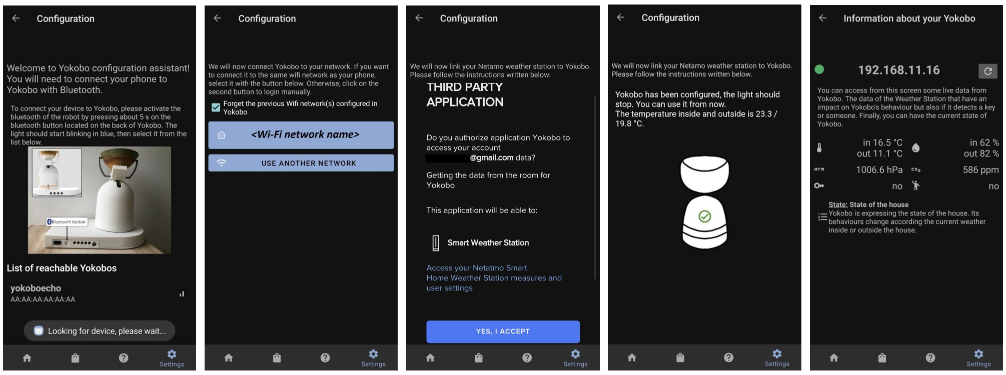
The app sets up the Bluetooth connection when the RPi is discoverable. Then it creates the messages with all the information (credentials or URL) and sends them to the RPi.

When the configuration is finished, the application can display some information to the user on a different screen (see Fig. 6.3e); and the developer can have more detailed ones. We use the NEP framework [110] for communication between the different scripts in Yōkobo, using the publisher-subscriber communication pattern. The application has to subscribe to appropriate topics to get the raw data, such as the temperature, humidity level, presence of a person (from ultrasonic sensors), body point (from the camera)... It allows the user to have an inner understanding of Yōkobo or to know some information at a distance, within the limit of the local area network.

The benefice for the developers is also significant since one can see the inner data of Yōkobo. It allows them to check if everything is working fine and to understand the situation in case of a problem. The data are displayed in a user-friendly format, and it avoids the need to use a computer to do simple checks. Especially during the first configuration of Yōkobo, to check if everything is working fine. This information is only accessible from the user's local network and is not available on the Internet for privacy reasons.

6.4 Design of the Application

The application is designed for Android and uses the language Kotlin which became in 2019 the new official language to develop Android applications. The application is using the MVC architecture, in order to have a better-built app by separating the data from the way to display them, see Figure 6.4. By doing so, the same *model* can be used by different controllers to access the same data. Moreover, if the representation of data in the memory has changed, the core (i.e. the *controller*) does not need to be modified, only the access to the data (the *model*). Finally, several *views* could be used, to adapt, for example, to a device with a big screen (e.g. a tablet).



(a) Bluetooth **(b)** Wi-Fi **(c)** WS **(d)** End **(e)** Data

Figure 6.3: Screens of the application, that displays: (a) the instructions to pair the smartphone to the robot. (b) the instructions to enter the Wi-Fi credentials. (c) the instructions to link the weather station. (d) the current temperature to confirm the success of the configuration. (e) some data from Yōkobo's sensors or the weather station.

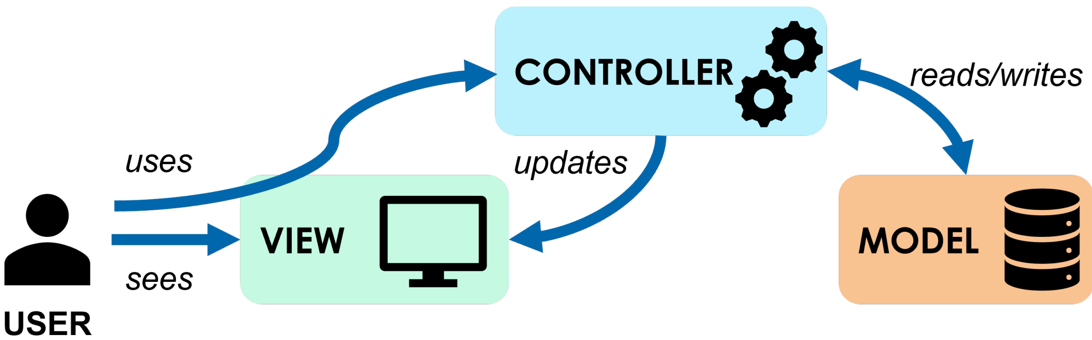


Figure 6.4: Schematic diagram of the MVC architecture

6.5 Experimental Tests, Results and Discussion

The research questions that guided the experiment are:

Q1 Is the developed configuration method easy to use, even by people with little knowledge about Yōkobo?

Q2 Can the users manage to configure the RPi solely with a smartphone?

One way to answer the first question is to use the SUS developed by Brooke J. (q.v. section 2.2.8). Hence, we conducted a user study in the lab to check the usability of the configuration process, see Fig. 6.5. The experimental protocol is as follows: first, we explained to the participants the goal of the experiment and asked them to follow the instruction of the application (in English) to execute a complete configuration of the robot, i.e. connecting the RPi to the Wi-Fi network and linking the weather station. Then, the researchers analysed the behaviour of the participants while they are using the application. When they finished the configuration, we asked them to fill out the SUS questionnaire, to which we added three custom questions. The custom questions' answers consist of a 5-value Likert scale. See the results in Table 6.3 and 6.4.

The experiment involved 11 participants (8/3 m/f; 27/11 μ/σ years old; 73% Japanese people). According to [191], it is possible to retrieve 80% of the usability problems with as few as five users, so with 11 participants, we expect a very good understanding of possible usability issues. To reply to Q1, we use the SUS tool, which gives a 0-100 rating about the system's usability thanks to 10 questions, with positive (odd questions) and negative (even ones) questions. The score of our system is **69.77**/100, which is above the SUS average score (68) [70], meaning our systems has a close to *good* usability [192]. We can also compare it to famous applications such as Facebook, which obtain a score of 71.39 [193], which is close to our app, considering that Facebook is a commercial product. We can also notice that the first question (*I think I would like to use this system frequently*) can also decrease the score in our case. Indeed, users are not expected to configure their devices frequently. We can also see that most participants are strongly familiar with smartphones (4.36/5). However, they are less familiar with smart devices (smartwatches, vocal assistants...) or robots (score closer to 3/5). Then, smartphones are a good tool to use something that is familiar to the user and makes it easier to apprehend.

We can see from Table 6.4 that all participants managed to configure Yōkobo without help in a time much shorter (less than 3 min) than the planned time (10 min). Answering Q2, it confirms that our solution is robust and fulfils its goal: configuring an RPi with a smartphone, without SMK. From the behavioural analysis, we can point out some other elements. First, most of the participants were surprised when the configuration ended, that it is already finished. To address this problem, we can use the recommendations of designers in UX, by indicating the list of the steps [194]. Nevertheless, it confirms the choice discussed in section 6.3.2 of having a few steps. The user did not get bored by the configuration process. Second, one participant took too much time to read the first screen and did not push the Bluetooth button on the robot in time, and reached the scanning timeout. S/he took some time to understand s/he had to click on the *retry* button to rerun a scan. A solution to this problem would be to use less text and more pictures, to make clearer and faster-to-read instructions. Third, since we rely on external services, the configuration timer has to be long enough, even if the participants usually finish quicker.

Lastly, we tested our solution mostly with young people, but since smartphones are a well-known tool as discussed in 6.1.2, and our target is young retired people (around 55-65 y.o.), we can expect similar results.

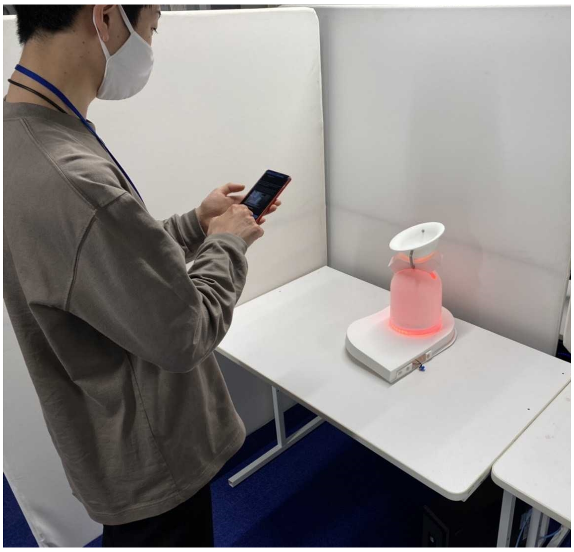


Figure 6.5: Experimental setup

Table 6.3: SUS results with 3 custom questions. Scores from *Strongly Disagree* (1) to *Strongly Agree* (5). The positive questions are in bold.

	Questions	μ	σ
SUS	**1. I think I would like to use this system frequently**	3.27	0.90
	2. I found the system unnecessarily complex	2.64	1.29
	3. I thought the system was easy to use	4.27	1.01
	4. I think I would need the support of a technical person to be able to use this system	2.27	1.35
	5. I found the various functions in this system were well integrated	3.36	0.81
	6. I thought there was too much inconsistency in this system	2.00	1.00
	7. I would imagine that most people would learn to use this system very quickly	4.64	0.50
	8. I found the system very cumbersome to use	1.82	0.60
	9. I felt very confident using the system	3.91	0.83
	10. I need to learn a lot of things before I could get going with this system	2.82	1.33
Custom	I am familiar to use smartphone	4.36	0.92
	I am familiar to use robots	3.18	1.40
	I am familiar to use smart devices	3.45	1.29

Table 6.4: Result of the behavioural analysis

Participants managing to finish	100%
Finishing time	Less than 180 s

Remarks:

- Most participants were surprised when they reached the end of the process.
- One participant took too much time to read the first screen and press the Bluetooth button, then the app searching time ends (10 s) and s/he had to push on the *retry* button to start scanning again. But this participant took more time to understand it.
- Some participants had to wait longer to get the acknowledgement to connect to the weather station.

Summary

- To integrate Yōkobo smoothly into the home, we developed a smartphone application to set up the RPi remotely without connecting any devices.
- We developed our original protocol to exchange messages between the robot and the phone.
- We proved the usability with a user study.
- **Achievement:** Siméon Capy, Enrique Coronado, Pablo Osorio, Shohei Hagane, Dominique Deuff, and Gentiane Venture. Integration of a presence robot in a smart home. *International Conference on Computer, Control and Robotics (ICCCR) 2023, Shanghai, China*, (accepted) [3].

Chapter 7

How Can We Have Long Term Interactions With Robots?

After creating Yōkobo and developed the first algorithm described in chapter 4, the goal will be to push further the interactivity of the robot with the user. Yōkobo has the aim to interact for a long time with people, on a daily basis, then a kind of special relationship has to be built. And one thing that makes the interactions between humans different from HRI is the unpredictability, the spontaneity. We can generally detect a pattern in the reaction of the robot, that follows a specific algorithm. It is the case of Yōkobo, where all its interactions are *hard coded*: *if the temperature is like that do that, otherwise do this.* The amount of different parameters is what makes the pattern difficult to discover for the user. If we want the users to enjoy Yōkobo, and get over the novelty effect, they have to be *curious* about it: *how will it react today?*

In this chapter, we will try to add some unpredictability to Yōkobo's behaviour by applying the principle of **cognitive robotics**. Yōkobo will learn from its interactions with the user to know what is the *best* reaction to achieve to have a positive relationship with his master. It will first create a representation of the world that will help him to predict what could be the future, and then find the best action to achieve, and learn from it.

Source code
`https://github.com/GVLaboratory/bewoda`

The source code of the algorithm presented in section 7.2.

7.1 Literature Review

7.1.1 Creating a Representation of the World

If we have agents that are able to understand the world around them, ie their environment, they could adapt their behaviours and then evolve naturally [195]. The idea of reasoning under uncertainty and finding casual relationships takes this notion even further. The goal is for non-biological agents to develop their own way to represent the world. One can cite the work of Ellis et al. [196] show that those systems are not static but can adapt, even to symbolic tasks. With a Bayesian approach, the agent used the concepts learnt before to create relationships and then uses them to extrapolate new generalisations that the model never saw before. Thus, it can generate new symbolic mappings that are interpretable and transferrable to new tasks while still growing scalability and flexibility with experience. The two-step cycle of the wake and sleep program learning of Ellis et al. allows the agent to create new hypotheses, test its program and then create a new one, similar to humans that are consolidating their memories and daily experiences while sleeping. Similarly, de Avila Belbute-Peres et al. [197] introduce a differentiable physics engine that can be integrated into any learning model to determine the laws ruling the world surrounding the agent.

Through a data-driven and expertise model, they can represent in a reduced space the physics foundation for a variety of tasks in a robot's task space. Thereby, thanks to this model, the robot can perform its tasks efficiently without needing previous expert knowledge. Even if those two examples are different in their objectives, they show the utility of environment understanding and then using it to achieve the agent tasks.

7.1.2 Creating Adaptable Actions

Even if the agent has created its own reasoning about the world, it would still need to have adaptability in its own actions or expressions. One can cite Reinforcement Learning (RL) which is a common tool from the last decades, that has the capability to adapt to task-specific approaches, either from simulation or the real world. Now, with a dynamic setting, robots can learn to succeed in various tasks simultaneously by themselves, Kulkarni et al. [198] and Silver et al. [199] provide a set of tools to exploit RL algorithms. In [198], they use a Deep Q-Learning algorithm, with different temporal scales, to solve tasks posed by the environment; the exploration and adaptation over time are set as intrinsic behaviours. While in [199] and [200], by using operator knowledge, sequential representation of tasks in motion planning and the symbolic links of the task, their algorithm can infer the correct transition clauses over time that a robot would need to succeed in different tasks. Furthermore, this algorithm can infer the correct action for any given environment. By mixing these approaches with the casual understanding described before, the agent is robust enough to act in its world and develop a proper plan of action, evolve over time and create links about what effect its actions have on its world.

7.1.3 Intention estimation

Yōkobo will have to estimate the next action of the user in order to find the best action to achieve like a human adapts their own behaviours to what s/he expects from their listener. The intention estimation is a domain that is important in autonomous driving regarding pedestrians' and cyclists' (P&C) avoidance. Indeed the autonomous car has to anticipate the actions of P&Cs if it wants to stop in time. Then we can see how it is achieved there, and adapt the algorithm used for Yōkobo.

The behaviour of P&Cs is harder to predict, as they have the tendency to less follow the rules, which forces AVs to estimate the action of P&Cs, like a human driver, will do [201, 202, 203]. Another point is the difference between each country regarding traffic code and the culture-related behaviour of each person. Then, even if some international treaties exist to standardise the rules, e. g. the *Vienna Convention on Road Signs and Signals*, not all countries ratified it, notably the USA or Japan. Apart from the driving side, signs and crossroad design can vary; in addition, the way the law punishes some misbehaviour of P&Cs could have an impact on them. For example, Hell et al. showed that the presence of a Pedestrian Flashing Green light (PFG) has an impact on the behaviour of Japanese pedestrian speed, contrary to their German counterpart [204].

Then, we achieve a systematic review of the current literature about the intention estimation for pedestrians and cyclists for autonomous driving.

a) Methodology

We followed the guidelines of systematic reviews described in [205, 206, 207].

Research question The main focus of this article is to analyse the algorithms and sensors used in recent research to perform P&C intention estimation, for the purpose of autonomous driving.

Search process The main concepts used to compose the search string are *intention estimation*, *movement prediction*, *pedestrian* and *cyclist*. Then, we define the search string as *('Pedestrian' OR 'Cyclist') AND ('intention estimation' OR 'movement prediction')*. We use this string to search potential articles in relevant digital databases (IEEE Xplore, Science Direct, ACM digital library, Springer Link, MDPI and Web of Science). The systematic search performed includes publications from the last 10 years, until February 2022. The number of papers after the process is 197.

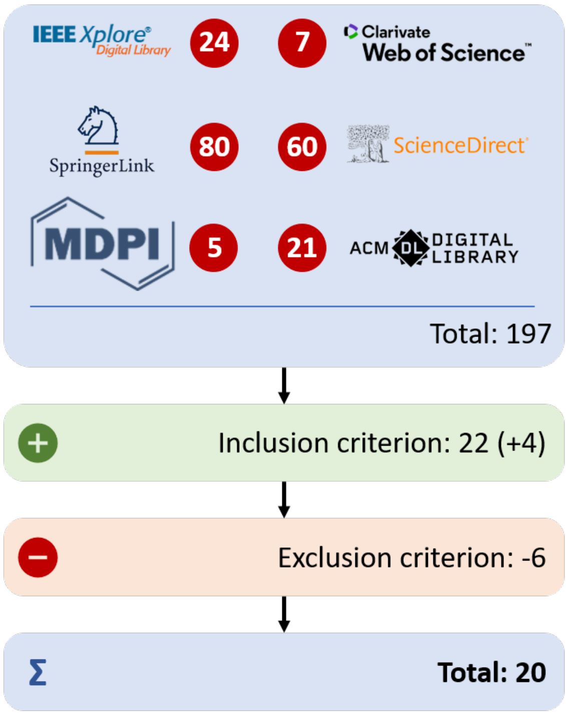


Figure 7.1: Process of the search strategy

Inclusion and exclusion criteria We considered the following inclusion criterion: *Papers that propose a solution to estimate cyclist and pedestrian intention estimation by relying, at least, on one vehicle sensor.* We have also included 4 articles that the search did not find and that experts recommended: [208, 209, 210, 211].

We then applied the following exclusion criterion : (i) if the article is not in English. (ii) the paper is not accessible. (iii) the study is only doing detection of the P&C.

Selection of papers The summary of the methodology is presented in Fig. 7.1. A total of 197 articles were found by the search engines. We then applied the following steps: *Step 1:* we skimmed the list of off-topic articles after reading the title and abstract. The list was reduced to 41 papers. *Step 2:* we applied the inclusion criterion after reading the articles. After the application of it, the number of fetched papers is 22. Some articles were not selected because the solutions used external cameras (such as UAVs or traffic CCTVs), or pedestrian-embedded sensors. We can also mention the paper [202] that is not proposing a technical solution but provides guidelines to achieve the intention estimation for cyclists. *Step 3:* we eliminated 6 papers, one because it was written in Spanish, one was doing only detection of the pedestrian, and the 4 others were not accessible. The final list counts 20 articles, with the 4 added articles, see Table 7.1.

b) Results

There are two common outputs for the intention estimation: compute the Crossing/Not Crossing probability ($\mathbb{P}[C/NC]$) or estimate the trajectory of the P&C. For cyclists, the most common task is trajectory prediction. Indeed, cyclists can be roadway users as well and will not cross the road like a pedestrian. In our case, with Yōkobo, the probability of crossing would have no sense, or even if it is adapted. Indeed, we do not have a binary option. Then we will look for the results with a trajectory prediction output.

Clues Some cues are essential to achieve a good estimation, but most of them are similar between pedestrians and cyclists. Then, [212] pointed out that the *orientation towards the road* is a key point regarding the intention estimation. The *gaze* and the *head orientation* as well, and [213]

Table 7.1: Comparison of the studies to estimate pedestrians/cyclists' intention. The detail of the acronyms is given in the List of Acronyms section. P: pedestrian; C: cyclist; V: vehicle. When the word *and* is used to describe the method, it means the authors developed several algorithms to compare them; the sign + is used to describe the several components of the same algorithm. $\mathbb{P}[C/NC]$ means *probability of crossing/not crossing*.

Study	**P/C/V**	**Sensor**	**Method**	**Output**	**Dataset**	**Remarks**
Rasouli et al. (2019) [218]	P	Video VS	RNN encoders-decoders (LSTM)	Trajectory est. $\mathbb{P}[C/NC]$	PIE	Creation of the dataset
Völz et al. (2016) [222]	P	Lidar	CNN (LSTM) and DNN	$\mathbb{P}[C/NC]$	*self*	Focusing on crosswalk
Wu et al. (2021) [214]	P	Video	Extend TPB with MLP + RNN	$\mathbb{P}[C/NC]$	PIE	
Hashimoto et al. (2015) [223]	P	Video TLS	DBN	$\mathbb{P}[C/NC]$	*self*	Focusing on crosswalk; TLS is set manually
Alvarez et al. (2020) [219]	P	3D video VS	Stacked RNN (GRU)	$\mathbb{P}[C/NC]$	PIE	Tested with real vehicle
Sun et al. (2019) [224]	P	Video	CGAN (LSTM encoders-decoders)	Future position estimation	Daimler [225] Context [226]	
Stolz et al. (2018) [220]	C	Radar	Hough transformation and RANSAC	DoM	*self*	
Girase et al. (2021) [227]	P/C/V	Video Lidar	CVAE + Scene graph + GRU	Trajectory est.; Intention	LOKI	Creation of the dataset
Kataoka et al. (2018) [228]	P	Video	Convnet + DeCAF + SVM	$\mathbb{P}[C/NC]$	NTSEL NDRDB	Creation of the datasets
Muscholl et al. (2021) [215]	P	Video	DBN	$\mathbb{P}[C/NC]$	OpenDS-CTS2	Focus of the interaction between 2 pedestrians across the street; Creation of the dataset
Poibrenski et al. (2020) [229]	P	Video	CVAE + RNN	Trajectory est.	JAAD [230] ETH UCY	
Kalatian and Farooq (2022) [216]	P	Video Lidar	RNN (Aux-LSTM)	Trajectory est.	NuScenes [231] Lyft [232] Waymo [233] PIE	Tested with virtual environment as well
Kooij et al. (2019) [217]	P/C	3D video	DBN (SLDS)	Trajectory est.	Context	
Ferguson et al. (2015) [234]	P	Lidar	Changepoint-DPGP	Trajectory est.	*self*	Tested with rover

Table 7.1 Continued: Comparison of the studies to estimate pedestrians/cyclists' intention

Wu et al. (2019) [221]	P	Video Lidar IMU VS	DBN + DSF	Trajectory est. $\mathbb{P}[C/NC]$	*self*	
Kotseruba et al. (2020) [212]	P	Video VS	RNN encoder-decoder (GRU)	Future position estimation	PIE	
Fang and López (2019) [208]	P/C	Video	Faster RCNN (P) Mask R-CNN (C) + RF	$\mathbb{P}[C/NC]$ (P) $[Arm\ pos]$ (C)	JAAD (P) CASR (C)	Creation of CASR
Saleh et al. (2018) [209]	C	Video	RNN (U- and B-LSTM)	Trajectory est.	CTD [235]	
Schulz and Stiefelhagen (2015) [210]	P	3D video VS	LDCRF	$\mathbb{P}[C/NC]$	Daimler	
Møgelmose et al. (2015) [211]	P	Video GPS	Munkres + IPM	Trajectory est.	*self*	Use the GPS to know if P. enters hazardous zone

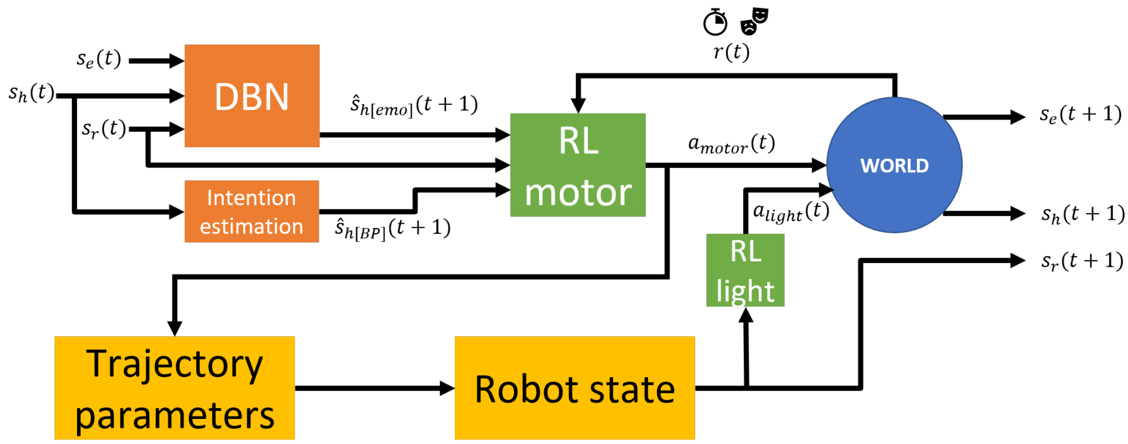


Figure 7.2: General architecture of the proposed solution.

mention it as well for the cyclists. Those parameters could also be valuable in the case of Yōkobo, with as well the *body points*. Five studies are explicitly using the head orientation of the P&C to determine their intention [214, 215, 216, 217, 210]. Five additional studies check the body posture/orientation [218, 219, 220, 221, 208]. We cannot notice any correlation with the output of each methodology, studies looking for both, $\mathbb{P}[C/NC]$ and trajectory estimation, can use this data.

Sensors used Using video is the most common input used with 85% of the studies using it, but less than half (8) solely rely on it. They can complete either with a lidar (3), using the vehicle speed (5) or GPS (1). In addition, the solution of [221] is the most complete by using the data of 4 different sensors. Finally, one can notice that three solutions solely rely on remote sensing methods (lidar or radar). Since Yōkobo only use a camera, we will focus on the studies using that input.

The methods One of the first steps of all of those methods is to identify the P&C, that we are already doing with Yōkobo with the HPEA.

About half of the studied methods are using RNN as the main method to achieve intention recognition. As mentioned by [216], the high correlation, temporally and spatially, of the pedestrian pattern makes RNN a good candidate. Indeed, they are good to deal with trajectories and dynamical systems [236]. The way to achieve the network is mostly using LSTM with some variation in the application since the network needs to remember previous states to deal with the trajectory.

The other common method is to use a DBN. They are an extension of Bayesian networks, made for dynamic systems. They can determine the probability for a specific event to happen, according to its evolution through time. That makes it suitable to obtain the probability of the event *crossing* to happen according to the evolution of the pedestrian through time.

7.2 BeWoDa's Architecture

We propose an architecture called Behaviours from World Data (BeWoDa) to adapt the behaviours of the robot according to the world's data and learn from them. The goal is to find the best *action* to achieve according to the current *state of the world*. This architecture is made for Yōkobo, but is not limited to it, and can be adapted according to the input and output of the system.

The general architecture is presented in Figure 7.2, and each block will be explained in the next subsections.

Table 7.3: The state space of the environment. The classes help to reduce the complexity of the problem. Hence the domain size of the Bayesian network decreases from 10^{12} to 800,000.

Data	Class	Meaning
Temperature (inside/outside)	$\{\leq 15 ; \leq 25 ; \leq 70\}$ °C	{cold ; warm ; hot}
Humidity (inside/outside)	$\{\leq 30 ; \leq 50 ; \leq 100\}$ %	{dry ; normal ; humid}
Atmospheric pressure	$\{\leq 1015 ; > 1015\}$ hPa	{low-pressure ; anticyclone}
CO_2 concentration	$\{\leq 400 ; \leq 1000 ; \leq 1200\}$ ppm	{good ; medium ; bad}
Time	{0-6 ; 6-12 ; 12-18 ; 18-24 } h	{night ; morning ; afternoon ; evening}

7.2.1 Inputs

At each time t Yōkobo reads the *state of the world* $s_w(t)$ which is divided into 3 different states:

State of the environment $s_e(t)$**:** it includes the data of the weather station (temperature in-/out, humidity in/out, atmospheric pressure and CO_2 level). In order to simplify the problem by reducing the size of the networks, the data will be discretised into classes. The networks have to run on the RPi of Yōkobo, then we need to be efficient. Indeed, the behaviour of Yōkobo does not need to be really different for each degree or pascal. See Table 7.3 for the definition of the classes.

State of the robot $s_r(t)$**:** it represents the PAD of the robot calculated in the *Robot state* block, see section 7.2.5 for more details.

State of the human $s_h(t)$**:** represent the user. It is divided into two states. With $s_{h[emo]}$ for the emotion (neutral, happy, sad, surprise, or anger) and $s_{h[BP]}$ for the body position of the user. Both of them are given by the OpenVINO toolkit [162].

7.2.2 Outputs

The actions $a(t)$ that Yōkobo can do are divided into two groups. First, the **motors' actions** $a_{motor}(t)$ for the command of each motor M_{1-3} knowing that the position of the motor M_3 is linked with the position of M_2. To simplify the output of the RL, the algorithm will not have to select a precise position in degree for each motor but instead decide to increase, decrease from a step the position of the motor, or stay at the same position. Then, the number of different actions is reduced from about 46.7 million (360^3) to 27 (3^3).

The second group of action is the **light** $s_{light}(t)$, it is composed of 4 components, one for each colour (Red, Green and Blue) and one for the luminosity. Those values are a number from 0 to 255 (one byte). In order to simplify the output, a palette of 10 colours will be selected. The algorithm will have to select a colour from that palette and then increases or decreases the luminosity for a step, or keep the same one. Hence, the number of possible actions for the light is 30 (10×3). Table 7.4 shows the palettes of colour. They have been selected to correspond to the seven traditional colours of the rainbow, plus pink, white and black (light off), hence 10.

The choice of the two steps has been made empirically, with 10° for the motors and 10 for the luminosity. Table 7.5 sums up the output.

7.2.3 User's Intention Estimation

The first blocks of the algorithm, in orange in Figure 7.2, will estimate the user's intention, and then use this when Yōkobo will decide the action it will perform (in the green blocks). This prediction is based on two points, on one hand, the emotion of the user and on the other hand the position of the user. We based our analysis on the systematic literature review done for the intention estimation for pedestrians.

Table 7.4: The palette of colours of the LED. The way the colours are rendered by the LED strip is not perfect, then the orange and yellow seem to be wrong but are rendered in the correct colour.

Colour	Name	RGB				
	off	0	;	0	;	0
	Red	255	;	0	;	0
	Orange	255	;	55	;	0
	Yellow	255	;	150	;	15
	Green	110	;	230	;	40
	Cyan	0	;	255	;	255
	Pink	255	;	50	;	160
	Blue	0	;	0	;	255
	Violet	120	;	30	;	250
	White	255	;	255	;	255

Table 7.5: The actions space, are the output of the reinforcement learning algorithms to command the motors and the LED of Yōkobo.

Component	Action	Classes
Motor 1 Motor 2 Motor 3	{-10 ; 0 ; +10}°	27 classes
LED Red LED Green LED Blue	10-colour palette	30 classes
LED Luminosity	{-10 ; 0 ; +10}	

a) Emotion Intention Estimation

This block's objective is to find the intention of the user in the future, and then Yōkobo will be able to use this estimation to find the best behaviour for Yōkobo. Since the Reinforcement Learning algorithm (see section 7.2.4) is focusing on the user's emotion as a reward, we will try to guess the likelihood user's emotion. Based on the literature review done on intention estimation for pedestrians, we decided to use a Dynamic Bayesian Network (DBN). Bayesian Network (BN) are a good tool to build the causal relationship between the variables [237], here we want to have the relationship of the future user's emotion between the weather data, Yōkobo's emotion and the current user's emotion.

Bayesian Network The BN is a probabilistic graphical model using a Directed Acyclic Graph (DAG) to represent the association of variables, see the Figure 7.3. They are used to predict the likelihood of a variable to happen according to the evidence, i.e. the value of the other variables (but it is not required to know all of them). The network will return the probability. It follows the joint probability function, given by equation 7.1.

Notations:

The probability function will be denoted as $\mathbb{P}()$ and the following simplification will be used for the conjunction of several variables:

$$\begin{aligned}\mathbb{P}\left(\bigwedge_{i=1}^{n} X_i = x_i\right) &= \mathbb{P}(X_1 = x_1 \wedge X_2 = x_2 \wedge ... \wedge X_n = x_n) \\ &= \mathbb{P}(x_1, ..., x_n)\end{aligned}$$

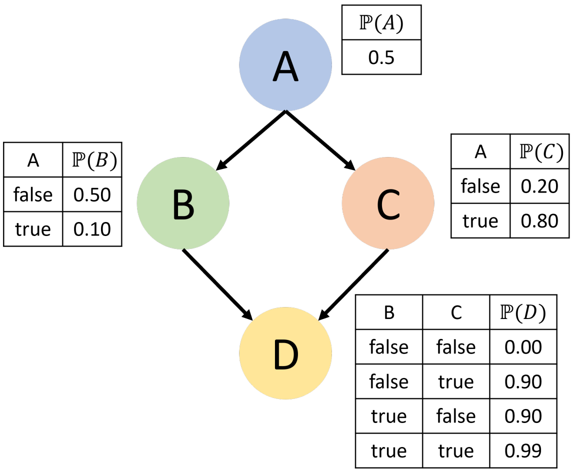


Figure 7.3: Example of a simple BN with four nodes $V = \{A, B, C, D\}$ connected by arcs. The tables are called CPTs and they store the probabilities of each node according to the parents. In this example each node is Boolean, but they could have more states than just $true$ and $false$, such as $cold, warm, hot$ for a temperature.

$$\mathbb{P}\left(x_1, ..., x_n\right) = \prod_{i=1}^{n} \mathbb{P}(x_i | parents(X_i)) \tag{7.1}$$

$$= \prod_{i=1}^{n} \mathbb{P}(x_i | x_{i-1}, ..., x_1) \tag{7.2}$$

Where $X_n \subset \mathbf{V}$ are the nodes of the networks. This is because, the BN follows the local Markov property, meaning that a node is conditionally independent of its non-descendants given its parents. To calculate the probability of a node, we will use the theorem of Bayes:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)}{\mathbb{P}(B)} \tag{7.3}$$

Let us say we want to know the probability of $X_1 = x_1$, then, using the theorem of Bayes, we can calculate the probability according to the $n - 1$ parents node, we have:

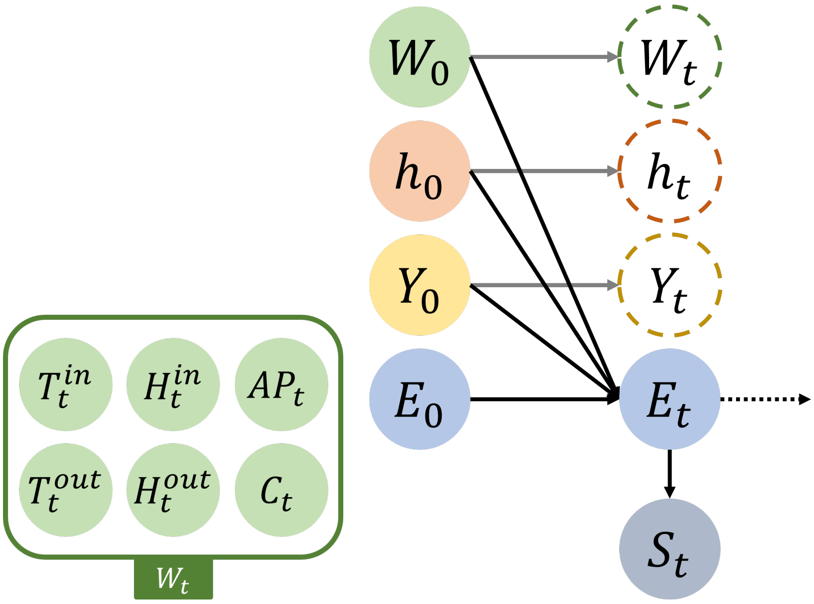
$$\mathbb{P}(x_1 | x_2, ..., x_n) = \frac{\mathbb{P}(x_2, ..., x_n | x_1) \cdot \mathbb{P}(x_1)}{\mathbb{P}(x_2, ..., x_n)} \tag{7.4}$$

$$posterior = \frac{likelihood \cdot prior}{evidence} \tag{7.5}$$

The *evidence* is what we know, the current states of the parents' nodes of X_1, the *prior* is the probability of the node X_1 to be in the state x_1. Finally the *likelihood*, the probability of the parents' nodes to be in this state, knowing that $X_1 = x_1$. All the probabilities, for each node, are stored inside what is called the Conditional Probability Table (CPT), they are the probabilities of the node for each state of its parents (if they are no parents, it is the probability of this node to happen).

Dynamic Bayesian Network It is an adaptation of the BN to describe a *dynamic* system, the network is not dynamic. It is simply the duplication (unrolling) of the same BN over a number of steps we aim to predict. In our case, we will only predict one step in the future, then the network will be composed of two slices: t and $t + 1$.

The input of our network are the weather data (see Table 7.3), the user's emotion (see the list in Table 7.6) and the robot's emotional state. For the latter, the PAD of the robot is converted into an emotion, according to Table 7.8. The structure of the network is described and explained



(a) Weather data **(b)** First slice of the DBN

Figure 7.4: The Dynamic Bayesian Network used to map the causal relationship from the world's data. (a) to simplify the graph, the data from the weather station are grouped inside a node called W_t, but in the real network, each weather node is independent. (b) represents the DBN, the prior states are, in addition to the weather, the current hour (h_0), the emotion of Yōkobo (Y_0, the PAD data are converted with the same mapping using in 7.8), and the user's emotion (E_0). Only the posterior of the user's emotion (E_t) is valuable in our case, then for simplification, the others (W_t, h_t, Y_t) are discarded. Finally, the evidence (S_t) to map the probability between $E_t^{T=1}$ (the estimation of the future emotion) and $E_0^{T=2}$ (the real future emotion). The algorithm is looped with a period T when a new set of data are read from the sensors.

in Figure 7.4. The network is following two time periods, one denoted with small t representing the DBN slices, the observation period. In our case, we only want to predict one step in the future, then $t \in \{0; 1\}$. The second one, with a capital T represents the sampling rate of the BeWoDa, $T \in \mathbb{N}$. See Figure 7.5 for a visual explanation.

The initial probabilities for the priors (for $T = 0$) are initialised empirically and they will be updated while the algorithm is running. The temperatures, humidity, atmospheric pressure and CO_2 level are assumed to follow a normal distribution. Each class is equiprobable for the user's and robot's emotional state and time. Concerning the weather data, the algorithm will update the mean and standard deviation of the normal distribution from the last N data. The goal is the have the most accurate probabilities, which is why the past data are discarded since the data in the distant past have little impact on the current weather. Concerning the emotion's data (from Yōkobo and user), the past data have still some relevance. That is why we decided to use the opposite of the arctan function, in order to give more weight to the close past data than the distant past one. The cutoff point[1] (p_c) can be set, as well as the weight of the distant past data (w_p).

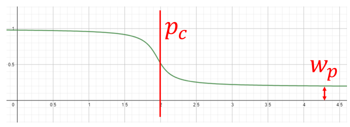


Figure 7.6: The parameters of the filter

The DBN is developed using the PyAgrum[2] library.

Particle Filtering (Dynamic) Bayesian Networks implies knowing the CPTs for each variable, thanks to a database. However, in our case, this is not possible to do, since the probability is specific to the user, we cannot ask to train the model for each user. Then, we will use **particle filtering** to build the parameters of the model over the time [195]. In our case, it will be used to calculate the probability of the posterior ($\mathbb{P}(E_t|W_0, h_0, Y_0, E_0)$). The filter will generate N random

[1] By analogy with the cutoff frequency of a filter
[2] `https://pyagrum.readthedocs.io/en/1.4.0/`

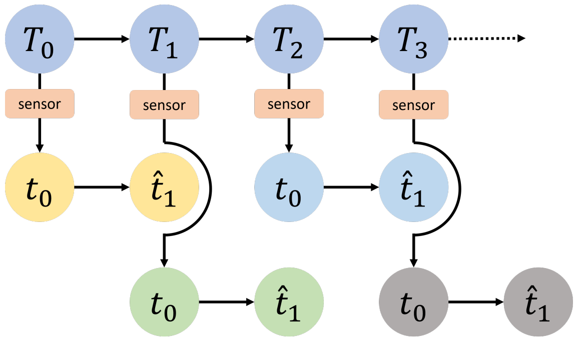


Figure 7.5: Representation of the difference between the two periods of the algorithm. The hat ($\hat{t}_1$) symbolises the estimation the algorithm is calculating on the future slice.

data (the particles) with the data from the sensors (the priors) and the sensor model (S_t), and from this sample, the CPTs of the posterior will be calculated. The sensor model S_t is updated during the next period T with the new data from the sensors.

b) Trajectory Estimation

The goal of this block is to estimate the next body position of the user ($\hat{s}_{h[BP]}(t+1)$) from the current position ($s_{h[BP]}(t)$). For this preliminary work, we focused more on the user's emotions than the body position. Then, the body points will be reduced to one unique position (centre of the hip), and then the trajectory will be approximated as a linear function using the two last positions. This hypothesis is relevant since the people are mostly doing sidewalk movements. We will also consider the user's velocity as constant because the place of Yōkobo is at the entrance, a place that is generally narrow. Then, the user will probably not run. Then, the future position computation follows the equation 7.6.

$$\left.\begin{aligned} y_{t-1} &= a \cdot x_{t-1} + b \\ y_t &= a \cdot x_t + b \\ dt &= x_t - x_{t-1} \end{aligned}\right\} \Rightarrow \begin{pmatrix} x_t + dt \\ a \cdot (x_t + dt) + b \end{pmatrix} = \begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} \tag{7.6}$$

7.2.4 Reinforcement learning algorithms

The RL algorithms observe the environment and read some **states** (s_t) from it. From them, the algorithm finds the best **action** (a_t) to achieve and do it. Then, the environment *replies* with a new state (s_{t+1}) and a **reward** (r_t). The algorithm uses these rewards to adapt its behaviour and try to maximise the reward it can obtain from a state/action couple.

In our solution, we have two different RL algorithms, one for the motors and one for the light. We decided to split it to reduce the number of different actions. Moreover, movement is the primary way of communication, and the light *supports* it. The block for the light selects the colour based on the generated motion.

The algorithms are developed using the PyTorch[3] and Gym[4] libraries.

a) For the motor action

The goal of this part of the algorithm is to find the best movement for Yōkobo according to the estimation of the future user's emotion ($\hat{s}_{h[emo]}(t+1)$), the current Yōkobo's state ($s_r(t)$) and the estimation of users' intention (trajectory, $\hat{s}_{h[BP]}(t+1)$), see Figure 7.2 and Table 7.6. The rewards are the user's emotions and the interaction's duration. Yōkobo will train to make the user happy and avoid making him angry or sad. Moreover, it will try to have a long enough interaction, but this reward will stop being used after 2 min. Indeed, the goal of Yōkobo is to interact with the person, and then try to catch his/her attention. But it is not necessary to catch the user for hours.

[3]`https://pytorch.org`
[4]`https://www.gymlibrary.dev`

Table 7.6: The states and actions for the reinforcement learning algorithm that create the trajectories of Yōkobo

States	**Values**	**Combinations**
Emotion	`["neutral", "happy", "sad", "surprise", "anger"]`	5
PAD	`{-2, -1, 0, 1, 2}`	$5^3 = 125$
Trajectory	Minimum 2 points: `(xA, yA)*(xB, yB) // 640*480`	$640^2 * 480^2 \approx 94\text{E}9$
Actions	**Values**	**Combinations**
Motors	`{-1, 0, 1}`	$3^3 = 27$

Q-Learning One of the most common algorithms for RL is called Q-Learning. To determine the best action to achieve according to the current state, it uses a table (called Q-Table) to store, for each (s, a) couple a Q-value. This value is calculated with the Bellmann equation (see equation 7.7) [238] after the action (a_t) has been done and the new state (s_{t+1}) and reward (r_t) observed. It depends on several parameters, first the current value stored in the Q-Table for the current state/action $(Q(s_t, a_t))$[5]. Then, there is the **learning rate** ($\alpha \in]0; 1]$), the smaller α is, and the longer it will take for the algorithm to learn. Finally, there is the maximum reward that could be obtained according to the new state (s_{t+1}): $\max_a Q(s_{t+1}, a)$. It is weighted by the $\gamma \in]0; 1]$ discount factor, which gives more weight to the previous tries.

$$Q^{new}(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \cdot \left(r_t + \gamma \cdot \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right) \tag{7.7}$$

As said before, all those values are stored in the Q-Table. The algorithm will then select the action a that maximises the Q value according to the current state s, and at this step, there is a last parameter: the *curiosity factor* ε (from the ε-greedy strategy) that is set to a small value. The algorithm will select the action according to the Q value (100-ε)% of the time, but sometimes will try another action **randomly**, by *curiosity*.

The Q-Table is a $(N_s \times N_a)$ table, where N_s is the number of possible states, and N_a is the number of possible actions. If we want the algorithm to perform well, during its training phase it has, at least, to try every (s_i, a_i) possible couple to calculate the Q-value. Hence, we can see a limitation here when the table is big.

In our case, if we refer to Table 7.6, we can calculate the dimension of the Q-Table (D_Q):

$$\begin{aligned}
N_s &\approx 5 \cdot 125 \cdot 94 \cdot 10^9 \\
&\approx 58.7 \cdot 10^{12} \\
N_a &= 27 \\
D_Q &= N_s \cdot N_a \\
&\approx 58.7 \cdot 10^{12} \cdot 27 \\
&\approx 1.6 \cdot 10^{15}
\end{aligned}$$

That means the Q-Table contains about one and a half quadrillions cases, which is too much and would take ages to complete it. That is why we decided to use a Deep Q-Learning approach instead.

Deep Q-Learning The DQL [239] is similar to the vanilla one, albeit the Q-Table is replaced by a Neural Network (NN). The input layer dimension is here the **number of states** and not the number of possible states. In our case, we have the emotion (1 value), the PAD (3 values) and the trajectory (4 values), hence the input dimension is **8**. The output of the NN is the Q-value for each possible action, in our case **27**, see Figure 7.7. The number and size of the inner layers are parameters that have to be tuned. Finally, the Bellmann equation (equation 7.7) is used here as a loss function for the backpropagation. Because we are using a DQL algorithm, we do not need

[5] At the beginning, the Q-Table is initialised at a specific value, generally 0.

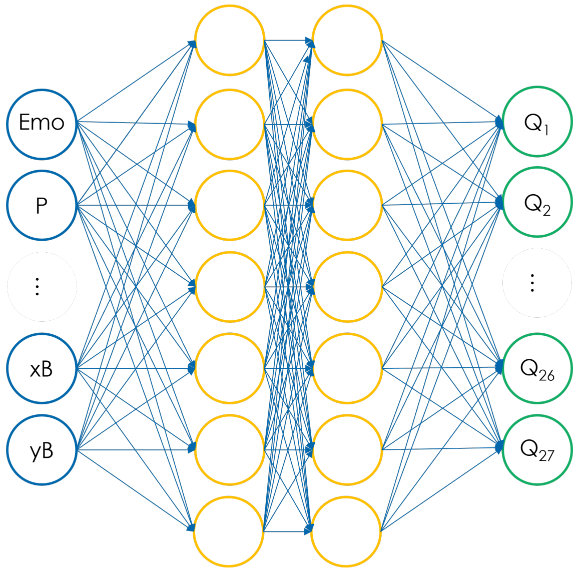


Figure 7.7: Schematic diagram of the Neural Network (NN) of the Deep Q-Learning algorithm. The inputs correspond to each component of the state (*Emo* for the emotion, *P* for the pleasure in PAD and x_B, y_B for the coordinate of the second point of the trajectory) and the output to the Q-values of each action Yōkobo can do. The number and size of the inner layers (in yellow) is only for illustrative purpose.

anymore to reduce the dimension of the PAD by using only 5 values, then we can use as an input a value from -1 to 1.

The algorithm is described in pseudocode in the appendix E (Alg. E.1 to E.4). It is divided into several classes called by the main function which is looping to generate `nbr_trial` episodes[6], each episode is composed of several steps that are done as follow:

1. The DQL choose an action to perform according to the current state (Alg. E.1, ℓ. 9).
2. Yokobo performs a step (Alg. E.3, ℓ. 12),
   (a) the motors are moved;
   (b) the PAD is calculated (see section 7.2.5);
   (c) the light are set (see section b));
   (d) the new state is read;
   (e) the reward is calculated;
   (f) the `done` Boolean is set.
3. The DQL is updated with the reward, new state and the `done` variable (Alg. E.1, ℓ. 11).
4. If we are in a terminal state (`done` is true), then the current episode is stopped.

Regarding the `done` flag, it is set to true when the user stops the interaction with Yōkobo, i.e they are moving out of the frame. The core of the algorithm is done in the step function (Alg. E.3, ℓ. 12), where after doing the action provided by the DQL algorithm, it checks the result on the world by observing it. Thanks to this observation, we can calculate the reward for the algorithm. We decided on two main points for them, first, Yōkobo should try to set the user in a **good mood**, and second it should try to have a *reasonable* **interaction time**. For the first point, we check the emotion of the user, if it is positive (i.e happy) the reward is positive as well, and vice versa for a negative emotion (i.e. sad or angry). Concerning the interaction time, the goal of Yōkobo is to maximise it until a certain amount, as mentioned before, the goal of Yōkobo is not to monopolise the user. Then, in a specific time range, the reward will be positive, and outside it will be negative.

[6]For training and test purpose. When the algorithm is implemented in the real robot, it will be an infinite loop

Table 7.7: The reward of the DQL algorithm that commands the motors. If the interaction time is below 10 s, we assume that the user was just passing in front of Yōkobo. That is why we do not penalise it. A continuous reward is added at each step, and a terminal one is only added at the terminal step.

User's emotion	**neutral**	**happy**	**sad**	**anger**	**surprise**
Continuous reward value	0	1	-1	-1	0

Time (in seconds)	**< 10**	**< 30**	**< 60**	**< 120**	**≥ 120**
Terminal reward value	0	-5	2	5	-5
Continuous reward value	0	0	1	1	0

This reward is calculated only when the `done` flag is true (indeed, the interaction time cannot be calculated before the end of the interaction), however, a small positive reward is given at each step if the interaction duration is inside the range. Table 7.7 sums up the different rewards and gives their value.

b) For the light action

The algorithm to select the colour of the light use similar architecture to the one for the motors. It is also a DQL algorithm, but the inputs-outputs and the rewards are different. The goal of this block is to fit the light colour and luminosity (the brightness) with Yōkobo's emotional state (PAD, see section 7.2.5). Then, the input is Yōkobo's PAD (three values) and the output is one colour of the palette (see table 7.4) with the luminosity; the action was described in table 7.5. When the action is selected, the colour is converted to an RGB tuple and the luminosity value is calculated.

For the training of this algorithm, we do not use the same rewards, based on the user's reaction. We want, here, the colours to match the current state of Yōkobo. Thus, we used the literature on psychology studies about the association by people between colours and emotions. First, we used the book of Mehrabian, which created the PAD scale, to convert the PAD values to emotions [240]. As a first approach, we selected the 5 same emotions we got from the user, plus some additional ones, see table 7.8. Then, based on the work of [241, 242, 243, 244], we mapped each emotion to a colour, even if some difference of perception can appear for example because of cultural background [245]. We used the colour that comes frequently for each emotion, as a first approach (we will see in section 7.4 proposition of optimisation). Then, if the colour chosen by the algorithm is the same as the one in the table, the reward is positive, otherwise negative. Moreover, to avoid Yōkobo to be flashing like a disco ball, a negative reward is given if the colour changes each step from the last 5 steps, but also if the colour stayed the same. Finally, a positive reward is added if the new colour is close to the previous one (by close, we mean in the colour spectrum, for example from *green* to *cyan*, see Table 7.4). Table 7.9 sums up the rewards.

7.2.5 Robot state

The next block of code has to evaluate the current emotional state of Yōkobo, and we decide to use the PAD to *put a number* on it [240], it is a common tool in robotics to express the emotion of a robot (q.v. section 2.2.6). In order to get the robot state, we decided to extract it from the **trajectory parameters** like the velocity, acceleration or jerk of the generated motion from the **RL motor** algorithm. We decided to use the movement of the end effector (i.e. Yōkobo's head) to define the emotional state. Based on the work of Claret et al. [246] where they achieve the inverse operation: getting the movement parameters from the PAD values. We decided to use the same relations:

Pleasure (P) is inversely proportional to the **jerk** in the motion.

Arousal (A) is proportional to the **kinematic energy** of the motion.

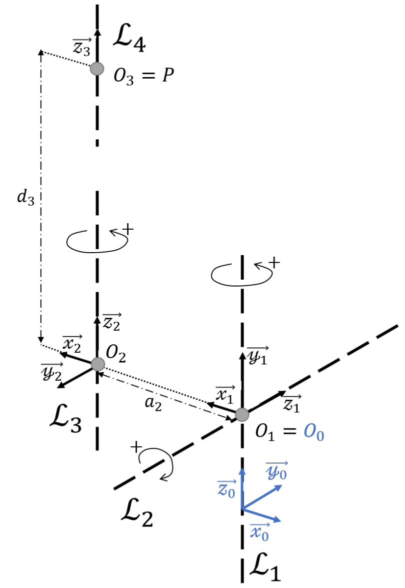
Dominance (D) is linked to the gaze of the robot. Even if Yōkobo does not have eyes, we used the answers from the questionnaire. Indeed, most of the participants identified the apex as

Table 7.8: The mapping table between emotion, PAD and colour. Based on [240, 241, 242, 243, 244], except for *neutral* and *curious* emotions. For the first one, we decided to set it for the null PAD value, with white which is a neutral colour. Regarding *curiosity*, we could not find any mapping with colour in the literature. From non-scientific mapping, it came out that green and yellow are often associated with this emotion. We selected green to have two different emotions for yellow, and two for green.

Emotion	**P**	**A**	**D**	**Colour**
angry	-0.51	0.59	0.25	RED
bored	-0.65	-0.62	-0.33	WHITE
curious	0.22	0.62	-0.01	GREEN
elated	0.50	0.42	0.23	YELLOW
hungry	-0.44	0.14	-0.21	ORANGE
loved	0.87	0.54	-0.18	PINK
sleepy	0.20	-0.70	-0.44	BLUE
violent	-0.50	0.62	0.38	RED
sad	-0.63	-0.27	-0.33	VIOLET
happy	0.81	0.51	0.46	YELLOW
surprised	0.40	0.67	-0.13	WHITE
fearful	-0.64	0.60	-0.43	VIOLET
relaxed	0.68	-0.46	0.06	GREEN
neutral	0.00	0.00	0.00	WHITE

Table 7.9: The rewards of the DQL algorithm for the light

	Continuous reward
The colour matches with the PAD	3
The colour does *not* match with the PAD	-3
The new colour is close to the previous one	2
The colour has changed each step for the last 5	-1
The colour stayed same each step for the last 5	-2



	a_i	α_i	d_i	θ_i
1	0	$\pi/2$	0	θ_1
2	a_2	$-\pi/2$	0	θ_2
3	0	0	d_3	θ_3

Figure 7.8 & Table 7.10: Reminder of the Denavit–Hartenberg parameters of Yōkobo. In this chapter, for simplification, a_2 will be changed in a and d_3 to d. P is located at the top of the head/bowl.

the head of Yōkobo, and the vent as its *eyes*. Then we decided to use the position of the head. If the **position of motor 2** is low, the dominance will be high (Yōkobo is looking down) and vice versa.

a) Arousal

Concerning Arousal, we have the following relationship:

$$A = \frac{2}{E_{k,max}} \cdot E_k - 1 \tag{7.8}$$

With E_k the kinematic energy and E_k, max the maximum of the kinematic energy. This value is calculated using the forward kinematic. The Figure 7.8 reminds the DH parameters. First, we calculate the expression of the position of the end effector (P) in the world frame ($\mathcal{R}_0$) thanks to the transformation matrix. For simplification, in this chapter, $C_i = \cos(\theta_i)$ and $S_i = \sin(\theta_i)$.

$$\begin{aligned}
{}^0A_3 &= {}^0A_1 \cdot {}^1A_2 \cdot {}^2A_3 && (7.9)\\
&= \begin{bmatrix} C_1 & 0 & S_1 & 0 \\ S_1 & 0 & -C_1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} C_2 & 0 & -S_2 & aC_2 \\ S_2 & 0 & C_2 & aS_2 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} C_3 & -S_3 & 0 & 0 \\ S_3 & C_3 & 0 & 0 \\ 0 & 0 & C_3 & d \\ 0 & 0 & 0 & 1 \end{bmatrix} && (7.10)\\
&= \begin{bmatrix} C_1C_2C_3 - S_1S_3 & -C_1C_2S_3 - C_3S_1 & -C_1C_3S_2 & \mathbf{aC_1C_2 - dC_1S_2} \\ C_1S_3 + C_2C_3S_1 & C_1C_3 - C_2S_1S_3 & -C_3S_1S_2 & \mathbf{aC_2S_1 - dS_1S_2} \\ C_3S_2 & 0 - S_2S_3 & C_2C_3 & \mathbf{aS_2 + dC_2} \\ 0 & 0 & 0 & 1 \end{bmatrix} && (7.11)
\end{aligned}$$

We can extract, from the top right of the matrix, the position of the end effector:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix}_P = \begin{pmatrix} aC_1C_2 - dC_1S_2 \\ aC_2S_1 - dS_1S_2 \\ aS_2 + dC_2 \end{pmatrix}_{\mathcal{R}_0} \tag{7.12}$$

Then, using the forward kinematic, we know that $\vec{v} = \mathcal{J}(\vec{\theta}) \cdot \dot{\vec{\theta}}$ with $\mathcal{J}$ the Jacobian matrix and

θ_i the parameters of the robot. The Jacobian matrix is defined as follows:

$$\mathcal{J} = \begin{bmatrix} \frac{\partial x}{\partial \theta_1} & \frac{\partial x}{\partial \theta_2} & \frac{\partial x}{\partial \theta_3} \\ \frac{\partial y}{\partial \theta_1} & \frac{\partial y}{\partial \theta_2} & \frac{\partial y}{\partial \theta_3} \\ \frac{\partial z}{\partial \theta_1} & \frac{\partial z}{\partial \theta_2} & \frac{\partial z}{\partial \theta_3} \end{bmatrix} \tag{7.13}$$

$$= \begin{bmatrix} -S_1(aC_2 - dS_2) & -dC_1C_2 - aC_1S_2 & 0 \\ C_1(aC_2 - dS_2) & -dS_1C_2 - aS_1S_2 & 0 \\ 0 & aC_2 - dS_2 & 0 \end{bmatrix} \tag{7.14}$$

We can calculate the velocity of Yōkobo's head:

$$\vec{v} = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix}_P = \begin{pmatrix} -S_1(aC_2 - dS_2)\dot{\theta}_1 - (dC_1C_2 + aC_1S_2)\dot{\theta}_2 \\ C_1(aC_2 - dS_2)\dot{\theta}_1 - (dS_1C_2 + aS_1S_2)\dot{\theta}_2 \\ (aC_2 - dS_2)\dot{\theta}_2 \end{pmatrix}_{\mathcal{R}_0} \tag{7.15}$$

The maximum of the kinematic energy is defined by $E_{k,max} = \frac{1}{2} \cdot m \cdot \|\vec{v}\|^2$, the energy is maximum when the velocity is maximum. We need to find which angular positions maximise the velocity of the head. Concerning the angular velocity, we will use the maximum of the motor indicated in the documentation[7]: $\Omega = 234.27\,\text{rev/min}$ ($24.53\,\text{rad}\cdot\text{s}^{-1}$).

$$V_{max} = \max_{\theta_1,\theta_2} \left(\sqrt{v_x^2 + v_y^2 + v_z^2} \right) \tag{7.16}$$

$$= \max_{\theta_1,\theta_2} \left(\Omega \cdot \sqrt{(-aS_{1+2} - dC_{1+2})^2 + (aC_{1+2} - dS_{1+2})^2 + (aC_2 - dS_2)^2} \right) \tag{7.17}$$

$$= \max_{\theta_1,\theta_2} \left(\Omega \cdot \sqrt{a^2(1 + C_2^2) + d^2(1 + S_2^2) - 2ad \cdot C_2S_2} \right) \tag{7.18}$$

We can see that the maximum velocity of the end-effector is independent of the angle of the first motor (θ_1). Then, we need to find the angle θ_2 that maximises the velocity:

$$\left(\frac{\partial V}{\partial \theta_2} = 0 \right)_{[0;2\pi]} \Rightarrow \theta_2 = \frac{\pi}{2} \tag{7.19}$$

Then, the maximum velocity is for $\theta_2 = \pi/2$, however, this position is physically unreachable for Yōkobo, since the amplitude of the second motor is limited between $\pm 0.31\,\text{rad}$ around 0 (for $\theta_2 = 0$ the velocity is minimal, and the head is vertically aligned with the base). Due to the eccentric of the third motor (a), the maximum is for $\theta_2 = -0.31\,\text{rad}$:

$$V_{max} = 2.65\,\text{m}\cdot\text{s}^{-1} \quad \forall \theta_1 \text{ and } \theta_2 = -0.31\,\text{rad} \tag{7.20}$$

We can obtain the maximum energy with $E_{k,max} = \frac{1}{2} \cdot m \cdot V_{max}^2$. The mass of the bowl is about 50 g, then:

$$E_{k,max} = 0.18\,\text{J} \tag{7.21}$$

b) Dominance

To calculate the Dominance, we just need to shit the position of motor 2 in a $[-1; 1]$ scale when the motor is at its minimum (*looking* up), the dominance is -1, otherwise to 1; then, we have:

$$D = \frac{2}{MAX - MIN} \cdot \theta_2 - \frac{MAX + MIN}{MAX - MIN} \tag{7.22}$$

[7] It is the same for the three motors.

Table 7.11: The parameters of the two DQL algorithms.

	DQL Motor	DQL Light
Learning rate α	0.003	0.003
Discount factor γ	0.9	0.9
Curiosity factor ε	0.9	0.9
Number of hidden layers	2	2
Size of the hidden layers	(256, 256)	(256, 256)

c) Pleasure

Like [246], we are using the jerk, but contrary to their project, we have to *read* it and not generate a jerked motion. We could have done something similar to the Arousal by calculating J_{max}, but that would require complex calculation, and in addition, need the maximum acceleration and jerk of the motors, which are not provided by the documentation. Then, we will define an arbitrary maximum value of the jerk ($1\,\mathrm{m{\cdot}s^{-3}}$), and if a bigger value is found, it will become the new maximum value. Thus, Yōkobo, will adapt from the time. Then, the formula is similar to the Arousal, besides the minus sign, since the Pleasure is inversely proportional to the jerk:

$$P = 1 - \frac{2}{J_{max}} \cdot j \tag{7.23}$$

7.3 Training of the algorithm

In this section, we will observe the result of BeWoDa in simulation, to see the movement of the motors and the change of the colour. This simulation is done to check the viability of the algorithm. Since BeWoDa is using the user's reaction from Yōkobo's behaviour as input, we cannot use data from recorded videos. Then, here we will generate fake data to evaluate the system, before testing it in real conditions.

7.3.1 Foreword

The algorithm has a lot of parameters that can be tuned to optimise the result of the algorithm. In addition to the rewards presented in Tables 7.7 and 7.9, we have also the parameters of the two DQL that can be changed. By lack of time, the value has been selected empirically and described in Table 7.6, the way those parameters are set is one of the future works.

Moreover, the terminal state of the algorithm is when the user stops the interaction with Yōkobo, in the simulation we decided to replace it with a finite number of steps: 300. Concerning the number of episodes, it as been set to 500.

7.3.2 Data Generation

To train the algorithm before testing it on Yōkobo, the data are manually generated. Concerning the weather data, since the weather station updates the data only each 10 min we do not need to generate a lot of data and try with different sets of data.

Regarding the user's emotions, we will simulate them by generating randomly them. To avoid the emotion changing at each step, the algorithm will keep the same emotion 70% of the time, and if the emotion of Yōkobo (given by the PAD) is one of the possible emotions of the emotion recognition algorithm (see Table 7.3), then the user will take the same emotion 40% of the time, to simulate empathy, see algorithm 7.1.

Algorithm 7.1: Generation of user's emotion data

```
FUNCTION generateEmotion
    input currentEmotion
    output newEmotion

    userPossibleEmotion ← ["neutral", "happy", "sad", "surprise", "anger"]

```

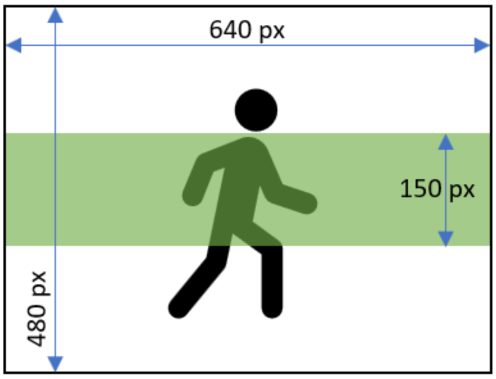


Figure 7.9: The green band symbolises the position where the training algorithm will generate the user's position

```
    newEmotion ← currentEmotion

    alea ← generate random number between 0 and 1
    IF alea < 0.3
        newEmotion ← select a random emotion
    END IF

    alea ← generate random number between 0 and 1
    IF Yokobo.emotion in userPossibleEmotion AND alea < 0.4
        newEmotion ← Yokobo.emotion
    END IF
END FUNCTION
```

For the user's position, the coordinate is also generated randomly. The horizontal position (x) is generated in the whole width of the screen (640 pixels, 480 px for the height), and the vertical coordinate (y) is set on a band of 150 px around the central line of the screen, see Figure 7.9. This constraint is done because the user will most likely remain on the depth, and then the y component be quite constant. In addition, when a new point is generated (being in the green band), it will randomly set in a square of 20 px size around the previous position. Again, we are assuming that the user is not moving too fast.

7.3.3 Results and Discussions

The graphs of the Figures E.1 and E.2 shows the positions of the motors and the PAD of Yōkobo over the steps for different episodes, two episodes at the beginning, one at the middle, and the last one. In addition, Figures E.3 and E.4 display moments of the animation of Yōkobo for episodes 0 and 498, respectively. A 3D animation of the movements has also been made with Blender. The angular position of each motor and the light colour and luminosity has been saved into a file, and then it is read by Blender to animate Yōkobo, see a screenshot on Figure 7.10.

We can see the evolution of Yōkobo's movement over time, while it is learning. In the beginning, for episodes 0 and 10 (Figures E.1a and E.1b), the movements are erratic and shaky with large amplitude. But those movements are not suitable for Yōkobo, because it should be a discreet robject in one's home. By doing huge movements, it will start to become noisy, moreover, it might eject the objects located inside the bowl. After some episodes (see Figures E.1c and E.1d), the movement is smoother, and the motors stay in a close range. For example, in the 498th episode, motor 3 mostly stays in a 30° range, while in the first episode, it is a 60° range.

Similar comments can be done for the PAD, especially for the Pleasure, which gets smoother over time. Since the goal of Yōkobo is to keep the user in a good mood, it is good that Yōkobo has a positive Pleasure (associated with good emotions); considering the movement of Yōkobo gets less shaky over time (the jerk decrease). One can also notice that Yōkobo's Arousal also stabilises around -1, meaning that Yōkobo's end-effector velocity is far from the maximum one. Finally, the Dominance follows the apex movement (M_3).

Concerning the light, following the rewards, after some episodes, Yōkobo gets smoother tran-

Figure 7.10: Screenshot of the animation generated by Blender. The 3D model of Yōkobo is using the second version.

sitions between each colour. In the beginning, the colours changes at each step (see Figure E.3), making it hardly legible for the users. While the learning progresses, the transitions are more natural and the colours are staying close to the previous one (see Figure E.4). One can also mention also that the RL for the light learns to match the colour with Yōkobo's emotion, indeed after 5 episodes more than 200/300 steps match colour and emotions (with a mean of 265 for the 500 episodes).

7.4 Future work and Improvement

As preliminary work, several upgrades could be done to the algorithm to enhance the interactions better. The algorithm has been made for Yōkobo, and some of the simplifications and assumptions are suitable for it since its way of expression is more limited than, for example, an android robot.

First, we focused on the face emotional side of the user only and simplified considerably the user's position and posture. However, this information is useful, since emotion can also be transmitted by body posture, as we saw in chapter 4. To integrate this part, we can take advantage of the review done on pedestrian intention estimation and use similar technologies. One way would be to estimate the future body posture of the user with an RNN algorithm, before sending it to the RL algorithm. Moreover, another block could be added to translate the body posture into emotion and use it in addition to the face one.

Second, we can improve the DBN by changing the particle filtering by the solution developed in [247]. They used instead Gaussian Particle Filtering (GPF) and improved it with kernel smoothing. The GPF can be used for continuous approximation, instead of discrete one.

Third, concerning the PAD-to-colour blocks, we simply assigned a colour to each emotion for the reward as a first try. However, as mentioned by [245], the perception of colour regarding emotion is different regarding culture, and even in the same culture, the perception can be different among people (q.v. the graph in [241] that do not have uniforms answer). Then, instead of using strict mapping, we could use a gradient of colour for each emotion to add more flexibility. This flexibility could be added to the palette of colour as well. We decided to use a palette instead of selecting directly the R, G and B values to avoid non-visible change and to limit the number of possible actions, but we could use a bigger palette to have a better association with the emotion. In addition, we selected only 14 emotions from the PAD values, but [240] gives a list of 151[8].

Fourth, the selection of the different parameters of the algorithms can be tuned to find the best combination. In addition, we could also imagine using the rewards of the RL algorithms to create different personalities for Yōkobo. Indeed, we decided to avoid big changes in the light, but

[8]Even if most of them would not apply to Yōkobo, but we could use a longer list

we could have a more *eccentric* Yōkobo that would not hesitate to be *funky*. On the other hand, we could increase the value of the reward to make it calmer. Those reflections apply also to the motion of Yōkobo.

Last, of course, in addition to simulation, the algorithm has to be tested with Yōkobo. Firstly, with short interactions to test the technical viability of BeWoDa, before doing a long-term experiment with people on daily-basis interaction. The technical test is important to see if the RPi can manage the algorithm, otherwise, it could need some code improvement or translation from Python to C++ in order to have better performance.

Summary

— Conception of an algorithm to create adaptative behaviours for long-term interactions.

— Creation of a DBN to represent the causal relationship of the world's data.

— Development of RL algorithms to command the robot actions.

— **Achievement:** Siméon Capy, Gentiane Venture, and Pongsathorn Raksincharoensak. Pedestrians and cyclists' intention estimation for the purpose of autonomous driving. *International Journal of Automotive Engineering*, 14(1): 10–19, 2023. doi: 10.20485/jsaeijae.14.1_10 [4].

Chapter 8

Conclusion

STUDYING HRI was the centre of this thesis, and since their appearance in the 20th century, robots have been closer and closer to humans. Nowadays, they are common in factories or warehouses, but still rare in homes, and lack social capabilities. With the aim to have robots able to understand and use human social skills, we explored in this thesis some ways to develop robotics systems closer to humans, and by using environmental data to adapt their behaviours. Then, the users will have the feeling that the robot is *part of their world* if it changes its behaviour because, for example, the weather is cold today, or if it becomes more comprehensive regarding the user if this one is looking sad.

We first used a dual-arm industrial robot to do HRI at a fair. This application was the opportunity to try those robots outside their usual environments: factories and laboratories. They are generally isolated from humans for safety reasons, or, if they interact with them, require specific training for the operator. In our project, the general public could interact with the robot without prior training, through their emotions and body movements; and they enjoyed it. One can imagine those kinds of applications in a shop, as a clerk, where the robot can understand if the client is OK with his choice or not. The validation of our scenario in an open-to-public exhibition was essential to have the opinion of people that are not familiar with robots, or who usually do not interact with them.

After this preliminary work, where the interactions are fast and in an unfamiliar environment for the user, we wanted to explore the impact of a robot on a daily basis, at home. Then, we created a presence robot, called Yōkobo. In addition to the body movements of the user, it also uses the weather conditions to change its behaviour, and then it expresses the *state of the house*. By doing so, it becomes more unpredictable for the users and then becomes less annoying with time. Yōkobo was made in the *slow tech* trend and then was to be "*understand*" over time, and not after the first use. Moreover, as mentioned in the literature, only movement can still transmit social cues, then Yōkobo uses solely non-verbal behaviours as means of communication (movement and light). We also developed and experimented with a new concept in HRI called HRHI where the robot is bound between two persons, as a medium for communication. Yōkobo has the aim to be used for a long time in the user's home, on a daily basis. Hence, we proved the technical stability of Yōkobo in our office, before the real experiment in people's homes that is currently taking place in France (second semester of 2022). We also proved that Yōkobo can be used in more environments than just at home.

Researchers generally try their robot in a controlled environment, like a laboratory, and then, they can adapt the parameters to their needs. But, since Yōkobo has to be used at home, it brings some constraints that we do not have in laboratories, and one of them is the configuration. If social robots want to be part of the home, they have to be connected to the Internet, to take advantage of it. Moreover, in the case of smart homes, the robot can use the data of other devices, like Yōkobo, or even command some of them. To be accepted by everyone, it should be simple, and intuitive to use. Then, we proposed a new method to configure the RPi with a smartphone, a device that is really common nowadays. We selected the RPi because it is a common tool for prototyping or building projects at a low-cost, that would require a PC of small size, like a robot. Unfortunately, the native configuration process of the RPi requires computing knowledge. We experimented with our method which was designed to have a minimum of steps, in our laboratory. All of them managed to configure Yōkobo successfully in less time than the configuration timeout, proving the usability of our method.

Finally, to validate the relevance of our projects, we also developed several experimental protocols assembling tools from the literature to create questionnaires. We choose to validate our solutions in the *wild*, with unscripted scenarios, as the general public would use social robots. Indeed, they can have more natural, authentic and spontaneous interactions with the robots in their daily environment (home/work) or without expecting it (in exhibitions), than in a controlled environment by following a specific scenario. We hope that, in addition to the developed solution, our experimental protocols could help future researchers to evaluate the HRI of their projects.

Ultimately, the goal of this thesis is to see how social robots can be improved to adapt their behaviours according to their environment, and how people would perceive them. We summarised the proposed solutions. We could prove their efficiency or usability during our experiments, the participants enjoyed the scenario at IREX, or felt welcomed by Yōkobo. The perception of the users was positive. We demonstrated that social robots can share space with people and be used without prior training from the users, the participants could use their prior knowledge of social interaction. It confirms that the creation of such robots, with capacities for adaptation is a domain that has a promising future, that needs more exploration.

8.1 Limitations

Social interactions are something that takes place on a long time basis, thus we do not become instantaneously friends/enemies with people but after we have learnt to know each other. It is similar to robots, if we want to evaluate how people perceive them, we need time. Even if the experimentation was considered a long-term experiment, two weeks is still short to have a better evaluation of the perception of the participants. It would have been interesting to evaluate Yōkobo over months and check at regular intervals the feeling of each participant. Such an experiment is taking place now, but the results will be known later. It is also a good opportunity to evaluate the acceptance of such robots in the home if people see them more as an intruder or a pet, or even a friend.

Another important point that should be explored is cultural acceptance. During our experiments, we did not have enough diversity in the participants, most of them being Japanese. We can expect that Japanese, French, British, Brazilian... people will not react and perceive the social robots in the same way. We can draw a parallel with human–human interactions, and the way each culture is doing them is different.

8.2 Open Challenges and Future Works

We proposed new scenarios in the field of HRI to enhance them, by having adaptable behaviours. However there is still a lot of work to do, and this was only one stone on the bridge that have been started decades ago. We developed a solution that uses the person's facial emotion to adapt the robot's behaviour, but that is just a first step. Indeed, humans can also transmit social cues with body gestures [248, 249][1], then the robot should be able to understand them and adapt its own behaviour. For example, if a person is angry by his gesture, the robot should try to calm him/her down.

The understanding of the social language by the robot would be one condition of their spread to the general public, people could get frustrated if they have to interact explicitly with them to be understood. Another condition, that is an open challenge, is ethical acceptance. *Do we* really *want social robots wide-spreading? Is it beneficial for society?* Those questions were not the main goal of this thesis but should be taken into account when the robots will become *smart* enough to be used, for example, as a companion.

[1] Like we saw for robots, in chapter 4

List of Acronyms

AP Atmospheric Pressure. 35, 36

B-LSTM Bidirectional LSTM. 68

BeWoDa Behaviours from World Data. viii, 69, 73, 81, 84, 120, 122, 124

BN Bayesian Network. 71, 72

CGAN Conditional Generative Adversarial Nets. 67

CNN Convolutional NN. 67, 68

CPT Conditional Probability Table. 72, 73

CVAE Conditional Variational AutoEncoder. 67

DBN Dynamic Bayesian Network. 67, 68, 71, 73, 83, 84

DeCAF Deep Convolutional Action Feature. 67

DNN Dense NN. 67

DoM Direction of Movement. 67

DPGP Gaussian process mixture model. 67

DQL Deep Q-Learning. xi, 75–78, 81, 120

DSF Driving Safety Field. 68

E1 Experiment 1. 44, 46–49

E2 Experiment 2. 44, 46–49, 52

GRQ General Research Question. 6, 7

GRU Gated Recurrent Unit. 67, 68

HPEA Human Position Estimation Algorithm. 38, 43, 44, 69

HRHI Human–Robot–Human Interaction. 31–33, 85

HRI Human–Robot Interaction. ix, 1, 4–6, 8, 9, 31, 33, 44, 64, 85, 86

IOS Inclusion of Others in Self. vi, ix, 11–13

IPM Inverse Perspective Mapping. 68

LDCRF Latent-Dynamic Conditional Random Fields. 68

LSTM Long Short Term Memory. 67

MC Motor Control. 44

MLP Multi-Layer Perceptron. 67

MVC Model-View-Controller. x, 59, 60

NARS Negative Attitude towards Robots Scale. vi, ix, 10, 11, 13, 22

NN Neural Network. x, 75, 76

NVB Non-Verbal Behaviour. 33, 34

PAD Pleasure-Arousal-Dominance. 11, 12

PCB Printed Circuit Board. 37

RA Robot Assistant. 32

RANSAC RANdom Sampling And Consensus. 67

RF Random Forest. 68

RFID Radio Frequency IDentification. 37

RL Reinforcement Learning. 65, 70, 74, 75, 83, 84

RNN Recurrent NN. 67, 68

SAM Self-Assessment Manikin. vi, 11–13

SLDS Switching Linear Dynamical System. 67

SP Selected Participant. 45, 46

SR Social Robot. 3–6

SUS System Usability Scale. vi, viii, 13, 47, 106

SVM Support Vector Machine. 67

TLS Traffic Light State. 67

TPB Theory of Planned Behaviour. 67

US Ultrasonic Sensor. 39

VA Vocal Assistant. 4, 6, 32

Bibliography

[1] Siméon Capy, Liz Rincon, Enrique Coronado, Shohei Hagane, Seiji Yamaguchi, Victor Leve, Yuichiro Kawasumi, Yasutoshi Kudou, and Gentiane Venture. Expanding the frontiers of industrial robots beyond factories: Design and in the wild validation. *Machines*, 10(12), 2022. ISSN 2075-1702. doi: 10.3390/machines10121179. URL `https://www.mdpi.com/2075-1702/10/12/1179`.

[2] Siméon Capy, Pablo Osorio, Shohei Hagane, Corentin Aznar, Dora Garcin, Enrique Coronado, Dominique Deuff, Ioana Ocnarescu, Isabelle Milleville, and Gentiane Venture. Yōkobo: A robot to strengthen links amongst users with non-verbal behaviours. *Machines*, 10(8), 2022. ISSN 2075-1702. doi: 10.3390/machines10080708. URL `https://www.mdpi.com/2075-1702/10/8/708`.

[3] Siméon Capy, Enrique Coronado, Pablo Osorio, Shohei Hagane, Dominique Deuff, and Gentiane Venture. Integration of a presence robot in a smart home. *International Conference on Computer, Control and Robotics (ICCCR) 2023, Shanghai, China*, (accepted).

[4] Siméon Capy, Gentiane Venture, and Pongsathorn Raksincharoensak. Pedestrians and cyclists' intention estimation for the purpose of autonomous driving. *International Journal of Automotive Engineering*, 14(1):10–19, 2023. doi: 10.20485/jsaeijae.14.1_10.

[5] Dominique Deuff, Isabelle Milleville-Pennel, Ioana Ocnarescu, Dora Garcin, Corentin Aznar, Siméon Capy, Shohei Hagane, Pablo Felipe Osorio Marin, Enrique Coronado Zuniga, Liz Rincon Ardila, and Gentiane Venture. Together alone, yōkobo, a sensible presence robject for the home of newly retired couples. In *Designing Interactive Systems Conference*, DIS '22, page 1773–1787, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450393584. doi: 10.1145/3532106.3533485.

[6] Siméon Capy. Social robotics: Human-robot interaction using eeg signals and head motion. Master's thesis, Institut National des Sciences Appliquées de Strasbourg, 2015.

[7] Francesco Grillo. *Hero of Alexandria's Automata: a critical edition and translation, including a commentary on Book One.* PhD thesis, University of Glasgow, 2019.

[8] Philip Steadman. *The automata of Hero of Alexandria*, chapter 3, pages 111–132. UCL Press, 2021. ISBN 9781787359161. URL `http://www.jstor.org/stable/j.ctv18msqmt.10`.

[9] Adelheid Voskuhl. Producing objects, producing texts: accounts of android automata in late eighteenth-century europe. *Studies in History and Philosophy of Science Part A*, 38(2):422–444, 2007. ISSN 0039-3681. doi: https://doi.org/10.1016/j.shpsa.2007.03.004. URL `https://www.sciencedirect.com/science/article/pii/S0039368107000192`. Objects, texts and images in the history of science.

[10] Elizabeth Stephens. The mechanical turk: a short history of 'artificial artificial intelligence'. *Cultural Studies*, pages 1–23, 2022.

[11] Ylva Fernaeus, Martin Jonsson, and Jakob Tholander. Revisiting the jacquard loom: Threads of history and current patterns in hci. In *Proceedings of the SIGCHI Conference on Human*

Factors in Computing Systems, CHI '12, page 1593–1602, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450310154. doi: 10.1145/2207676.2208280. URL https://doi.org/10.1145/2207676.2208280.

[12] Allan G Bromley. Charles babbage's analytical engine, 1838. *Annals of the History of Computing*, 4(3):196–217, 1982.

[13] Anna Henschel, Guy Laban, and Emily S. Cross. What makes a robot social? a review of social robots from science fiction to a home or hospital near you. *Current Robotics Reports*, 2(1):9–19, Mar 2021. ISSN 2662-4087. doi: 10.1007/s43154-020-00035-0. URL https://doi.org/10.1007/s43154-020-00035-0.

[14] ShuzhiSam Ge and MajaJ. Matarić. Preface. *International Journal of Social Robotics*, 1(1): 1–2, 2009. ISSN 1875-4791. doi: 10.1007/s12369-008-0010-2. URL http://dx.doi.org/10.1007/s12369-008-0010-2.

[15] Justinas Mišeikis, Pietro Caroni, Patricia Duchamp, Alina Gasser, Rastislav Marko, Nelija Mišeikienė, Frederik Zwilling, Charles de Castelbajac, Lucas Eicher, Michael Früh, et al. Lio-a personal robot assistant for human-robot interaction and care applications. *IEEE Robot. and Autom. Lett.*, 5(4):5339–5346, 2020.

[16] Masahiro Fujita. Aibo: Toward the era of digital creatures. *The Int. J. of Robotics Research*, 20(10):781–794, 2001. doi: 10.1177/02783640122068092.

[17] Intuition Robotics. URL https://elliq.com.

[18] Heather Knight. Eight lessons learned about non-verbal interactions through robot theater. In *Int. Conf. on Social Robotics*, pages 42–51. Springer, 2011.

[19] Gentiane Venture and Dana Kulić. Robot expressive motions: a survey of generation and evaluation methods. *ACM Transactions on Human-Robot Interaction (THRI)*, 8(4):1–17, 2019.

[20] RoboCup. Robocup standard platform league, 1999. URL https://spl.robocup.org/history/.

[21] Toshiyo Tamura, Satomi Yonemitsu, Akiko Itoh, Daisuke Oikawa, Akiko Kawakami, Yuji Higashi, Toshiro Fujimooto, and Kazuki Nakajima. Is an Entertainment Robot Useful in the Care of Elderly People With Severe Dementia? *The Journals of Gerontology: Series A*, 59(1):M83–M85, 01 2004. ISSN 1079-5006. doi: 10.1093/gerona/59.1.M83. URL https://doi.org/10.1093/gerona/59.1.M83.

[22] Takanori Shibata and Kazuyoshi Wada. Robot therapy: a new approach for mental healthcare of the elderly–a mini-review. *Gerontology*, 57(4):378–386, 2011.

[23] Markus Kolstad, Natsu Yamaguchi, Ankica Babic, and Yoko Nishihara. Integrating socially assistive robots into japanese nursing care. *Studies in health technology and informatics*, 270: 1323–1324, 2020.

[24] A. K. Pandey and R. Gelin. A mass-produced sociable humanoid robot: Pepper: The first machine of its kind. 25(3):40–48, 2018. doi: 10.1109/MRA.2018.2833157.

[25] Elena Knox and Katsumi Watanabe. Aibo robot mortuary rites in the japanese cultural context. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2020–2025, 2018. doi: 10.1109/IROS.2018.8594066.

[26] Astrid Weiss, Daniela Wurhofer, and Manfred Tscheligi. "i love this dog"—children's emotional attachment to the robotic dog aibo. *International Journal of Social Robotics*, 1 (3):243–248, Aug 2009. ISSN 1875-4805. doi: 10.1007/s12369-009-0024-4. URL https://doi.org/10.1007/s12369-009-0024-4.

[27] Enrique Coronado, Xela Indurkhya, and Gentiane Venture. Robots meet children, development of semi-autonomous control systems for children-robot interaction in the wild. In *2019 IEEE 4th International Conference on Advanced Robotics and Mechatronics (ICARM)*, pages 360–365, 2019. doi: 10.1109/ICARM.2019.8833751.

[28] Gentiane Venture, Bipin Indurkhya, and Takamune Izui. Dance with me! child-robot interaction in the wild. In Abderrahmane Kheddar, Eiichi Yoshida, Shuzhi Sam Ge, Kenji Suzuki, John-John Cabibihan, Friederike Eyssel, and Hongsheng He, editors, *Social Robotics*, pages 375–382, Cham, 2017. Springer International Publishing. ISBN 978-3-319-70022-9.

[29] I. Npochinto Moumeni and F. Mourey. Intérêt en ehpad du robot émotionnel pepper dans les troubles neurocomportementaux de la maladie d'alzheimer. *NPG Neurologie - Psychiatrie - Gériatrie*, 21(121):11–18, 2021. ISSN 1627-4830. doi: https://doi.org/10.1016/j.npg.2020.09.005. URL `https://www.sciencedirect.com/science/article/pii/S1627483020301598`.

[30] SoftBank Robotics. URL `https://www.softbankrobotics.com/emea/en/nao`.

[31] Mustapha Kessous. Portrait-robot de... jean-mi. *Le Monde*. URL `https://www.lemonde.fr/televisions-radio/article/2014/10/24/portrait-robot-de-jean-mi_4512201_1655027.html`.

[32] Syamimi Shamsuddin, Hanafiah Yussof, Luthffi Ismail, Fazah Akhtar Hanapiah, Salina Mohamed, Hanizah Ali Piah, and Nur Ismarrubie Zahari. Initial response of autistic children in human-robot interaction therapy with humanoid robot nao. In *2012 IEEE 8th International Colloquium on Signal Processing and its Applications*, pages 188–193, 2012. doi: 10.1109/CSPA.2012.6194716.

[33] Adriana Tapus, Andreea Peca, Amir Aly, Cristina Pop, Lavinia Jisa, Sebastian Pintea, Alina S Rusu, and Daniel O David. Children with autism social engagement in interaction with nao, an imitative robot: A series of single case experiments. *Interaction studies*, 13(3):315–347, 2012.

[34] Stefan Vikkelsø, Tuyet-Hoa Hoang, Fransine Carrara, Karl Damkjær Hansen, and Birthe Dinesen. The telepresence avatar robot orihime as a communication tool for adults with acquired brain injury: an ethnographic case study. *Intelligent Service Robotics*, 13(4):521–537, Oct 2020. ISSN 1861-2784. doi: 10.1007/s11370-020-00335-6. URL `https://doi.org/10.1007/s11370-020-00335-6`.

[35] Double Robotics, 2012. URL `https://www.doublerobotics.com`.

[36] Pablo Osorio and Gentiane Venture. Control of a robot expressive movements using non-verbal features. 2022.

[37] Akiyoshi Hayashi, Liz Katherine Rincon-Ardila, and Gentiane Venture. Improving hri with force sensing. *Machines*, 10(1), 2022. ISSN 2075-1702. doi: 10.3390/machines10010015. URL `https://www.mdpi.com/2075-1702/10/1/15`.

[38] Matteo Spezialetti, Giuseppe Placidi, and Silvia Rossi. Emotion recognition for human-robot interaction: Recent advances and future perspectives. *Frontiers in Robotics and AI*, 7, 2020. ISSN 2296-9144. doi: 10.3389/frobt.2020.532279. URL `https://www.frontiersin.org/articles/10.3389/frobt.2020.532279`.

[39] Iolanda Leite, Carlos Martinho, and Ana Paiva. Social robots for long-term interaction: A survey. *International Journal of Social Robotics*, 5(2):291–308, Apr 2013. ISSN 1875-4805. doi: 10.1007/s12369-013-0178-y. URL `https://doi.org/10.1007/s12369-013-0178-y`.

[40] GROOVE X. URL `https://lovot.life/en/`.

[41] NTT Disruption. URL `https://jibo.com`.

[42] Kwangmin Jeong, Jihyun Sung, Hae-Sung Lee, Aram Kim, Hyemi Kim, Chanmi Park, Yuin Jeong, JeeHang Lee, and Jinwoo Kim. Fribo: A social networking robot for increasing social connectedness through sharing daily home activities from living noise data. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '18, page 114–122, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450349536. doi: 10.1145/3171221.3171254. URL https://doi.org/10.1145/3171221.3171254.

[43] Violet. URL https://www.nabaztag.com.

[44] Naoto Yoshida, Shuto Yonemura, Masahiro Emoto, Kanji Kawai, Naoki Numaguchi, Hiroki Nakazato, Shunsuke Otsubo, Megumi Takada, and Kaname Hayashi. Production of character animation in a home robot: A case study of lovot. *International Journal of Social Robotics*, 14(1):39–54, Jan 2022. ISSN 1875-4805. doi: 10.1007/s12369-021-00746-0. URL https://doi.org/10.1007/s12369-021-00746-0.

[45] Birthe Dinesen, Helle Kidde Hansen, Gry Bruun Grønborg, Anne-Kirstine Dyrvig, Sofie Dalskov Leisted, Henrik Stenstrup, Cathrine Skov Schacksen, Claus Oestergaard, et al. Use of a social robot (lovot) for persons with dementia: Exploratory study. *JMIR Rehabilitation and Assistive Technologies*, 9(3):e36505, 2022.

[46] Matthew Rueben, Shirley A. Elprama, Dimitrios Chrysostomou, and An Jacobs. *Introduction to (Re)Using Questionnaires in Human-Robot Interaction Research*, pages 125–144. Springer International Publishing, Cham, 2020. ISBN 978-3-030-42307-0. doi: 10.1007/978-3-030-42307-0_5. URL https://doi.org/10.1007/978-3-030-42307-0_5.

[47] Benjamin Knopp, Dmytro Velychko, Johannes Dreibrodt, and Dominik Endres. Predicting perceived naturalness of human animations based on generative movement primitive models. *ACM Trans. Appl. Percept.*, 16(3), sep 2019. ISSN 1544-3558. doi: 10.1145/3355401. URL https://doi.org/10.1145/3355401.

[48] Miskatyas Putri Aransih and Rizki Edmi Edison. The naturalness of biological movement by individuals with autism spectrum conditions: Taking neurotypical individuals' viewpoint. *Open Access Macedonian Journal of Medical Sciences*, 7(16):2574, 2019.

[49] Christina Lichtenthäler and Alexandra Kirsch. Legibility of Robot Behavior : A Literature Review. working paper or preprint, April 2016. URL https://hal.archives-ouvertes.fr/hal-01306977.

[50] B. de Gelder, A.W. de Borst, and R. Watson. The perception of emotion in body expressions. *WIREs Cognitive Science*, 6(2):149–158, 2015. doi: https://doi.org/10.1002/wcs.1335. URL https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcs.1335.

[51] Shouyo Ishida. Robot arm's expressive movement generation using optimal control with time-varying weights. Master's thesis, Tokyo University of Agriculture and Technology, 2022.

[52] Shouyo Ishida, Tatsuki Harada, Pamela Carreno-Medrano, Dana Kulić, and Gentiane Venture. Human motion imitation using optimal control with time-varying weights. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 608–615. IEEE, 2021.

[53] Ankur Joshi, Saket Kale, Satish Chandel, and D Kumar Pal. Likert scale: Explored and explained. *British journal of applied science & technology*, 7(4):396, 2015.

[54] Rensis Likert. A technique for the measurement of attitudes. *Archives of psychology*, 1932.

[55] Hagen Lehmann, Joan Saez-Pons, Dag Sverre Syrdal, and Kerstin Dautenhahn. In good company? perception of movement synchrony of a non-anthropomorphic robot. *PLOS ONE*, 10(5):1–16, 05 2015. doi: 10.1371/journal.pone.0127747. URL https://doi.org/10.1371/journal.pone.0127747.

[56] Mitsuo Nagamachi and Anitawati Mohd Lokman. *Innovations of Kansei engineering.* CRC Press, 2016.

[57] Mitsuo Nagamachi. *Kansei/affective engineering.* crc press, 2016.

[58] Anca D. Dragan, Kenton C.T. Lee, and Siddhartha S. Srinivasa. Legibility and predictability of robot motion. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 301–308, 2013. doi: 10.1109/HRI.2013.6483603.

[59] Tatsuya Nomura, Takayuki Kanda, Tomohiro Suzuki, and Kennsuke Kato. Psychology in human-robot communication: An attempt through investigation of negative attitudes and anxiety toward robots. In *RO-MAN 2004. 13th IEEE international workshop on robot and human interactive communication (IEEE catalog No. 04TH8759)*, pages 35–40. IEEE, 2004.

[60] Tatsuya Nomura, Tomohiro Suzuki, Takayuki Kanda, and Kensuke Kato. Measurement of negative attitudes toward robots. *Interaction Studies*, 7(3):437–454, 2006.

[61] Stanislava Naneva, Marina Sarda Gou, Thomas L Webb, and Tony J Prescott. A systematic review of attitudes, anxiety, acceptance, and trust towards social robots. *International Journal of Social Robotics*, 12:1179–1201, 2020.

[62] Tatsuya Nomura, Tomohiro Suzuki, Takayuki Kanda, and Kensuke Kato. Altered attitudes of people toward robots: Investigation through the negative attitudes toward robots scale. In *Proc. AAAI-06 workshop on human implications of human-robot interaction*, volume 2006, pages 29–35, 2006.

[63] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1(1):71–81, Jan 2009. ISSN 1875-4805. doi: 10.1007/s12369-008-0001-3. URL `https://doi.org/10.1007/s12369-008-0001-3`.

[64] Hiroko Kamide, Yasushi Mae, Koji Kawabe, Satoshi Shigemi, Masato Hirose, and Tatsuo Arai. New measurement of psychological safety for humanoid. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 49–56. IEEE, 2012.

[65] Yue Hu, Mehdi Benallegue, Gentiane Venture, and Eiichi Yoshida. Interact with me: An exploratory study on interaction factors for active physical human-robot interaction. *IEEE Robotics and Automation Letters*, 5(4):6764–6771, 2020.

[66] Michelle Karg, Ali-Akbar Samadani, Rob Gorbet, Kolja Kühnlenz, Jesse Hoey, and Dana Kulić. Body movements for affective expression: A survey of automatic recognition and generation. *IEEE Transactions on Affective Computing*, 4(4):341–359, 2013. doi: 10.1109/T-AFFC.2013.29.

[67] Margaret M Bradley and Peter J Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25 (1):49–59, 1994.

[68] Lewis R Goldberg. An alternative" description of personality": the big-five factor structure. *Journal of personality and social psychology*, 59(6):1216, 1990.

[69] J Brooke. Usability evaluation in industry, chap. sus: a "quick and dirty" usability scale, 1996.

[70] John Brooke. Sus: a retrospective. *J. of usability studies*, 8(2):29–40, 2013.

[71] S Camille Peres, Tri Pham, and Ronald Phillips. Validation of the system usability scale (sus) sus in the wild. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 57, pages 192–196. SAGE Publications Sage CA: Los Angeles, CA, 2013.

[72] Antonio Martinez-Molina, Paola Boarin, Isabel Tort-Ausina, and José-Luis Vivancos. Assessing visitors' thermal comfort in historic museum buildings: Results from a post-occupancy evaluation on a case study. *Building and Environment*, 132:291–302, 2018.

[73] Brendan Burchell and Catherine Marsh. The effect of questionnaire length on survey response. *Quality and quantity*, 26(3):233–244, 1992.

[74] Renier Steyn. How many items are too many? an analysis of respondent disengagement when completing questionnaires. 2017.

[75] Jon A Krosnick. Questionnaire design. In *The Palgrave handbook of survey research*, pages 439–455. Springer, 2018.

[76] Kourosh Darvish, Francesco Wanderlingh, Barbara Bruno, Enrico Simetti, Fulvio Mastrogiovanni, and Giuseppe Casalino. Flexible human–robot cooperation models for assisted shop-floor tasks. *Mechatronics*, 51:97–114, 2018.

[77] Sarah L Müller-Abdelrazeq, Kathrin Schönefeld, Max Haberstroh, and Frank Hees. Interacting with collaborative robots—a study on attitudes and acceptance in industrial contexts. In *Social Robots: Technological, Societal and Ethical Aspects of Human-Robot Interaction*, pages 101–117. Springer, 2019.

[78] Stanislava Naneva, Marina Sarda Gou, Thomas L Webb, and Tony J Prescott. A systematic review of attitudes, anxiety, acceptance, and trust towards social robots. 2020.

[79] Thomas M Wendt, Urban B Himmelsbach, Matthias Lai, and Matthias Waßmer. Time-of-flight cameras enabling collaborative robots for improved safety in medical applications. *International Journal of Interdisciplinary Telecommunications and Networking (IJITN)*, 9 (4):10–17, 2017.

[80] Tal Oron-Gilad and Peter A Hancock. From ergonomics to hedonomics: Trends in human factors and technology—the role of hedonomics revisited. In *Emotions and affect in human factors and human-computer interaction*, pages 185–194. Elsevier, 2017.

[81] Mayumi Fukuyama. Society 5.0: Aiming for a new human-centered society. *Japan Spotlight*, 27:47–50, 2018.

[82] Bzhwen A Kadir, Ole Broberg, and Carolina Souza da Conceicao. Current research and future perspectives on human factors and ergonomics in industry 4.0. *Computers & Industrial Engineering*, 137:106004, 2019.

[83] Tyge-F Kummer, Kerstin Schäfer, and Neda Todorova. Acceptance of hospital nurses toward sensor-based medication systems: a questionnaire survey. *International journal of nursing studies*, 50(4):508–517, 2013.

[84] S Shyam Sundar, T Franklin Waddell, and Eun Hwa Jung. The hollywood robot syndrome media effects on older adults' attitudes toward robots and adoption intentions. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 343–350. IEEE, 2016.

[85] Jenay M Beer, Akanksha Prakash, Tracy L Mitzner, and Wendy A Rogers. Understanding robot acceptance. Technical report, Georgia Institute of Technology, 2011.

[86] Christina Bröhl, Jochen Nelles, Christopher Brandl, Alexander Mertens, and Christopher M Schlick. Tam reloaded: a technology acceptance model for human-robot cooperation in production systems. In *International conference on human-computer interaction*, pages 97–103. Springer, 2016.

[87] Malte Jung and Pamela Hinds. Robots in the wild: A time for more robust theories of human-robot interaction. *ACM Transactions on Human-Robot Interaction (THRI)*, 7(1):2, 2018.

[88] Selma Šabanović, S Reeder, and Bobak Kechavarzi. Designing robots in the wild: In situ prototype evaluation for a break management robot. *Journal of Human-Robot Interaction*, 3(1):70–88, 2014.

[89] Mary Ellen Foster, Rachid Alami, Olli Gestranius, Oliver Lemon, Marketta Niemelä, Jean-Marc Odobez, and Amit Kumar Pandey. The mummer project: Engaging human-robot interaction in real-world public spaces. In *International Conference on Social Robotics*, pages 753–763. Springer, 2016.

[90] Paul Bremner, Miriam Koschate, and Mark Levine. Humanoid robot avatars: an 'in the wild'usability study. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 624–629. IEEE, 2016.

[91] Elin A Björling, Kyle Thomas, Emma J Rose, and Maya Cakmak. Exploring teens as robot operators, users and witnesses in the wild. *Frontiers in Robotics and AI*, 7:5, 2020.

[92] Paulina Zguda, Anna Kołota, Gentiane Venture, Bartlomiej Sniezynski, and Bipin Indurkhya. Exploring the role of trust and expectations in cri using in-the-wild studies. *Electronics*, 10 (3):347, 2021.

[93] Kawada Robotics. `http://nextage.kawada.jp/en/open/`.

[94] Tatsuya T Nomura, Dag Sverre Syrdal, and Kerstin Dautenhahn. Differences on social acceptance of humanoid robots between japan and the uk. In *Procs 4th int symposium on new frontiers in human-robot interaction*. The Society for the Study of Artificial Intelligence and the Simulation of ..., 2015.

[95] Kerstin Sophie Haring, Céline Mougenot, Fuminori Ono, and Katsumi Watanabe. Cultural differences in perception and attitude towards robots. *International Journal of Affective Engineering*, 13(3):149–157, 2014.

[96] Sonia Brondi, Monica Pivetti, Silvia Di Battista, and Mauro Sarrica. What do we expect from robots? social representations, attitudes and evaluations of robots in daily life. *Technology in Society*, 66:101663, 2021.

[97] Lizzy Sinnema and Maryam Alimardani. The attitude of elderly and young adults towards a humanoid robot as a facilitator for social interaction. In *International Conference on Social Robotics*, pages 24–33. Springer, 2019.

[98] Matthijs HJ Smakman, Elly A Konijn, Paul Vogt, and Paulina Pankowska. Attitudes towards social robots in education: Enthusiast, practical, troubled, sceptic, and mindfully positive. *Robotics*, 10(1):24, 2021.

[99] Shu-Chuan Chen, Cindy Jones, and Wendy Moyle. Health professional and workers attitudes towards the use of social robots for older adults in long-term care. *International Journal of Social Robotics*, pages 1–13, 2019.

[100] Shirley Adriane Elprama, Charlotte IC Jewell, An Jacobs, Ilias El Makrini, and Bram Vanderborght. Attitudes of factory workers towards industrial and collaborative robots. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 113–114, 2017.

[101] Iina Aaltonen and Timo Salmi. Experiences and expectations of collaborative robots in industry and academia: barriers and development needs. *Procedia Manufacturing*, 38:1151–1158, 2019.

[102] BVSA Elprama, Ilias El Makrini, and A Jacobs. Acceptance of collaborative robots by factory workers: a pilot study on the importance of social cues of anthropomorphic robots. In *International Symposium on Robot and Human Interactive Communication*, 2016.

[103] Min-Gyu Kim, Jaeryoung Lee, Yuusuke Aichi, Hiroki Morishita, and Munehiro Makino. Effectiveness of robot exhibition through visitors experience: A case study of nagoya science hiroba exhibition in japan. In *2016 International Symposium on Micro-NanoMechatronics and Human Science (MHS)*, pages 1–5. IEEE, 2016.

[104] Denis Vidal and Philippe Gaussier. Visitor or artefact! an experiment with a humanoid robot at the musée du quai branly in paris. In *Wording Robotics*, pages 101–117. Springer, 2019.

[105] Daphne Karreman, Geke Ludden, and Vanessa Evers. Visiting cultural heritage with a tour guide robot: A user evaluation study in-the-wild. In *International Conference on Social Robotics*, pages 317–326. Springer, 2015.

[106] Michiel Joosse and Vanessa Evers. A guide robot at the airport: First impressions. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 149–150, 2017.

[107] Takayuki Kanda, Masahiro Shiomi, Zenta Miyashita, Hiroshi Ishiguro, and Norihiro Hagita. A communication robot in a shopping mall. *IEEE Transactions on Robotics*, 26(5):897–913, 2010. doi: 10.1109/TRO.2010.2062550.

[108] Sandra Drolshagen, Max Pfingsthorn, Pascal Gliesche, and Andreas Hein. Acceptance of industrial collaborative robots by people with disabilities in sheltered workshops. *Frontiers in Robotics and AI*, 7:173, 2021.

[109] Chiara Rossato, Patrik Pluchino, Nicola Cellini, Giulio Jacucci, Anna Spagnolli, and Luciano Gamberini. Facing with collaborative robots: The subjective experience in senior and younger workers. *Cyberpsychology, Behavior, and Social Networking*, 24(5):349–356, 2021.

[110] Enrique Coronado and Gentiane Venture. Towards iot-aided human–robot interaction using nep and ros: A platform-independent, accessible and distributed approach. *Sensors*, 20(5): 1500, 2020.

[111] Yuyuan Shi, Yin Chen, Liz Rincon Ardila, Gentiane Venture, and Marie-Luce Bourguet. A visual sensing platform for robot teachers. pages 200–201, 09 2019. ISBN 978-1-4503-6922-0. doi: 10.1145/3349537.3352764.

[112] Joseph A Gliem and Rosemary R Gliem. Calculating, interpreting, and reporting cronbach's alpha reliability coefficient for likert-type scales. Midwest Research-to-Practice Conference in Adult, Continuing, and Community . . . , 2003.

[113] Aimee L Drolet and Donald G Morrison. Do we really need multiple-item measures in service research? *Journal of service research*, 3(3):196–204, 2001.

[114] Adamantios Diamantopoulos, Marko Sarstedt, Christoph Fuchs, Petra Wilczynski, and Sebastian Kaiser. Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. *Journal of the Academy of Marketing Science*, 40(3):434–449, 2012.

[115] Lars Bergkvist. Appropriate use of single-item measures is here to stay. *Marketing letters*, 26(3):245–255, 2015.

[116] Larry Freed. *Innovating Analytics: How the next generation of net promoter can increase sales and drive business results.* John Wiley & Sons, 2013.

[117] Jeff Sauro. Is a single item enough to measure a construct? `https://measuringu.com/single-multi-items/`, 2018.

[118] Hui Jiang and Lin Cheng. Public perception and reception of robotic applications in public health emergencies based on a questionnaire survey conducted during covid-19. *International Journal of Environmental Research and Public Health*, 18(20), 2021. ISSN 1660-4601. doi: 10.3390/ijerph182010908. URL `https://www.mdpi.com/1660-4601/18/20/10908`.

[119] Christina Bröhl, Jochen Nelles, Christopher Brandl, Alexander Mertens, and Verena Nitsch. Human–robot collaboration acceptance model: Development and comparison for germany, japan, china and the usa. *International Journal of Social Robotics*, 11(5):709–726, Dec 2019. ISSN 1875-4805. doi: 10.1007/s12369-019-00593-0. URL https://doi.org/10.1007/s12369-019-00593-0.

[120] Yuhua Liang and Seungcheol Austin Lee. Fear of autonomous robots and artificial intelligence: Evidence from national representative data with probability sampling. *International Journal of Social Robotics*, 9(3):379–384, Jun 2017. ISSN 1875-4805. doi: 10.1007/s12369-017-0401-3. URL https://doi.org/10.1007/s12369-017-0401-3.

[121] Tatsuya Nomura. Cultural differences in social acceptance of robots. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 534–538, 2017. doi: 10.1109/ROMAN.2017.8172354.

[122] Jaime Romero and Nora Lado. Service robots and covid-19: exploring perceptions of prevention efficacy at hotels in generation z. *International Journal of Contemporary Hospitality Management*, 33(11):4057–4078, Jan 2021. ISSN 0959-6119. doi: 10.1108/IJCHM-10-2020-1214. URL https://doi.org/10.1108/IJCHM-10-2020-1214.

[123] Guy Hoffman and Wendy Ju. Designing robots with movement in mind. *J. of Human-Robot Interaction*, 3(1):91–122, 2014.

[124] Enrique Coronado, Gentiane Venture, and Natsuki Yamanobe. Applying kansei/affective engineering methodologies in the design of social and service robots: A systematic review. *Int. J. of Social Robotics*, pages 1–11, 10 2020. doi: 10.1007/s12369-020-00709-x.

[125] Martin G Helander and Halimahtun M Khalid. Underlying theories of hedonomics for affective and pleasurable design. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 49, pages 1691–1695. SAGE Publications Sage CA: Los Angeles, CA, 2005.

[126] Carlos Eduardo Díaz, Roemi Fernández, Manuel Armada, and Felipe García. A research review on clinical needs, technical requirements, and normativity in the design of surgical robots. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 13 (4):e1801, 2017.

[127] Sebastian Glende, Isabel Conrad, Livia Krezdorn, Susann Klemcke, and Carola Krätzel. Increasing the acceptance of assistive robots for older people through marketing strategies based on stakeholder needs. *International Journal of Social Robotics*, 8(3):355–369, 2016.

[128] Laura Alzola Kirschgens, Irati Zamalloa Ugarte, Endika Gil Uriarte, Aday Muniz Rosas, and Víctor Mayoral Vilches. Robot hazards: from safety to security. *arXiv preprint arXiv:1806.06681*, 2018.

[129] Grandys Frieska Prassida and Ully Asfari. A conceptual model for the acceptance of collaborative robots in industry 5.0. *Procedia Computer Science*, 197:61–67, 2022. ISSN 1877-0509. doi: https://doi.org/10.1016/j.procs.2021.12.118. URL https://www.sciencedirect.com/science/article/pii/S1877050921023425. Sixth Information Systems International Conference (ISICO 2021).

[130] William T. Odom, Abigail J. Sellen, Richard Banks, David S. Kirk, Tim Regan, Mark Selby, Jodi L. Forlizzi, and John Zimmerman. Designing for slowness, anticipation and revisitation: A long term field study of the photobox. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, CHI '14, page 1961–1970, New York, NY, USA, 2014. ISBN 9781450324731. doi: 10.1145/2556288.2557178.

[131] Randy Gomez, Deborah Szapiro, Kerl Galindo, and Keisuke Nakamura. Haru: Hardware design of an experimental tabletop robot assistant. In *the ACM/IEEE Int. Conf. on human-robot interaction*, pages 233–240, 2018.

[132] Michal Luria, Guy Hoffman, and Oren Zuckerman. Comparing social robot, screen and voice interfaces for smart-home control. In *the Conf. on human factors in computing systems*, pages 580–628, 2017.

[133] Anastasia K. Ostrowski, Vasiliki Zygouras, Hae Won Park, and Cynthia Breazeal. Small group interactions with voice-user interfaces: Exploring social embodiment, rapport, and engagement. In *the ACM/IEEE Int. Conf. on Human-Robot Interaction*, page 322–331, 2021. ISBN 9781450382892. doi: 10.1145/3434073.3444655.

[134] Florian Vaussard, Michael Bonani, Philippe Rétornaz, Alcherio Martinoli, and Francesco Mondada. Towards autonomous energy-wise robjects. In *Conf. Towards Auton. Robotic Sys.*, pages 311–322. Springer, 2011.

[135] L. Anderson-Bashan, B. Megidish, H. Erel, I. Wald, G. Hoffman, O. Zuckerman, and A. Grishko. The greeting machine: An abstract robotic object for opening encounters. In *IEEE Int. Symp. on Robot and Human Interactive Communication*, pages 595–602, 2018. doi: 10.1109/ROMAN.2018.8525516.

[136] Gabriele Trovato, Massimiliano Zecca, Salvatore Sessa, Lorenzo Jamone, Jaap Ham, Kenji Hashimoto, and Atsuo Takanishi. Cross-cultural study on human-robot greeting interaction: acceptance and discomfort by egyptians and japanese. *Paladyn, J. of Behavioral Robotics*, 4 (2):83–93, 2013.

[137] Brandon Heenan, Saul Greenberg, Setareh Aghel-Manesh, and Ehud Sharlin. Designing social greetings in human robot interaction. In *Conf. on Designing interactive systems*, pages 855–864, 2014.

[138] Hedwig Anna Theresia Broers, Jaap Ham, Ron Broeders, P. Ravindra de Silva, and Michio Okada. Goal inferences about robot behavior: Goal inferences and human response behaviors. In *ACM/IEEE Int. Conf. on Human-Robot Interaction*, page 91–92. IEEE Press, 2013. ISBN 9781467330558.

[139] Francesco Mondada, Julia Fink, Séverin Lemaignan, David Mansolino, Florian Wille, and Karmen Franinović. Ranger, an example of integration of robotics into the home ecosystem. In *New Trends in Medical and Service Robots*, pages 181–189. Springer, 2016.

[140] Hadas Erel, Yoav Cohen, Klil Shafrir, Sara Daniela Levy, Idan Dov Vidra, Tzachi Shem Tov, and Oren Zuckerman. Excluded by robots: Can robot-robot-human interaction lead to ostracism? In *the ACM/IEEE Int. Conf. on Human-Robot Interaction*, page 312–321, 2021. ISBN 9781450382892. doi: 10.1145/3434073.3444648.

[141] Heike Brock, Selma Šabanović, and Randy Gomez. Remote you, haru and me: Exploring social interaction in telepresence gaming with a robotic agent. In *the ACM/IEEE Int. Conf. on Human-Robot Interaction*, page 283–287, 2021. ISBN 9781450382908. doi: 10.1145/3434074.3447177.

[142] Daniel J Ricks and Mark B Colton. Trends and considerations in robot-assisted autism therapy. In *Int. Conf. on robotics and automation*, pages 4354–4359. IEEE, 2010.

[143] Danielle Rifinski, Hadas Erel, Adi Feiner, Guy Hoffman, and Oren Zuckerman. Human-human-robot interaction: robotic object's responsive gestures improve interpersonal evaluation in human interaction. *Human-Computer Interaction*, pages 1–27, 02 2020. doi: 10.1080/07370024.2020.1719839.

[144] Riccardo Campa. The rise of social robots: a review of the recent literature. *J. of Evolution and Technology*, 26(1), 2016.

[145] Valentina Ramírez, Dominique Deuff, Xela Indurkhya, and Gentiane Venture. Design space survey on social robotics in the market. *Journal of Intelligent & Robotic Systems*, 105(2):25, May 2022. ISSN 1573-0409. doi: 10.1007/s10846-022-01624-1. URL `https://doi.org/10.1007/s10846-022-01624-1`.

[146] M. Mori, K. F. MacDorman, and N. Kageki. The uncanny valley [from the field]. 19(2): 98–100, 2012. doi: 10.1109/MRA.2012.2192811.

[147] Dingjun Li, PL Patrick Rau, and Ye Li. A cross-cultural study: Effect of robot appearance and task. *Int. J. of Social Robotics*, 2(2):175–186, 2010.

[148] T. Paschal, M. A. Bell, J. Sperry, S. Sieniewicz, R. J. Wood, and J. C. Weaver. Design, fabrication, and characterization of an untethered amphibious sea urchin-inspired robot. *IEEE Robot. and Autom. Lett.*, 4(4):3348–3354, 2019. doi: 10.1109/LRA.2019.2926683.

[149] Rita Latikka, Tuuli Turja, and Atte Oksanen. Self-efficacy and acceptance of robots. *Computers in Human Behavior*, 93:157–163, 2019. ISSN 0747-5632.

[150] David Feil-Seifer and Maja J Matarić. Socially assistive robotics. 18(1):24–31, 2011.

[151] K. S. Haring, K. Watanabe, and C. Mougenot. The influence of robot appearance on assessment. In *ACM/IEEE Int. Conf. on Human-Robot Interaction*, pages 131–132, 2013. doi: 10.1109/HRI.2013.6483536.

[152] Florent Levillain and Elisabetta Zibetti. Behavioral objects: The rise of the evocative machines. *JHRI*, 6(1):4–24, 2017.

[153] Nuno Ferreira Duarte, Mirko Raković, Jovica Tasevski, Moreno Ignazio Coco, Aude Billard, and José Santos-Victor. Action anticipation: Reading the intentions of humans and robots. *IEEE Robot. and Autom. Lett.*, 3(4):4132–4139, 2018.

[154] H. A. T. Broers, J. Ham, R. Broeders, R. De Silva, and M. Okada. Goal inferences about robot behavior goal inferences and human response behaviors. In *IEEE Int. Conf. on Human-Robot Interaction*, pages 91–92, 2013. doi: 10.1109/HRI.2013.6483516.

[155] Alisha Bevins and Brittany A. Duncan. Aerial flight paths for communication: How participants perceive and intend to respond to drone movements. In *the ACM/IEEE Int. Conf. on Human-Robot Interaction*, page 16–23, 2021. ISBN 9781450382892. doi: 10.1145/3434073.3444645.

[156] Hagen Lehmann, Joan Saez-Pons, Dag Sverre Syrdal, and Kerstin Dautenhahn. In good company? perception of movement synchrony of a non-anthropomorphic robot. *PloS one*, 10(5):e0127747, 2015.

[157] Guy Hoffman. Dumb robots, smart phones: A case study of music listening companionship. In *IEEE Int. Symp. on Robot and Human Interactive Communication*, pages 358–363. IEEE, 2012.

[158] Sondra Ashmore and Kristin Runyan. *Introduction to agile methods*. Addison-Wesley Professional, 2014.

[159] Dominique Deuff, Ioana Ocnarescu, Luis Enrique Coronado, Liz Rincon-Ardila, Isabelle Milleville, and Gentiane Venture. Designerly way of thinking in a robotics research project. *J. of Robotics Soc. of Japan*, 38(8):692–702, 2020.

[160] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *CoRR*, abs/1812.08008, 2018.

[161] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, et al. Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172*, 2019.

[162] Intel. Openvino, 2019. URL `https://docs.openvinotoolkit.org/`.

[163] Intel. URL `https://www.intel.com/content/www/us/en/developer/tools/neural-compute-stick/overview.html`.

[164] B. Martin, B. Hanington, and B.M. Hanington. *Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions*. Rockport Publishers, 2012. ISBN 9781592537563.

[165] Jan Stoklasa, Tomáš Talášek, and Jana Stoklasová. Semantic differential for the twenty-first century: scale relevance and uncertainty entering the semantic space. *Quality & Quantity*, 53(1):435–448, 2019.

[166] Jun Liang, Deqiang Xian, Xingyu Liu, Jing Fu, Xingting Zhang, Buzhou Tang, Jianbo Lei, et al. Usability study of mainstream wearable fitness devices: feature analysis and system usability scale evaluation. *JMIR mHealth and uHealth*, 6(11):e11066, 2018.

[167] Wolfgang Palaver. *René Girard's mimetic theory*. MSU Press, 2013.

[168] Christopher Cannon, Kelly Goldsmith, and Caroline Roux. A self-regulatory model of resource scarcity. *J. of Consumer Psychology*, 29(1):104–127, 2019.

[169] Muhammad Asadullah and Ahsan Raza. An overview of home automation systems. In *Int. Conf. on Robotics and Artificial Intelligence*, pages 27–31, 2016. doi: 10.1109/ICRAI.2016.7791223.

[170] Cristina Stolojescu-Crisan, Calin Crisan, and Bogdan-Petru Butunoi. An iot-based smart home automation system. *Sensors*, 21(11), 2021. ISSN 1424-8220. doi: 10.3390/s21113784.

[171] Laura Romeo, Antonio Petitti, Roberto Marani, and Annalisa Milella. Internet of robotic things in smart domains: applications and challenges. *Sensors*, 20(12):3355, 2020.

[172] iRobot. URL `https://www.irobot.com/roomba`.

[173] Amazon. URL `https://amzn.to/3QxgAov`.

[174] KEYi Tech. URL `https://keyirobot.com/en-asia`.

[175] ROYBI Robot. URL `https://roybirobot.com`.

[176] Ha Manh Do, Minh Pham, Weihua Sheng, Dan Yang, and Meiqin Liu. Rish: A robot-integrated smart home for elderly care. *Robotics and Autonomous Systems*, 101:74–92, 2018. ISSN 0921-8890. doi: https://doi.org/10.1016/j.robot.2017.12.008.

[177] Michal Luria, Guy Hoffman, Benny Megidish, Oren Zuckerman, and Sung Park. Designing vyo, a robotic smart home assistant: Bridging the gap between device and social agent. In *IEEE Int. Symp. on Robot and Human Interactive Communication*, pages 1019–1025. IEEE, 2016.

[178] Michal Luria, Guy Hoffman, and Oren Zuckerman. Comparing social robot, screen and voice interfaces for smart-home control. In *Proc. of the 2017 CHI Conference on Human Factors in Computing Systems*, page 580–628, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450346559. doi: 10.1145/3025453.3025786.

[179] Stéphane Legleye, Amandine Nougaret, and Louise Viard-Guillot. 94% des 15-29 ans ont un smartphone en 2021. Technical Report 259, INSEE, 2022. URL `https://www.insee.fr/fr/statistiques/6036909`.

[180] Christina Bröhl, Peter Rasche, Janina Jablonski, Sabine Theis, Matthias Wille, and Alexander Mertens. Desktop pc, tablet pc, or smartphone? an analysis of use preferences in daily activities for different technology generations of a worldwide sample. In Jia Zhou and Gavriel Salvendy, editors, *Human Aspects of IT for the Aged Population. Acceptance, Communication and Participation*, pages 3–20, Cham, 2018. Springer International Publishing. ISBN 978-3-319-92034-4.

[181] Jonathan Anderson, John McRee, Robb Wilson, et al. *Effective UI: The art of building great user experience in software*. " O'Reilly Media, Inc.", 2010.

[182] Dwi Ely Kurniawan, Mohd Iqbal, John Friadi, Rohmat Indra Borman, and Rio Rinaldi. Smart monitoring temperature and humidity of the room server using raspberry pi and whatsapp notifications. *J. of Physics: Conference Series*, 1351(1):012006, nov 2019. doi: 10.1088/1742-6596/1351/1/012006.

[183] S. N. Ishak, N. N. N. Abd Malik, N. M. Abdul Latiff, N. Effiyana Ghazali, and M. A. Baharudin. Smart home garden irrigation system using raspberry pi. In *Malaysia Int. Conf. on Communications*, pages 101–106. IEEE, 2017. doi: 10.1109/MICC.2017.8311741.

[184] Yasirli Amri and Mukhammad Andri Setiawan. Improving smart home concept with the internet of things concept using RaspberryPi and NodeMCU. *IOP Conf. Series: Materials Science and Engineering*, 325:012021, mar 2018. doi: 10.1088/1757-899x/325/1/012021.

[185] Waheb A. Jabbar, Chong Wen Wei, Nur Atiqah Ainaa M. Azmi, and Nur Aiman Haironnazli. An iot raspberry pi-based parking management system for smart campus. *Internet of Things*, 14:100387, 2021. ISSN 2542-6605. doi: https://doi.org/10.1016/j.iot.2021.100387.

[186] Raspberry Pi Ltd. Raspberry Pi documentation. URL `https://www.raspberrypi.com/documentation/computers/configuration.html#configuring-networking`.

[187] GNOME Project. URL `https://wiki.gnome.org/Projects/NetworkManager`.

[188] Hugo Dupras. URL `https://pypi.org/project/pyatmo/`.

[189] Internet Engineering Task Force. URL `https://datatracker.ietf.org/doc/html/rfc6749`.

[190] Lucía Jiménez Iglesias, Carlos Aguilar Paredes, Lydia Sánchez, and Mario Pérez-Montoro. User experience and the media: the three-click rule on newspaper websites for smartphones. *Revista Latina de Comunicación Social, 2018, num. 76, p. 595-613.*, 2018.

[191] Jakob Nielsen and Thomas K. Landauer. A mathematical model of the finding of usability problems. In *the INTERACT and Conf. on Human Factors in Computing Systems*, page 206–213, 1993. ISBN 0897915755. doi: 10.1145/169059.169166.

[192] Aaron Bangor, Philip T. Kortum, and James T. Miller. An empirical evaluation of the system usability scale. *Int. J. of Human–Computer Interaction*, 24(6):574–594, 2008.

[193] Aycan Kaya, Reha Ozturk, and Cigdem Altin Gumussoy. Usability measurement of mobile applications with system usability scale (sus). In Fethi Calisir, Emre Cevikcan, and Hatice Camgoz Akdag, editors, *Industrial Engr. in the Big Data Era*, pages 389–400. Springer Int. Publishing, 2019. ISBN 978-3-030-03317-0.

[194] Raluca Budiu. Wizards: Definition and design recommendations, 2017. URL `https://www.nngroup.com/articles/wizards/`.

[195] Stuart J Russell. *Artificial intelligence a modern approach.* Pearson Education, Inc., 2010.

[196] Kevin Ellis, Catherine Wong, Maxwell Nye, Mathias Sable-Meyer, Luc Cary, Lucas Morales, Luke Hewitt, Armando Solar-Lezama, and Joshua B Tenenbaum. Dreamcoder: Growing generalizable, interpretable knowledge with wake-sleep bayesian program learning. *arXiv preprint arXiv:2006.08381*, 2020.

[197] Filipe de Avila Belbute-Peres, Kevin Smith, Kelsey Allen, Josh Tenenbaum, and J Zico Kolter. End-to-end differentiable physics for learning and control. *Advances in neural information processing systems*, 31, 2018.

[198] Tejas D Kulkarni, Karthik Narasimhan, Ardavan Saeedi, and Josh Tenenbaum. Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation. *Advances in neural information processing systems*, 29, 2016.

[199] Tom Silver, Rohan Chitnis, Joshua Tenenbaum, Leslie Pack Kaelbling, and Tomás Lozano-Pérez. Learning symbolic operators for task and motion planning. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3182–3189. IEEE, 2021.

[200] Aidan Curtis, Tom Silver, Joshua B Tenenbaum, Tomás Lozano-Pérez, and Leslie Kaelbling. Discovering state and action abstractions for generalized task and motion planning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 5377–5384, 2022.

[201] Shuchisnigdha Deb, Lesley Strawderman, Janice DuBien, Brian Smith, Daniel W. Carruth, and Teena M. Garrison. Evaluating pedestrian behavior at crosswalks: Validation of a pedestrian behavior questionnaire for the u.s. population. *Accident Analysis & Prevention*, 106:191–201, 2017. ISSN 0001-4575. doi: https://doi.org/10.1016/j.aap.2017.05.020. URL `https://www.sciencedirect.com/science/article/pii/S0001457517301835`.

[202] Roni Utriainen and Markus Pöllänen. How automated vehicles should operate to avoid fatal crashes with cyclists? *Accident Analysis & Prevention*, 154:106097, 2021.

[203] Nan Jiang, Mi Shi, Yilong Xiao, Kan Shi, and Barry Watson. Factors affecting pedestrian crossing behaviors at signalized crosswalks in urban areas in beijing and singapore. In *ICTIS 2011: Multimodal Approach to Sustained Transportation System Development: Information, Technology, Implementation*, pages 1090–1097. 2011.

[204] Lorena Hell, Janis Sprenger, Matthias Klusch, Yoshiyuki Kobayashi, and Christian Müller. Pedestrian behavior in japan and germany: A review.

[205] David Budgen and Pearl Brereton. Performing systematic literature reviews in software engineering. In *Proceedings of the 28th international conference on Software engineering*, pages 1051–1052, 2006.

[206] Barbara Kitchenham. Procedures for performing systematic reviews. *Keele, UK, Keele Univ.*, 33, 08 2004.

[207] Kai Petersen, Sairam Vakkalanka, and Ludwik Kuzniarz. Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64:1–18, 2015. ISSN 0950-5849. doi: https://doi.org/10.1016/j.infsof.2015.03.007. URL `https://www.sciencedirect.com/science/article/pii/S0950584915000646`.

[208] Zhijie Fang and Antonio M López. Intention recognition of pedestrians and cyclists by 2d pose estimation. *IEEE Transactions on Intelligent Transportation Systems*, 21(11):4773–4783, 2019.

[209] Khaled Saleh, Mohammed Hossny, and Saeid Nahavandi. Cyclist trajectory prediction using bidirectional recurrent neural networks. In Tanja Mitrovic, Bing Xue, and Xiaodong Li, editors, *AI 2018: Advances in Artificial Intelligence*, pages 284–295, Cham, 2018. Springer International Publishing. ISBN 978-3-030-03991-2.

[210] Andreas Th Schulz and Rainer Stiefelhagen. Pedestrian intention recognition using latent-dynamic conditional random fields. In *2015 IEEE Intelligent Vehicles Symposium (IV)*, pages 622–627. IEEE, 2015.

[211] Andreas Møgelmose, Mohan M Trivedi, and Thomas B Moeslund. Trajectory analysis and prediction for improved pedestrian safety: Integrated framework and evaluations. In *2015 IEEE intelligent vehicles symposium (IV)*, pages 330–335. IEEE, 2015.

[212] Iuliia Kotseruba, Amir Rasouli, and John K Tsotsos. Do they want to cross? understanding pedestrian intention for behavior prediction. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, pages 1688–1693. IEEE, 2020.

[213] Paul E Hemeren, Mikael Johannesson, Mikael Lebram, Fredrik Eriksson, Kristoffer Ekman, and Peter Veto. The use of visual cues to determine the intent of cyclists in traffic. In *2014 IEEE International Inter-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, pages 47–51. IEEE, 2014.

[214] Haoran Wu, Sifa Zheng, Qing Xu, and Jianqiang Wang. Applying the extended theory of planned behavior to pedestrian intention estimation. In *2021 IEEE Intelligent Vehicles Symposium (IV)*, pages 1509–1514. IEEE, 2021.

[215] Nora Muscholl, Matthias Klusch, Patrick Gebhard, and Tanja Schneeberger. Emidas: explainable social interaction-based pedestrian intention detection across street. In *Proceedings of the 36th Annual ACM Symposium on Applied Computing*, pages 107–115, 2021.

[216] Arash Kalatian and Bilal Farooq. A context-aware pedestrian trajectory prediction framework for automated vehicles. *Transportation research part C: emerging technologies*, 134: 103453, 2022.

[217] Julian FP Kooij, Fabian Flohr, Ewoud AI Pool, and Dariu M Gavrila. Context-based path prediction for targets with switching dynamics. *International Journal of Computer Vision*, 127(3):239–262, 2019.

[218] Amir Rasouli, Iuliia Kotseruba, Toni Kunic, and John K Tsotsos. Pie: A large-scale dataset and models for pedestrian intention estimation and trajectory prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6262–6271, 2019.

[219] Walter Morales Alvarez, Francisco Miguel Moreno, Oscar Sipele, Nikita Smirnov, and Cristina Olaverri-Monreal. Autonomous driving: Framework for pedestrian intention estimation in a real world scenario. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, pages 39–44. IEEE, 2020.

[220] Martin Stolz, Mingkang Li, Zhaofei Feng, Martin Kunert, and Wolfgang Menzel. Direction of movement estimation of cyclists with a high-resolution automotive radar. In *2018 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM)*, pages 1–4. IEEE, 2018.

[221] Renfei Wu, Xunjia Zheng, Yongneng Xu, Wei Wu, Guopeng Li, Qing Xu, and Zhuming Nie. Modified driving safety field based on trajectory prediction model for pedestrian–vehicle collision. *Sustainability*, 11(22):6254, 2019.

[222] Benjamin Völz, Karsten Behrendt, Holger Mielenz, Igor Gilitschenski, Roland Siegwart, and Juan Nieto. A data-driven approach for pedestrian intention estimation. In *2016 ieee 19th international conference on intelligent transportation systems (itsc)*, pages 2607–2612. IEEE, 2016.

[223] Yoriyoshi Hashimoto, Yanlei Gu, Li-Ta Hsu, and Shunsuke Kamijo. Probability estimation for pedestrian crossing intention at signalized crosswalks. In *2015 IEEE International Conference on Vehicular Electronics and Safety (ICVES)*, pages 114–119. IEEE, 2015.

[224] Yasheng Sun, Tao He, Jie Hu, Haiqing Huang, and Biao Chen. Intent-aware conditional generative adversarial network for pedestrian path prediction. In *2019 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, pages 155–160. IEEE, 2019.

[225] Nicolas Schneider and Dariu M. Gavrila. Pedestrian path prediction with recursive bayesian filters: A comparative study. In Joachim Weickert, Matthias Hein, and Bernt Schiele, editors, *Pattern Recognition*, pages 174–183, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. ISBN 978-3-642-40602-7.

[226] Julian Francisco Pieter Kooij, Nicolas Schneider, Fabian Flohr, and Dariu M. Gavrila. Context-based pedestrian path prediction. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 618–633, Cham, 2014. Springer International Publishing. ISBN 978-3-319-10599-4.

[227] Harshayu Girase, Haiming Gang, Srikanth Malla, Jiachen Li, Akira Kanehara, Karttikeya Mangalam, and Chiho Choi. Loki: Long term and key intentions for trajectory prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9803–9812, 2021.

[228] Hirokatsu Kataoka, Yutaka Satoh, Yoshimitsu Aoki, Shoko Oikawa, and Yasuhiro Matsui. Temporal and fine-grained pedestrian action recognition on driving recorder database. *Sensors*, 18(2):627, 2018.

[229] Atanas Poibrenski, Matthias Klusch, Igor Vozniak, and Christian Müller. M2p3: multimodal multi-pedestrian path prediction by self-driving cars with egocentric vision. In *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, pages 190–197, 2020.

[230] Iuliia Kotseruba, Amir Rasouli, and John K Tsotsos. Joint attention in autonomous driving (jaad). *arXiv preprint arXiv:1609.04741*, 2016.

[231] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020.

[232] R Kesten, M Usman, J Houston, T Pandya, K Nadhamuni, A Ferreira, M Yuan, B Low, A Jain, P Ondruska, et al. Lyft level 5 perception dataset 2020, 2019.

[233] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2446–2454, 2020.

[234] Sarah Ferguson, Brandon Luders, Robert C Grande, and Jonathan P How. Real-time predictive modeling and robust avoidance of pedestrians with uncertain, changing intentions. In *Algorithmic Foundations of Robotics XI*, pages 161–177. Springer, 2015.

[235] Ewoud AI Pool, Julian FP Kooij, and Dariu M Gavrila. Using road topology to improve cyclist path prediction. In *2017 IEEE Intelligent Vehicles Symposium (IV)*, pages 289–296. IEEE, 2017.

[236] Samir B. Unadkat, Mãlina M. Ciocoiu, Larry R. Medsker, David G. Hagner, Mohamad H. Hassoun, and Paul B. Watta. *RECURRENT NEURAL NETWORKS, Design and Applications*, chapter 1, 10. CRC Press, 2001.

[237] Laura Uusitalo. Advantages and challenges of bayesian networks in environmental modelling. *Ecological Modelling*, 203(3):312–318, 2007. ISSN 0304-3800. doi: https://doi.org/10.1016/j.ecolmodel.2006.11.033. URL `https://www.sciencedirect.com/science/article/pii/S0304380006006089`.

[238] Christopher J. C. H. Watkins and Peter Dayan. Q-learning. *Machine Learning*, 8(3):279–292, May 1992. ISSN 1573-0565. doi: 10.1007/BF00992698. URL `https://doi.org/10.1007/BF00992698`.

[239] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.

[240] Albert Mehrabian. *Basic dimensions for a general psychological theory: Implications for personality, social, environmental, and developmental studies.* Cambridge : Oelgeschlager, Gunn Hain, 1980. ISBN 978-0-89946-004-8.

[241] Avery N Gilbert, Alan J Fridlund, and Laurie A Lucchina. The color of emotion: A metric for implicit color associations. *Food Quality and Preference*, 52:203–210, 2016.

[242] Naz Kaya and Helen H Epps. Relationship between color and emotion: A study of college students. *College student journal*, 38(3):396–405, 2004.

[243] Ümit Demir. Investigation of color-emotion associations of the university students. *Color Research & Application*, 45(5):871–884, 2020. doi: https://doi.org/10.1002/col.22522. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/col.22522.

[244] Tina M. Sutton and Jeanette Altarriba. Color associations to emotion and emotion-laden words: A collection of norms for stimulus construction and selection. *Behavior Research Methods*, 48(2):686–728, Jun 2016. ISSN 1554-3528. doi: 10.3758/s13428-015-0598-8. URL https://doi.org/10.3758/s13428-015-0598-8.

[245] Xiao-Ping Gao, John H. Xin, Tetsuya Sato, Aran Hansuebsai, Marcello Scalzo, Kanji Kajiwara, Shing-Sheng Guan, J. Valldeperas, Manuel José Lis, and Monica Billger. Analysis of cross-cultural color emotion. *Color Research & Application*, 32(3):223–229, 2007. doi: https://doi.org/10.1002/col.20321. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/col.20321.

[246] Josep-Arnau Claret, Gentiane Venture, and Luis Basañez. Exploiting the robot kinematic redundancy for emotion conveyance to humans as a lower priority task. *International Journal of Social Robotics*, 9(2):277–292, Apr 2017. ISSN 1875-4805. doi: 10.1007/s12369-016-0387-2. URL https://doi.org/10.1007/s12369-016-0387-2.

[247] Jinsong Yu, Yue Song, Diyin Tang, and Jing Dai. A digital twin approach based on nonparametric bayesian network for complex system health monitoring. *Journal of Manufacturing Systems*, 58:293–304, 2021. ISSN 0278-6125. doi: https://doi.org/10.1016/j.jmsy.2020.07.005. URL https://www.sciencedirect.com/science/article/pii/S0278612520301151. Digital Twin towards Smart Manufacturing and Industry 4.0.

[248] Donald Glowinski, Nele Dael, Antonio Camurri, Gualtiero Volpe, Marcello Mortillaro, and Klaus Scherer. Toward a minimal representation of affective gestures. *IEEE Transactions on Affective Computing*, 2(2):106–118, 2011.

[249] Ferdous Ahmed, A. S. M. Hossain Bari, and Marina L. Gavrilova. Emotion recognition from body movement. *IEEE Access*, 8:11761–11781, 2020. doi: 10.1109/ACCESS.2019.2963113.

Appendix A

Questionnaires Scale

A.1 The Big Five questionnaire

1. I like to talk.
2. I like being busy.
3. I am healthy and active.
4. I don't worry about many things.
5. I usually don't get hurt (emotionally).
6. I rarely feel nervous.
7. I have a lot of imagination.
8. I am flexible.
9. I am curious.
10. I like to plan.
11. I work hard on issues.
12. I respect rules and promises.
13. I think from other's perspectives.
14. I like to collaborate with others.
15. I like to communicate my feelings.

A.2 The SUS scale

1. I think I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.

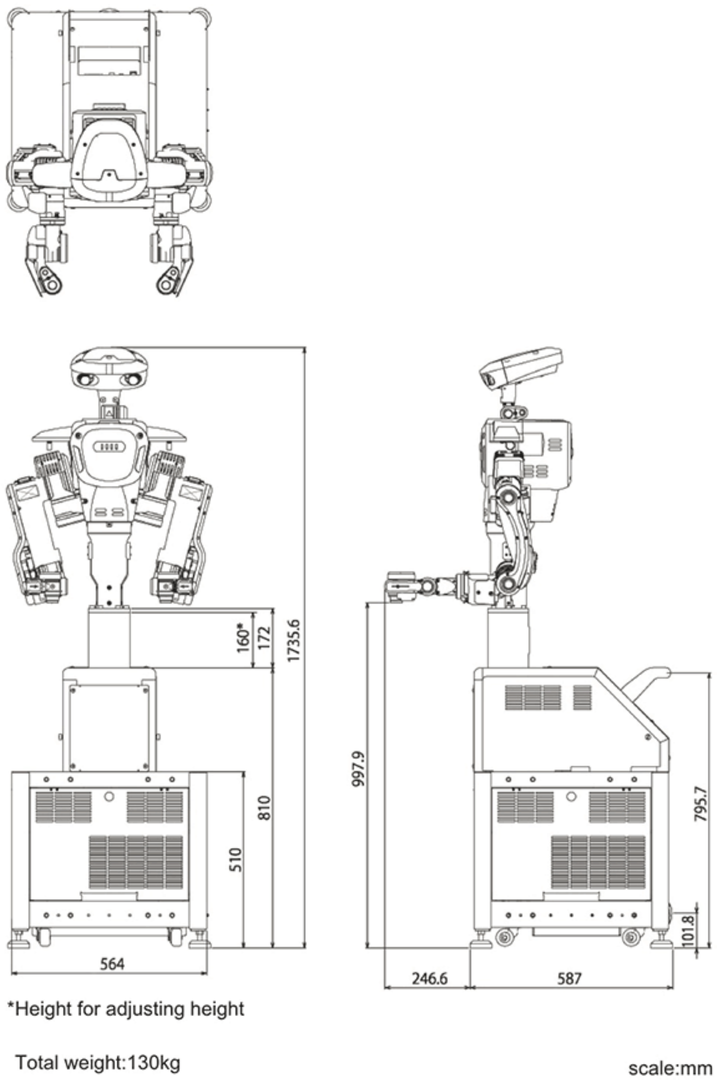
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I need to learn a lot of things before I could get going with this system.

Appendix B

NEXTAGE OPEN Characteristics

All the data come from the online documentation available in [93].

B.1 Dimensions



B.2 Specifications

Table B.1: Robot specifications

Degree of freedom		15 axes (6 for arms × 2, 2 for neck, 1 for waist)
Axes configuration	Arms	Shoulder Y - Shoulder P - Elbow P - Wrist Y - Wrist P - Wrist R
	Neck	Neck Y ? Neck P
	Body	Waist Y
Payload (TCP)		1.5 kg (one arm) 3.0 kg (two arms)
Position repeatability[1]		Within 0.5 sec
Positioning repeatability[1]		Within 30 μm
Max speed of each axis		Waist Y: 110 deg/s Neck Y: 150 deg/s Neck P: 250 deg/s Shoulder Y: 172 deg/s Shoulder P: 133 deg/s Elbow P: 215 deg/s Wrist Y: 263 deg/s Wrist P: 224 deg/s Wrist R: 300 deg/s
Operating environment	Ambient operating temperature	0 to + 40 °C
	Ambient operating humidity	20 to 80 % RH (no condensation shall occur)
	In storage	Same as the operating environment
Weight		29 kg

[1]The test conditions are:

— Load applied to hands: 1.5 kg

— Position coordinate to be measured: (x, y, z) = (315, ±100, 220) (+: Left arm, -: Right arm)

— Operation speed rated at 100%, 50% and 10%

— Number of measurements: 30 Error band: ±30 μm

Appendix C

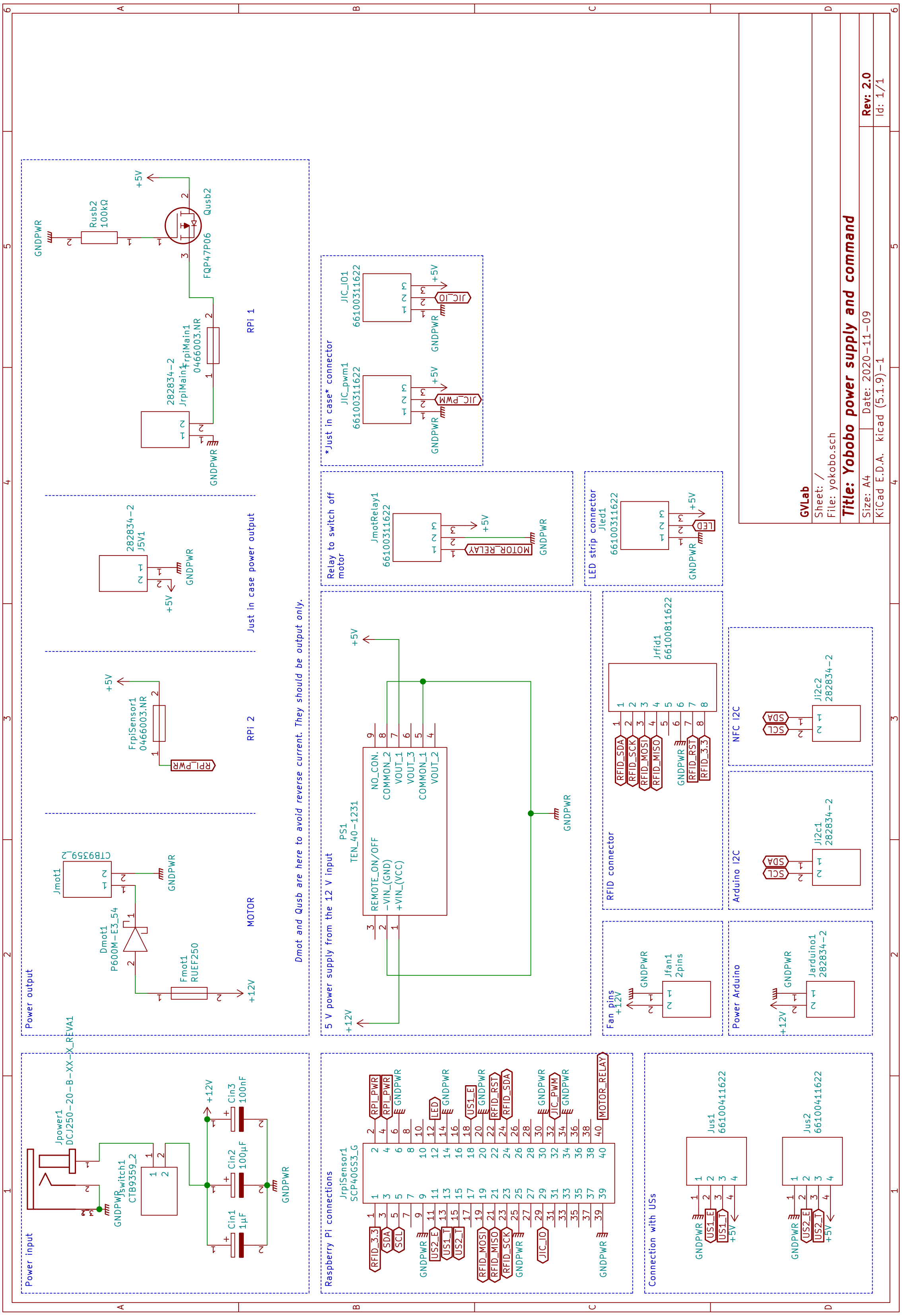
Yōkobo's Hardware

C.1 Bill of Material

Component	Ref.	Qty	Characteristics
Raspberry Pi 4B		2	8 GB (RAM) 32 GB (μSD)
Ultrasonic sensor		2	HC-SR04
Wide angle camera	-	1	5MP 1080p 720p
TRACO 5 V - 6 A	PS1	1	Vout: 5/$\pm$12V dc 9 $\rightarrow$ 18 V dc, 40W 12 V dc
Resistance	Rusb2	1	100 kΩ 0.25W $\pm$5%
Power female socket	Jpower1	1	DCJ250-20-B-K1-A
Power male plug	-	1	5.0A 12.0 V 2.5mm 5.5mm
PCB Terminal Blocks (screw)	Jarduino1, Ji2c1, Ji2c2, J5V1, JrpiMain1	5	2.54mm, 2 channels
PCB Terminal Blocks female	Jmot1, Jswitch1	2	5.08mm 2 channels
PCB Termial Blocks male	-	2	5.08mm 2 channels
Simple IDC connector female 3 pins (LED+JIC)	-	4	2.54mm 3 channels
Simple IDC connector male 3 pins (LED+JIC)	Jled1, JIC_pwm1, JIC_IO1, JmotRelay1	4	2.54mm 3 channels
Switch	-	1	SPST, On-Off
Fuse polyswitch	Fmot1	1	2.5A 30V dc RUEF250
Condensator (Cin1)	Cin1	1	1 μF 50V dc
Condensator (Cin2)	Cin2	1	100 μF - 100V dc
Condensator (Cin3)	Cin3	1	100nF - 50V dc
Diode Schottky	Dmot1	1	6A, 400V
IDC connector (RPi) male	JrpiSensor1	1	2.54mm 40 channels (2x20)
RPI Ribbon + IDC	-	1	40 Pin 20 cm, Female to Female
Simple IDC connector male 4 pins (US)	Jus1, Jus2	2	2.54mm 4 channels
Simple IDC connector female 4 pins (US)	-	4	2.54mm 4 channels
Fan pin	Jfan1	2	
Fan	-		
Relay	-	1	SRD-05VDC-SL-C
Fuse raspberry	FrpiSensor1, FrpiMain1	2	3A, 32V ac/dc

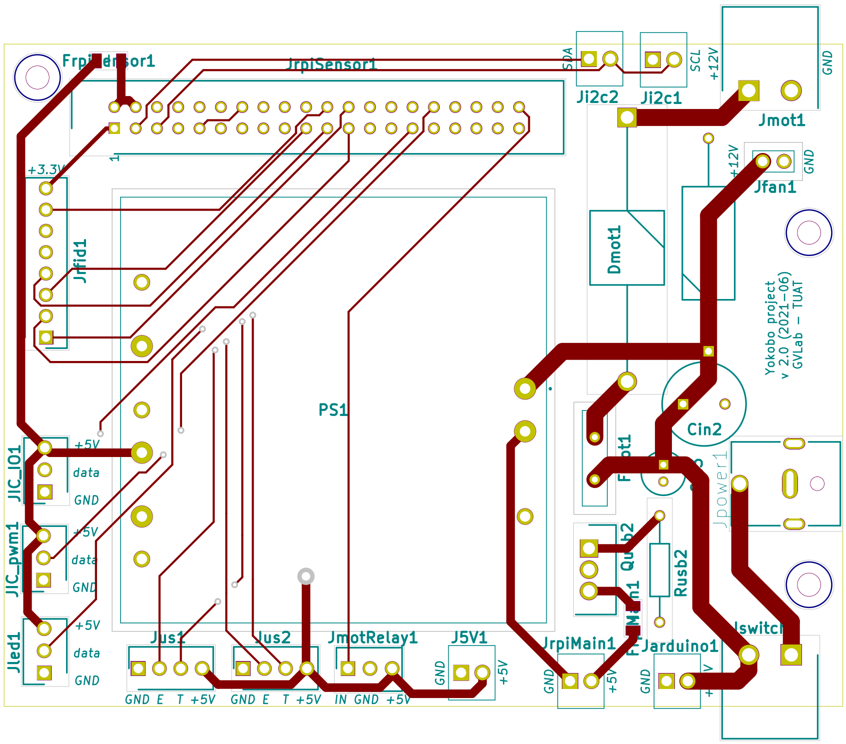
Simple IDC connector male 8 pins (RFID)	Jrfid1	1	2.54mm 8 channels
Simple IDC connector female 8 pins (RFID)	-	2	2.54mm 8 channels
Heatsink Raspberry Pi	-	2	
Ribbon cable	-	1	16 pins 20.3 mm
LED strip	-	1	WS2812B
Ribbon camera	-	1	Raspberry Pi camera ribbon cable 50 cm
Intel compute stick 2	-	1	
Motor driver	-	1	SMPS2DYNAMIXEL
Motor cables (serial cables)	-	1	DYNAMIXEL 4P 240mm
Motor USB device (motor command)	-	1	U2D2 INTL
Dynamixel Motor		1	MX-28
Dynamixel Motor		1	MX-64
Dynamixel Motor		1	MX-106
Ethernet cable	-	1	10 cm

C.2 Final PCB Circuit Diagram

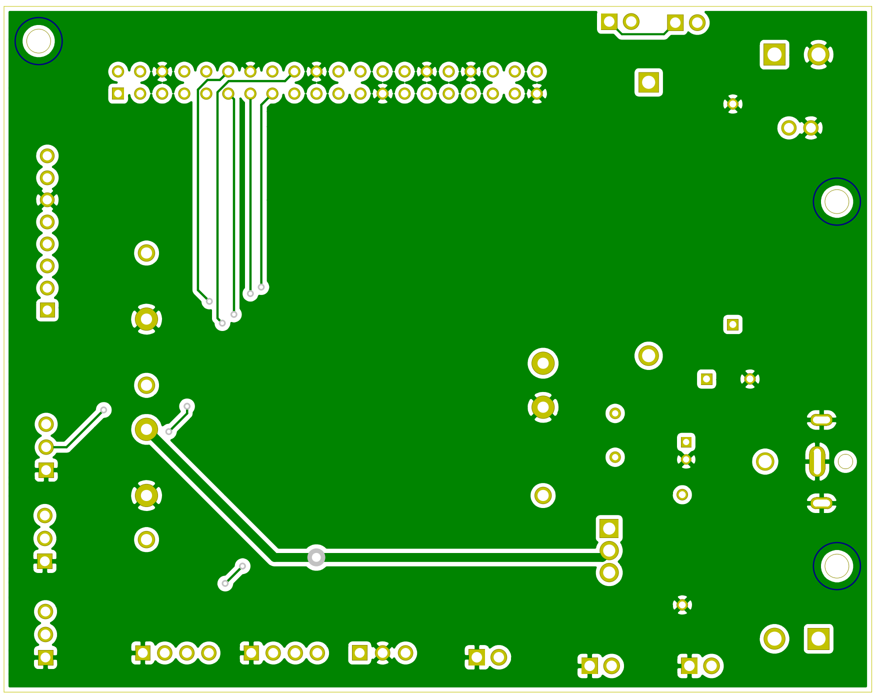


C.3 Final PCB Routing Diagram

C.3.1 Front



C.3.2 Bottom



Appendix D

Questionnaires and Data of the Technical Experiment

From the experiment described in Chapter 5.

D.1 Questionnaires

In this section the different questions asked with the three questionnaires (QS, QW1 and QW2) are presented in the Tables D.1, D.2 and D.3. The questions for the first one (QS) are asked in the future tense since the participants answer them before any interactions.

Table D.1: Questions asked in QS, QW1 and QW2. The text in square brackets is added for this article and was not present in the questionnaire.

Question	**Answer choice**
Personal questions	
Student ID[1]	*number*
How old are you?[2]	*number*
What is your gender?[2]	Female, Male, Prefer not to say
What is your nationality?[2]	*text*
What is your level of knowledge about robots?[2]	Not familiar (1) - Familiar (5)
What are your knowledge about Yōkobo?[2]	I was involved in the first experiment I already interact with Yōkobo on my own I know how it works and already saw it I only saw it I don't know about Yōkobo
About Yōkobo and its behaviour	
How much Yōkobo welcomed you? [Likert scale]	Not at all (1) - I felt welcomed (5)
Did you see intelligence in Yōkobo? [Likert scale]	Not at all (1) - Totally (5)
Did you see life in Yōkobo? [Likert scale]	Not at all (1) - Totally (5)
What is your feeling about Yōkobo? (1: not at all ; 5: totally [Likert scale])	Curious Happy Afraid Enthusiastic Confusion Friendly
Yōkobo is [*positive adj.* 1 2 3 4 5 *negative adj.* - semantic scale]	Smart (1) Stupid (5) Simple (1) Complicated (5) Dynamic (1) Static (5) Lifelike (1) Artificial (5) Responsive (1) Slow (5) Emotional (1) Emotionless (5) Useful (1) Useless (5) Familiar (1) Unknown (5) Desirable (1) Undesirable (5) Cute (1) Ugly (5) Modern (1) Old (5) Attractive (1) Unattractive (5)
I ______ Yōkobo [semantic scale]	Like (1) - Dislike (5)
Design	
Name with your word, in the next questions, the different parts of Yōkobo. Numbers 1 to 4 are pointing the whole part. Numbers 5 and 6 are pointing the holes.	
How do you will call the mark 1[3]	*text*
How do you will call the mark 2[3]	*text*
How do you will call the mark 3[3]	*text*
How do you will call the mark 4[3]	*text*
How do you will call the mark 5[3]	*text*
How do you will call the mark 6[3]	*text*
Interactions	
How many times did you interact with Yōkobo?[4]	*number*
Additional remarks	
Do you have any additional remarks or comment about Yōkobo?[5]	*text*

[1] The student ID is used to match the answers of the three questionnaires.

[2] Only for QS

[3] The number refers to the Figure D.1

[4] Only for QW1 and QW2

[5] Only for QW2

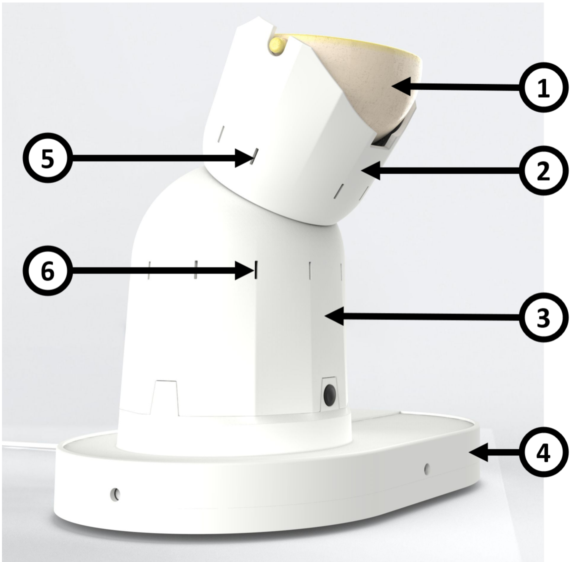


Figure D.1: Picture of Yōkobo used in the questionnaire to know the vocabulary used by participants to describe the robot.

Table D.2: Questions in QW2 for the selected participants. The text in square brackets is added for this article and was not present in the questionnaire.

Question	**Answer choice**
Messages	
How many messages did you received from your partner?	*number*
How many messages did you sent to your partner?	*number*
How would you rate the recording of a message? [semantic scale]	Hard (1) - Easy (5)
How much did you feel the existence of your partner during this week? [Likert scale]	Does not exist (1) - Exists (5)
Do you think Yōkobo helped you to feel your partner? [Likert scale]	Not at all (1) - A lot (5)
What does Yōkobo represent in the connection with your partner?	*text*
Did you have the impression that the movement of the robot was Yōkobo's behaviours or was coming from your partner?	*text*

Table D.3: SUS questions in QW2 for the selected participants, using a Likert scale

Question	Answer choice
1. I think I would like to use Yōkobo frequently 2. I found Yōkobo unnecessarily complex 3. I though Yōkobo was easy to use 4. I think I would need the support of a technical person to be able to use Yōkobo 5. I found the various functions in Yōkobo were well integrated 6. I thought there was too much inconsistency in Yōkobo 7. I would imagine that most people would learn to use Yōkobo very quickly 8. I found Yōkobo very cumbersome to use 9. I felt very confident using Yōkobo 10. I need to learn a lot of things before I could get going with Yōkobo	Strongly Disagree (1) - Strongly Agree (5)

D.2 Graffiti wall

Sample of the answers to the graffiti wall.

Question 1

What do you think about Yokobo? Can you give us your first impressions?

1. (E2) Friendly (Yokobo seems to greet me).
2. (E2) 見た目が可愛い、全体的に小動物っぽい（It looks cute, it looks like a small animal overall）ゆらめくLEDがきれい、周囲を動くと目で追ってきて愛らしい（The shimmering LED is beautiful, and when I move around, It can follow me with its eyes and it's adorable）
3. (E2) It feels that Yokobo notices you when you step close since the LED pattern and color change. It also follows you with the camera which is cute. However, I wasn't too sure on how to interact and play with Yokobo.

Question 2

For you, what Yokobo's movements express?

1. (E1) 水槽の中で泳ぐ魚みたい (It moves like a fish in an aquarium)
2. (E1) I felt being greeted since the robot motion looked a bit like waving or bowing (kind of a puppy welcoming their owner)
3. (E2) Exciting
4. (E2) stare at me, nod

Question 3

Just now you approached Yokobo and were close to it, what was your experience?

1. (E1) I was curious about what it would have done and waited a bit to see its "reaction"
2. (E1) The robot was tilting and moving its head in a offering kind of manner. I was able to feel that it detects me and can change how to behave wen the situation changes
3. (E2) ロボットとコミュニケーションをすること自体が初めてで新鮮かつ可愛かった（It was my first time to communicate with a robot and it was fresh and cute.）
4. (E2) 意外に動きに幅があり、驚いた。合わせて動くのが面白かった。（I was surprised that there was a wide range of movements. It was fun to move together.）

Question 4a (participants with a tag)

Do you feel connected to Yokobo and/or your partner? After 2 weeks, does Yokobo bring something to your daily life?

1. (E1) It's nice to have a welcoming "partner" when entering the lab. I think since I know there might be a message from my partner, I am not sure if it's Yokobo moving or my partner, but in principle I think more that it's my partner. Not knowing who he/she is, make me feel a bit less connected, I think. So more connected to Yokobo than my partner.
2. (E2) I couldn't feel the exist of partner (*sic*)
3. (E2) 相手がいるというのは分かった（I understand there was a partner）
4. (E2) 何かが起きた（something happened）

Question 4b (participants without a tag)

Do you feel close to Yokobo? After 2 weeks, does Yokobo bring something to your daily life?

1. (E1) 人の方を向いてくれるため、インタラクションを実感できた（I was able to feel the interaction because it turned to people）

Appendix E

BeWoDa

E.1 Algorithm

Algorithm E.1: DQL - main function

```
FUNCTION main()
    env ← new Environment()
    dql ← new Deep-Q-Learning()

    FOR EACH nbr_trial
        state ← env.reset()
        score ← 0
        REPEAT
            action ← dql.choose_action(state)
            reward, newState, done ← env.step(action)
            dql.learn(newState)
            state ← newState
            score ← score + reward
        UNTIL done
    END FOR EACH
END FUNCTION
```

Algorithm E.2: DQL - Deep-Q-Learning class

```
CLASS Deep-Q-Learning()
    method
        FUNCTION choose_action
            input state (list)
            output action (int)
        END FUNCTION
        FUNCTION learn
        # use the new state to update the NN
            input newState (list)
        END FUNCTION
END CLASS
```

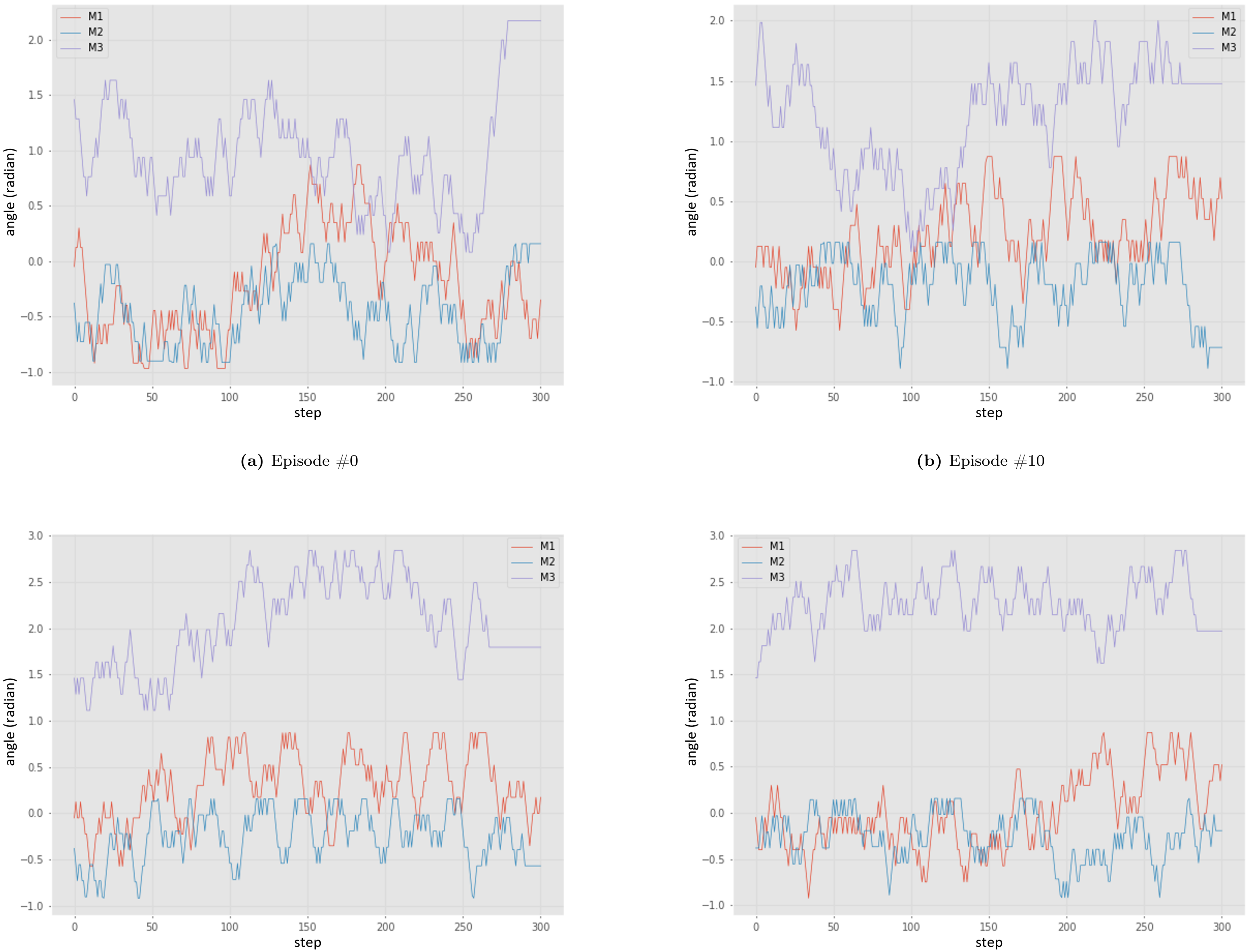
Algorithm E.3: DQL - Environment class

```
CLASS Environment()
    attribute
        robot ← new Yokobo()
        nep ← new Nep()

    method
        FUNCTION reset()
            output state (list)
            robot.reset()
            state ← nep.read_data() + robot.pad()
        END FUNCTION
        FUNCTION step
        # apply the action and observe new state, if it is a terminal state, then the step is 'done'
            input action (int)
            output reward (int), newState (list),
                   done (bool)
        END FUNCTION
END CLASS
```

Algorithm E.4: DQL - Yokobo class

```
CLASS Yokobo()
    attribute
        motors ← list of new Motor()

    method
        FUNCTION reset()
            motors.set_to_origin()
        END FUNCTION
        FUNCTION pad
        # Algorithm described in section 7.2.5
            output pad (int)
        END FUNCTION
        FUNCTION move
            input command (list)
            motors.set_command(command)
END CLASS
```

E.2 Results



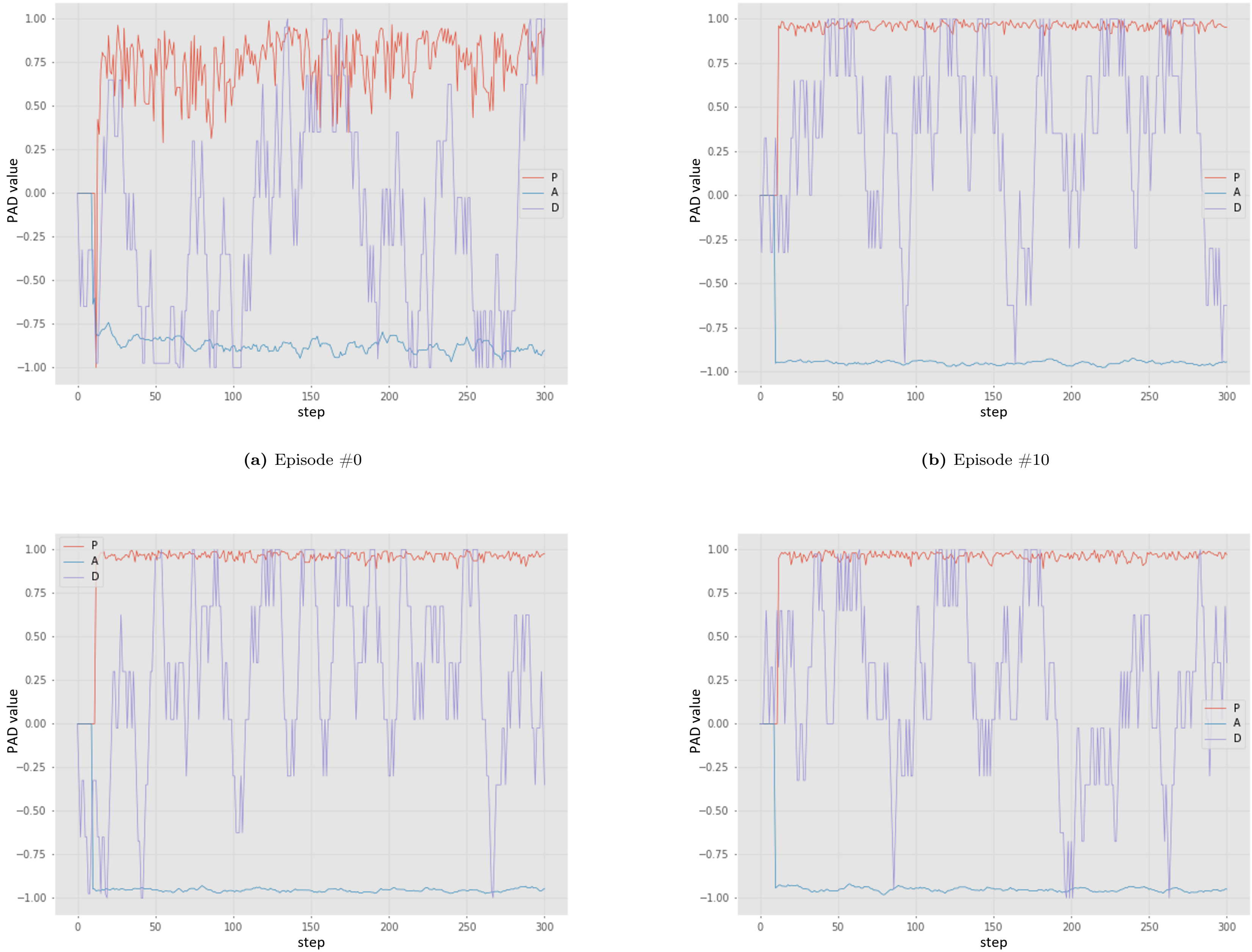
(a) Episode #0

(b) Episode #10

(c) Episode #250

(d) Episode #498

Figure E.1: The angular position (in radian) of each motor over steps, for some episodes



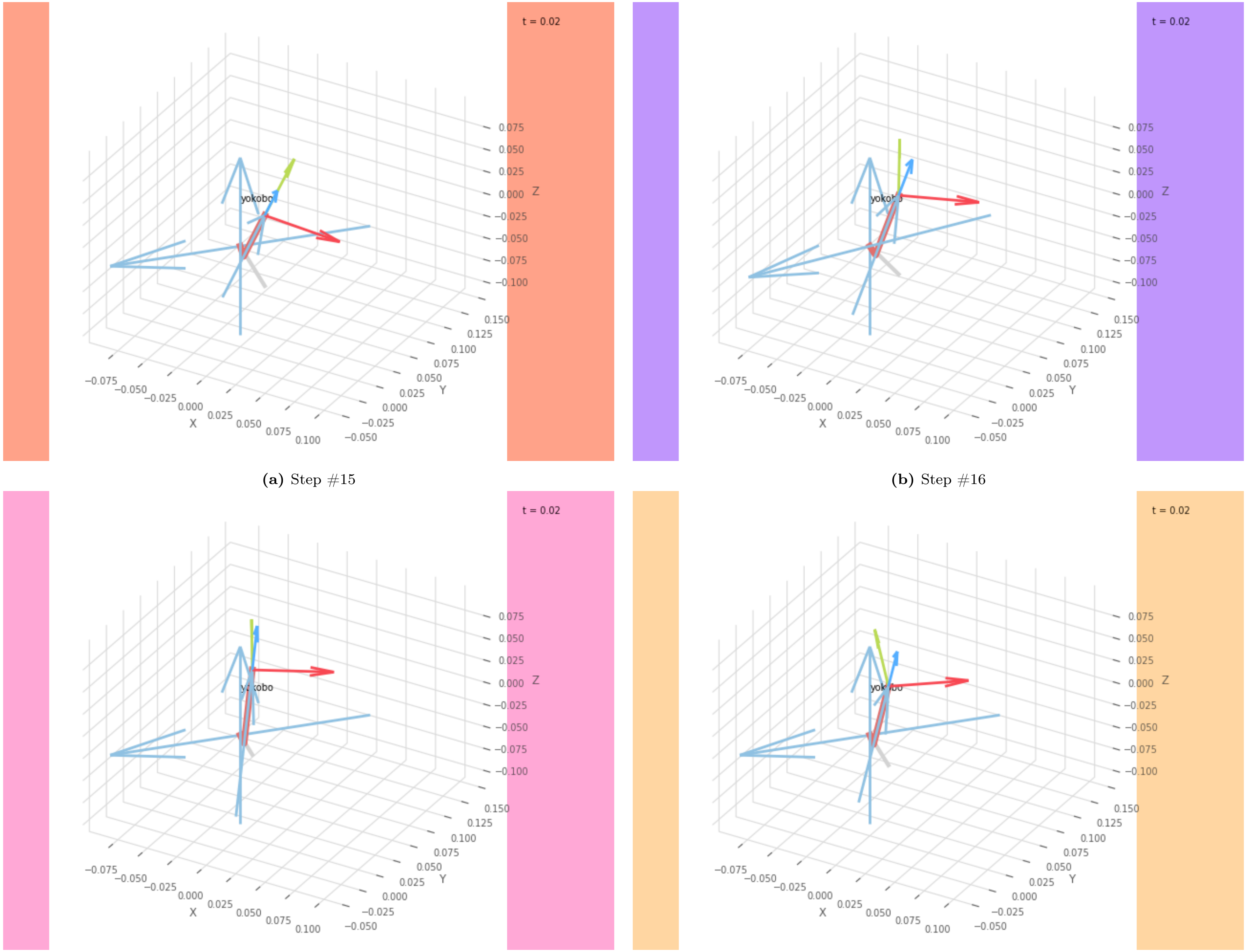
(a) Episode #0

(b) Episode #10

(c) Episode #250

(d) Episode #498

Figure E.2: The PAD of Yōkobo over steps, for some episodes



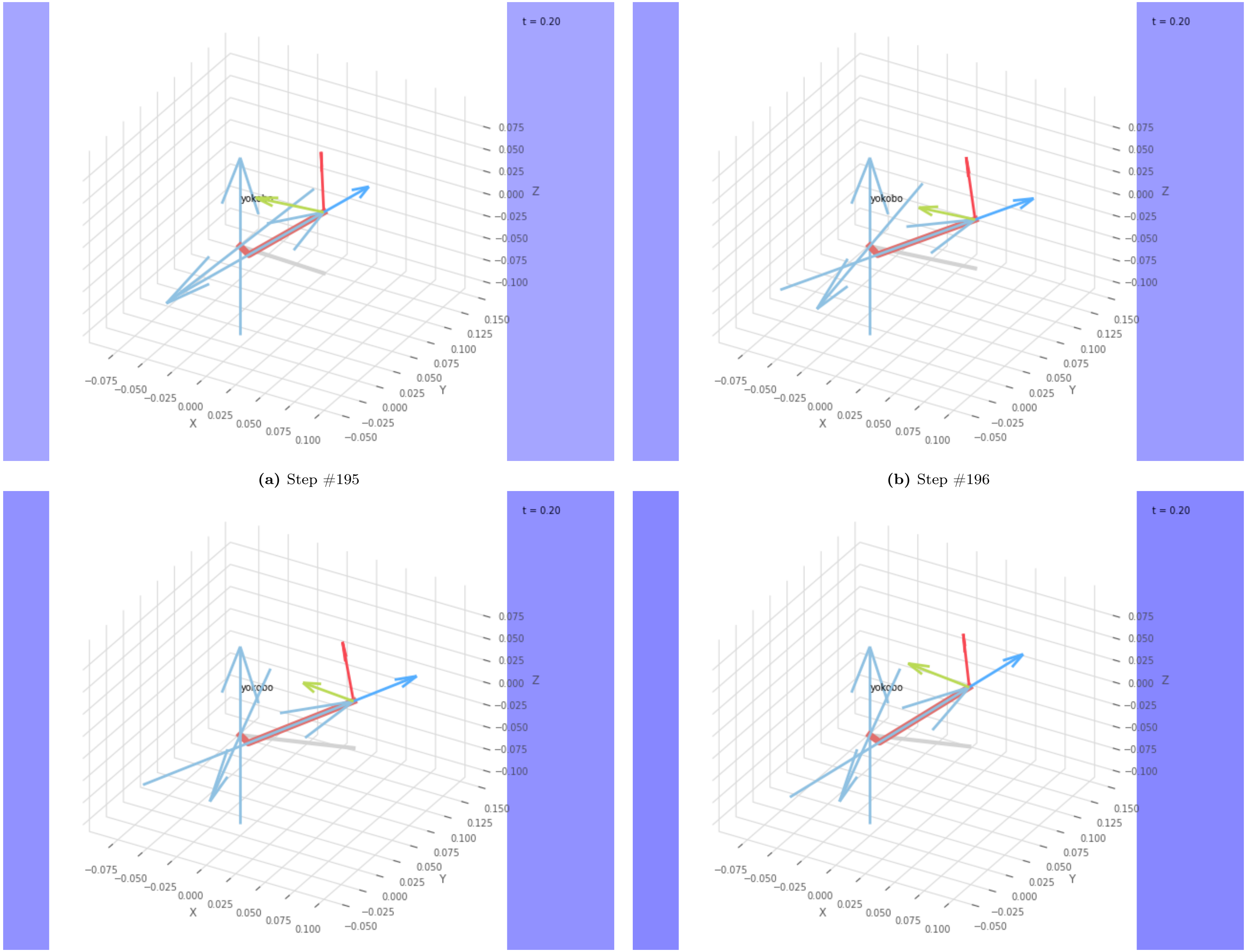
(a) Step #15

(b) Step #16

(c) Step #17

(d) Step #18

Figure E.3: Posture of Yōkobo with the colour of the light for four consecutive steps at episode 0



(a) Step #195

(b) Step #196

(c) Step #197

(d) Step #198

Figure E.4: Posture of Yōkobo with the colour of the light for four consecutive steps at episode 498