

Faculteit Letteren en Wijsbegeerte
Bachelorscriptie Taal- en Letterkunde
Bachelor Engels-Frans

Natural Language Processing Research with NAO Robots

Jessica De Smedt

Promotor: Prof. Dr. Walter Daelemans

Assessor: Dr. Guy De Pauw

Universiteit Antwerpen

Academiejaar 2014-2015

Ondergetekende, Jessica De Smedt, studente Taal- & Letterkunde Engels-Frans, verklaart dat deze scriptie volledig oorspronkelijk is en uitsluitend door haarzelf geschreven is. Bij alle informatie en ideeën ontleend aan andere bronnen, heeft ondergetekende expliciet en in detail verwezen naar de vindplaatsen.

Borgerhout, 08/05/2015

Preface

In the last couple of years, social robots have appeared more and more often in Belgian news broadcasts. NAO, as one of the most popular models, is increasingly used in medical settings as a therapeutic companion for autistic children and the elderly. His cute appearance and behaviour can easily capture the heart of anyone who meets him. As they did mine. Meeting the little guy at the university suddenly solved the problem of trying to come up with a suitable topic for this bachelor's thesis. Not only would it be the perfect combination of my two fields of interest (I earned a bachelor's degree in Applied Informatics – Software Management before I started to study Linguistics), it would also be a challenge. Natural language processing with NAO robots is a relatively new topic on which few studies have focussed, and thus sources could turn out to be rather scarce. Always loving a good academic challenge, and attracted by the prospect of writing a master's thesis on the same topic, the decision was easily made. All expectations were met: it has been a fascinating and enriching challenge, which I could not have completed without the help of several people. Therefore, I would first of all like to express my gratitude towards Professor Dr Walter Daelemans, supervisor of this bachelor's thesis, for spending so much of his time on helping me write and correcting my mistakes and for not imposing a limit on the amount of pages. Secondly, I would like to thank Dr Guy De Pauw for reading and evaluating my work. Last but not least, I would like to thank Philip Carels, my significant other, for proofreading my text and helping me in any way he could

Table of Contents

1.	Introduction	9
2.	Natural Language Processing (NLP).....	12
2.0	Introduction	12
2.1	Natural Languages and Artificial Languages	12
2.2	Computational linguistics.....	12
2.3	Challenges of Natural Language Processing.....	13
2.4	Deductive versus Inductive NLP systems	15
2.5	The Uncanny Valley.....	15
2.6	Recent Research on the Uncanny Valley	16
3.	NAO Robots.....	18
3.0	Introduction	18
3.1	A Family of Robots	18
3.2	Specifications	19
4.	Human-Robot Interaction through Natural Language.....	21
4.0	Introduction	21
4.1	Language Grounding.....	22
4.2	Natural Language Frameworks	23
4.2.1	A Frame-based Dialogue Framework	24
4.2.2	An Event-based Dialogue Framework	24
4.2.3	A Reward-based Meta-cognitive Framework.....	26
4.3	Turn-taking.....	28
4.3.1	NAO’s Turn-taking Behaviour.....	28
4.3.2	Kismet’s Turn-taking Behaviour.....	31
4.4	Problems with Dialogues.....	32
4.5	Open-Domain versus Closed-Domain Dialogues.....	34
4.6	Strict versus Flexible Hierarchical Dialogue Control (HDC).....	36
4.7	Cooperative Tasks	38
4.8	Semantic Gestures	41
4.9	End-user Programming.....	44
4.9.1	User-friendly Programming.....	44
4.9.2	Cybele : a Motion Description Language.....	45
4.9.3	RIOLA: A Robot Interaction Language	46
4.10	Conclusion.....	47
5.	Communication of Emotions.....	51
5.0	Introduction	51
5.1	Emotion Expression	52

5.1.1	NAO's Emotional Body Language	52
5.1.2	An Affect Space for NAO	53
5.1.3	The 2009 Library of Emotional Expressions for NAO	55
5.1.4	The 2011 Library of Emotional Expressions for NAO	55
5.1.5	ICat	59
5.1.6	Kismet	60
5.1.7	Brian	61
5.1.8	KOBIAN	64
5.2	Emotion Detection.....	66
5.2.1	NAO as a Detector of Human Emotions	66
5.2.2	Emotion Detection in the ROMEO Project.....	68
5.2.3	Brian as a Detector of Human Emotions	70
5.3	Conclusion.....	71
6.	Influence of Personality Traits	74
6.0	Introduction	74
6.1	Personality Types for Social robots.....	75
6.2	Personality Matching.....	75
6.3	The Effects of the Task on the Perceived Personality of a Robot	77
6.4	The Effects of Group Interactions on the Perceived Personality of a Robot.....	79
6.5	The Effects of Neighbouring Cultures on the Perceived Personality of a Robot	81
6.6	Conclusion.....	84
7.	Case Studies	85
7.0	Introduction	85
7.1	Influence of Embodiment.....	86
7.2	NAO in Autism Therapy	89
7.2.1	ASK NAO	89
7.2.2	NAO and the National Autism Society of Malaysia	90
7.2.3	A Customizable Platform for Robot Assisted ASD Therapy	93
7.2.4	Robot Assisted Pivotal Response Training	95
7.2.5	Conclusion.....	96
7.3	NAO in Diabetes Therapy: The ALIZ-E Project.....	97
7.3.1	Requirements for Robots in Diabetes Therapy.....	98
7.3.2	Two Robotic Companions.....	99
7.3.3	Children's Adaptation in Multiple Interactions with NAO	101
7.3.4	Conclusion.....	103
7.4	NAO as a Teaching Assistant for Sign Language	103
8.	Conclusion.....	106

9. Further Research.....	111
10. Bibliography.....	113
11. List of Figures.....	120
12. List of Tables.....	122
Appendix I: NAO versions.....	123
Appendix II: NAO version and body types diagrams.....	124
1. Versions.....	124
2. Body Types.....	126
Appendix III: NAO Evolution Datasheet.....	128
Appendix IV: A frame-based Dialogue System.....	131
Appendix V: An Event-based Dialogue System.....	132
Appendix VI: WikiTalk-based Open-dialogue with NAO.....	133
Appendix VII: High-level Architecture of a HDC System.....	135
Appendix VIII: PRT Scenario for <i>Boomgaardje</i> and prompting diagram.....	136
Appendix IX: Story with TSL for NAO.....	138

1. Introduction

When humans think about robots, they do no longer only think about mechanical arms that facilitate the production process of cars. They think about artificial creatures with a human-like appearance whose intelligent capabilities may one day very well grow beyond those of human beings. They think about science fiction films in which robots, initially created to assist humans, turn against their makers in an ultimate battle for world dominance. However, more and more, they also think about cute, pet- or child-like companions who help in hospitals or residential care homes or who teach autistic children social behaviours. The popularity of robots is increasing and they will mostly likely become an integrated part of human lives in a matter of years.

Therefore, many interdisciplinary studies are being set up, combining the expertise of multiple fields (such as robotics and linguistic) to develop these robotic companions of the future. The scientific discipline of natural language processing (NLP) will play an important role in this process, as robots will need to be able to deal with natural language. The function of the computational linguist, however, is not limited to making sure that the robot can use and understand language: NLP also becomes an issue in wider domains, such as the communication of emotions and personality traits. These two combine the knowledge of fields such as linguistics and psychology to create communicational patterns consisting of both body language and natural language.

Eventually, science hopes to develop an artificial kind of intelligence (AI) which will meet – or even surpass – the capabilities of the human brain. One of the steps to accomplish this difficult endeavour is to reach a profound understanding of natural language. However, although the field continues to advance rapidly, true AI remains out of reach. The importance of NLP can thus not be overestimated to move the field forwards in the direction of this complicated form of AI.

In the scope of this bachelor's thesis, I will try to summarize the state-of-the-art in the field of NLP with NAO robots, a humanoid created by the French company Aldebaran. NAO is one of the best-known robots on the market because of the many health care and teaching applications in which this humanoid has been used worldwide. Furthermore, NAO is also well known as the standard model for the yearly RoboCup, a soccer competition for robots.

After this brief introductory chapter, NLP and its challenges will be explained in more detail in chapter 2. The differences between natural and artificial languages will be presented, which will bring us to the most difficult problem for NLP: the notion of ambiguity. Computer systems need a way to deal with the ambiguous nature of natural language, which has proven to be a difficult obstacle. Furthermore, traditional views on the uncanny valley problem as described by Mori are compared to more recent findings.

In chapter 3, the robotic company Aldebaran and NAO are presented. In this chapter, NAO is presented as a member of a robotic family (together with two other humanoids developed by this company). This approach was chosen because it clearly shows the vision of the company and the goal of robotic companions. Next, NAO's specifications – and especially those relevant for NLP – are discussed.

The next three chapters are dedicated to different applications in which NLP plays an important part. Chapter 4 deals with human-robot interaction (HRI) through natural language; chapter 5 covers both the expression and the detection of emotions, and chapter 6 presents studies on the influence of personality traits on a human's perception of a robot. As mentioned before, these last two applications also profit from developments in NLP, although to a lesser extent than pure natural language based HRI. In chapter 4, the problem of language grounding is first explained, after which several natural language frameworks are discussed. In the following section, NAO's turn-taking behaviour – an aspect of natural language which is needed to make HRI feel natural – is compared to that of another robot, Kismet. Developed by MIT, she is one of the best-known early sociable robots. Her goal is to learn social behaviour through HRI. Next, several problems regarding dialogues in natural language are discussed. Systems working

with open-domain dialogues will be compared to others using closed-domain dialogues, and flexible hierarchic dialogue control will be contrasted with strict hierarchic dialogue control. In the next section, we will take a closer look at cooperative tasks. These are essential to NLP as controlling such tasks is one of the main functions – and advantages – of natural languages. In the future, robots will need to work together both with other robots as with humans. Then, the topic of semantic gestures will briefly be introduced. Unlike pure body language gestures, these are movements more closely related to sign language. This means that they do not express emotions or personality but transfer semantic meaning. For example, think about the typical thumbs-up gesture to say ‘*Well done*’. These kinds of gestures are important as well in communication between humans, and therefore, it should be studied to which extent they are transferable to robots. In the penultimate section, end-user programming will be described. As robot designers cannot expect every user to be a programmer, it is important to find ways in which humans will be able to communicate with robots without following a course in coding. This fourth chapter will be concluded by a discussion of RIOLA, a robot interactional language.

Chapter 5 consists of two main parts: the expression of emotions and the detection thereof. In the first part, we will examine NAO’s body language through a series of studies such as the development of an affect space (Beck et al., 2012). We will also briefly discuss two libraries of emotions that have been created for NAO. This part ends with the comparison of NAO’s emotional body language to four other robots: iCat (a research robot by Philips), Kismet, Brian (a Canadian humanoid companion for the elderly) and KOBIAN (a Japanese humanoid research robot for the study of robotic emotions). In the second part, different studies on NAO’s capabilities to detect human emotions are presented and compared to Brian’s capabilities.

In chapter 6, different common personality types for robots will be presented, after which the main problem of this field of study will be examined: what kind of personality should a robot have to optimise HRI? Should a robot’s personality match a user’s personality or complement it? Should robots possess a distinguishable human-like personality or should they be clearly robotic? Should robotic personalities be programmed or learned? Literature has not yet found a conclusive answer to these questions and therefore, chapter 6 contains various visions which do not always agree. Among these visions are three young voices, who took part in an IT conference for their university in the Netherlands. The results of their studies were not revolutionary, but their input is valuable, however, as they take a completely different position as compared to the established researchers. While literature is mainly divided into two opposing camps (those who believe the personality of a robot should match the user’s and those who believe it should complement it), they present a hypothesis in which neither vision is relevant. They favour a theory in which other, often external, factors are significant for personality matching, such as the particular task performed by the robot, the effect of group interactions and the effect of cultural background.

Next, in chapter 7, several case studies are discussed. These focus mainly on three different real-life situations in which NAO is used and in which NLP plays an important part. First of all, NAO as a companion for children with autism is discussed. In this section, the ASK NAO programme by Aldebaran is first introduced. This acronym stands for Autism Solution for Kids and is an initiative launched in order to support research into robot-aided therapies for Autism Spectrum Disorder (ASD). Next, we will take a closer look at studies conducted in context of the National Autism Society of Malaysia. As mentioned before, users should not be required to acquire programming skills to successfully communicate with robots, but neither should therapists. Therefore, platforms for robot-aided ASD therapies need to be developed that are user-friendly. These platforms should, moreover, be customizable, as autism is different for each child. These kinds of platforms are the topic of the penultimate section on ASD therapies. The final section is dedicated to robot Assisted Pivotal Response Training, an established method in ASD therapies.

Secondly, NAO as a companion for children with diabetes is presented, within the context of the ALIZE project. This project was an international collaboration, supported by Europe, between 2010 and 2014.

They wanted to create robotic companions and monitors to support hospitalized children (mainly suffering from diabetes). One of the topics studied in the scope of this project is the way in which children adapt to HRI in multiple interactions with NAO. This is important because, as a companion, NAO will need to become an integral part of the children's daily life. The robot should therefore feel familiar to the children, and the long-term HRI should be perceived as natural and comfortable. To this extent, it is also important to determine which features are necessary to include in the design of a robotic companion and which in the design of a robotic monitor. These two are entirely different functions for a robot and should thus be implemented in another manner.

Thirdly, NAO's usefulness in the context of sign language teaching will be examined. As sign language is often teacher-dependent, it could be useful to introduce a robot assistant to the classroom. This would limit the problems that arise when the human teacher needs to be replaced, as robots are able to endlessly repeat gestures in exactly the same way.

Chapter 8 contains the conclusion about the state-of-the-art of NLP with NAO robots, based on the selected studies as described above, and chapter 9 presents some possibilities for further research.

After the bibliography and the lists of figures and tables, several appendices can be found. These were included because they contain relevant information on the topics discussed, but they were considered too extensive to be integrated into the main body of this bachelor's thesis.

As this is a study of the available literature, many authors are cited. Whenever an extended block of text was dedicated to the work of a particular researcher or research team, a footnote was added to indicate the source and to limit the number of in-text references.

Finally, I would like to explain the use of pronouns in this bachelor's thesis. As robots can be seen as non-living objects, they are generally referred to as *it*. Here, *it* will be used when discussing robots as either commercial items or machines. However, as the robots discussed are specifically designed to be human *companions*, it seems fitting to refer to particular robots as *he/she* as this emphasises the emotional relationship between human and companion.

2. Natural Language Processing (NLP)

2.0 Introduction

In this second chapter, we will, first of all, explain the difference between natural languages and artificial languages. Then, we will take a closer look at one of the fields interested in natural languages, namely computational linguistics or natural language processing (NLP). Furthermore, we will touch briefly upon the different challenges of NLP, among which, dealing with ambiguity is the most problematic one. We will also shortly introduce two of the main approaches in NLP (inductive and deductive methods), after which we will conclude with a discussion of the uncanny valley issue. This phenomenon is widely known in the field of robotics as it seems to limit the freedom of the designers. However, new research indicates that these limitations might not be accurately depicted on the uncanny valley graph.

2.1 Natural Languages and Artificial Languages

Languages can be divided into two main categories: natural languages and artificial languages (Beardon et al., 1991).¹ Natural languages are those that have not been artificially created by humans but have evolved naturally into mother tongues. Their prime function is to allow humans to communicate with others, without there being any restriction on the possible topics of that communication, or on the situation in which the communication takes place.

By contrast, artificial languages have been consciously created by humans to fulfil specific functions. As Beardon et al. point out, this kind of languages (e.g. programming languages) usually impose restrictions on their use, for example, restrictions on ambiguity. Individual words, sentences and phrases can be (and often are) ambiguous in natural language, which poses one of the greatest challenges for natural language processing. Therefore, artificial languages impose rules to avoid ambiguity, for example by using words with a unique fixed meaning (reserved words) in programming languages.

2.2 Computational linguistics

Natural languages are studied in different fields: in linguistics in general, but also in computational linguistics (Daelemans, 2013)². This interdisciplinary field of study examines similar questions as linguistics (e.g. how can text be transformed into meaning?), but it shares its research method with artificial intelligence (AI), which belongs to the field of computer science. Computational linguistics create computer models, similar to those used in AI to develop intelligent systems. A key concept within the field of AI is the “Intelligent Agent”, a computer program that can observe and interact with its environment, solve problems and learn. These agents need to be capable of using natural language, which is the task of computational linguistics, or natural language processing (NLP) as it is called in AI.

It is important to keep in mind that AI does not limit itself to models of human intelligence. One of its main hypotheses, called the Physical Symbol Systems Hypothesis (PSSH), argues that intelligent behaviour can be described by abstract manipulation of symbols, independent of the implementation thereof in the human brain. This means that if NLP succeeded in defining knowledge and cognitive processes as representations and algorithms, a computer (or more specifically for this bachelor’s thesis, a robot) could be said to be intelligent as well. The PSSH, formulated by Allen Newell and Herbert Simon, allows algorithms to be represented as structures, so that they can be manipulated by other algorithms (Gillis et al., 1995). This recursion explains the concept of learning, as ‘the mind can change itself in useful ways by manipulating its own mental structures and program by means of a learning program (Gillis et al., 1995). In this hypothesis, the manipulation of symbols is the only necessary condition for intelligent behaviour (Gillis et al., 1995). Figure 1 shows a diagram of the PSSH.

¹ The section *Natural Languages and Artificial Languages* is based on (Beardon et al., 1991).

² The section *Computational Linguistics* is based on (Daelemans, 2013), unless otherwise indicated.

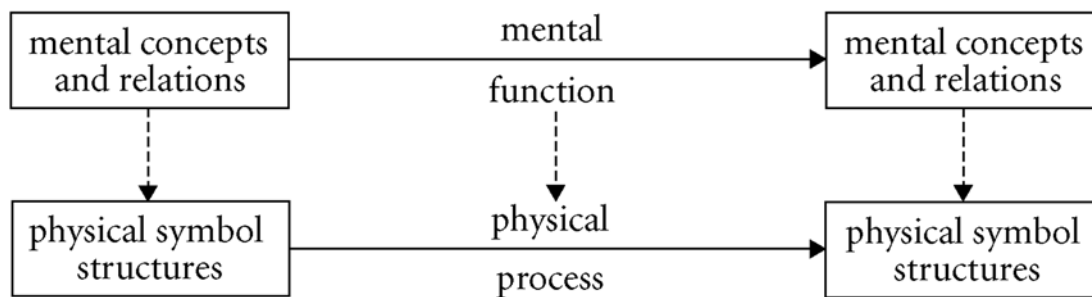


Figure 1 Physical structures and processes represent mental functions (Gillis, Daelemans & De Smedt, 1995)

2.3 Challenges of Natural Language Processing

As described above, the most difficult hurdle to cross in the field of natural language processing is the problem of ambiguity. Computational linguistics describes language processing as a series of transformations between symbolic linguistic representations (Daelemans, 2013)³. Two types of transformations are important when attributing meaning to text: segmentation and identification. Segmentation subdivides input text into smaller units, which are transformed by the process of identification into output elements. Both transformations are confronted with the problem of ambiguity, which interferes on all levels of language description, even though most users are unaware of its presence.

A first problem is lexical ambiguity, as most words can have multiple meanings. In a sentence like *‘Philip likes reading about stars’*, the word *star* is ambiguous, as it is unclear whether the subject of the sentence enjoys stargazing or reading gossip magazines.

Another kind of ambiguity that can be encountered is morphological, as demonstrated by the following sentence: *‘I shut the door’*. In this sentence, the verb *shut* is morphologically ambiguous because of the fact that the verb form is the same in the present tense and in the past tense.

At a higher level of language description, syntactic ambiguity poses large problems to computer systems, because some parts of speech can be attached to several other parts. This kind of ambiguity can be found in sentences like *‘Philip saw the man with the telescope’* (Inspired by: Kraf & Trapman, 2006). It is unclear whether Philip used a telescope to spot the man or whether he saw a man carrying a telescope.

Finally, there is also ambiguity at the level of the discourse, as shown in the following sentence: *‘The judge convicted the man because he feared he would kill again’*. This sentence is ambiguous, because it could either be the judge who feared the man would kill again or the murder himself.

Humans also have to solve ambiguity problems, but as stated earlier, they do this most of the time without even realising the sentence poses a difficulty in the first place. They are capable of reducing the number of possible meanings because they possess knowledge of the world they navigate. When they are confronted with sentences like *‘A Catholic priest married my son on Tuesday’*, they discard the possibility of the priest being wed to the son, as their knowledge of the world informs them that Catholic priests do not wed. A computer system, however, does not always have access to the same information, which is one of the most difficult problems in NLP.

Next to ambiguity, there are also other difficulties, such as the complexity of natural language (Beardon et al., 1991)⁴. These authors point out that ‘the structure of statements in artificial languages is usually kept very simple’ (Beardon et al., 1991). This stands in stark contrast to the structure of natural languages, which can be very complex. The complexity of these structures renders the development of

³ The information on ambiguity in this section is based on (Daelemans, 2013), unless otherwise indicated.

⁴ The information on other difficulties for NLP in this section is based on (Beardon et al., 1991).

a natural language parser much more demanding than it would have been if they had been as straightforward as artificial language constructions.

Furthermore, the fact that artificial languages are developed for a specific purpose entails that it is less difficult to find a single way to represent the meaning of everything a particular language can express. According to Beardon et al., the meaning of a fragment of programming code can be seen as ‘the machine code that it produces to run on a computer’ (Beardon et al., 1991). For natural language units, however, such definitions cannot that easily be found, as natural language can be used in a wide variety of situations (commanding, describing, asking, etc.).

A fourth difficulty for natural language processing arises when we separate the part of a system that processes the structure of an utterance from the part that processes its meaning. Artificial languages differ from natural languages because of the relationship between these two parts. To compile computer code, the system first determines whether or not the structure of the code is correct. Only when this step is completed satisfactorily, the meaning of the processed structure will be interpreted. To understand natural languages, however, structure and meaning cannot be separated this easily, as the meaning of an utterance is often needed to process its structure.

Table 1 summarizes the four most important differences between artificial and natural languages that lead to difficulties for natural language processing. All these problems need to be solved to create “conversational agents” or “dialogue systems”: programs that communicate with humans by using natural language (Jurafsky & Martin, 2008)⁵.

Crucial Differences		
	Natural Language	Artificial Language
Ambiguity	Plenty	Controlled
Complexity	High	Low
Representation of meaning	No simple universal way	Simpler
Relationship structure – meaning	Interconnected	Often separable

Table 1 Differences that lead to difficulties for NLP (Based on: Beardon et al., 1991)

Conversational agents not only need to attribute meaning to text, they also need to be able to decide how they should react. Different variants of sentences can be constructed which contain the same information, yet demand another reaction: a request (‘Close the window.’), a statement (‘The window is closed.’) or a question (‘Is the window closed?’). Furthermore, these agents should also know how to be polite. To accomplish these tasks, conversational agents should thus possess a certain kind of pragmatic or dialogue knowledge. Table 2 summarizes the different sorts of knowledge of language needed to create conversational agents.

Knowledge of language	
	Knowledge about:
Phonetics & Phonology	How words are pronounced in terms of sequences of sounds and how each of these sounds is realized acoustically.
Morphology	The meaningful components of words
Syntax	How words in a sentence are related to each other
Semantics	The meaning of words
Pragmatics	The relationship between the meaning of the words and the intention of the speaker
Discourse	Linguistic units that are larger than single utterances

Table 2 Different kinds of knowledge needed for conversational agents (Based on: Jurafsky & Martin, 2008)

⁵ The information on conversational agents in this section is based on (Jurafsky & Martin, 2008).

2.4 Deductive versus Inductive NLP systems

There are two important approaches in computational linguistics: a deductive and an inductive one (Daelemans, 2013).⁶ The deductive method focussed on rules and formal descriptions of language to transform input into output. This method was predominant up until the second part of the 90s, when it was replaced by the inductive method and its focus on general learning capacities. Inductive systems learn models by means of statistical pattern recognition techniques based on training data, which they use to calculate their output. Whichever system is opted for, it should have access to the kinds of knowledge described in Table 2.

2.5 The Uncanny Valley

By learning language, robots become more and more similar to human beings, which is one of the goals of robotics (Mori, 2012)⁷. However, it is important to keep in mind that the relation between “human likeness” and “affinity” is not a simple one. If you increase the human likeness of an object, the affinity felt by humans for it will increase as well – at first. Somewhere around 60% of human likeness, there is a sudden drop in the affinity for the object, which is called the “uncanny valley”. Objects located in this area are perceived as creepy or unsettling rather than pleasant. When the object reaches a human likeness of about 90%, it is no longer situated in this valley. Figure 2 shows the uncanny valley.

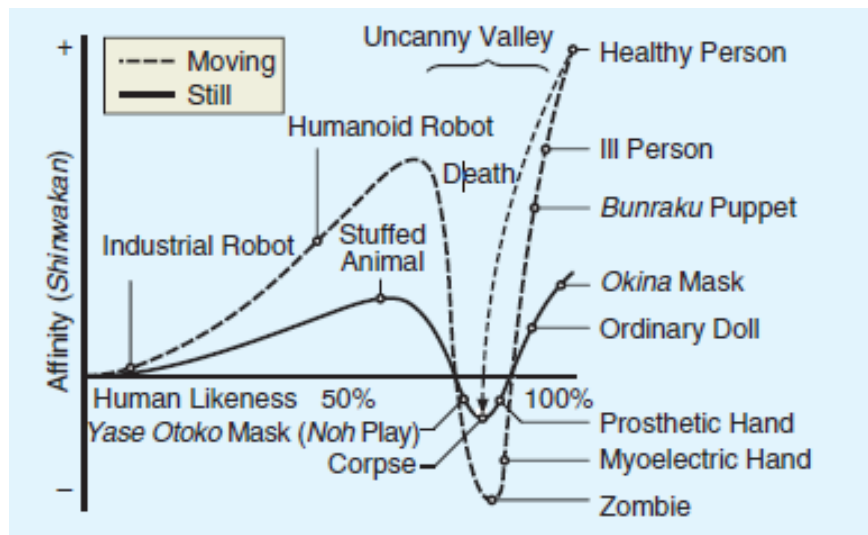


Figure 2 The Uncanny Valley (Mori, 2012)

As shown in Figure 2, movement has an important influence on the uncanny valley graph. Humanoids (like NAO) are situated right before the first peak. They are reasonably similar to humans who are thus inclined to feel affectionate towards them. However, experiments have shown that if a humanoid, one which has artificial muscles to simulate facial expressions, is programmed to smile at half the speed of a human, it is perceived as unsettling. According to Mori, a variation in movement can easily cause something, close to a human being in terms of appearance, to fall into the valley. While death causes the affinity felt towards a living, healthy person to tumble to the bottom of the valley of the graph representing non-moving subjects, we still feel more warmly towards a corpse than towards a zombie, which is the lowest point on the graph representing moving subjects. When designing robots, it is thus important to avoid the valley at all costs. Mori recommends designers to work towards the first peak of the graphic, as this would render a robot which moderately resembles a human and evokes a great feeling of affection. However, new research suggests that the model as designed by Mori might have been too simplistic to describe the effect of the uncanny valley accurately.

⁶ The section *Deductive versus Inductive NLP Systems* is based on (Daelemans, 2013).

⁷ The section *The Uncanny Valley* is based on (Mori, 2012).

2.6 Recent Research on the Uncanny Valley

First of all, it is important to realise that, according to Mori, robots are faced with the uncanny valley problem at different levels: the use of language, the display of intelligence and the way of using their bodies.

Being able to communicate with humans in natural language often causes a robot to be perceived as “intelligent” by its users. This immediately raises the question of the uncanny valley: can humans still feel affective towards a robot if this robot could interact with them in a human-like manner? Or would this make them seem too realistic and too human to be trusted? Judging from the uncanny valley graph, it could be concluded that language would make a robot tumble down into the valley; after all, even though speech is not mentioned on the axes, it is a typical human characteristic. However, many studies have indicated that this is not necessarily the case. For example, Hanson et al. have developed a social robot PKD (named after Philip K. Dick, the late science-fiction writer), which used several mechanisms to deal with natural language (Hanson et al., 2005). They found that they were not hindered by the uncanny valley effect: in fact, they believe that robots should become as human-like as possible, if we are to learn more about social intelligence (Hanson et al., 2005). According to them, the only way to get past the valley is to explore it entirely first (Hanson et al., 2005).

Yet, other researchers, like Becker-Asano et al., are more tentative in their findings. This team has conducted an experiment in which visitors of a festival were asked to interact with their android robot Geminoid HI-1 (Becker-Asano et al., 2010). The results were mixed: while some participants reported to enjoy talking with the robot, others thought its speech revealed the fact that it was not a real human (Becker-Asano et al., 2010). Furthermore, one of the participants mentioned that he liked the conversation at first, until he realised that he had been talking with a computer, after which he experienced ‘a weird feeling’ (Becker-Asano et al., 2010). This might indicate that the experience of the uncanny valley can indeed be increased by language, but it might also reveal an important element which cannot be seen on the graph: personal factors. Only some visitors reported an uncanny feeling, which might lead to the conclusion that the uncanny valley graph is too general, not taking into account the interpersonal differences in perception. We will now compare these findings with experiments in which the uncanny effect of the robot’s physical appearance was tested.

Recent developments in animation have increased visual realism of characters, which is a combination of both physical and behavioural realism (Beck et al., 2012)⁸. Creators of animated characters believed that this increase would also lead to an increase of believability, but in reality, these characters were, likewise, confronted with the uncanny valley. These observations were thought to impose limits on the extent to which humanoids, like NAO, would be able to mimic humans. After all, many of these humanoids were especially designed to be personal companions and therefore, they should never make humans uncomfortable. However, the concept of the uncanny valley was not based on systematic experiments. This might indicate that there are much more elements which influence the acceptability of a robot, outside of its resemblance to humans. Beck et al. suggest that the effect might also be due to the robot’s body language. This is supported by animation theories that imply that emotion should always be expressed through a combination of body and face, as the character would otherwise look unnatural to a viewer.

To explore their hypothesis, Beck et al. designed an experiment in which participants were confronted with three types of characters: a real actor, a realistic animation and a simplified one. Based on Mori’s uncanny valley, they made two predictions: (a): ‘A highly realistic character will be harder to interpret and will also be perceived as less emotional’ and (b): ‘As characters get more realistic, they will be subject to a drop in believability and naturalness’ (Beck et al., 2012). It was thus predicted that the participants would consider the actor better than the simplified character, which would in turn be

⁸ The section *Recent Research on the Uncanny Valley* is based on (Beck et al., 2012), unless otherwise indicated.

considered better than the realistic character. Furthermore, two personal factors were taken into account: the emotional intelligence (EQ) of the participants and their experience with games and animations.

The actor had to perform ten different emotions, of all of which two variants were made: normal emotions and exaggerated ones. These were closely mirrored by the two types of animations, except for small characteristics such as breathing. Participants were then asked to evaluate the videos they were shown: (a) Which emotion is being displayed; (b) How strong is the displayed emotion; (c) How natural and (d) how believable does it come across.

The results indicated that the type of character had no effect on the identification of the emotion. This means that when physical realism is simplified, this does not negatively affect the transmission of emotional information. However, the type of character did have an effect on the strength of the emotion: those performed by the actor were perceived as stronger than those performed by both kinds of animation. Surprisingly, there was no difference in perception of strength between the emotions expressed by the realistic and the simplistic animation. This might indicate that emotional strength is not solely created by physical realism, as that would have meant that the emotions performed by the realistic character would have been perceived as stronger than those by the simplified character as well. The gap between the perception of the actor and the perception of the animations might be explained by the fact that the animations did not display microgestures such as breathing and sighing.

Character type had also an effect on the believability of an emotion: participants were more inclined to believe the actor than the characters, and they perceived the emotions of the realistic character as more believable than those of the simplified character. Similarly, the emotions of the actor were considered to be more natural than those of the realistic characters, which were in turn perceived as more natural than those of the simplified character. As there was no difference in secondary cues and microgestures between the two animations, the different perception of the two seems to suggest that it was their physical realism that affected their believability and naturalness.

These results thus seem to contradict the uncanny valley theory, as characters are not considered less believable when they are more realistic. Furthermore, Mori's graph does not take into account any personal differences. The experiment showed that there was indeed no correlation between the EQ of the participant and the correct identification of emotions. However, there was a clear influence of the EQ on the perception of believability and naturalness of emotions expressed by the realistic character. Participants with a high EQ often considered the realistic character more believable. The realistic character was the one that was most likely to be affected by the uncanny effect. The results of this experiment, however, might indicate that individuals with a high EQ are less likely to experience the effect.

Moreover, the results indicated a correlation between experience with video games and the correct identification of the emotions displayed by the actor and by the simplified character, although experience with animated characters had no influence on the identification at all. According to Beck et al., that correlation might be due to the fact that when humans become used to realistic characters in video games, they start to consider them as increasingly believable. This might prevent the feeling of uncanniness from occurring, as this may well be linked to the novelty of being confronted with such levels of realism.

This experiment might thus indicate – like the experiment by Becker-Asano et al. did – that the uncanny valley graph is too simplistic, as it does not take into account personal factors. In reality, each user positions particular characters in other places on the graph, based on their own personal perception and experiences.

3. NAO Robots

3.0 Introduction

In the previous chapter, we introduced some of the general features of NLP. In this chapter, we will take a closer look at some of the robots which are used in this field. Our focus will be on NAO, the best-known robot of the French robotic company Aldebaran. First, we will shortly introduce NAO's family, Pepper and Romeo, in order to get a better view on the context in which NAO came to be. Then, we will zoom in on a more detailed overview of NAO's specifications (primarily the ones that are important to NLP).

3.1 A Family of Robots



Figure 3 The Aldebaran Robotic Family (Aldebaran Robotics, 2015)

In 2005, Aldebaran Robotics was founded in Paris by Bruno Maisonnier, the current CEO of the company. Their vision is to 'build humanoid robots, a new humane species, for the benefit of humankind' (Aldebaran Robotics, 2015)⁹. To accomplish this goal, the company is creating a family of companion robots, which currently consists of three members: NAO, Pepper and Romeo (See Figure 3).

One year after the foundation of Aldebaran, the company created its first NAO prototype. This model was not yet ready to be sold to the general public, but in 2008, NAO managed to position himself in the international spotlight by replacing Sony AIBO in the RoboCup Standard Platform League. This annual soccer competition for robots was originally only open for teams of AIBO robots, but when Sony decided to cancel the production in 2006, the organisation decided that NAO would become the new model (RoboCup, 2015). From thereon, NAO was developed further to become 'a standard in the academic world for research and education' (Aldebaran Robotics, 2015). In 2010, NAO was one of the main attractions at the World Expo and in 2011, a new version was launched. NAO Next Gen had

⁹ The chapter *NAO Robots* is based on (Aldebaran Robotics, 2015), unless otherwise indicated. Available information on the company or on NAO is usually created by Aldebaran itself or by (former) employees. Therefore, this information cannot be regarded as completely neutral. However, some other authors included one or two sentences on the performance or affordability of NAO in their papers. This new – and slightly less biased – information usually corresponded to the information provided in this chapter. When it did not (in case NAO failed the researcher's expectations on some points), the critique was added in the chapter in which these studies were described.

improved at the level of interaction, which allowed its market to be expanded to secondary schools. In 2014, the current version of NAO was released: NAO Evolution.

In 2009, Aldebaran joined the ROMEO project supported by Cap Digital¹⁰, which was continued in 2012 by the ROMEO 2 project. The goal of these projects is to unite different companies to develop a robot staff assistant, Romeo. He can be seen in the middle of Aldebaran's family picture (see Figure 3).

The latest member of the Aldebaran robotic family was introduced in 2014: Pepper. He is their first humanoid that was especially designed to share the lives of human beings. The company's goal is to develop Pepper one step at a time to transform him into a human's full-time companion.

3.2 Specifications



Figure 4 NAO Evolution (Aldebaran Robotics, 2015)

NAO is a humanoid, which means that he is a robot with the proportions of a human. He is 58 cm tall and comes in different colours.¹¹ He is especially designed to be a daily companion: he can recognise humans, communicate with them and help them in their activities. Although he is not entirely ready for use at home, he has become one of the most popular models of robots in educational environments. Nowadays, NAO is used in over 70 countries, from primary education up to university. Eventually, Aldebaran wants to transform NAO into an interactive daily companion who would be perceived as an endearing, living member of the family.

The robot is designed to function as a real companion and it thus needs the capacity to interact with its environment. First of all, NAO needs to be able to see what is happening around him, and therefore, he is equipped with two cameras. Furthermore, he needs to communicate with his users, which becomes possible through touch sensors and four directional microphones. These microphones receive the sound wave at different times, which can be processed to find out where the sound was produced, thereby enabling NAO to locate the source. This method is called "Time Difference of Arrival". NAO can also move freely, because he has 25 degrees of freedom and an inertial measurement unit to decide whether he is sitting down or standing up. The input of these technologies should then be interpreted, which is

¹⁰ Cap Digital is a business cluster which aims to develop innovative technologies in the Paris Region since 2006 (Cap Digital, 2015).

¹¹ This bachelor's thesis will concentrate on the latest version, NAO Evolution. For an overview of the versions, please see Appendix I. There are also several body types available. Please see Appendix II for the diagrams of different versions and types. The body type discussed above is the most complete one, H25.

done by the embedded software in his head. NAO is driven by NAOqi, an operating system especially designed for this robot. Thanks to his lithium-ion battery, NAO has about 1.5 hours of autonomy.

While collecting all this data from the environment is important, the most vital step is of course the interpretation of this data. Therefore, NAO has a set of algorithms that can process faces and shapes. This way, the robot can recognise with whom he is interacting or he can find the objects he needs. To complete this last task, NAO should of course be able to estimate distances. He does this by using a sonar range finder which allows him to detect objects located up to three metres further. NAO, however, does not receive distance information about objects that are closer than 15 cm.

NAO can also be connected to the internet, and on Aldebaran’s website, multiple examples of applications designed by NAO users that use the internet can be found. For example, NAO can use his IP address to locate himself and provide a weather report, or he can read Wikipedia to answer questions about specific topics. For more specifications, please see Table 3¹².

NAO Evolution (H25)	
Company	Aldebaran (France)
Date	2014
Focus	Companionship + Education + Autism
Type	Humanoid
Specifications	
- Height	58 cm
- Sensor Network	
o Cameras	2 (Forehead, mouth)
o Directional Microphones	4 (front, right, rear, left)
o Sonar Rangefinder	✓ (2 transmitters, 2 receivers)
o IR emitters & receivers	2
o Inertial Board	1
o Tactile Sensors	9 (Top of head, hands)
o Pressure Sensors	8
- Connectivity	
o Wi-Fi	✓
o Ethernet	✓
o Network compatibility	WPA / WEP
o Infrared	✓
- Degrees of freedom	25
- Communication Devices	
o Voice Synthesizer	✓
o LED lights	✓
o High-Fidelity Speakers	2
- CPU	
o Type	Intel ATOM 1.6ghz (in head)
o Location 1 st CPU	Head
o Location 2 nd CPU	Torso
- Operating System	NAOqi 2.0
o Kernel	Linux
- Battery	48.6-watt-hour battery
- Language	Up to 19 languages
o Text-to-Speech	✓
o Voice Recognition	✓

Table 3 NAO Evolution specifications (Based on: Aldebaran Robotics, 2015)

¹² More details on NAO Evolution’s specifications can be found in his datasheet, included as Appendix III.

4. Human-Robot Interaction through Natural Language

4.0 Introduction

In the previous chapter, NAO and his family have been presented; from this chapter onwards, we will take a closer look to NAO's competence with natural language. This competence is essential to human-robot interaction (HRI), which is one of the key domains of robotics. More and more robots are being developed in order to find suitable artificial companions for humans. These companions are meant to be used in a variety of functions. For example, robots could be used to assist the elderly or to care for the sick. Moreover, robots have proven to be excellent companions for children with illnesses or for autistic children.

In order to become such a companion, robots should be able to communicate with their users, and the most obvious way to do this is through natural language. As we have seen in chapter 2, NLP still poses many problems to robotic designers. However, progress is being made. Robots now learn to interact with all kinds of people, whether they are traumatised children or invalid senior citizens. They learn how to recognise human emotions and how to express their own. But most importantly, they learn how to communicate using natural language.

However, natural language is only one possible type of dialogue that can occur between a human and a robot. There are two other main types of dialogue, namely low-level¹³ and non-verbal (Fong et al., 2003). After all, when humans communicate, they use multiple para-linguistic social cues, such as facial expressions and body language to control their dialogues (Cassell, 1999), and these cues have proven to be effective for robots as well (Breazeal, 2003). This results in sociable¹⁴ robots that can be used in diverse situations, ranging from at home to at the hospital.

In this chapter, we will examine how HRI can be developed by using natural language, and in the next chapter, we will take a look at other modes of communication, more precisely, at the communication of emotions through body language and facial expressions (natural language will play a role therein as well, but to a lesser extent).

First, we will discuss the concept of language grounding. Next, we will take a closer look at some possible frameworks that can be implemented to allow a robot to deal with natural language. We will then continue with a discussion on turn-taking, one of the essential parts of human communication. In this section, we will compare NAO's turn-taking behaviour to Kismet's, a sociable robot developed by MIT.

In the fourth part of this chapter, we will discuss some of the problems that occur when trying to establish HRI with natural language dialogues. As we have seen previously, the uncanny valley might be an issue, but there are also other problems that need to be considered, such as the creation of faulty perceptions of robots and the repetitiveness of dialogues based on manually implemented templates. This last problem could be solved by using crowdsourcing to elaborate the set of dialogue templates.

Next, we will compare open-domain and closed-domain dialogue systems. We will zoom in on an open-domain system that uses WikiTalk to interact with humans. Thanks to this system, a robot can talk about any imaginable topic by using Wikipedia as its source of knowledge.

The dialogue systems mentioned above are based on a single main dialogue. Hierarchical Dialogue Control (HDC) systems, however, are being developed in which dialogues are divided into sub-

¹³ Low-level dialogues are pre-linguistic dialogues.

¹⁴ Based on (Breazeal, 2003), there are four classes of social robots (socially evocative, social interface, socially receptive and sociable), of which sociable robots are the most advanced. These robots are different from those of the three other classes because they have their own internal goals and motivations.

dialogues. In the sixth part of this chapter, we will take a closer look at the two possible types of such systems, namely flexible HDC and strict HDC.

In the seventh part of this chapter, we will discuss cooperation. One of the functions of natural language is to allow people to govern cooperative tasks. Therefore, if robots want to be full companions, they need to be able to use language to cooperate with humans.

Then, we will examine the use of semantic gestures in HRI. Human communication is always a combination of verbal and non-verbal behaviour. In chapter 5, we will take an extended look at the non-verbal communication of emotions; in this chapter, we will limit ourselves to gestures that convey a semantic meaning.

In the final part of this chapter, we will discuss end-user programming. As robots are ultimately meant to be used by people with non-technical backgrounds, technologies need to be developed which allow these people to control the robot without coding. Systems which require minimal programming are an important first step to this end. Yet, robot programming in natural language would even be better. Therefore, we will take a look at both possibilities in this final section. Some researchers, however, do not believe that natural language will ever be a suitable medium for HRI. They think that NLP will not succeed in creating efficient natural language based systems in time for the arrival of millions of robots into our lives – if it will ever succeed at all (ROILA, 2015). Therefore, the Eindhoven University of Technology has created an artificial language, ROILA, to replace natural language in daily HRI. We will conclude this chapter with a brief introduction to this Robot Interaction Language.

4.1 Language Grounding

HRI can only take place when the robot and the human share a language that is “grounded”, which means that they each use the same symbols to describe common objects (Fong et al., 2003). If they do not share these, one of them (most likely the robot) will need to receive information about the symbols used by the other and learn based on this information (Fong et al., 2003).

When a robot acquires his “native” language, he is confronted with several problems (Dindo & Zambuto, 2010)¹⁵. First of all, he needs to identify the meaning of words. The words used in this experiment are grounded in non-linguistic perceptual data, which means that they refer to concepts in reality. Examples of such words include colours (e.g. *red*, *blue*) and geometrical shapes (e.g. *rectangle*, *circle*). Secondly, the robot needs to match these discovered meanings to lexical units. Lastly, he needs to be able to infer a basic grammar from the relations that exist between the different words in the utterance.

Dindo & Zambuto have conducted an experiment in which a NAO robot was taught new words. The teacher first attracted the robot’s attention by fixing his gaze on a to-be-learned-object or by pointing towards it. This creates an atmosphere of joint attention, which is an important condition for learning. The robot uses these visual cues to determine which area is most salient. All the objects which are located in this area are then stored into the robot’s memory. Once the teacher has attracted the robot’s attention, he will describe the object. This description is stored with the salient objects into the training set. This is an example of multi-instance learning: a label is not assigned to a specific instance, but to a group of instances. For example, if the word *red* is discovered in a description, it applies to all objects found in the associated salient area at this point in the learning process. To learn the meaning of the word, all groups of instances are being divided into two categories (positive or negative) based on the presence or absence of the word in the description. The robot then tries to pin the meaning on a specific instance through statistical methods. Figure 5 shows a schematic representation of the system.

¹⁵ The section *Language Grounding* is based on (Dindo & Zambuto, 2010), unless otherwise indicated.

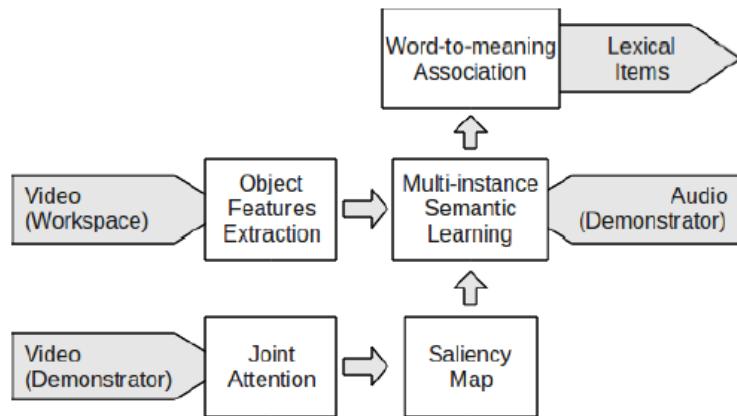


Figure 5 Diagram of System (Dindo & Zambuto, 2010)

In the experiment, NAO was presented with a set of objects on a table. These objects only differed in shape, colour, size and position. Participants were asked to use simple utterances when describing these objects, never referring to anything but the target object. After having learned the descriptions, NAO was given instructions, such as: ‘Grasp the object to the left of the blue one’. In these instructions, recently learned words (indicating size, colour or shape) were combined with hard-coded words (e.g. *to grasp*, *to point*) and special relationships. Figure 6 shows NAO trying to figure out which object was intended by the user. In [a], NAO points to the yellow rectangle, asking the user if that was the desired object. He received a negative answer and therefore, he chose another object that met the description (being located to the left of the blue object). He thus pointed to the blue circle in [b] and asked if that was the target object. As this was the case, he grabbed the blue circle in [c]. The results of this experiment indicate that joint attention and multi-instance learning can indeed be used to let a robot acquire a native language, but the method seems still limited to simple concepts with a restricted number of variable features. Dindo & Zambuto believe the method can be improved by building more complex concepts through a combination of simple ones.

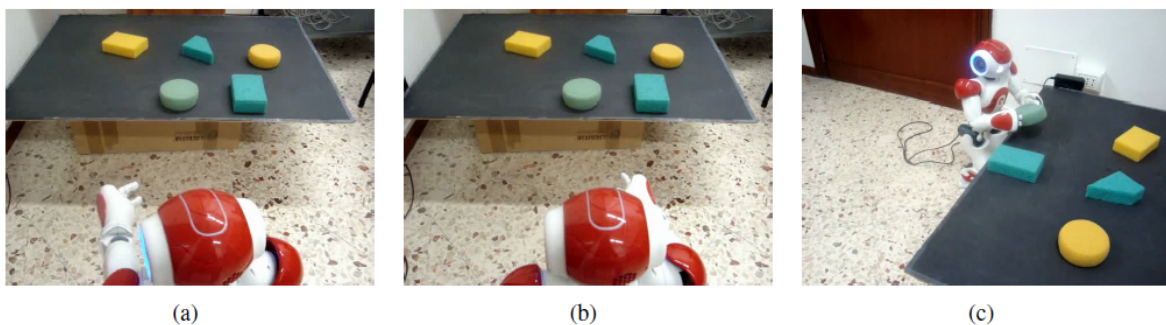


Figure 6 NAO Following Instructions (Dindo & Zambuto, 2010)

4.2 Natural Language Frameworks

In the previous section, we have discussed the grounding of language that can be used to help a robot to learn a language. However, a robot needs first and foremost to be able to manage interaction. Therefore, architectures need to be developed in which the different components needed for HRI are integrated. There are many different possible frameworks for natural language HRI. In this section, we will take a look at three possibilities: a frame-based dialogue framework, an event-based dialogue framework and a reward-based meta-cognitive framework.

4.2.1 A Frame-based Dialogue Framework

When humans interact with computers, they usually do so by typing instructions. The user-interface used in this kind of communication tends to be user-friendly, but it never feels as natural as talking (Barabás et al., 2012)¹⁶. HRI would thus be improved if robots could be controlled in natural language. Such systems are called “spoken dialogue systems” (SDS) and come in three different variants. The simplest type is a state-based SDS, which can be used for simple tasks. The system enters in a predefined dialogue with the user, during which several states are reached. In each state, the system will ask input from the user, which will be used to calculate the final output of the dialogue system. Frame-based SDSs are a slightly more complex variant of these systems, in which frames are seen as tasks with slots. The system will ask questions to fill the slots with the required information, after which it will complete the task and provide the desired output. Thirdly, there are also agent-based systems. These systems are far more complicated as they require a collaboration between users and system and an exchange of knowledge to come to the final result.

Barabás et al. have designed a frame-based dialogue system¹⁷ based on two principles: domain-adaptivity and language-adaptivity. Domain-adaptivity means that the system should be usable in multiple domains without changing source code. Language-adaptivity means that the system should be capable of processing different languages. However, frameworks that can work with any language do not exist yet. Usually, language-adaptivity means that a system supports a limited list of languages, which can be extended later. In this architecture, two modules remain language-dependent: the text-cleaner and the morphology module. The text cleaner cannot be made language-independent because there are different alphabets and text directions. The morphology module is language-dependent because of the fact that each language has its own vocabulary and grammar. Next to these two modules, there is one other module that is semi-language-dependent: the domain ontology module. This module could become language-independent if a list of word-code pairs would be implemented in the morphology module. This list would link translations of a word to an abstract string, which could then be used in the domain ontology module in a language-independent manner (for example, the words *dog*, *hond*, *chien* and *hund* could all be mapped to a code string like “#dog”).

In this experiment, a Nuance speech recogniser was used to convert spoken Hungarian to text, which became the input of the frame-based dialogue system. 18 functions were implemented into NAO, ranging from basic commands like ‘*sit down*’ to more complex commands such as ‘*turn 15 degrees to the right*’. Response times show that speech recognition is the slowest step in the process. Once the spoken language had been converted to text, the architecture allowed for quick responses, resulting in almost real-time action by the NAO robot. This might indicate that the designed architecture is suitable for robot controlling in natural language, although it must be kept in mind that the functionality of this system was very limited during the experiment (18 functions only), which might have influenced its performance.

4.2.2 An Event-based Dialogue Framework

In order for robots to interact with humans, many processes need to be managed. As mentioned before, this can be realised in many different architectures. One of the possibilities is an event-based conversational system in which the various components needed for HRI are integrated through the open source Urbi SDK (Kruijff-Korbayová et al., 2011)¹⁸. The architecture of the system discussed in this section can be found in Appendix V. In the experiment proposed by Kruijff-Korbayová et al., three games were implemented on this system: a dance, a quiz and an imitation game of arm movements.

¹⁶ The section *A Frame-based Dialogue Framework* is based on (Barabás et al., 2012), unless otherwise indicated.

¹⁷ The layered architecture of the resulting NLP engine is included as Appendix IV.

¹⁸ The section *An Event-based Dialogue Framework* is based on (Kruijff-Korbayová et al., 2011), unless otherwise indicated. This framework was developed in the larger context of the ALIZ-E Project. Please see chapter 7.3 for more information.

At the heart of the system, the Urbi framework combines and manages all other components into an integrated system. The dialogue manager is the component responsible for the robot's behaviour during the interaction. At first, this component was designed as a finite state machine that can enter three different states: dialogue, action or call. These states were used to control the flow of the interaction. However, based on the results of experiments with this architecture in 2011, the finite state machine was exchanged for a more flexible model in 2012 (Kruijff-Korbayová et al., 2012). This was needed because children's behaviour turned out to be too unpredictable for a finite state machine (Ros Espinoza et al., 2011) and too dependent on the individual child. Therefore, a spoken dialogue management method was chosen which used probabilistic methods and optimisation of dialogue policies based on reinforcement learning (Kruijff-Korbayová et al., 2012). Furthermore, as dialogues should be adapted to their users, online learning of policies was integrated, which allowed the system to create flexible interactions, much in the same way as humans adapt their own behaviour to their conversational partners (Kruijff-Korbayová et al., 2012).

The dialogue manager receives information about the user (such as name and game scores) from the user model component. Quiz questions are made available to the dialogue manager by the quiz question database. Next to this information, the dialogue manager also needs to follow the interaction. Therefore, the NLU (Natural Language Understanding) component parses the human speech detected by the robot's audio system and sends it to the dialogue manager.

The NLU component uses two different methods to interpret human speech. Quiz questions and answers are processed by using fuzzy matching of content words against the quiz database entries. This technique (also called approximate string matching) is used to find key words in databases when there might be spelling mistakes or other errors (Hall & Dowling, 1980). The second technique used is partial parsing. This technique is used to interpret any other speech input.

The system can generate output in two different ways. The dialogue manager can ask the NLG (Natural Language Generation) component to send canned text to the TTS (Text-to-Speech) component which transforms its text-input into audio-output. The dialogue manager can also specify a communicative goal, which can then be used in utterance content planning to create deeper, less repetitive outcome.

The interaction needs to be kept interesting for children, repetitiveness should thus be avoided (Kruijff-Korbayová et al., 2012). Furthermore, child-robot interaction greatly improves when the robotic voice sounds child-like, therefore, the research team chose to implement the open-source Mary TTS platform, rather than the Acapela TTS system that is standard available on NAO (Kruijff-Korbayová et al., 2012). Moreover, the speech output of the robot was created in such a way that familiarity with the child was explicitly expressed to create a stronger bond between child and robot (Kruijff-Korbayová et al., 2012).

In order to manage the imitation game, the architecture also needs a GRU (Gesture Recognition and Understanding) component, to detect the user's face, four types of body movements used in this game (left hand up or down, right hand up or down) and the combination thereof.

As mentioned before, children who participated in the experiment could choose one out of three possible games to play with NAO which are made possible through the above described framework. The first option was to learn a dance routine (Ros Espinoza et al., 2011)¹⁹. This experiment is part of a healthcare project (ALIZ-E), which explains the importance of physical activity in the chosen games. Furthermore, dance is considered to be a social activity that allows children to express themselves emotionally and creatively. To increase familiarity, NAO uses the name of the children when giving verbal feedback throughout the dance sessions. The game starts by NAO greeting the child and performing a sample dance, after which NAO starts to show the child the different moves, one at a time. A wizard is used to

¹⁹ The information on the three games in this section is based on (Ros Espinoza et al., 2011), unless otherwise indicated.

evaluate the child’s execution of the dance moves, which can lead to repetition or to adaptation of certain difficult moves. Once the child has mastered all the moves, the robot creates a dance by combining them.

The second option was a Simon Says game, adapted to be played by two players. The robot and the child take turns inventing arm movements which the other should repeat in the right sequence. When a mistake is made, the other player begins a new series of movements. During this game, the child and the robot become more familiar with each other. This is supported by NAO’s speech: the robot tells the child they are secret agents who have to learn a sign language to complete a secret mission. NAO continuously motivates the child to keep trying, which supports the goal of the project that states that children should be taught to persist in their endeavours.

The third and final option was a quiz in which the children had to answer questions asked by quizmaster NAO. The children received a point for each correctly answered multiple-choice question. This game was mainly used to examine NAO’s capacity to help children learn about their medical condition (all questions were related to health).

4.2.3 A Reward-based Meta-cognitive Framework

Another framework has been developed with the particular aim to support linguistic creativity (Pipitone et al., 2014)²⁰. In order for robots to be creative, they need to be able to perform very complex meta-cognitive behaviours such as having intuitions, experiencing and reading emotions and self-reflexion. Linguistic creativity is needed to interact with humans in an interesting way. Robots should thus be able to manage open-ended dialogues on all kinds of subjects²¹.

Pipitone et al. have proposed an architecture based on the unified management of uncertainty in Markov Decision Processes (MDP). MDPs are mathematical frameworks that model sequential decision making with an uncertain outcome (Puterman, 2005). It consists of decision moments (called epochs), states, actions, transactions and rewards (Puterman, 2005). When an action is chosen in a particular state, a reward is generated which determines the state in which the next decision will have to be made (Puterman, 2005). As shown in Figure 7, the agent consists of two MDP layers, each of which contains three nodes: perception, action and state. Perceptions of the environment are sent to the cognitive MDP layer. The state of this layer describes the agent’s model of the environment. The actions available to this layer pass perceptual data to the meta-cognitive MDP layer. This layer can also receive perceptions through self-reflexivity.

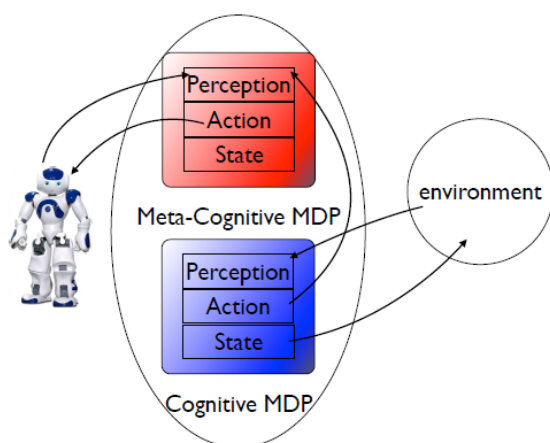


Figure 7 Cognitive and Meta-cognitive MDPs (Pipitone et al., 2014)

²⁰ The section *A Reward-based Meta-Cognitive Framework* is based (Pipitone et al., 2014), unless otherwise indicated.

²¹ Please see chapter 4.5 for a more elaborate discussion of open-domain dialogues.

Based on this schematic representation of the MDP layers, a meta-cognitive architecture was designed as shown in Figure 8.

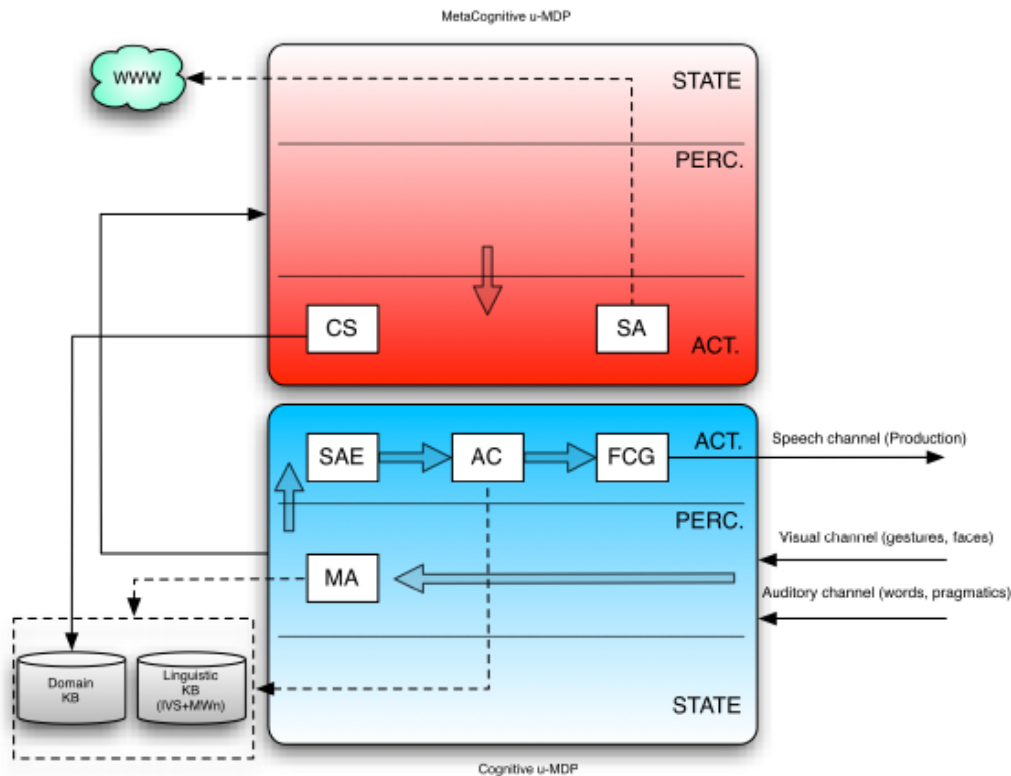


Figure 8 A Meta-cognitive Architecture based on Two MDPs (Pipitone et al., 2014)

In Figure 8, three different types of arrows are used. Continuous arrows represent sensory input and/or output and the functional connections between the system's components. Dashed arrows indicate transactions of internal and external information. Thick arrows indicate the perception-action cycles. Furthermore, there are also white rectangles, which represent software components.

As mentioned above, MDP is reward based: in this case, dialogue rewards are associated to the human's interest degrees as perceived by a robot. Robots, in this case NAO, will aim at receiving these dialogue rewards and will thus try to keep their conversational partner interested. They have different methods to do this: they can change the topic, search for detailed information based on the human's interest or limit the duration of their speech turns. Dialogue reward levels will increase when the human shows interest in the interaction and will decrease when no interest is perceived. When these levels are very low, the robot will decide to propose a change of topic to save the conversation.

To manage this complicated dialogue behaviour, three main dialogue tasks are implemented: understanding and producing natural language; searching new information; and switching context. Furthermore, the system has access to two different knowledge bases. The Domain Knowledge Base contains the internal representation of the dialogue domain. The Linguistic Knowledge Base is the source of the robot's lexicon. In this case, the Linguistic Knowledge Base consists of two lexical databases: MultiWordnet (MWn) and Italian Verbs Source (IVS). Verbs are retrieved from the IVS while the other parts of speech come from the MWn.

Visual and auditory sensory data arrives at the perception node of the cognitive MDP layer and triggers the Meaning Activator component (MA). This component compares the query-graph to the conceptual-graph by using the Graph Edit Distance method (Zeng et al., 2009). This is a method to determine the

similarity of two graphs that is often used for pattern recognition (Zeng et al., 2009). The output of this component is then enriched by the knowledge of the Linguistic Knowledge Base and sent to the Speech Act Execution component (SAE). This is the start of the process responsible for the understanding and production of natural language that happens in the action node of the cognitive MDP layer. The SAE component annotates the sub-graphs it receives from the MA component. These annotations are strings (e.g. *negative*, *interrogative negative*) that correspond to the planned response. These annotated sub-graphs are then passed on to the Answer Composer component (AC), which is responsible for the production of lexical and grammatical descriptions of the domain terms. The next component, the Fluid Construction Grammar (FCG), then transforms these descriptions into constructions consisting of two poles. The Form Pole contains the syntax properties and the Semantic Pole the term's meaning. There are two types of constructions created by the FCG component: lexical constructions (related to individual words) and grammatical constructions (conjunctions of lexical constructions). The output of this component is then produced as speech.

When the human conversational partner explicitly indicates that he is bored with the conversation, the perception node of the meta-cognitive MDP layer is addressed. Based on the understanding of the MA component, this can either lead to a topic switch or to a search for new information. In the first case, the Context Switching component (CS) is triggered to use the Domain Knowledge Base to determine a new dialogue topic. In the second case, the Semantic Annotator component (SA) will retrieve external contents from the Internet to improve the robot's answers.

While the research team that has developed this architecture does not provide the results of the experiments conducted with NAO, the article concludes with the statement that the framework performed well in an Italian context and was very versatile when it came to tutoring and managing open-ended dialogues.

4.3 Turn-taking

Once a framework has been chosen, the different components needed for HRI can be installed. One of the features which should be provided is a way to manage turn-taking. When interacting, humans automatically apply turn-taking patterns to their conversation. From an early age onwards, humans learn the ability of waiting for their turn in conversation, and expect others to do this as well. Therefore, it is important that robots should also be able to do this when interacting with humans. After all, humans automatically expect and apply it when talking to a robot, even though they do so unconsciously (Baxter et al., 2013). Baxter et al., showed this in their experiment in the Science Museum of Milan, in which 15 children were asked to sort food images into two categories by using a touch screen together with a NAO robot (Baxter et al., 2013). Even though the children were not given the instruction to obey turn-taking rules, the experiment shows that they did, simply because they considered NAO to be a social agent (Baxter et al., 2013). This clearly shows that humans expect turn-taking to be a part of HRI, just as much as they expect it to be a part of their daily human-to-human communications.

4.3.1 NAO's Turn-taking Behaviour

If robots are to become human companions, HRI should also concern itself with long-term relationships between humans and robots (Kruijff-Korbayová et al., 2013)²². Therefore, Kruijff-Korbayová et al., conducted a series of experiments in which children were exposed several times to a NAO robot. The results showed that children adapted various aspects of their communicative behaviour (both verbal and non-verbal) to the robot over time. One of these adapted aspects is turn-taking. When the children and the robot become more accustomed to each other, they are less likely to talk at the same time or to ignore

²² The section *NAO's Turn-taking Behaviour* is based on (Kruijff-Korbayová et al., 2013), unless otherwise indicated.

each other. This once again shows that humans like to project human characteristics on robots²³, and in this case, it actually improves HRI because it makes the interaction feel more natural as humans automatically adapt themselves to their robotic conversational partner, just like they would to a human one. Furthermore, the results also indicated that children were more willing to adapt their turn-taking patterns when the robot gave explicit signals of familiarity (like using their names or referring to earlier interactions).

In the experiment²⁴, 19 children each had three sessions with NAO on three different days. During the first session, the children could choose a main activity: a quiz, dancing or an imitation game. This activity would be the first interaction of each next session. If enough time remained, a second activity could be chosen for each session separately. At the beginning of the very first interaction, NAO introduced himself to the children, asking for their names. This is an important step to build a bond of familiarity between robot and child. Figure 9 shows the three main activities.



Figure 9 Children Playing with NAO. Left to right: Quiz, Dancing, Imitation Game (Kruijff-Korbayová et al., 2013).

During the experiment, NAO could be in either of two states: familiarity-display or neutral display. When in the familiarity-display condition, NAO addressed the children by their names and referred to their shared history by sentences such as *'I am happy to see you again'* and *'The next question should sound familiar'*. These verbal signs of familiarity were accompanied by non-verbal signs such as nodding. When NAO was in the neutral display condition, he simply greeted the children in a general way and used far more neutral sentences such as *'I am happy to see you'* and *'The next question'*.

The NLG component (Natural Language Generation) and TTS component (Text-to-Speech) are responsible for these verbal outputs of the system. To avoid repetitiveness, many variations were implemented. These can be selected either randomly or controlled by certain selection criteria. Among these criteria are the details of the content being transferred (e.g. number of options of a quiz question), details of the context (e.g. the child's gender, the number of questions that are already asked) and the familiarity-display condition.

The results of the experiment show that children increasingly respected turn-taking over the three sessions. Furthermore, they forced themselves twice as much to wait for the robot to finish his sentence in the familiarity-display condition than in the neutral display condition (see Figure 10). Overlap between children and robots decreased from 14.15% to 7.63% in the familiarity-display condition, and from 19.93% to 12.82% in the neutral display condition.

²³ This is shown by many researchers and theories, among which the Media Equation Theory, which is a communication theory that states that humans usually treat computers or robots as if they were human beings (Reeves & Nass, 1996).

²⁴ This experiment was conducted in the context of the ALIZ-E project. For more information on this project, please turn to chapter 7.3.

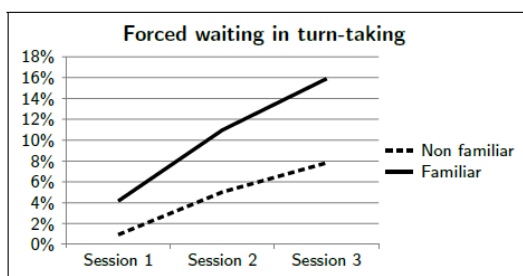


Figure 10 Forced Waiting during Turn-taking over 3 Sessions (Kruijff-Korbayová et al., 2013)

The results also showed that the number of utterances ignored by the robot decreased across the three sessions from 23.05% to 9.05% in the familiarity-display condition and from 28.2% to 12.89% in the neutral display condition (see Figure 11).

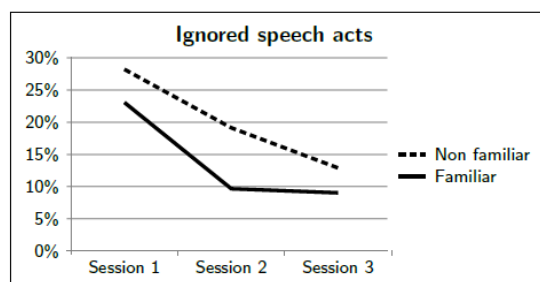


Figure 11 Ignored Speech Acts over 3 Sessions (Kruijff-Korbayová et al., 2013)

Figure 10 and Figure 11 show that the main adaptation happens between the first and the second sessions. This means that children adapt their communicational behaviour rather quickly after starting to interact with a robot and that this new behaviour persists in later interactions.

Turn-taking behaviour is regulated by non-verbal cues such as eye-gaze shifts (Meena et al., 2012)²⁵. It is the speaker rather than the listener who influences this behaviour. Possible gestures to implement in NAO to regulate turn-taking can be found in Table 4 and seen on Figure 12.

Turn-taking Gestures	
Head nod up	Turn-yielding: End of a sentence where NAO expects the user to provide an explicit response. The speaker’s gaze at the listener indicates a possibility for the listener to grab the conversational floor.
Speaking-to-listening	Turn-yielding: Listening mode. NAO goes to standing posture from the speaking pose and listens to the user.
Listening-to-speaking	Turn-accepting: NAO goes to speaking posture from the standing posture to prepare for presenting information to the user.

Table 4 Gestures for Turn-taking (Meena et al., 2012)

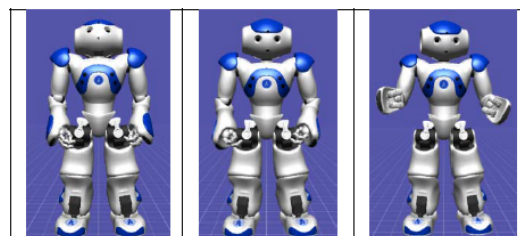


Figure 12 NAO's Turn-taking Gestures. A. Head Nod Up; B. Listening Key Pose; C. Speaking Key Pose (Meena et al., 2012)

²⁵ The information on non-verbal cues for turn-taking in this section is based on (Meena et al., 2012), unless otherwise indicated.

The results of an experiment in which twelve users were asked to interact with NAO indicate that these gestures considerably improve NAO's expressivity and turn-taking abilities. However, most work is still done by the human, as NAO's speech recogniser does not allow interruptions from the user (Csapo et al., 2012). Users are therefore forced to wait for a beep before they can respond to the robot, which makes them responsible for waiting an appropriate amount of time (Csapo et al., 2012).

4.3.2 Kismet's Turn-taking Behaviour

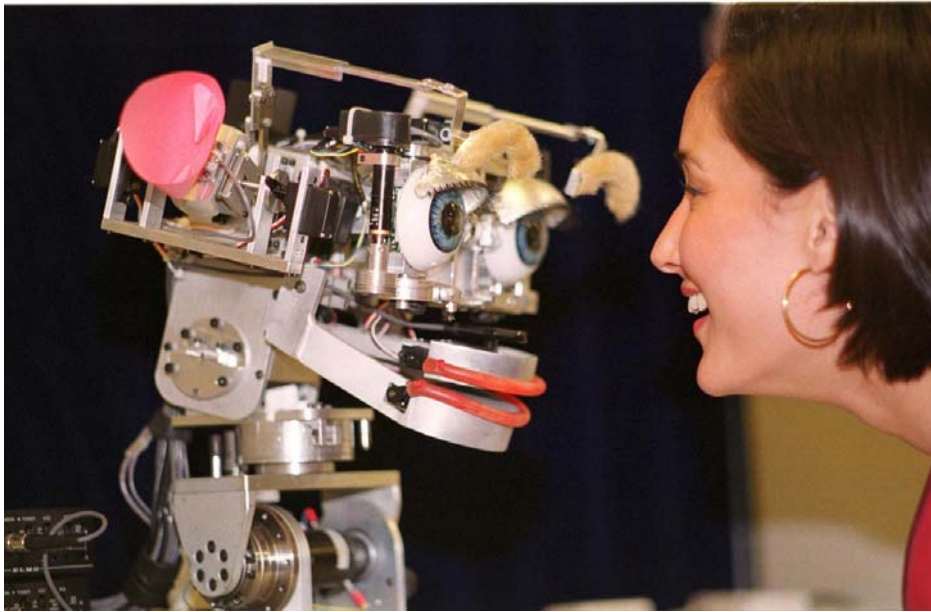


Figure 13 Kismet (Plasticpals, 2015)

In the 1990s, creators of robots realised that it would no longer be enough for robots to accomplish human tasks independently (Breazeal, 2015). Humanoid robots were needed to fulfil tasks in new environments such as hospitals and schools, because they would be able to communicate naturally with humans due to their similar morphology (Breazeal, 2015). Cynthia Breazeal developed the sociable robot Kismet for her doctoral research at MIT. Kismet can communicate with humans in a natural way, not only by perceiving social cues (both visual and auditory) but also by expressing social cues (through facial expressions, including gaze direction, emotional body postures and language) (Breazeal, 2015). The ultimate goal of Kismet is for her to learn through her interactions with human beings (Breazeal, 2003)²⁶. Kismet's face is highly expressive to match her infant-like personality (Breazeal, 1999). Because of this, humans automatically interact with her as if she were a 6-months-old baby (Breazeal, 1999).

As said before, humans use several paralinguistic cues to regulate turn-taking behaviour, such as blinking or facial expressions (Cassell, 1999). These “envelope displays” are likewise important to HRI because of speech processing limitations. Kismet has the ability to engage in turn-taking, but this happens slower than in human-to-human interactions. Humans can change turns 0.25 seconds after the other person has stopped speaking, while Kismet needs at least 0.5 seconds. Kismet's conversational partners, however, read the paralinguistic cues displayed by the robot and automatically establish a conversational pace in which both partners can function adequately. To avoid making the conversation look too rehearsed, Kismet does not use these cues in a rigidly. Please see Table 5 for Kismet's envelope displays.

²⁶ The section *Kismet's Turn-taking Behaviour* is based on (Breazeal, 2003), unless otherwise indicated.

Kismet's Envelope Displays	
To acquire the floor	Break eye contact and/or lean back a bit.
To start a speaking turn	Vocalize a Kismet-esque babble. ²⁷
To stop a speaking turn	Stop vocalizing and re-establish eye contact.
To hold the floor	Look to the side.
To relinquish the floor	Raise brows and/or lean forward a bit.
End of vocalisation	Blinking

Table 5 Kismet's Envelope Displays (Breazeal, 2003)

In the experiment, four participants were asked to converse with Kismet (even though Kismet spoke no English). Each speaking turn was considered to consist of four different phases: acquire the floor, start speaking, stop speaking and relinquish the floor. When a speaker maintained the speaker role after a pause, this was called "holding the floor". At the beginning of each conversation, the participants used long sentences and expected the robot to display turn-taking behaviour in the same way and at the same pace as a human being. This resulted in quite a lot of problems, such as Kismet interrupting the participant. However, when time went by, participants shortened their sentences and waited longer for Kismet to respond to them. They also started to observe Kismet's paralinguistic cues to regulate the conversation. This led to a decrease in interruptions by the robot and a smoother conversation in general.

Kismet's pro-active role in turn-taking management allows the robot to receive the right amount of stimuli to create a learning-environment, which suits Kismet's ultimate goal. The results of the experiment indicate that in order to create natural HRI, turn-taking should be implemented.

This experiment, like the one discussed in chapter 4.3.1, indicates that humans automatically expect turn-taking behaviour when interacting with robots. Moreover, they impose it on the conversation, eventually creating a situation of balance between their expectations and the robot's capabilities. This leads to a natural conversational rhythm which allows for smooth interactions between human and robot.

4.4 Problems with Dialogues

As shown in chapter 2.6, dialogues between humans and robots might cause the humans to experience an uncanny feeling, even though this is largely dependent on the background of the individual instead of on the interaction itself. There are, however, several other problems regarding HRI which need to be addressed to optimise these kinds of dialogues.

First of all, dialogues can create false perceptions (Fong et al., 2003)²⁸. If a robot talks in a stereotypical way, for example, its user might be misled and might imagine characteristics for the robot that it does not actually possess. This might also lead them to wrong impressions about what the robot can or cannot do or about how it works.

Second, dialogue systems often work with manually created templates to which dynamic information is added at runtime (Mitchell et al., 2014)²⁹. If the system only has access to a limited number of templates, interaction will quickly become repetitive. This also has a negative effect on the naturalness of the communication. Mitchell et al., have therefore conducted an experiment with a NAO robot to explore if crowdsourcing could be used to enrich this type of dialogue systems. After all, crowdsourcing has become a popular tool in several fields of computational linguistics because it offers an economic way

²⁷ Kismet communicates through proto-dialogue in this experiment, because the content of the conversation is inferior to the way in which paralinguistic cues are used to regulate the turn-taking patterns.

²⁸ The information on false perception in this section is based on (Fong et al., 2003).

²⁹ The information on crowdsourcing in this section is based on (Mitchell et al., 2014).

to have access to human intelligence. Their methodology for the development of crowd-based natural language generation templates can be seen in Figure 14.

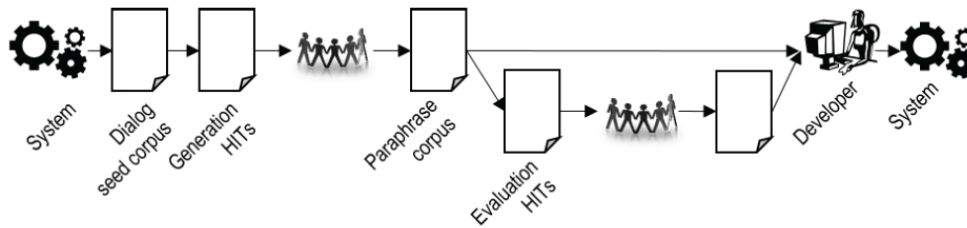


Figure 14 Pipeline for crowd-based development of natural language generation templates (Mitchell et al., 2014)

To get the initial data, NAO was placed in front of an elevator for several months. His task was to give directions to humans who asked him to. NAO fulfilled his task by using rule-based natural language generation based on two modules: one module designed for giving directions based on computed paths, and one containing 38 standard templates for the other parts of the dialogues. These dialogues were then collected in a dialogue seed corpus. Based on this corpus, generations HITs were automatically constructed. Each of these contained a highlighted phrase in a context of variable length. An example of highlighted sentences with and without context can be found below:

Without context

System: *Sorry, that was Ernestine Patrick's office you wanted, correct?*

With context

System: *You said Ernestine Patrick's office, right?*

User: *Nop...*

System: *Pardon me?*

User: *...No*

System: *I'm sorry! I still didn't get that. **Sorry, that was Ernestine Patrick's office you wanted, correct?***

User: *...No*

The phrase printed in bold is the highlighted phrase, generated based on a template and completed at runtime by the part that is not printed in italics, based on the previous utterances in the current dialogue. Crowd-workers were then hired via the Universal Human Relevance System, which is a general-purpose crowdsourcing marketplace. These people had to paraphrase the highlighted phrases, staying as close as possible to the original meaning and making the new paraphrase fit as well as possible in the context. These paraphrases were then edited to form a corpus of crowd templates (correcting spelling, punctuation, capitalization etc.). Based on this corpus, evaluation HITs were constructed, which were afterwards evaluated by the crowd. These contained variants of the highlighted sentences such as: '*I apologize, are you looking for [Place]?*', and were evaluated against several criteria. People in the crowd first of all needed to decide whether or not a new paraphrase had the same meaning as the original sentence. Then they needed to rate the naturalness of both the original and the new variant in the context of the dialogue. The results show that 90% of the new paraphrases were considered to have the same meaning as the original phrases. This kind of evaluation, however, is not enough to accept new templates as standard templates. Therefore, additional evaluation by developers was conducted, to check, for example, if the register and style were appropriate. Of the original amount of paraphrases that remained after the evaluation by the crowd, only 33% were accepted after the second evaluation. This means that even though crowdsourcing could become an important tool to make HRI less repetitive, manual intervention is still needed afterwards to ensure the quality of the output. Mitchell et al., suggest that the evaluation criteria for the crowd evaluation should be updated to incorporate more of the issues found by the developer. This would lower the number of manual interventions needed afterwards, which would speed up the entire process of developing larger sets of templates.

4.5 Open-Domain versus Closed-Domain Dialogues

In the previous section, some of the problems that need to be resolved when dealing with dialogues have been discussed. Now we will take a closer look at the two main kinds of dialogue that are possible in HRI: closed-domain and open-domain (Wilcock, 2012)³⁰. Closed-domain dialogues are easier to implement, but they are also very restricted. As the name already indicates, this kind of dialogues can only be used within the limits of a single domain. For example, a robot can be taught to ask a series of questions to fill in a form to book a flight. It is not hard to alter the questions in this kind of system, but it is very difficult to change a particular system to another domain. Please see Figure 15 for an example of a finite-state machine using a closed-domain dialogue.

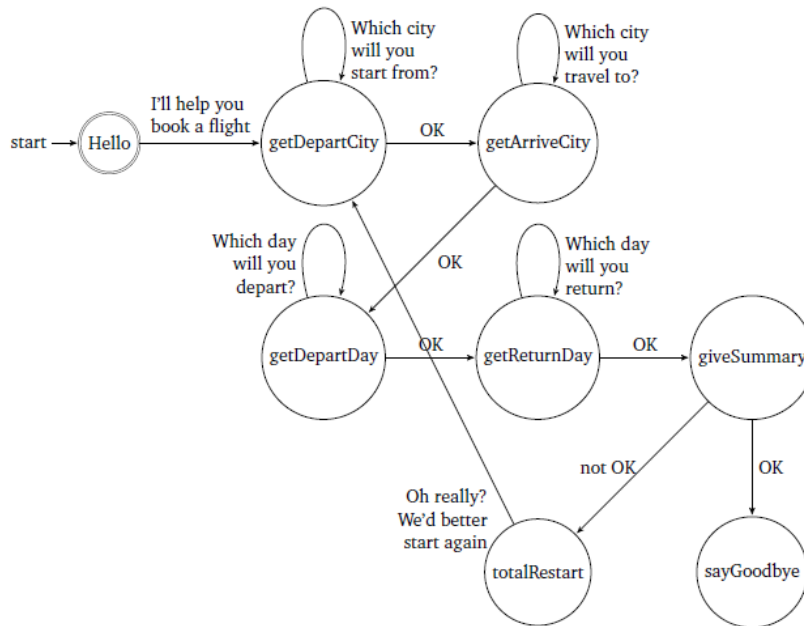


Figure 15 Closed-Domain Dialogue Flight Reservation System (Wilcock, 2012)

Much more interesting, however, are open-domain dialogue systems, which are not restricted to particular topics. In this kind of systems, Wikipedia can be used as a source of knowledge, which enables robots to talk about any topic imaginable. WikiTalk is an example of such an open-domain knowledge access system; it has been developed to be integrated into other systems, like into NAO for example. NAO is an embodied agent, which means that WikiTalk will need to be integrated with other modules such as face-tracking, nodding and gesturing and proximity recognition as well. Through these, NAO can observe whether or not a user is interested in the current topic. Furthermore, beat gestures (small hand movements) can be used to highlight certain parts of the conversation. WikiTalk also enables NAO to make smooth topic-shifts, by using the hyperlinks on the pages of the wiki. Please see Figure 16 for an example of an open-dialogue finite-state machine.

³⁰ The section *Open-Domain versus Closed-Domain Dialogues* is based on (Wilcock, 2012), unless otherwise indicated.

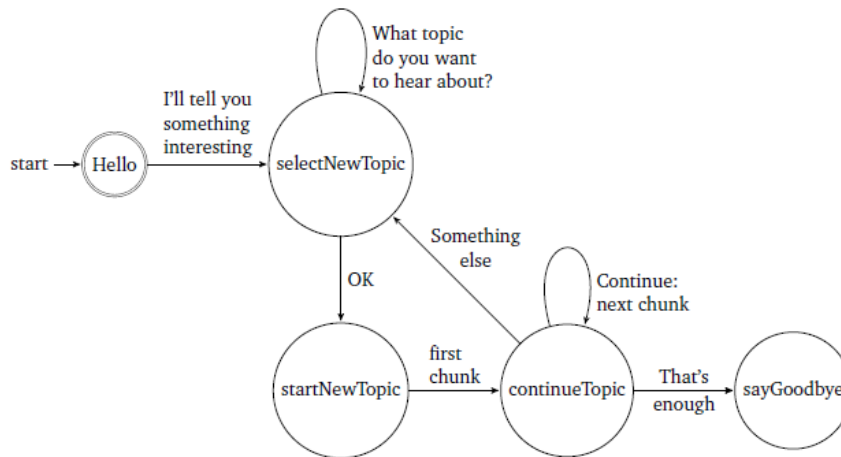


Figure 16 Open-Dialogue WikiTalk System (Wilcock, 2012)

While the finite states were closely linked to a particular topic in closed-dialogue systems, they are essentially interaction managers responsible for topic-tracking and topic-changing in open-dialogue systems. We will now take a closer look at the different states of the open-dialogue finite-state machine as shown in Figure 16.

As it can be seen on Figure 15 and Figure 16, both systems need ways to start and to finish the interaction. These are managed by the “hello” state and the “goodbye” state, which can be implemented according to the needs of the particular system. In the next state, a new topic should be chosen, which can be done in different ways. For example, WikiTalk can provide a list of favourite topics from which the robot can choose one, or it might present the list of recently talked-about topics. Once the new topic has been established, the “new topic” state is entered and WikiTalk retracts the desired topic from Wikipedia. Wikipedia is designed to be visually appealing in a browser and therefore, the text needs a certain amount of refactoring before it can be passed to the speech synthesizer module. Furthermore, the text needs to be divided, because NAO has to enter in a dialogue with a human, he does not have to give a speech. The desired size depends on several factors, among which the quality of the interrupt mechanism. If the user can easily interrupt the robot, it can use larger blocks of text. When NAO starts talking, the “continue topic” state is entered. As long as the user shows interest in the current topic, NAO will keep adding new pieces of information to the interaction, if not, a new topic can be chosen. When he reaches the end of the Wikipedia page, NAO will inform the user that there is no more information on the current topic and will ask for a new one. The user can also intervene after each block of information: he can ask to repeat a block, to return to the previous one or to skip to the next one. Furthermore, the user can interrupt the robot, after which NAO will enter the “interrupt” state and remember where he left off.

What is missing in the diagram above, is a way to make a smooth topic change. This is a topic change to a related theme, rather than to a completely new topic. In the past, manually created topic trees were often used in dialogue systems to organise topics into branches of related topics. The hyperlinks on the Wikipedia pages can be used as an alternative way to create such topic clusters, as they provide a ready-made organisation of the knowledge within the domain of the current topic. The blocks of text used by NAO contain “new infos”, annotated with hyperlinks to related articles. The robot accompanies these new infos with beat gestures, as a part of the multimodal conversation management. After all, the human is unaware of the structure of the article, and does thus not know which hyperlinks are present in the article (Meena et al., 2012). Examples of such gestures are a rhythmic up and down movement of a vertical open palm or the head nodding down (Meena et al., 2012). When a user hears something on which he would like more information, he can simply repeat the word. NAO will then smoothly shift back into the “new topic” state and start talking about this related topic, as can be seen in Figure 17.

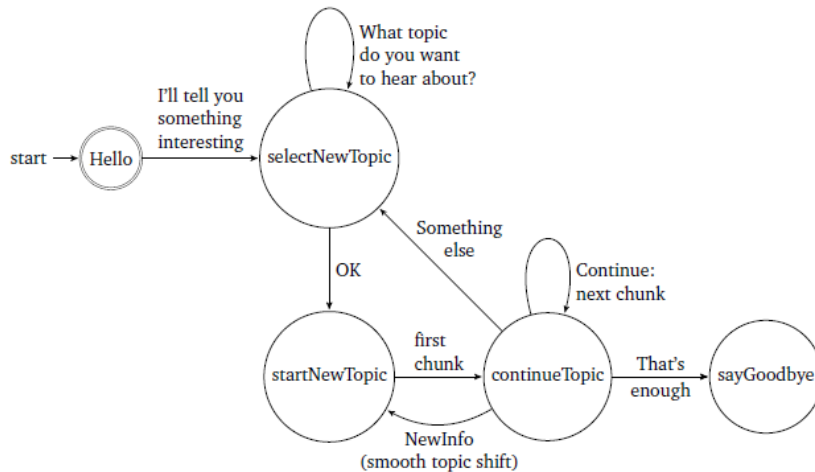


Figure 17 Open-Dialogue WikiTalk System with Smooth Topic-shifts (Wilcock, 2012)

There are two important problems when using hyperlinks to change the topic however. First, sometimes a key word is only once marked as a hyperlink on a page, even though it appears more often. One possible solution would be to draw up a list with all the hyperlinks on the page when a new topic has been selected. Second, Wikipedia grows every day and some people believe that one day, each word will be a hyperlink. This would mean that the WikiTalk speech recogniser would need open vocabulary speech recognition. Yet, this is not currently possible, as the enormous amount of words would drastically reduce the effectiveness of a speech recogniser. Therefore, this system now only recognises a limited set of commands and a varying list of hyperlinks. The speech recognition is assisted by confidence scores. If the module is certain about having recognised the right word, it will proceed immediately. If it is not 100% confident, but thinks that it might have recognised “X”, it will ask ‘*Did you mean X?*’ to make sure. If the module cannot recognise anything, it will ask the user to repeat himself.

Next to these smooth topic-changes, dialogues also contain awkward topic changes to a completely unrelated topic. Using WikiTalk, this can be done by spelling the first few letters of the new topic (by using the standard phonetic letter names). Wikipedia speeds up this process by suggesting topics that start with the selected letters. Another advantage is that Wikipedia can help by providing new possible topics if the user cannot decide on a particular one. This can be done, for example, by using the daily *Did you know?* Section of the English Wikipedia main page. NAO will randomly pick one of the topics from that list and ask the user ‘*Did you know...?*’ This continues until NAO has found a topic that seems interesting to the human. For an example of a WikiTalk-based open-dialogue with a NAO robot, please see Appendix VI.

4.6 Strict versus Flexible Hierarchical Dialogue Control (HDC)

The two types of dialogue systems described in 4.5 are only concerned with the management of one main dialogue. However, it might be interesting to create systems in which a dialogue agent can invoke sub-dialogue agents (Cuayáhuitl & Kruijff-Korbyová, 2012)³¹. These kinds of systems are called hierarchical dialogue control systems (HDC). HDC has several advantages compared to systems based on only one dialogue. First of all, sub-dialogues are easier to specify than main dialogues, as they are by definition shorter and less general. Second, only relevant dialogue knowledge is needed to govern sub-dialogues, which means that the management thereof is less complex. Third, sub-dialogues can easily be reused when the system learns a new behaviour.

³¹ The section *Strict versus Flexible Hierarchical Dialogue Control* is based on (Cuayáhuitl & Kruijff-Korbyová, 2012), unless otherwise indicated.

There are two main types of HDC, namely strict and flexible, which are shown in Figure 18.

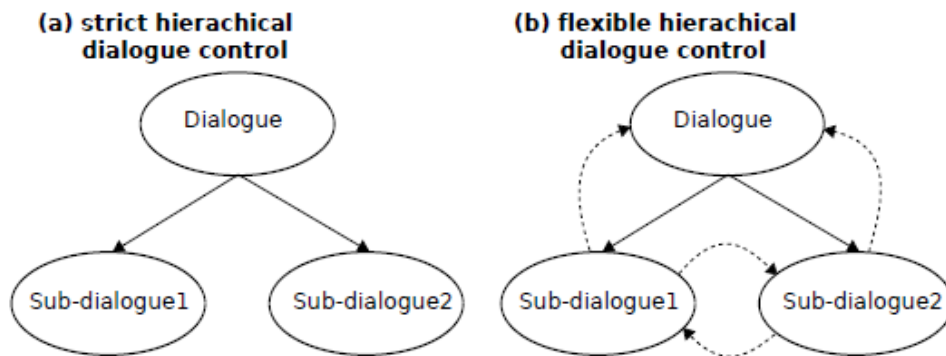


Figure 18 Strict (a) versus Flexible (b) HDC (Cuayáhuítl & Kruijff-Korbayová, 2012)

Strict HDC is more limited than flexible HDC because it creates rigid interactions in which the user has no influence over the dialogue structure. Flexible HDC results therefore in interactions which are perceived as more natural by the users. Furthermore, the dialogue structures used in flexible HDC cannot only be driven by the user: the machine can likewise learn to use different structures and to initiate their use.

In the HDC system as proposed by Cuayáhuítl & Kruijff-Korbayová, an algorithm uses a stack of subtasks which are executed as shown in Figure 19. The fourth operation depicted can only be implemented in flexible HDC.

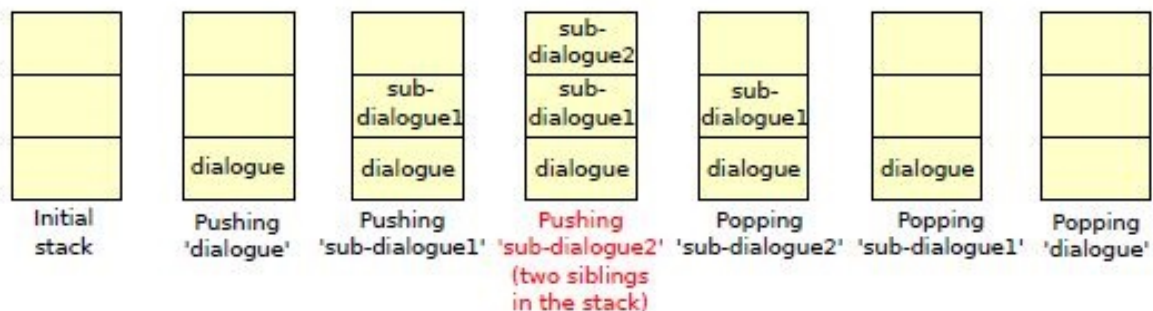


Figure 19 Operations on Stack-based HDC (Cuayáhuítl & Kruijff-Korbayová, 2012)

In flexible HDC, a sub-dialogue can be interrupted in two different ways. If the new sub-dialogue is already on the stack (which would mean that it was a parent of the previous sub-dialogue), all intervening sub-dialogues would be terminated until the parent sub-dialogue is reached. However, if the new sub-dialogue is not on the stack yet, it is pushed onto the stack after which it is executed first. To manage these interruptions, the system needs to maintain knowledge about the dialogue history.

To illustrate their flexible HDC model, an experiment with a NAO robot was conducted in which a hierarchy consisting of a parent agent and two children agents was used. These children were “robot asks” and “user asks”. The architecture of NAO’s integrated system has been included as Appendix VII.

In the experiments, children were asked to interact with NAO in the context of a quiz game. The sample dialogue below shows four groups of utterances, which each represent a sub-dialogue in a flexible HDC system. The users can override the structure, for example, by saying that they want to be the quizmaster next. These interruptions lead to switches in the dominant agent.

System: *Hi, my name is NAO, and you?*
User: *I am Peter*
System: *Nice to meet you Peter. Do you want to play a Quiz game?*
User: *Ask me questions. [Jump to “robot asks”]*

System: *The first question is: ‘What is a typical magic instrument?’ Your choices are: One, guitar. Two, drums. Three, violin. Four, flute. What is your choice?*
User: *Number two.*
System: *That is not quite right. Do you have another choice?*
User: *Maybe number three.*
System: *The correct answer is number four, flute. The next question is: ‘Which one of these is an important artificial channel of Central America?’ Your choices are: One, channel five. Two, Panama Canal. Three, Suez Canal. Four, English Channel. What is your choice?*
User: *Now I want to ask you questions. [Jump to “user asks”]*

System: *Ask me the first question.*
User: *What does water become at the temperature of zero degrees?*
System: *What are my choices?*
User: *Your choices are: One, condensation. Two, ice. Three, steam.*
System: *I choose number one.*
User: *That is incorrect.*
System: *So what is the right answer?*
User: *The correct answer is number two.*
System: *Okay, ask me another question.*
User: *The next question is: ‘Through which breach did the Italian army enter into Rome in 1870?’ Your choices are: One, Porta Venezia. Two, Porta Pia. Three, Porta Paziienza. Four, Porta Nuova. What is your choice?*
System: *Paris*
User: *Let’s stop playing. [Jump to “root”]*

System: *Did you like playing the Quiz Game?*
User: *Yes.*
System: *I am glad to hear that. It was nice playing with you, see you soon. Bye!*

The above-described system can be used in situations in which topic-shifts are user-driven (as shown in the dialogue above) and in situations in which these are machine-driven. For example, NAO can decide based on the situation or on the perceived emotional state of the user to stop the game or to start another game. This results in a very flexible system which can be used in many different HRI contexts.

4.7 Cooperative Tasks

Humans always use dialogues to serve a particular goal. One of the functions for which human beings use natural language is to guide cooperation (Petit et al., 2013)³². The ability to cooperate is considered to be one of the characteristics that differentiate human cognition from nonhuman cognition. Simple forms of cooperation can occur without the aid of language (e.g. small children can accomplish some tasks together without using words, as long as the tasks remain easy), but the moment the cooperation turns complex, language is needed to guide it. In order to cooperate, humans need to be able to establish a shared plan and to modify it during the execution thereof. Petit et al., designed an experiment in which NAO was taught to cooperate with users. In this experiment, the robot learned the steps of the task in three different ways: through imitation, through kinaesthetic teaching³³ and through spoken language programming. Language was used to coordinate these three learning modalities. The system architecture

³² The section *Cooperative Tasks* is based on (Petit et al., 2013), unless otherwise indicated.

³³ Kinaesthetic teaching is a form of demonstration in which the user guides a passive robot through the desired actions.

used is a BASSIS architecture (Biomimetic Architecture for Situated Social Intelligence Systems), as shown in Figure 20³⁴.

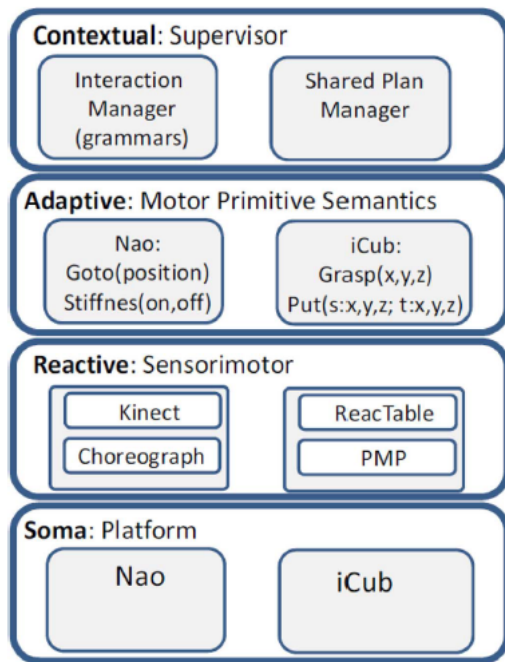


Figure 20 BASSIS Architecture (Petit et al., 2013)

The architecture consists of three levels of control that are implemented in a physical platform (the soma layer). The reactive and adaptive layers are platform dependent while the contextual layer is not. The reactive layer deals with sensorimotor issues, such as perception and movement. These are implemented in NAO respectively by Kinect and Choregraphe³⁵. The adaptive layer controls the adaptive motor capabilities, while the contextual layer fulfils a supervisor role over the system. This last layer is separated into two main parts: an interaction manager and a shared plan manager. The interaction manager controls the HRI in general by using state-based dialogue management techniques. The process is subdivided into the following four states: specifying the plan, modifying it, teaching steps of the plan and executing the plan. Furthermore, this component enables the robot to tell the user when an error has occurred. The shared plan manager, on the other hand, governs the learning and executing of shared plans. As discussed earlier, three different learning modalities were used in this experiment, which were grouped into two complementary categories: pure spoken language programming and a combination of language and perceptual action recognition. In the first category, the user will only use language to describe the actions the robot should undertake to accomplish the shared plan. This method is only used for actions which were already taught to the robot in the past. When errors occur, there are several “spoken language programming” functions which can be used. For example, if a user has used a wrong command, he or she can use an editing function by repeating the faulty command, directly followed by the right one. The user can also choose to delete or insert commands. In the second category, the user will use a combination of language and motion to teach the robot new actions. This can be done either through kinaesthetic demonstration or through imitation.

³⁴ Petit and al also conducted an experiment with iCub, a humanoid developed by the EU consortium RobotCub that resembles a 3-year-old. In this bachelor’s thesis, however, only the NAO experiment will be described.

³⁵ Kinect is a motion sensing input device by Microsoft which is often used in robotic research even though it was originally meant for gaming purposes. Choregraphe is a desktop application created by Aldebaran which allows users to control robots without coding. According to the Aldebaran website, it is called *Choregraphe*, not *Choreograph*.

In the experiment, NAO's task is to cooperate with a human user to clean a table that has some objects on it. There is a chest next to the table which needs to be opened first to put the objects in, after which it should be closed again. In order to do this, NAO will have to perform two sets of actions: first learn the plan, then execute it. The first step of learning the shared plan is the establishment thereof:

PART 1: Establishing the plan

1. **NAO:** *Tell me what I have to do.*
2. **Human:** *Pedro, help me to clean the table.*
3. **NAO:** *The shared plan is clean. Please explain me.*
4. **Human:** *I reach the toy, then I reach the trash box, then you open the trash box, then I put the toy in the trash box, then you close the trash box.*
5. **NAO:** *I reach the toy, you reach the trash box, then I open the trash box, you put the toy in the trash box, I close the trash box. Do you agree with this plan?*

The user first tells the robot (called Pedro in this experiment) his intentions [1-2]. NAO uses his shared plan manager to check if he already knows this plan. As he does not, the manager creates an empty shared plan *Clean the table*, after which NAO asks for an explanation of the plan [3]. The human then uses language to explain the steps of the plan [4], which are segmented by the spoken language interface of the robot. NAO then repeats the instructions to give the user the chance to verify his interpretation [5]. NAO, however, made a mistake (indicated in red). Therefore, the user does not agree and a second step in the learning phase is launched, namely the correcting phase:

PART 2: Correcting the plan

6. **Human:** *No, I do not.*
7. **NAO:** *So, what did I misunderstand?*
8. **Human:** *I reach the toy.*
9. **NAO:** *What do you want me to do instead?*
10. **Human:** *I reach the toy.*
11. **NAO:** *So where I reached the toy, instead, you will reach the toy. Do you agree with this plan?*
12. **Human:** *Yes I do*

As the human does not agree with the plan [6], he can alter it in several ways (as specified above). In this example, the user chooses to specify the faulty command and replace it with the right one [8-10]. NAO now possesses the same plan as the human user. In the next step, they should check if the robot has learned all the steps in the process:

PART 3: Learning component actions

13. **NAO:** *Behaviour 'open' is unknown. I am going to learn 'open'. Tell me what to do.*
14. **Human:** *Move to Kinect init.*
15. **NAO:** *I am going to move to Kinect init. Are you sure?*
16. **Human:** *Yes, I am.*
17. **NAO:** *Tell me what to do.*
18. **Human:** *Imitate me.*
19. **NAO:** *Imitation activated. Tell me when it is finished.*
20. **Human:** *It is finished*
21. **NAO:** *The behaviour 'close' is unknown. I am going to learn 'close'. Tell me what to do.*
22. **Human:** *Let me move your arms.*
23. **NAO:** *Stiffness changed. Tell me when it is finished.*
24. **Human:** *It is finished.*

NAO will iterate through all the steps of the shared plan until he encounters one for which he does not find the appropriate action stored in the adaptive layer. When he does, he will inform the user and ask for help [13]. The contextual layer will then enter the learning state. The human can then choose one or more of the three possible learning mechanisms to explain NAO how to perform the unknown action. In this example, NAO first needs help opening the chest and the user decides to teach him through imitation [18]. The movement and positions of the user's body are perceived through the Kinect device,

allowing NAO to learn the action ‘open’. The iteration then continues until the robot encounters another action with which he is not familiar, in this case, the action ‘close’. This time, the user chooses to demonstrate the action by moving NAO’s limbs. The command issued in [22] changes the contextual layer to a state in which NAO’s arms have zero stiffness, so that it would be easier for the human to bring them into the right position. This finishes the learning stage in this example.

In the execution stage, NAO and the human can now start the shared task:

PART 4: Execution stage

25. **NAO:** *The current step is you reach toy. You are supposed to reach. Tell me when it is done.*
26. **Human:** *It is done.*
27. **NAO:** *The current step is I open trash box.*
28. **NAO:** *The current step is you put the toy trash box. You are supposed to put. Tell me when it is done.*
29. **Human:** *It is done.*
30. **NAO:** *The current step is I close trash box.*
31. **NAO:** *That was the last command.*

The results of this experiment show that language can be used in two ways in the context of cooperative tasks. First of all, it can fulfil a controlling function: constructing and modifying the shared plan. Second, it can also be one of the learning modalities through which new actions can be inserted into the cooperation. There is, however, an important limitation: the user can only use simple grammatical constructions. Unconstrained natural language cannot yet be processed in this type of scenario. This means that humans should not use relative clauses, for example, or should be careful with the use of pronouns. Yet, this still allows this method to work for rather complex shared plans.

4.8 Semantic Gestures

When humans interact with each other, they do not only use verbal communication. In fact, any kind of human-to-human interaction is always a combination of verbal communication and non-verbal communication (Zheng & Meng, 2012)³⁶. It is therefore logical to include both components into HRI as well. Humans usually are not consciously aware of the body language they are displaying during normal interaction. Yet, to the careful observer, non-verbal behaviour can reveal a lot about a person, for example, about his emotional state. This is why body language is such a popular topic in recent HRI research, as will be explored in chapter 5. That kind of non-verbal communication has almost no connection to semantics. This is not true for any kind of body language, however. There are also “semantic gestures”, which are used to support or emphasize verbal communication.

When semantic gestures are designed for robots, these are usually created by imitating human gestures. Zheng & Meng, however, were concerned about the perception of these gestures. They thought it could not be simply assumed that users would react to a gesture in the same way when it was being performed by a robot as when it was being performed by a human, as a robot has fewer degrees of freedom than a human being. Therefore, they created a framework for NAO to generate and evaluate these kinds of gestures. After all, if robotic companions need to make users feel comfortable around them, they must use gestures that are accurately perceived. Non-verbal behaviour can be categorised into five different classes, as shown in Table 6.

Non-verbal gestures with semantic meanings are either emblems or illustrators, and to Zheng & Meng’s research, emblems are the most important ones. Emblems are typically used in social situations in which speech is not an option. For example, distance or noise can drown out words and therefore, humans tend to use easy-to-recognise emblems. Yet, it is important to keep in mind that the interpretation of these emblems is culturally defined. For example, when an American makes a forefinger-thumb circle gesture,

³⁶ The section *Semantic Gestures* is based on (Zheng & Meng, 2012), unless otherwise indicated.

he means ‘*Okay*’; A Frenchman doing the same would mean ‘*worthless*’. This is important to remember when designing robotic gestures. In the experiment by Zheng & Meng that is being discussed here, Chinese participants were chosen. The research team thus had to be careful when using typically American gestures.

Non-verbal Categories		
Category	Explanation	Example
Emblems	= non-verbal act with direct verbal translation	- Thumb up = approval - Headshake = negation
Illustrators	= non-verbal acts directly tied to speech to serve as illustration	- Pointing to topic
Affect Displays	= facial expressions and body movements associated with emotions	- Smiling
Regulators	= non-verbal acts that regulate two-way conversation	- Nodding = continue talking
Adapters	= unintentional habits	- Adjusting glasses = tense - Self-touching = anxiety

Table 6 Non-Verbal Categories (Ekman & Friesen, 1969)

Zheng & Meng suggest to evaluate the perception of particular gestures when performed by a human first. Afterwards, these gestures should be evaluated when performed by NAO. The comparison of these results should then lead to an improved version of the robotic semantic gestures.

Participants were first shown nine gestures performed by a human. They had to decide whether or not each gesture had a semantic meaning and if it did, what that meaning might be. Then, two months later, they were asked the same questions about these gestures performed by NAO. Furthermore, they were also asked to indicate which gestures seemed the most useful in a daily HRI-context and which one seemed the most useless. Figure 21 shows three of the nine key poses for NAO.

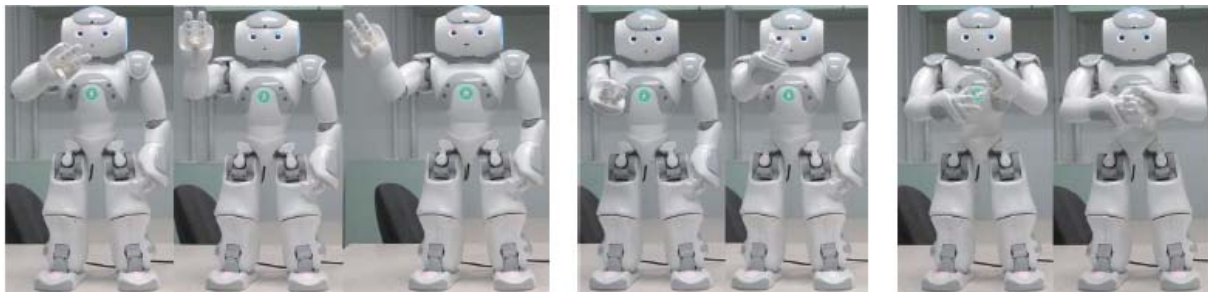


Figure 21 Key Poses for NAO; A. Wave with Forearm; B. Beckon with Palm up; C. Clap (Zheng & Meng, 2012)

Table 7 shows the interpretation participants gave to the nine gestures performed by the human. Beckoning gestures were always correctly interpreted. Waving with the forearm shows the influence of culture on the perception of semantic gestures. While Americans would only interpret this movement as a way of saying ‘*bye*’ or ‘*hello*’, Chinese participants recognised it as a common Chinese gesture for ‘*no*’. Cultural background also explains the fact that one participant interpreted waving with a whole arm as a way of saying ‘*I am here*’.

Perception of Semantic Gestures by Human			
Nr	Gesture	Meaning	Rate
1	Beckon (palm up)	Come here	16/16
2	Beckon (palm down)	Come here	16/16
3	Wave (forearm)	Bye Hello No	10/16 4/16 2/16
4	Wave (whole arm)	Bye Hello I'm here	7/16 8/16 1/16
5	Direct (head and one arm direct to the same orientation)	This way please Introduce somebody / something	13/16 3/16
6	Hand scratch head	Let me think Wonder / Doubt / Puzzle I don't know	8/16 5/16 3/16
7	Bow	Hello / Show respect / Thanks	16/16
8	Clap	Encourage / Praise / Welcome / Joy Not recognised as semantically meaningful	14/16 2/16
9	Shake head and both palms facing up	I don't know Helplessness Not recognised as semantically meaningful	11/16 4/16 1/16

Table 7 Perception of Gestures Performed by Human (Zheng & Meng, 2012)

When comparing these results with the perception rates of the gestures performed by NAO, important differences can be noticed. In general, all recognition rates were lower when NAO was observed. Moreover, five out of nine gestures were considered semantically empty by some of the participants. The gestures that were recognised were usually interpreted in a similar way, except for the beckoning gestures. When NAO beckoned with his palm up, some participants perceived it as a way of saying 'look at me'. The second beckoning gesture (with the palm down) had low recognition rates and was not even considered semantically meaningful by nearly half of the participants. This was caused by the limited degree of freedom of NAO's hands and arms. Table 8 shows the perception of the gestures made by NAO.

When asked to vote for the most useful gesture, seven participants chose the directing gesture. The beckoning gesture with the palm down was nominated the most useless gesture by five participants (they did not even attribute it any semantic meaning whatsoever).

The results of this experiment indicate that the interpretation of semantic gestures is indeed culturally dependent and that even within one culture, interpretations can differ based on the situation. For example, the waving gestures can have various meanings, determined by the context. Furthermore, sometimes multiple gestures can be used to express the same idea. For example, both scratching the head as headshaking with palms up indicate uncertainty. Moreover, there is indeed an important difference between the perception of human gestures and the perception of robotic gestures. It is therefore crucial to first test gestures on the intended population before implementing them into the robot to avoid misinterpretation.

Perception of Semantic Gestures by NAO			
Nr	Gesture	Meaning	Rate
1	Beckon (palm up)	Come here Look at me	13/16 3/16
2	Beckon (palm down)	Come here Not recognised as semantically meaningful	10/16 6/16
3	Wave (forearm)	Bye Hello No	10/16 4/16 2/16
4	Wave (whole arm)	Bye Hello I'm here Not recognised as semantically meaningful	9/16 5/16 1/16 1/16
5	Direct (head and one arm direct to the same orientation)	This way please Introduce somebody / something Not recognised as semantically meaningful	11/16 4/16 1/16
6	Hand scratch head	Let me think Wonder / Doubt / Puzzle I don't know	6/16 8/16 2/16
7	Bow	Hello / Show respect / Thanks / Welcome	16/16
8	Clap	Encourage / Praise / Welcome / Joy Not recognised as semantically meaningful	14/16 2/16
9	Shake head and both palms facing up	I don't know Helplessness Not recognised as semantically meaningful	10/16 5/16 1/16

Table 8 Perception of Gestures Performed by NAO (Zheng & Meng, 2012)

4.9 End-user Programming

4.9.1 User-friendly Programming

Although robotic research has made enormous progress in the last decade, use of social robots in real-life situations remains rare (Lourens & Barakova, 2011)³⁷. This can be explained partly by the fact that controlling robots still requires too much technical knowledge. If robots are to be used in health care, for example, people with a medical background need to be able to program a robot according to the needs of the situation. This requires user-friendly programming environments which allow non-technical users to build complex social interactions. Lourens & Barakova have proposed such a framework for NAO: it allows end-users to use minimal programming to create social behaviours for the robotic platform. The system is based on the concept of re-using modules, which can be combined in different ways to simulate a range of behaviours. For example, a command could be built as follows: '[a|b] & c|d & e'. This command consists of five modules: modules a and b will be executed in parallel (indicated by the "|"), while d and e will be executed subsequently (indicated by the "&"). This leads to the execution of all five modules following the pattern shown in Figure 22.

³⁷ The section *User-friendly Programming* is based on (Lourens & Barakova, 2011), unless otherwise indicated.

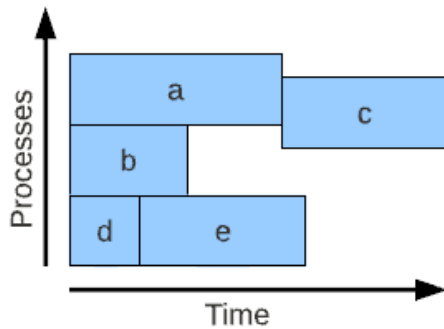


Figure 22 Graphical Representation of the Execution Order (Lourens & Barakova, 2011)

Such minimalist coding makes it much easier for people without a technical background to control robotic behaviour. An experiment was conducted in which five participants with a non-technical medical background were asked to program NAO to conduct four tasks (Barakova et al., 2013). The results show that it was possible to complete this programming within 25 minutes, which indicates that coding-skills are no longer required to control robots (Barakova et al., 2013).

4.9.2 Cybele : a Motion Description Language

In the above framework, however, some minimal programming is still required. It would be useful if the end-user could program robotic behaviour in natural language as well. For this reason, Shukla & Choi have developed a descriptive language and framework that allows users to specify motions for NAO through dialogues (Shukla & Choi, 2013).³⁸ Their motion description language consists of four different layers that are used to provide syntactic and semantic structures. The Joint Angle Layer is the most detailed layer and is used to control the robot’s movements directly. The Path Layer describes a motion in a slightly more abstract way, as a series of dots connected by a line. The Motion Primitive Layer is more abstracted still, describing basic motions which can be used to build more complex ones. Finally, the Motion Sequence Layer provides a high-level description of a motion as a whole motion block. Examples of motions described at each layer can be found in Table 9.

Motion Description Layers	
Motion Sequence	Walk, Run, Jump, turn
Motion Primitive	Raise, Lower, Forward, Backward
Path	Hand(v1), Foot(v1, v2)
Joint Angle	Knee Joint 30, Elbow Joint 45

Table 9 Motion Description Layers (Shukla & Choi, 2013)

Based on this framework, Shukla & Choi developed Cybele, a motion description language. This object-oriented language provides a simple syntax to combine semantic elements (motions) into parallel or sequential patterns of execution.

The framework and language were then integrated in a dialogue system, which allows users to communicate with NAO in natural language. This system consists of two parts: a dialogue module to manage the HRI and an action module to perform the desired motions. In order for users to teach NAO a particular behaviour, they first need to direct the robot to enter the learning mode. Using his text-to-speech module, NAO will then ask his user to give him instructions. Next, he will transform these instructions into text by using his speech recognition module. This text is then converted by the framework into motion descriptions. If the instructions given by the user were unclear or insufficient, NAO will ask for more information, all the while updating the motion descriptions. When NAO has

³⁸ The section Cybele: a *Motion Description Language* is based on (Shukla & Choi, 2013), unless otherwise indicated.

fully understood the instructions, the action module will execute the motion description. Figure 23 shows some of the instructions that were given to two NAO robots during the testing of the dialogue system.

In [A], two NAO robots were instructed to greet each other. In [B], one robot was told to sit down while the other had to remain standing during the greeting process. The robots were also trained perform a series of motions when recognising an apple, as can be seen in [C]. Furthermore, the robots were taught to interact with other robots, such as Sony's AIBO robotic dog in [D]. In [E], the robots were instructed to grab the tail of the robotic toy dinosaur PLEO and in [F], they tried to reach for each other while avoiding the I-SOBOT robot next to them.



Figure 23 NAO Interactions. A. Greeting; B. Sitting Down; C. Apple Recognising; D. Petting AIBO; E. Playing with PLEO Dinosaur; F. Reaching (Shukla & Choi, 2013)

4.9.3 RIOLA: A Robot Interaction Language

As mentioned before, the ultimate goal of NLP with robots is to allow HRI in natural language. However, although progress is being made, the current technology is not yet able to provide easy-to-use and faultless speech recognition (ROILA, 2015)³⁹. Therefore, the Eindhoven University of Technology has developed ROILA (Robot Interaction Language), an artificial language that robots can understand without problems and that is not very hard for humans to learn. It consists of a grammar without exceptions and the vocabulary is easy to pronounce as it consists only of phonemes that appear in most natural languages. When pronounced, these words differ more significantly than natural language words, which means that robots have less difficulty selecting the right word from their integrated vocabulary. Moreover, the research team of the university argues that as robots will become a new community in our society, they deserve their own language. We will now shortly introduce the characteristics of this alternative to natural language HRI.

Eleven consonants and five vowels make up the set of available letters in ROILA. These can be combined into three different word types (CVCV, CVCVC and CVCVCV). These types were selected because they were easier recognised during speech recognition tests.

³⁹ The section *RIOLA: A Robot Interaction Language* is based on (RIOLA, 2015), unless otherwise indicated.

The meanings of these words are based on simple English words. For example, consider the subsequent English sentence:

I really liked to know you.

This sentence would be translated to ROILA in the following way:

Pito loki jifi bati bama

As this sentence shows, there is little or no resemblance between ROILA and English. A word-by-word translation might shed some light on the way in which ROILA functions:

English	<i>I</i>	<i>really</i>	<i>liked</i>	<i>to</i>	<i>know</i>	<i>you</i>
ROILA	<i>Pito</i>		<i>loki</i>	<i>jifi</i>	<i>bati</i>	<i>bama</i>
Literal	<i>I</i>		<i>love</i>	<i><Past</i>	<i>know</i>	<i>you</i>
Meaning				<i>marker></i>		

As this translation shows, word markers are used to fulfil the function of affixes. Instead of using a past form of the verb *to know*, the present (and only) tense is used followed by a past marker. This, together with other characteristics such as the absence of gender, makes the grammar quite easy to understand.

This language thus provides a – relatively – simple alternative for HRI based on NLP. The communication indeed proves smoother because of the lower number of miscommunications, but it should be kept in mind that not every user will be willing to learn an artificial language to interact with a robot. After all, many robots are designed to assist the elderly, and they might be less eager to adopt ROILA.

4.10 Conclusion

In this chapter, we have discussed different possibilities to integrate natural language in HRI. We have started out by discussing an experiment in which joint attention and multi-instance learning were used to teach a “native” language to a NAO robot (Dindo & Zambuto, 2010). When presented with a set of objects on a table, NAO was able to identify correctly the desired object after having learned descriptions of subsets of these objects (Dindo & Zambuto). This proves that joint attention is indeed an important factor to allow robots to learn language. However, it must be kept in mind that NAO was only able to learn grounded words this way. The to-be-learned words were all words that referred to objects in the surroundings. Not every word of natural languages is grounded, however. Abstract words such as *love*, *beauty* or *compassion*, can therefore not be taught in this way. Furthermore, not all words used during the experiment were learned either and thus needed to be hardcoded, such as the verbs *to point* and *to grasp*. This limits the effectiveness of such learning techniques, as they can only be applied to a limited set of words.

In the second part of this chapter, we have presented three possible frameworks which allow robots to use natural language. Of course, these are only three out of many more possibilities, but these seemed the more interesting variants as they presented a lot of differences. Barabás et al. have created a frame-based architecture in which 18 basic commands were implemented (Barabás et al., 2012). The experiment showed that frame-based dialogue systems can indeed be used for NLP, but further research into optimizing the speech recognition process should be carried out.

Kruijff-Korbayová et al. created an event-based dialogue architecture in the context of the ALIZ-E project, an international initiative that aims to create suitable robotic companions for diabetic children (ALIZ-E will be discussed in more detail in section 7.3). They first tried to build their system based on a finite-state machine dialogue manager, but they soon found that this would not allow for the required flexibility (Kruijff-Korbayová et al., 2012). Therefore, they switched to dialogue manager that used probabilistic methods and optimisation of dialogue policies based on reinforcement learning (Kruijff-

Korbayová et al., 2012). This allowed them to successfully implement three different games that would increase the physical activities of the hospitalized children.

The reward-based meta-cognitive framework as proposed by Pipitone et al., was especially created to stimulate linguistic creativity. This system will incite robots to aim for dialogue rewards, which they will get when they manage to capture the interest of the human (Pipitone et al., 2014). Robots can either achieve this by changing the topic when the human loses interest, by searching more information about the interests of the human or by limiting the duration of speech turns (Pipitone et al., 2014). To do this, they need to extract sources from two different knowledge bases, which either contain lexical units or the internal representation of the dialogue domain (Pipitone et al., 2014).

Turn-taking, the topic of the third part of this chapter, has proven to be an important aspect of human communication which humans automatically insert into and expect of HRI (Baxter et al., 2013). As it is thus important for robotic designers to integrate this into HRI, we have compared the turn-taking behaviour of two different robots, namely NAO and Kismet. The NAO experiment, conducted by Kruijff-Korbayová et al. in the context of the ALIZ-E project, consisted of several sessions in which children were allowed to play three different games with the robot (Kruijff-Korbayová et al., 2013). During these activities, NAO could optionally display signs of familiarity. The results indicate that children increasingly partake in turn-taking behaviour after several sessions, and that familiarity has an important influence on their communicative behaviour (Kruijff-Korbayová et al., 2013). They also showed that children adapt rather quickly, as the largest adaptation happens between the first and second session (Kruijff-Korbayová et al., 2013). Furthermore, other research has shown that NAO's turn-taking abilities can be improved by implementing non-verbal behaviour to regulate turn-taking (Meena et al., 2012).

NAO's turn-taking abilities were then compared to those of Kismet, a sociable robot developed at MIT. Kismet can engage in turn-taking, but at a much slower pace than humans, making the HRI feel less active and dynamic (Breazeal, 2003). However, as Kismet uses paralinguistic cues to regulate turn-taking, humans automatically adapt to her rhythm (Breazeal, 2003). Eventually, this led to a balanced conversation in which the human's expectations were met and the robot was capable of keeping track of the conversation at her own speed.

In the fourth part of this chapter, we have discussed several problems of HRI. In sections 2.5 and 2.6, we had already seen that an uncanny feeling can be experienced when conversing with a robot, but there are other problems that need to be solved to advance HRI. First of all, humans can sometimes feel misled by the robot because of the fact that its behaviour gave them the wrong impression about its capabilities or functions (Fong et al., 2003). Secondly, HRI is often based on manually created templates into which dynamic information can be inserted at runtime (Mitchell et al., 2014). This often makes the conversation seem repetitive and unnatural to the human, which is something which should be avoided at all costs. Mitchell et al. have therefore conducted experiments in which crowdsourcing was used to generate a more diverse set of templates. The results indicate that this is indeed a possible method to be used in this context, but, it should be kept in mind that manual intervention is still needed afterwards to check the new templates (Mitchell et al., 2014).

Next, the difference between open-domain and closed-domain dialogues was discussed. Open-domain dialogues are much more interesting for NLP, as they are not restricted to a particular domain and are thus much wider applicable. Therefore, an experiment with WikiTalk by Wilcock was discussed in which NAO could use Wikipedia to communicate with users on an almost infinite number of topics. In this system, both smooth and awkward topic changes (to a related or unrelated topic respectively) are implemented, as is the possibility for interrupting NAO (Wilcock, 2012). To enable this, NAO uses beat gestures to accompany the words that are hyperlinks in the Wikipedia articles, to indicate that the users can repeat that particular word to get more information about it (Wilcock, 2012). This has proven to be

an effective way to manage smooth topic changes at the moment, even though this might become less efficient in the future when all words on Wikipedia will be changed into hyperlinks (Wilcock, 2012).

The above-mentioned dialogue control systems only deal with one main dialogue. However, it would be useful to create systems in which the main dialogue can be subdivided over multiple sub-dialogues, as this would allow for re-use and for the creation of more specific dialogues (Cuayáhuitl & Kruijff-Korbayová, 2012). We have discussed two types of hierarchical dialogue control (HDC) systems in section 4.6, namely strict and flexible HDC. Flexible HDC can be seen as more useful in the context of HRI as it provides flexible dialogues of which the form can be controlled by the user (Cuayáhuitl & Kruijff-Korbayová, 2012). The most important advantage of this kind of system is that it can easily be used in different HRI scenarios which would greatly broaden the range of situations in which a robot can be used.

In section 4.7, one of the main functions of natural language was discussed: the management of cooperation. In an experiment, conducted by Petit et al., NAO's capabilities to cooperate with a human being were tested (Petit et al., 2013). The task was to clear a table, based on a shared plan which was instructed to the robot by the human participant (Petit et al., 2013). The cooperation consists of different stages. First, NAO and the user need to establish a shared plan which they will execute together (Petit et al., 2013). Then, NAO needs to learn all the desired actions which can be taught by the user in three different ways: through imitation, through kinaesthetic teaching or through spoken language (Petit et al., 2013). Finally, the plan needs to be executed (Petit et al., 2013). This experiment shows that natural language can perform two very different functions in the context of cooperation: controlling the cooperation (to create or edit the shared plan) and helping NAO to learn an action (as one of the three described learning modalities) (Petit et al., 2013). One thing needs to be kept in mind, however. The user is limited to the use of fairly simple sentences, as complex ones cannot yet be processed efficiently by this kind of systems (Petit et al., 2013). Although this is indeed a limiting factor, this does not seem to impose great difficulties as complex tasks can also be coordinated by simple language (Petit et al., 2013).

Section 4.8 briefly introduced the topic of semantic gestures. This is important because human communication seldom only consists of verbal aspects. Usually verbal and non-verbal behaviour are mixed to enrich the interaction. In chapter 5, body language will be dealt with extensively, while this chapter dealt with another kind of non-verbal language: gestures that transfer semantic meanings. These are considered to be an essential part of natural interaction between humans, and therefore, many designers decide to implement this functionality into robots as well. However, as Zheng & Meng point out, it should be examined first whether or not these gestures are perceived in the same way when performed by a robot as to avoid undesired results (Zheng & Meng, 2012). Furthermore, these gestures are greatly dependent on culture and should thus be designed specifically for the envisioned audience of the robot (Zheng & Meng, 2012). Their experiment showed that some gestures might indeed be performed by the robot, but that the perception rates were much lower than when they were performed by a human (Zheng & Meng, 2012). Moreover, many gestures were interpreted differently when the robot performed them, which proves that this kind of gestures should indeed be tested on a human audience before they are implemented on a robot (Zheng & Meng, 2012).

Section 4.9 dealt with one of the most important reasons why robots are not yet commonly part of human lives: programming and controlling them. Nowadays, most robotic programming and controlling tasks still need an understanding of coding. However, most robots are to be used by non-technical people who might not want to learn how to write programming code. Eventually, natural language would be the best means of completing these tasks, but as this is not yet entirely possible, intermediate solutions are proposed by various research teams. One of these solutions is a user-friendly programming environment which would only require basic coding skills (Lourens & Barakova, 2011). Such an environment might be based on re-usable modules that can easily be re-arranged and re-combined to build complex behaviours without knowing having extensive programming skills (Lourens & Barakova, 2011).

As mentioned before, it would be even more convenient for users to program robots through natural language. Therefore, research in motion description languages that allow end-users to define motions for the NAO robot through dialogue (Shukla & Choi, 2013). Shukla & Choi developed such a language, Cybele, which they used in combination with a four-layered framework and a dialogue system (Shukla & Choi, 2013). This allowed users to give NAO instructions in natural language for the execution of certain motions (Shukla & Choi, 2013). This is an important first step in the direction of natural language end-user programming, but it has not yet come close to fulfilling the eventual dream of complete natural language control. Some researchers, however, remain sceptic of this possibility: they believe that science will not succeed in solving this problem in time (as robots are expected to become an integral part of our daily lives soon). Worst yet, some of these scientists do not even believe that the problem will ever be entirely solved at all. Therefore, they concentrate on alternative solutions, such as RIOLA. This robot interaction language was created to provide end-users with an easy-to-learn artificial language that would increase the performance of HRI (as its artificial words are less likely to be interpreted wrongly by the robot's speech recognition system) (RIOLA, 2015). The Eindhoven University of Technology, who is responsible for the development of this artificial language, believes that this would be a perfect alternative to the out-of-reach natural language based HRI (RIOLA, 2015). Moreover, they believe that robots deserve their own language, as they will most likely become a new class in society (RIOLA, 2015). As good as their argumentation might sound, it is important to keep in mind that not all future users will be as eager to learn a new artificial language as they seem to expect. After all, many robots are designed to be companions for the elderly, often even for the demented. Whether or not RIOLA will prove to be effective in these contexts remains to be seen.

5. Communication of Emotions

5.0 Introduction

In the previous chapter, we discussed natural language in the context of HRI. One of the functions of natural language is to communicate emotions. Speech is a very effective and efficient medium through which emotions can be conveyed and consists of several parameters that can be adjusted to express the desired emotion, such as volume, pitch and prosody (Fong et al., 2003). Furthermore, different speakers use very similar vocal effects to express particular emotions (Murray & Arnott, 1993), which means that this could be easily interpreted by robots. Yet, it is more difficult for robots to use natural language themselves to express emotions, as the quality of synthesized speech is still inferior to the quality of synthesized facial and bodily expressions (Bartneck, 2002). Therefore, robots are more likely to be designed to express themselves through their body (or through a combination of body and voice) instead of through their voice alone.

Humans, likewise, do not only express their emotions through speech; they also use body language. To the careful observer, our body betrays a great deal about our emotional state. Therefore, it is important to examine not only the perception and display of emotions through natural language, but also through other media, such as the body.

The importance of emotion detection and expression to a conversation should not be underestimated. First of all, sentiment is an important factor to interpret speech acts correctly. The emotional state of a person gives an extra dimension to their words. For example, if someone's employer comes into the office smiling and looking pleased, a sentence like '*Come see me in my office later!*' might lead to positive things such as compliments or a raise. However, if he storms into the office with a face like thunder, that same sentence would suddenly sound as a dark omen. Emotional body language and other factors such as speech volume and tone thus help the interlocutor to determine the full meaning of sentences correctly. This leads to a second important function of emotional communication. Because of the fact that the conversational partner can accurately assess the emotions behind a particular speech act, it becomes easier to adjust their own behaviour appropriately. The employee in the previous example should know to react differently to a promise than to a threat if he wants the communication to be successful and efficient.

As robots are designed to interact with humans through natural language, emotion becomes important to them as well. They have to be capable to both display emotions and correctly detect them. In the first part of this chapter, we will take a closer look at some experiments that were conducted in order to develop accurate systems for robots to display emotions. We will first present the research by Beck et al., a research team that developed an affect space for the generation of NAO's emotional body language. Then we will explore two other sets of emotions for NAO, the first one created by Monceaux et al., the second one by Häring et al., We will also consider a study by Cohen et al. in which NAO's body language is compared to the facial emotions of the robot iCat. Then, we will compare the results of the NAO studies with those with other robots such as Kismet, Brian and KOBAN.

In the second part of this chapter, we will explore the possibilities of detecting human emotions. We will present a study by Zhang et al. in which a NAO robot uses a combination of semantic and facial cues to determine the emotion of the human with whom he is interacting. This research thus makes the link between the two major media of emotional expression we have mentioned before: natural language and body language.

Furthermore, we will also discuss an experiment with Brian by McColl et al. in which the Davis Nonverbal States Scale (DNSS) is applied for the first time to HRI. This scale allows for the classification of body language into levels of accessibility, which is an important emotion to evaluate the success of a particular interaction between humans and robots.

5.1 Emotion Expression

5.1.1 NAO's Emotional Body Language

In order for humans to accept robots into their lives and empathize towards them, robots need to be able to display emotions (Beck et al., 2012)⁴⁰. Research into the believability of artificial agents has shown that ‘emotion is one of the primary means to achieve this believability, this illusion of life, because it helps us know that [they] really care about what happens in the world, that they truly have desires’ (Bates, 1994). NAO does not have facial expression and therefore, body language is the best alternative. NAO's eyes can change colour, which can be used to support his emotional expressiveness, but research by Häring et al. has indicated that these led lights cannot be the only elements responsible for the expression of emotions (Häring et al., 2011). Beck et al. have conducted several experiments which showed that humans were better than chance at identifying NAO's emotions, which proves that body language, on the other hand, is a suitable alternative to facial emotions. Please see Table 10 for the recognition rates.

Recognition rates					
Anger	Sadness	Fear	Pride	Happiness	Excitement
88%	85%	92%	88%	73%	73%

Table 10 Recognition rates (Based on: Beck et al., 2010a)

These experiments also indicate that the position of the head is important to express and identify emotions. When NAO's head was up, positive emotions such as happiness, pride and excitement were more easily recognised; the same held for a downward head-position and negative emotions, such as anger and sadness. The perception of fear, however, seems not to be influenced by the position of the head. Figure 24 shows these emotions as expressed by NAO through body language. These poses are based on the performances of a human actor, as they found that there is no difference between the interpretation of emotional body language of artificial agents or of human beings.

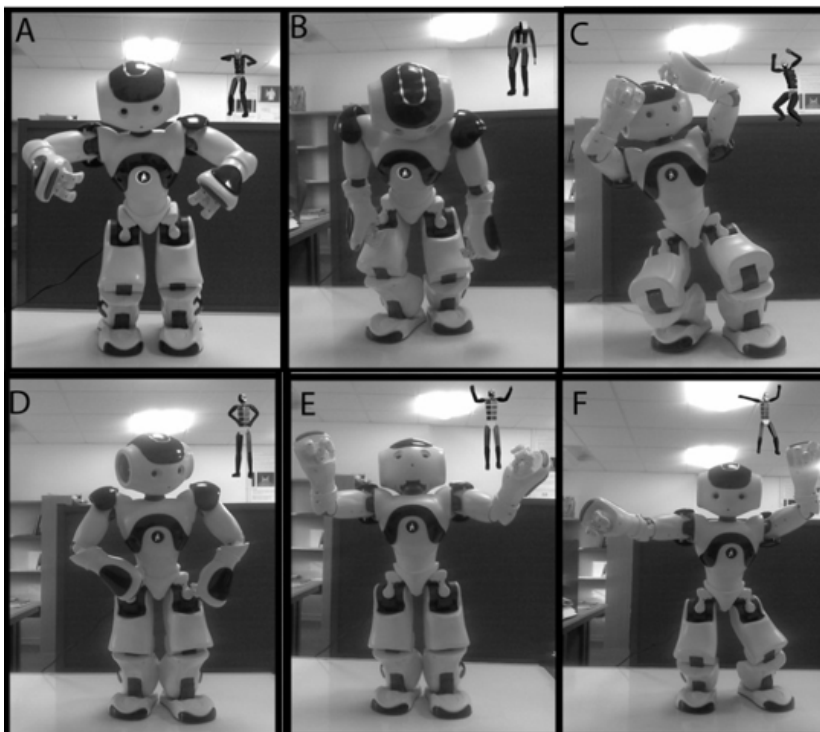


Figure 24 Static poses expressing emotions. A: Anger. B: Sadness. C: Fear. D: Pride. E: Happiness. F: Excitement. (Beck et al., 2010a)

⁴⁰ The section *NAO's Emotional Body Language* is based on (Beck et al., 2012), unless otherwise indicated.

5.1.2 An Affect Space for NAO

Beck et al., have studied the creation of an “affect space” for the generation of emotional body language of robots (Beck et al., 2012)⁴¹. This affect space is ‘generated by blending different emotional expressions to create new ones’ and can be used ‘to improve the expressiveness of humanoid[s]’ (Beck et al., 2010b). The emotions of the humanoid are blended on three dimensions: arousal, valance and stance. Arousal indicates the level of energy used; valance indicates the positivity or negativity of the stimulus; and stance indicates its approachability. This method had only been used in the past for robots with facial expressions, such as Kismet (See chapter 5.1.6). For robots that use body language to express emotions, it is of course vital that this should not interfere with their other functions, such as walking or carrying things (Beck et al., 2010a). This can be realised by only using a minimal set of body parts rather than full body postures to create the affect space (Beck et al., 2010a).

Beck et al. created an experiment in which an algorithm blended between a defined set of key poses to generate new ones (Beck et al., 2010b). These key poses are ‘static posture[s] modelled so that [they] clearly describe the emotion displayed’ (Beck et al., 2010a). They started out with these static positions rather than with movement because it is a well-known method in the field of animation to create believable characters.

Based on earlier research (Beck et al., 2010a) they chose four emotions to represent the extremities of the arousal and valance scale: happiness, pride, fear and sadness. The axes, as shown in Figure 25, were built based on these four emotions in combination with a neutral position. Each emotion is blended with its neighbours at three different levels, as symbolised by the dots: 100%, 70%/30% & 50%/50%.

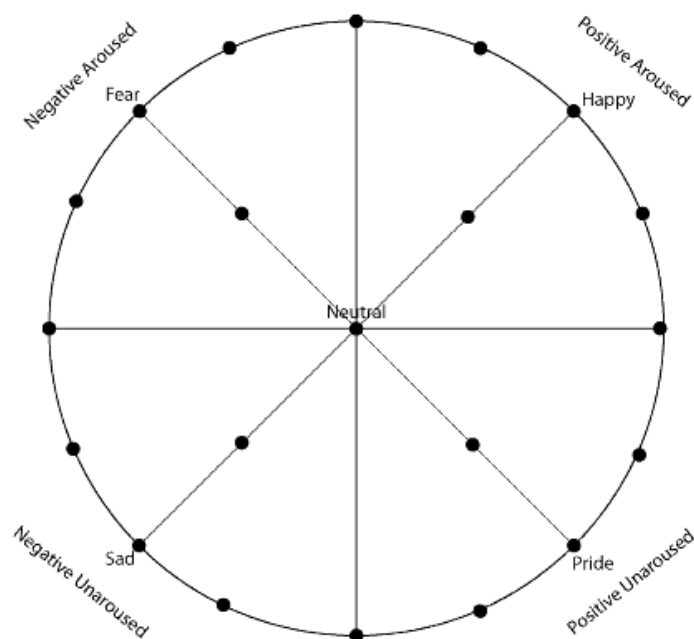


Figure 25 Resulting system based on 2 dimensions: Arousal & valence (Beck et al., 2010b)

The goal of the experiment was to test if the interpretation of the key poses displayed was consistent with their position in the affect space (Beck et al., 2010b). Participants were asked to assess of 20 poses: which emotion (happiness, pride, excitement, fear, anger, sadness, or neutral) is being expressed, which valence and which arousal (Beck et al., 2010b). Please see Figure 26 for five examples of such poses.

⁴¹ The section *An Affect Space for NAO* is based on (Beck et al., 2012), unless otherwise indicated.



Figure 26 Five generated key poses: A: 100% sadness; B: 70% Sadness & 30% Fear; C: 50% Sadness & 50% Fear; D: 30% Sadness & 70% Fear; E: 100% Fear (Beck et al., 2010b)

The results of this experiment showed that a humanoid could display emotions without using its face and that these emotions were interpretable by humans. The participants were able to recognise the four basic emotions and they were likewise capable of interpreting the generated expressions. This means that an affect space can be used to generate automatically different expressions for an emotion. Key poses created by using a 50/50 blend were more difficult to interpret, even though their valence and arousal values were usually correctly assessed. According to Beck et al., these results indicated that the affect space they have created could greatly improve NAO's expressiveness. The robot would no longer be forced to keep repeating the same body language for a particular emotion, and at the same time, this variation would not lower the effectiveness of its emotional communication.

There were some unexpected results regarding valence, however. A blend of fear and sadness was interpreted as negative, but less negative than 100% fear or 100% sadness. This might be explained by the fact that the 100% versions of the emotions are considered prototypical and are thus regarded as more negative.

The results of the experiments show that the perceived arousal can be increased or decreased by blending in an aroused or un-aroused posture. A blend of 50% fear and 50% sadness was interpreted as neutral, but its arousal was rated higher than that of sadness and lower than that of fear. Emotions expressed with the head held up were always interpreted as more highly aroused than those expressed with the head straight or down (Beck et al., 2010a). Furthermore, according to research by Andry et al., arousal is related to the speed of movements (Andry et al., 2001). Beck et al. therefore expect that variations in speed, determined by NAO's arousal, would be beneficial to their model (Beck et al., 2010b).

The experiments also showed that the position of the head also considerably affected stance (Beck et al., 2010a). Poses with the head held up were evaluated as "more approaching" (higher stance) than those with the head held straight (Beck et al., 2010a). Poses with the head held straight were in turn regarded as more approaching than those with the head down (Beck et al., 2010a).

The head position can thus be said to influence the expressiveness of a pose: moving the head up increases the perceived arousal, stance and valence; moving the head down decreases these dimensions. These findings indicate that intuitive signals could be sent by changes in the head position during HRI, which could evaluate the success or failure of a particular communicative act (Beck et al., 2010a).

These experiments did not take into account the effect of culture, and because of the fact that most participants were British, the results of the study might not be applicable to participants of all cultures (Beck et al., 2010a). Furthermore, the participants in these studies were all adults. Therefore, Beck et al. decided to conduct further research to include the interpretation of a robot's body language by children

(Beck et al., 2011)⁴². This is especially important for projects that focus on children, such as the ALIZ-E project (See chapter 7.3). Earlier studies indicate that while children, aged 8 and older, are able to recognise happiness, sadness, anger and fear in naturally generated dance expression (Boone & Cunningham, 1998), their emotional recognition is not yet fully developed until they are mature (Tonks, 2007). Furthermore, it is found that adults and children do not perceive robots in the same way (Woods et al., 2005). Therefore, it cannot be simply assumed that they would interpret the body language of a robot in the same way.

Beck et al. set up an experiment to test how children (aged 11 to 13) perceived the emotions expressed by NAO. For comparative reasons, the experiment was designed to be very similar to the earlier experiment with the adults. Children had to identify six emotions (anger, sadness, fear, pride, happiness and excitement), but due to their age, they did not have to evaluate valence, stance or arousal. The results showed that the children, like the adults, were able to identify the expressed emotions. However, unlike in the first experiment, there were strong variations between the recognition rates of the different emotions: ranging from 58% for anger to 100% for pride. The results also showed that children were likewise better at interpreting positive emotions when the head was up and negative ones when the head was down, similar to the way adults used this visual clue. Furthermore, the children's perception of fear was not influenced by the position of the head, as had already been concluded from the experiment with adults. Overall, it can be concluded that the results of these experiments were more or less consistent with one another.

5.1.3 The 2009 Library of Emotional Expressions for NAO

In 2009, Monceaux et al. developed a library of emotional expressions, to allow researchers to create more complicated forms of behaviour for NAO (Monceaux et al., 2009)⁴³. These expressions are based on a combination of three elements: joint movements, eye colours and sounds. Monceaux et al. created 40 different behaviours, related to 15 emotional states and ordered from pleasure to neutrality to displeasure. They also developed variations of duration and intensity. These motions were combined with different sounds. The researchers, concerned about consistency, chose 'sounds similar to those a being the size of [NAO] would produce, such as high-pitched samples, as the text-to-speech cannot produce onomatopoeia' (Monceaux et al., 2009). These sounds also vary with the duration and intensity of NAO's emotional behaviour. To this combination of motion and sound, they added different eye colours.

Monceaux et al. point out that researchers who want to use their library should consider two important limitations. Firstly, as emotion is culturally defined, it is possible that certain behaviours could be wrongly interpreted by users of some cultures. Secondly, NAO does not possess certain physical characteristics, such as eyebrows, to express subtle behaviours, which might make them difficult to guess out of context.

5.1.4 The 2011 Library of Emotional Expressions for NAO

A few years later, in 2011, Häring et al. created a comparable set of emotions, using a NAO V3+ Academic Edition (Häring et al., 2011)⁴⁴. This set consisted of four emotions (fear, joy, sadness and anger) of which two versions were developed each time. These emotions were once again combinations of body movements, sounds and eye colours. The motion used by NAO to express his emotions was created as to approximate human behaviour. Then, sounds were added to these movements and the

⁴² The information on the experiment with children in this section is based on (Beck et al., 2011), unless otherwise indicated. This study was conducted in the context of the ALIZ-E Project. For more information, please see chapter 7.3.

⁴³ The section *The 2009 Library of Emotional Expressions for NAO* is based on (Monceaux et al., 2009), unless otherwise indicated.

⁴⁴ The section *The 2011 Library of Emotional Expressions for NAO* is based on (Häring et al., 2011), unless otherwise indicated.

research team chose to use emotion related human and/or animalistic sounds in most cases, fear being the only exception. The sounds that accompany joy are both based on human expressions, but they are created in completely different ways. Version 1 is generated by the text-to-speech module of NAO, while version 2 is the recording of a male voice. Furthermore, eye colours were added, resulting in the combinations as shown in Figure 27 and Table 11.



Figure 27 Combined emotional expressions: Anger 1, Anger 2, Fear 1, Fear 2, Joy 1, Joy 2, Sadness 1, Sadness 2 (Häring et al., 2011)

Combined emotional expressions				
		Motion	Sound	Eye colour
Anger	Version 1	Furiously gesticulating with arms Leaning forwards Raising arms & shaking	[Frantic noises] Ranting in gibberish voice Bleeb sounds (censoring curse words) Rolling thunder	Glowing red
	Version 2	Turning head to left (fixating someone) Turning body left & clenching & shaking fist (restraining itself)	[Frantic noises] Growling like dog	Glowing red
Joy	Version 1	Dance of joy (like a cheerleader)	[cheering] 'Jippie Yay!'	Bright yellow

	<i>Version 2</i>	Slowly raising arms over head then pulling them down fast. Bending knees a bit (Winning pose)	[cheering] 'Yehaa!' (like a cowboy)	Bright yellow
Sadness	<i>Version 1</i>	Hands before face Moving head from side to side (wiping away tears)	[crying] Crying like a small child	Dark violet
	<i>Version 2</i>	Going limp Crying in armpit Sighing Lifting arms and knees (taking a deep breath)	[crying] Crying like a woman but alienated to sound like a robot	Dark violet
Fear	<i>Version 1</i>	Raising arm before head (protecting itself) Shying away with upper body	[non-human/animal sounds] Loud metallic bang Echo	Dark green
	<i>Version 2</i>	Cowering & trembling	[non-human/animal sounds] Modulated sound that grows louder	Dark green

Table 11 Combined emotional expressions (Based on: Häring et al., 2011)

Eye colour is not an element of natural communication, as humans cannot change their eye colour. However, colours can have an effect on our emotions and therefore, they might be used to support our perception of some emotions. Choosing colours to match emotions is a difficult task, as most theories on this subject are either artistic or esoteric. The research team therefore used commonly known examples: the red eyes of aggressive robots in science fiction films, the dark violet tones of sad Disney scenes, the bright and warm colours associated with positivity and dark colours associated with negativity.

During an open lab day, a pre-test was conducted. 67 Participants were asked to fill in a questionnaire in which they had to identify each of the eight presented emotions as either fear, sadness, joy, anger, neutral or other. The recognition rates (as shown in Table 12) were satisfying in general.

Recognition rates				
	Anger	Fear	Sadness	Joy
Version 1	82.1%	82.1%	95.5%	73.1%
Version 2	94.0%	85.1%	91.0%	74.6%

Table 12 Recognition rates (Based on: Häring et al., 2011)

Some recognition rates are lower than they should actually be, because of the strict way in which the label “other” was used: many people gave synonyms, which should have been considered correct. To the surprise of the researchers, however, the second version of joy was sometimes labelled as anger, probably because of the final position of the movement. It was expected that the second version of fear would be confused with the first version of sadness, as these were very similar. However, while Fear 2 was indeed often mistakenly identified as sadness, Sadness 1 was only occasionally labelled as fear. Because of the fact that some participants indicated that they were able to identify emotions based on a single modality, an experiment was conducted to measure the expressivity of each cue separately.

In this experiment, the eight body movements, eight sounds and four eye colours were separated, resulting in 20 expression cues. Participants were asked to assign a specific value for each element on the Pleasure-Arousal-Dominance (PAD) scale (Mehrabian & Russell, 1974), as this would allow the researchers to investigate to which degree each modality contributed to these three dimensions.

To get a global view of the ranking of each emotion on the pleasure-arousal-dominance scale, the three cues were once again combined into expressions of one single emotion. Please see Figure 28 for the results.

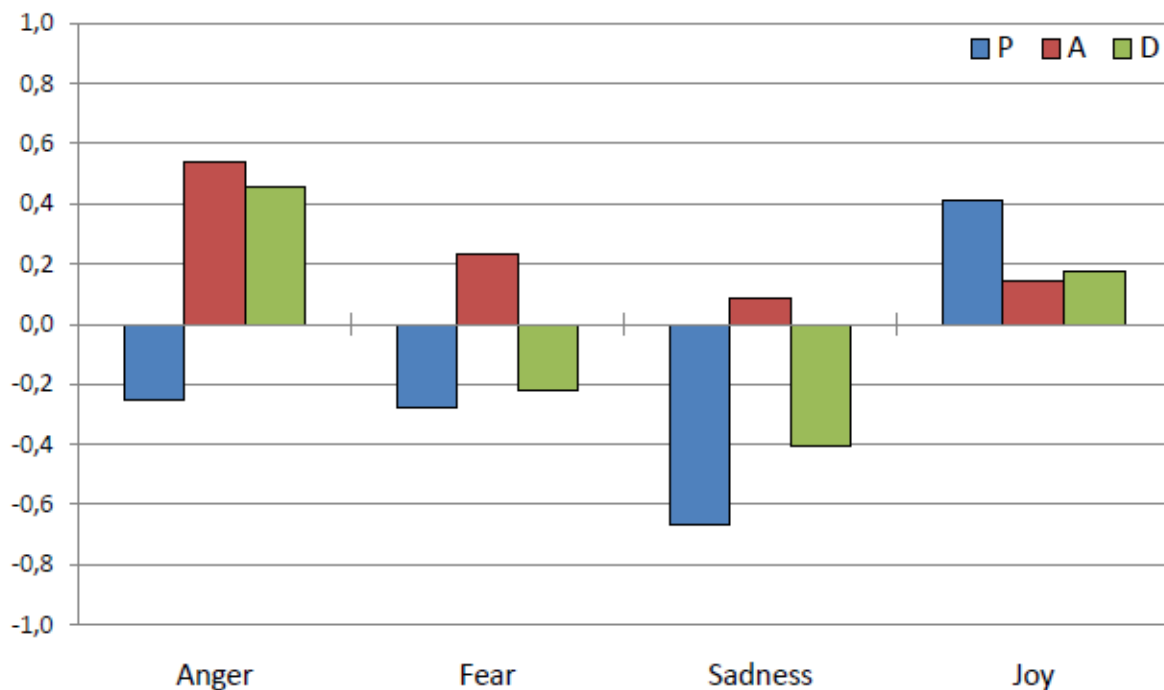


Figure 28 Results for the Pleasure – Arousal – Dominance (PAD) measurements (Häring et al., 2011)

Figure 28 shows that there are significant differences between the emotions on the pleasure scale, except for fear and anger. This was expected because it is generally accepted that anger and fear cannot be distinguished based on pleasure. Within the arousal dimension, anger stood out by being significantly different from the three others. All emotions differed meaningfully within the dominance dimension, except for fear and sadness.

The research team then examined the results for each emotional cue separately. They found that the separated expression cues for anger fell, as expected, within the octant of hostility (-P+A+D), except for the sounds used in Anger 1. The gibberish sounds were interpreted as excitement (+P+A+D) instead of hostility, so they indicated that future research would be needed to solve this ambiguity.

The cues for fear were all categorised as anxious (-P+A-D), as would be expected by the PAD octant, except for the eye colour, which was interpreted as docile (+P-A-D) and the second sound, which was

identified as hostile. Häring et al. assume that the wrong interpretation of the eye colour might be due to the lighting of the laboratory, which leads to an important problem. If colour is perceived differently depending on the lighting, this would prove to be a relatively unreliable factor. After all, no designer could have any influence over the conditions in which the robot would be used as a daily companion. This might thus result in completely undesired and unexpected perceptions on the part of the human user.

Sadness proved to be a difficult emotion to create expression cues for. It was expected that the results should be situated within the bored (-P-A-D) octant of the temperament space, but only the motion and sound of the second version were classified as such. The eye colour was identified as disdainful (-P-A+D), while the motion and sound of the first version were interpreted as anxious.

Results indicated that all cues for joy were correctly perceived as exuberance (+P+A+D), except for the sounds of Joy 1 and the colour of the eyes. Those two cues were considered to represent docility instead. The Joy 1 sound was created by NAO's text-to-speech module and was perceived as rather monotone, which explains why the participants did not consider it aroused nor dominant.

Based on these results, Häring et al. concluded that most of the chosen eye colours did not match with the other cues, except for the red eyes expressing anger. Furthermore, half of the sounds did not prove efficient to represent the chosen emotions, as they usually differed in one dimension, leading to a wrong perception. In addition, NAO's text-to-speech module had clearly failed to express joy accurately. The body movements, however, can be considered to be accurate, except for the first variant of sadness.

5.1.5 iCat

All of the above studies focused on body language as the sole medium of emotional expression. However, it would be interesting to compare NAO's expressiveness to a robot which does have the ability to express his feelings through moveable facial features. Such an experiment was conducted by Cohen et al., because they wanted to investigate which type of robot was best suited to support children with chronic diseases (Cohen et al., 2011).⁴⁵ Participants (Dutch children between eight and nine years old) were asked to evaluate NAO's expressions and those of iCat⁴⁶, which is a research robot that can move its eyes, eyelids, eyebrows and lips. Both robots expressed the following five emotions: fear, anger, surprise, happiness and sadness. The results, which can be seen in Table 13, indicate that there were no important global differences between the recognition rates of the two modes of expression. The only emotion which was perceived a great deal better when expressed by iCat was sadness. Furthermore, participants were asked to interact twice with each robot: once within the context of a story and once without. This showed that context, as was expected, did increase the recognition rates. Moreover, children seemed to become better at interacting with the robot: their second results were always superior to those of the first time they interacted with the robot.

Recognition rates					
	Anger	Fear	Sadness	Happiness	Surprise
NAO	96.43%	87.5%	67.86%	89.28%	68.75%
iCat	99.11%	88.39%	94.64%	73.21%	69.64%

Table 13 Recognition rates of NAO and iCat (Based on: Cohen et al., 2011)

⁴⁵ The comparison made in this section between iCat and NAO is based on (Cohen et al., 2011), unless otherwise indicated.

⁴⁶ For more information about iCat, please see (Philips, 2014).

5.1.6 Kismet

In chapter 4.3.2, the robot Kismet has been introduced. As discussed in that chapter, people automatically adapt their turn-taking behaviour to Kismet’s abilities. This is exactly what Breazeal et al. hoped, as Kismet’s task is to engage people in one-on-one conversations and to learn more about social behaviour through these interactions (Breazeal, 1999)⁴⁷.

As Kismet is developed as a sociable robot, it is important that she is able to express and perceive emotions. In chapter 5.1.2, an experiment by Beck et al. was discussed in which they created an affect space for NAO (Beck et al., 2010a). This affect space is based on Kismet’s expressive behaviour: her expressions are based on nine prototypical facial expressions, which are blended along the same three axes as those later used by Beck et al., namely arousal, valance and stance (Beck et al., 2010a).

Figure 29 shows the architecture on which Kismet is based and which allows her to combine perception, attention, internal drives, emotions and motor skills to perform complex social interactions. As Kismet is supposed to be a robotic infant, she mimics the behaviour of human children (e.g. highly interested in faces and in moving objects). Therefore, Kismet has access to three basic feature detectors: face finding, motion detection and colour saliency analysis. Her attention system provides high-level influences such as motivation which are combined with these low-level perceptions.

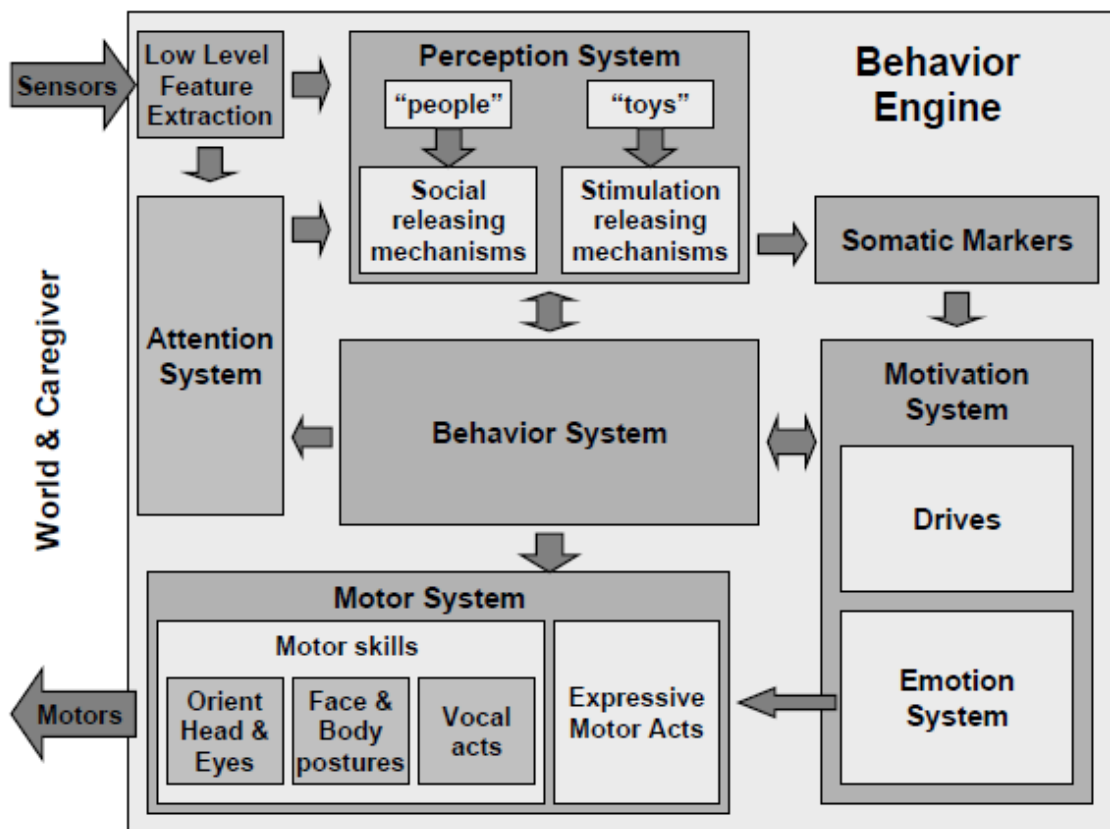


Figure 29 Kismet's Architecture (Breazeal, 1999)

The attention system classifies perceptual stimuli as either social (i.e. moving people with faces) or non-social (i.e. moving, colourful toys). These perceptions are then transformed into releasing mechanisms which contain the nature of the stimulus (social or non-social) and the quality (e.g. presence or absence).

⁴⁷ The section *Kismet* is based on (Breazeal, 1999), unless otherwise indicated.

Based on these releasing mechanisms, particular behaviours are selected. Somatic markers are then used to tag these mechanisms with values for arousal, stance and valence. This way, the affective state of the robot is influenced.

These somatic markers then address Kismet's motivational system, consisting of drives (basic needs) and emotions. Drives can be either under-stimulated (when they are ignored) or over-stimulated (by too many stimuli) and both conditions motivate the robot into taking action to restore the desired drive-levels (the so-called homeostatic levels).

When a drive is in the homeostatic region, Kismet's arousal and valence levels will be moderate, keeping her calm. When they are in the under-stimulated region, however, arousal and valence will drop, making her sad. In the over-stimulated region, these levels will rise causing her to become agitated. Kismet's emotions are thus influenced by her arousal, valence and stance levels. The stronger the emotion, the more her behavioural and attentional focus systems will be influenced. Once a particular behaviour is selected, it influences the actions of the robot and her facial expression.

Breazeal developed learning mechanisms which allow users to train Kismet through emotive channels of communication. When Kismet does something undesired, the caretaker will put her in a negative affective state (which thus mirrors the emotional state of the caretaker). She will learn how to avoid this type of behaviour because she wants to avoid experiencing negative feelings.

This research shows how robotic emotions can influence the social environments they navigate. By showing her emotions, Kismet can encourage the user to take care of her by responding to her needs. The user can likewise influence Kismet, by teaching her behaviours through manipulating her emotions. As humans are intentional creatures, they also expect their conversational partners to be driven by intentions. Therefore, Kismet's architecture is important to guarantee natural HRI.

5.1.7 Brian



Figure 30 Brian 2.0 (ASB Lab, consulted: 4/03/2015)

Brian 2.0 is a human-like robot developed by the University of Toronto to assist elderly persons in their daily tasks (ASB Lab, 2015). Brian's upper body and head have been modelled after an adult male, but he does not have a lower body (McColl & Nejat, 2014)⁴⁸. The research team, led by Goldie Nejat, has conducted experiments in order to determine a body language for Brian 2.0 which would be appropriate for one-one-one HRI.

Like Beck et al. (See 5.1), they believe that 'non-verbal communication [...] convey[s] a human's intent better than verbal expressions, especially in representing changes in affect' (McColl & Nejat, 2014).

⁴⁸ The section *Brian* is based on (McColl & Nejat, 2014), unless otherwise indicated.

Therefore, they created an experiment in order to investigate whether or not a human would be capable of recognising the emotions expressed like the robot with similar recognition rates as for emotions expressed by human beings. The participants were asked to identify the following eight emotions (all of which were judged to be realistic emotions in context of social HRI): sadness, elated joy, anger, interest, fear, surprise, boredom and happiness. Brian 2.0 can use different combinations of movements and postures to express these emotions, as shown in Table 14 and Table 15.

Participants were instructed to watch videos of Brian 2.0, after which they had to choose the emotion that was being displayed according to them. They then had to repeat this procedure with videos of an actor who performed the same emotions while keeping a neutral facial expression (as Brian’s facial expressions had not been activated either) and using the same descriptors for the emotions.

Positive & neutral emotions				
	Elated Joy	Interest	Surprise	Happiness
Trunk	Stretching	Stretching	Stretching	Stretching
Head	Tilted back	/	/	Forward
Arms	Opening	Opening	/	Hanging
Motions	Overall upward	Overall upward & forward	Overall backward	/
Movement Dynamics⁴⁹	High	Low	High	Low
Movement Activity⁵⁰	High	/	/	/
Movement Expansion	Expansive	/	/	/

Table 14 Body language descriptors for positive and neutral emotions (Based on: McColl & Nejat, 2014)

Negative emotions				
	Sadness	Anger	Fear	Boredom
Trunk	Bowing	Bowing	Bowing	Bowing
Head	Forward	Downward	Downward	Tilted back
Arms	Hanging	/	Closing	Hanging
Motions	/	/	Overall backwards	/
Movement Dynamics	Low	High	High	/
Movement Activity	/	High	/	Low
Movement Expansion	Unexpansive	/	/	unexpansive

Table 15 Body language descriptors for negative emotions (Based on: McColl & Nejat, 2014)

As with the research conducted by Beck et al., the emotions used in the experiment were chosen based on their level of arousal and valance. Table 16 shows the recognition rates for the emotions displayed by Brian 2.0. These indicate that participants were better than chance level at identifying emotions.

Recognition rates (Brian 2.0)							
Elated Joy	Interest	Surprise	Happiness	Sadness	Anger	Fear	Boredom
72%	38%	82%	20%	84%	76%	26%	56%

Table 16 Recognition rates for emotions displayed by Brian 2.0 (Based on: McColl & Nejat, 2014)

⁴⁹ Movement dynamics refers to the energy used (McColl & Nejat, 2014).

⁵⁰ Movement activity refers to the amount of movement (McColl & Nejat, 2014).

These results show interesting similarities to those of Beck et al. (See Table 10). In both studies, happiness scored the lowest recognition rate, although the two numbers differ a great deal (NAO: 73%, Brian 2.0: 20%). Both research teams argue that this low percentage is due to confusion with another particular emotion (NAO: excitement 73%, Brian 2.0: interest 38%). McColl & Nejat suggest that other descriptors should be found to distinguish these emotions better.

There is also a remarkable difference between the two sets of results: the perception of fear. While participants of the NAO experiment by Beck et al. scored highest on the perception of fear (92%), those of the Brian 2.0 experiment by McColl and Nejat often failed at identifying fear (26%). In fact, fear has even the second-lowest recognition rate in the latter experiment.

Recognition rates (Actor)							
Elated Joy	Interest	Surprise	Happiness	Sadness	Anger	Fear	Boredom
60%	56%	66%	2%	34%	100%	70%	86%

Table 17 Recognition rates for emotions displayed by the actor (Based on: McColl & Nejat, 2014)

As can be seen in Table 17, the recognition rates for emotions displayed by the actor are not at all similar to those displayed by the robot. Statistically, participants were better at recognising anger, fear and boredom, when these were being expressed by the actor. Elated joy, surprise, and interest were identified with similar recognition rates. Sadness, however, was identified considerably better when expressed by the robot than when expressed by the actor. Happiness scored the lowest recognition rates, both with Brian 2.0 as with the actor. This might again be due to the fact that the descriptors were too similar to those of other emotions.

Furthermore, the researchers point out that when Brian 2.0 displayed fear, participants identified the emotion both as fear and boredom with the same frequency. According to them, the difficulty of this emotion is due to the rigidity of its body. Brian 2.0 cannot easily curl his shoulders, which makes the perception of fear far more difficult. This explains why the recognition rate of fear displayed by the actor is much higher.

McColl & Nejat also suggest that the lack of facial emotions may have contributed to the low recognition rates of happiness and to the confusion between fear and boredom.

This experiment shows that some of the descriptors (those for sadness, elated joy, anger, surprise and boredom) were successful as indicators of emotions and could therefore be considered for HRI settings. It would be necessary, however, to develop better descriptors for happiness, fear and interest.

The robot Brian 2.0 could be further improved by adding other natural communication modes, such as facial expressions and vocal intonation, to his body language to create a multi-modal communication system.

Another variant of the robot, Brian 2.1, was tested in a long-term care facility (McColl et al., 2013). The experiment consisted of observing Brian 2.1 interact with elderly individuals during meal-times and during a card game (McColl et al., 2013). Brian's intended functionality – to assist people with weakening cognitive capabilities – require strong social abilities (McColl et al., 2013). Afterwards, the participants were also given a questionnaire (McColl et al., 2013). Results showed that most participants were engaged in the interaction and listened to the robot's suggestions. Furthermore, Brian 2.1 scored high on the level of sociability as the participants were very enthusiastic about his emotional expressiveness.

5.1.8 KOBIAN

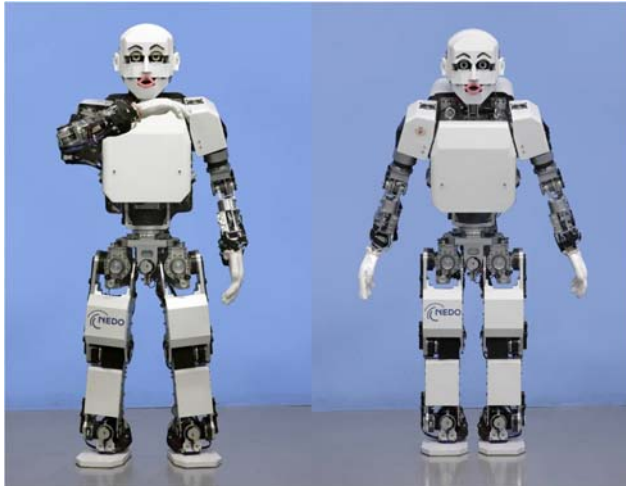


Figure 31 KOBIAN (Takanishi Laboratory, 2011)

Researchers at the Waseda University in Tokyo have created a humanoid, KOBIAN, to study the expression of human-like emotions by robots. KOBIAN is a combination of two earlier robots: WABIAN and WE-4RII. According to Zecca et al., the elderly-dominated Japanese society leads to an increasing need for services to help elderly people at home, both on a physical as on a psychological level (Zecca et al., 2009)⁵¹. They state that a robot should closely resemble a human being in order to achieve a human-like communicative level. This leads to their proposition that ‘humanoids should be designed to balance *human-ness*, to facilitate social interaction, and *robot-ness*, to avoid false expectations about the robots’ abilities’ (Zecca et al., 2009). According to these researchers, robots should especially be able to display happiness and perplexity, as these two elements are considered to be essential to natural HRI.

Zecca et al. conducted several experiments with KOBIAN, the first of which was meant to assess the importance of the body and the face in emotional expression. Participants were asked to recognise emotions (anger, fear, disgust, sadness, happiness, surprise and perplexity) in seven pictures of KOBIAN showing only body pictures and seven more showing a combination of body postures and facial expressions. The results indicate that the combination of both elements is far more effective than the use of either element on its own. Please see Table 18 for the recognition rates.

Recognition rates		
Facial expression	Combination	Bodily expression
44.6%	70%	33.8%

Table 18 Recognition rates first experiment (Based on: Zecca et al., 2009)

These results show that the combination of both facial and bodily expression is the most effective way of displaying emotions. Anger and surprise were much better recognised when a combination was used compared to when only the face was involved in the emotional expression (an increase of 61.7% and 68.5% respectively). Sadness was found to be the only emotion which could be accurately expressed by the body alone.

In a second experiment, Zecca et al. asked a professional photographer and a professional cartoonist to create original emotional poses for the chosen emotions. They designed these, based on the experience they had gained after being allowed to play with the robot for a full day. This led to two sets of postures, some of which were quite similar and some of which were completely different. All of these turned out to be very different from the poses used in the first experiment (created in the university lab). Motion

⁵¹ The section KOBIAN is based on (Zecca et al., 2009), unless otherwise indicated.

patterns for KOBIAN were then created based on the movements of a professional actor performing these 14 expressions. Participants were then asked to watch videos based on the postures created by the lab, by the photographer and by the cartoonist.

In general, the recognition ratio of the professionals was almost identical to the one of the lab (70.5% and 70% respectively). However, there were some important differences on the level of individual emotions. These rates can be found in Table 19⁵².

Recognition rates							
	Anger	Happiness	Surprise	Disgust	Sadness	Fear	Perplexity
Lab	52.3%	43.3%	93.3%	NA	100%	NA	86.7%
Photographer	83.9%	71.0%	90.3%	NA	38.7%	NA	93.5%
Cartoonist	80.6%	29.0%	0%	NA	93.5%	NA	93.5%

Table 19 Recognition rates 2nd experiment (Based on: Zecca et al., 2009)

These results show great – yet inconclusive – differences between the perceptions of the emotions display in the three video sets. The recognition rates of disgust and fear remain low, while the other emotions seem to vary strongly between the three sets. The perception of perplexity, however, remained more or less stable across all three variants.

Zecca et al. then organised a third experiment, in which they asked the photographer and the cartoonist to alter the postures which scored very low recognition rates. These were added to the former set of videos and presented to the participants.

The results of this experiment were extremely negative, as the average recognition ratio was very low. However, the recognition ratio of the set created by the lab itself had also dropped from 70% to 57.5%. Furthermore, while perplexity had been recognised easily in the previous experiment, its recognition rates now fell to 76.4% (cartoonist) and to 70.5% (photographer). Zecca et al. thus concluded that the negative results were mainly caused by low recognition rates of the participants themselves, instead of by inefficient videos.

Zecca et al. conclude from this series of experiments that KOBIAN was limited (due to its hardware) by only being able to express symmetrical facial expressions, while humans strongly rely on unsymmetrical facial cues (such as the raising of eyebrows or the movement of the lips) to express their emotions. Furthermore, problems were also caused by the timing of the movements of the robot, which is also an important factor in the expression of human emotions. These limitations in the hardware of the robot caused KOBIAN to tumble in the uncanny valley (which could not have been predicted based on pictures of the static robot).

When comparing these results to the results of the experiment conducted by Beck et al. (see Table 10), there are some remarkable differences. For example, the recognition rates of the four shared emotions (anger, happiness, fear and sadness) seem to be a lot higher in general when expressed by NAO than when expressed by KOBIAN, even though NAO can only express his emotions through body language.

In 2012, a new version of KOBIAN was developed by the Waseda University: KOBIAN-R. This robot was used to examine the effect of asymmetrical facial expressions on the perception of emotions (Trovato et al., 2012)⁵³. The experiment consisted of a web survey in which participants had to indicate the most appropriate expression for one of four different emotions (disgust, disbelief, annoyance and incomprehension). They could choose each time between a symmetrical and an asymmetrical version. Furthermore, they were asked to do the same thing for different versions of happiness. The participants also had to indicate whether the asymmetrical version concealed an additional meaning.

⁵² The article does not mention the recognition rates of disgust and fear.

⁵³ The information on KOBIAN-R in this section is based on (Trovato et al., 2012), unless otherwise indicated.

The results (as shown in Table 20) indicate that asymmetrical expressions scored higher recognition rates in general. Asymmetry had a completely different effect on the perception of happiness, however, as 80.2% of the participants preferred the symmetrical version, as the asymmetrical versions were considered to hide some negative meaning.

Symmetrical – Asymmetrical Preference			
	Symmetrical	No preference	Asymmetrical
Disbelief	4.00%	17.30%	78.70%
Annoyance	10.70%	37.30%	52.00%
Disgust	24.00%	20.00%	56.60%
Incomprehension	33.30%	13.30%	53.30%

Table 20 Symmetrical - Asymmetrical Preference (Based on: Trovato et al., 2012)

5.2 Emotion Detection

According to McColl and Nejat, robots designed for social HRI ‘need to be socially intelligent in order to engage in natural bi-directional communication with humans’ (McColl & Nejat, 2014). This means that robots should not only be capable of expressing emotions; they should also be able to react adequately to complex human emotions (McColl & Nejat, 2014). Zhang et al. state that HRI would be greatly enhanced if robots could automatically recognise facial expressions, intentions or communicative goals (Zhang et al., 2013). Human emotions, however, are the results of complex psychological processes which cannot be easily recognised, as they differ based on experience, context and individual differences (Zhang et al., 2013). Therefore, it is difficult to use a single modal recognition system to perceive emotions (Zhang et al., 2013).

5.2.1 NAO as a Detector of Human Emotions

Based on the above-described observation, Zhang et al. conducted an experiment in which a NAO NextGen H25 used both facial and semantic clues to interpret the emotions of a human being (Zhang et al., 2013)⁵⁴. The architecture of the system used in the experiment is shown in Figure 32.

NAO’s vision APIs collect facial data, which is analysed by upper and lower facial action analysers, based on neural networks (NN). These analysers can recognise 17 action units, which are contractions or relaxations of facial muscles. These action units are then interpreted as one of six basic emotions (happiness, anger, disgust, fear, sadness or surprise) or a neutral state. Facial data is not enough information for a robot to assess human emotion accurately and therefore, NAO also uses its speech recognition API. Latent Semantic Analysis (LSA) was used to search for semantic similarities between a particular utterance and the training corpus, to discover conversational themes. NAO will then generate an appropriate response, based on the combination of the detected topic and the perceived emotion.

NAO has a great vision of its environment because of the two cameras integrated into his head. However, in the experiment by Zhang et al., the robot only had to recognise facial emotions from frontal views of the participants using its face detection API (AlFaceDetection API). Using this API, NAO is able to process information about the features of the face and to link the face to a previously stored name. In this particular experiment, NAO was only confronted with posed facial expression instead of with spontaneous ones.

⁵⁴ The section *NAO as a Detector of Human Emotions* is based on (Zhang et al., 2013), unless otherwise indicated.

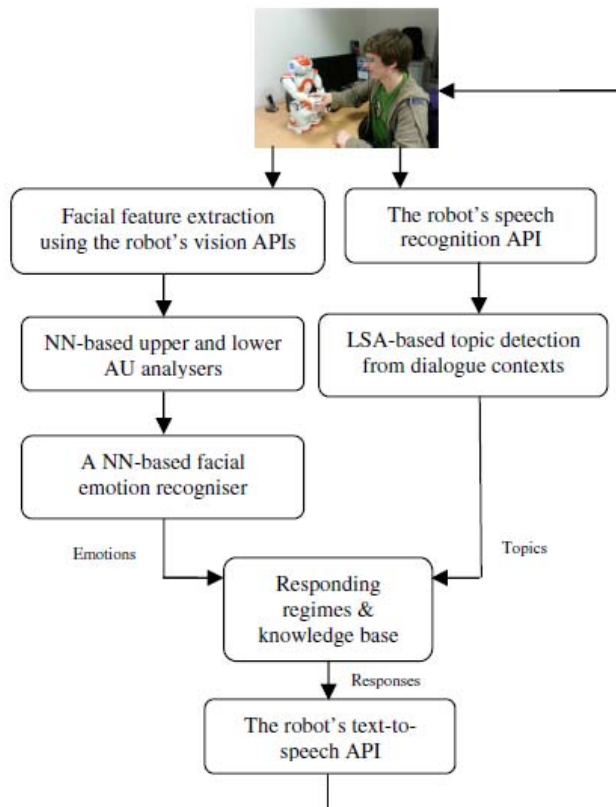


Figure 32 System Architecture (Zhang et al., 2013)

NAO was trained for the experiment in two different ways. For the first training, participants were asked to show him five different versions of a selected emotion or of a neutral state. NAO asked them after each pose to indicate the intensity of the emotion displayed, ranging from 0 to 1. During the second training, NAO was sat in front of a computer screen on which images from a database were shown. This was necessary to enhance the training set with a larger amount of data.

During the experiment, NAO first greeted participants after having recognised their faces. He then explained the experiment and asked them to display a particular emotion. After having determined which emotion was being expressed, NAO used his speech synthesis engine to formulate his findings. The participant was then asked to inform the robot whether or not his statement was accurate.

After NAO had detected the emotion displayed, he started a conversation with the participants, asking why they were experiencing that particular emotion. This open-ended dialogue based interaction used LSA to detect the topic of the conversation. This NLP technique is a method for ‘extracting and representing the contextual-usage meaning of words by statistical computations applied to a large corpus of text’ (Landauer et al., 1998). LSA is based on the idea that the difference or similarity of one word to another is based on restraints on the contexts in which these words can or cannot be used (Landauer et al., 1998). Zhang et al. chose LSA because it goes beyond the limits of formal linguistics and because it proved to be a reliable method for interpreting and managing dialogues. LSA compares meanings by situating words and documents into a ‘concept space’, where they are then compared. To accomplish this, Zhang et al. integrated a semantic vector package into NAO’s platform.

NAO was trained for LSA by providing him with training documents which dealt with topics such as bullying, diseases and school life. These were specifically chosen to fit the ultimate goal of Zhang and al., which was to develop a conversational theme detection system for sensitive topics that might drastically influence the lives of children and teenagers. Furthermore, the researchers included four types of metaphorical examples, because linguistic systems usually experience severe problems when trying

to recognise metaphorical expressions automatically. When a participant said something to NAO, he used his speech recognition engine to transform speech into text. The text was then analysed to determine the topic and to identify metaphorical phenomena.

Based on the detected facial emotion and conversational topic, NAO then generated a response. This was guided by ‘eighty pattern matching appraisal rules and knowledge base’ (Zhang et al., 2013). NAO’s response is then randomly generated from several suitable stored responses. The robot can also decide to use a part of the input in his output, as shown in the following example:

User: *I’m your best mate.*

NAO: *Why are you my best mate?*

The results of the experiment show that the neural network-based facial recogniser performed reasonably well (71.3% accuracy rate). For the recognition rates, please see Table 21.⁵⁵

Recognition rates						
Anger	Disgust	Fear	Surprise	Sadness	Happiness	Neutral
90%	83%	65%	NA	NA	80%	NA

Table 21 Recognition rates (Based on: Zhang et al., 2013)

These results indicate that negative emotions (anger and disgust) were more easily recognised than the other four emotions. The most difficult emotion to perceive was sadness, as it was often classified as anger. Fear and negative surprise were also often confused, as were happiness and positive surprise. The LSA-based topic detection scored likewise reasonably well: 76% accuracy rate for topic classification and 83% accuracy for the recognition of metaphorical expressions. It can thus be concluded from these results that an LSA-based method can successfully and efficiently be used to detect topics in open-ended dialogues.

5.2.2 Emotion Detection in the ROMEO Project

As mentioned in chapter 3.1, Aldebaran joined the ROMEO project in 2009. The goal of this project is to design a robotic companion for elderly people that can likewise interact naturally with children (for example, play with the grandchildren of the user) (Tahon et al., 2011)⁵⁶. To do this, ROMEO will need to be able to express emotions and to detect the emotions of his human companion. Non-verbal and verbal cues will be used to this end. The experiment conducted by Tahon et al. involves a NAO robot (as a test platform for the to-be-developed ROMEO) and the processing of audio cues.

As described above, ROMEO will have to interact with both adults and children. This means that he will need to process two very different speech sounds. It then remains to be examined whether or not a single emotion detection model would be sufficient to deal with these two types. To test this, the research team has chosen to conduct several cross-corpora experiments. They chose the existing AIBO corpus for children’s speech to be used as a well-known reference to test their own corpora against. This corpus contains audio materials of 51 children interacting with the Sony AIBO robot. The AIBO corpus is split into two smaller corpora for this experiment, based on the two schools which are represented: AIBO-Ohm and AIBO-Mont. This corpus is then compared to two corpora created by the research team. The first corpus is the NAO-HR corpus, containing the speech sounds of ten children playing games with NAO (Delaborde & Devillers, 2010). The second corpus is the IDV-HR corpus which consists of audio materials of interactions between 22 visually-impaired elderly people and NAO.

The research team believes that a multi-level processing of audio non-verbal cues is necessary to detect emotions in human speech accurately. As seen in Figure 33, low level cues are used to derive multi-

⁵⁵ The article, while mentioning most recognition rates, does not provide those of surprise, sadness and the neutral state.

⁵⁶ The section *Emotion Detection in the ROMEO Project* is based on (Tahon et al., 2011), unless otherwise indicated.

level markers such as emotion type (positive or negative), activation (active or non-active), emotional labels (e.g. “joy” or “sadness”), rhythm, etc. These data can then be processed on a higher level to reveal emotional and social tendencies of the conversational partner (e.g. a shy person or an aggressive person).

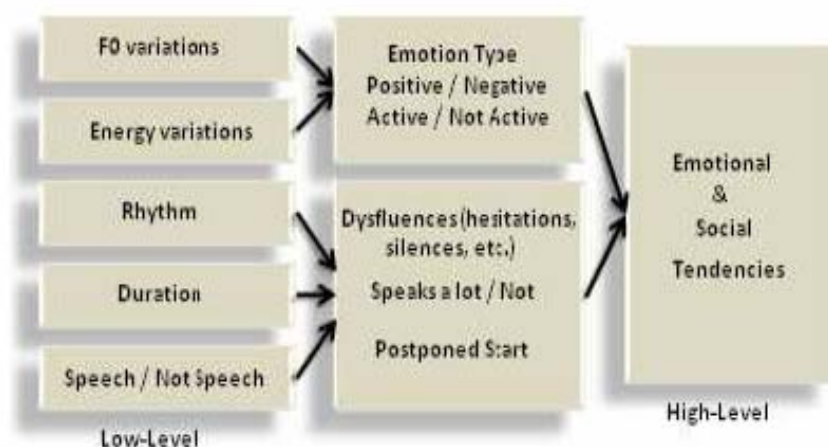


Figure 33 Multi-level Detection of Emotional and Interactional Cues (Delaborde & Devillers, 2010)

This means that the detection of emotion needs to happen on two different levels. On the lowest level, the emotion of a particular speech act is examined. On the highest level, the history of perceived emotion is used to draw conclusions about the user. Based on these two levels, the robot will decide how to behave around this particular human being.

In order to test whether or not two separate models are needed to detect emotions in children speech and adult speech, cross-corpus tests between NAO-HR, IDV-HR, AIBO-Ohm and AIBO-Mont have been conducted. Table 22 shows the Unweighted Average Recall (UAR) performances of the four corpora.

Cross-Corpus Classification				
Training Corpora	Test Corpora			
	AIBO-Mont	AIBO-Ohm	NAO-HR	IDV-HR
AIBO-Mont	/	50.20%	40.25%	33.24%
AIBO-Ohm	53.18%	/	47.06%	31.29%
NAO-HR	38.75%	34.68%	/	29.02%
IDV-HR	30.30%	30.29%	25.05%	/

Table 22 Performance in Cross-corpus Classification (Based on: Tahon et al., 2011)

These results indicate that cross-corpus tests between AIBO-Mont and AIBO-Ohm score high performance levels. This was expected by the research team, as both corpora actually belong to one larger corpus and share a lot of characteristics (e.g. language, annotation protocol and same age group). They also indicate that when the system is trained on AIBO, it performs better when tested with NAO-HR than with IDV-HR. This can be explained by the fact that speakers in the NAO-HR corpus are far younger than those in the IDV-HR corpus (even though they share the same language and the same annotation protocol). This also holds for training with NAO-HR. The system trained with NAO-HR performs better when tested with AIBO than with IDV-HR. In fact, any combination featuring IDV-HR scores below (or on) the random guess level of 33%.

We can thus conclude that it seems far more complex to perform cross-corpora studies with corpora of speakers of different age groups than with corpora of peers. Therefore, two different models will be needed to detect emotions in children speech and adult speech.

5.2.3 Brian as a Detector of Human Emotions

In chapter 5.1.7, Brian’s expressiveness was discussed. However, research has also been conducted in which Brian had to identify correctly emotions expressed by his conversational partner (McColl et al., 2011)⁵⁷. In order for robots to keep a conversation interesting, they need to be able to interpret the human’s body language, as this gives a lot of information about the human’s affective state. McColl et al. therefore propose an automated upper body language identification and classification technique based on a person’s accessibility. Based on the perceived accessibility, the robot can then adapt his own behaviours to suit the context of that particular conversation.

For this experiment, Brian was aided by three separate cameras that record the body language of the human. These recordings were then classified using the Davis Nonverbal States Scale (DNSS), a method for analysing body postures during one-on-one human interactions (Davis, 1991). According to McColl et al., their work was the first in which this model was applied to HRI.

Based on the body poses of the human, the system determined how accessible a human was during his or her interaction with the robot. There are four different levels of accessibility, ranging from I (least accessible) to IV (most accessible). Please see Table 23 for more details of these levels.

Accessibility levels		
	Trunk orientation: Upper/Lower trunk	Arm orientation
Level I	A/A	A, N (if not A or T), T
Level II	N/N, A/N, N/A, T/A, A/T	A, N (if not A or T), T
Level III	T/N, NT Except: positions that involve upright or forward leans	A, N (if not A or T), T
Level IV	T/N, N/T: Combined with upright or forward leans, T/T	A, N (if not A or T), T

Table 23 Accessibility Levels. T: toward; A: Away; N: Neutral (McColl et al., 2011)

Two steps are needed to determine the accessibility of a human automatically. First, the orientation of the upper and lower trunk is determined and expressed using the A, T and N parameters. Furthermore, the forward or upright orientation of the upper trunk is also recorded. This allows for the classification of the trunk in the DNSS model. Second, the accessibility level is made more detailed by recording the orientation of the arms. The arms are then likewise classified in the DNSS model.

This system was tested by an experiment consisting of one-on-one interactions between humans and Brian. During this experiment, Brian expressed himself both verbally and through body language. It was important that participants displayed varying body language and accessibility levels and therefore, each interaction consisted of different parts: an introduction stage, a storytelling stage, a repetitive stage and a silent stage.

During the introduction stage, Brian introduced himself and his capabilities to the participant, after which he asked a set of questions to get acquainted with the human. In the storytelling stage, Brian could either be happy or angry, which was expressed through Brian’s speech features. When he was angry, he talked in a stern and loud fashion. When he was happy, he spoke softer and far more energetically. Next, during the repetitive stage, Brian kept repeating certain phrases or gestures. Finally, in the silent stage, Brian refused to interact with the human.

After the interactions, participants were asked to evaluate the videos of their conversation with the robot. They had to indicate how they would describe the body postures they displayed based on their own experience with Brian. To do this, they used a scale ranging from 1 (least accessible) to 3 (most

⁵⁷ The section *Brian as a Detector of Human Emotions* is based on (McColl et al., 2011), unless otherwise indicated.

accessible). The results of this self-study were then associated to the results of the DNSS system developed by the researchers. This association can be found in Table 24.

Association of self-study results and DNSS results			
	Level 1	Level 2	Level 3
Level I	19	10	3
Level II & III	0	24	9
Level IV	0	0	34

Table 24 Association of the Results of the Self-study and the Results of the DNSS System (McColl et al., 2011)

This association indicates that the DNSS system reached 78% recognition rates for the accessibility levels of the participants. Two arm patterns (“arms crossed” and “arms on the hips”) caused some of the recognitions to fail. “Arms crossed” was usually classified by the participants as level 1, regardless of the orientation of the trunk. “Arms on the hips” was usually considered level 2 if the trunk was either oriented towards the participant or in a neutral state. 45 of the body poses were then randomly selected to be evaluated by a DNSS specialist. These results were compared to the results of the automated system, which turned out to be a relative good match. This comparison can be seen in Figure 34.

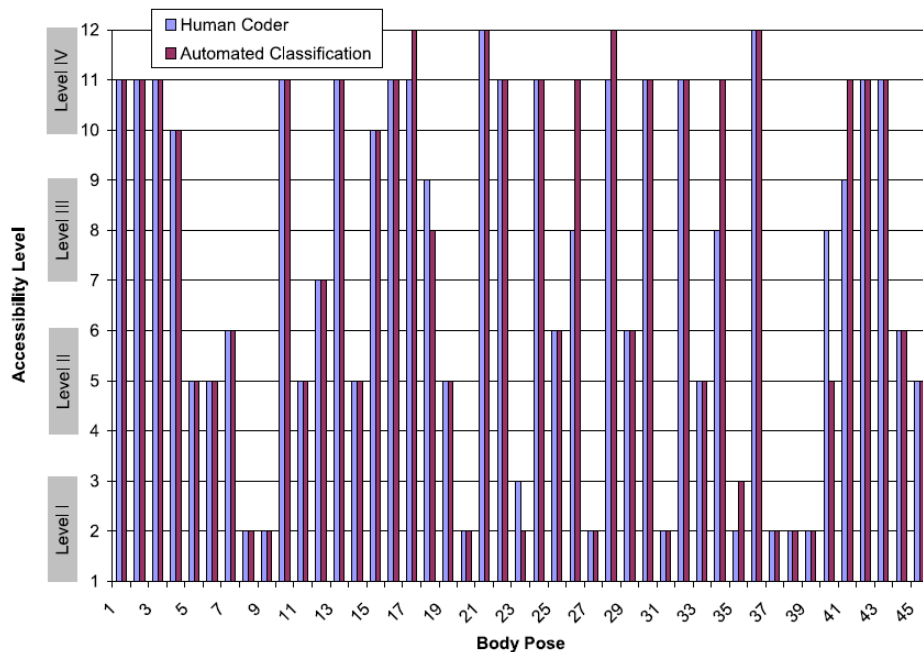


Figure 34 Comparison between Results of DNSS Specialist and the Automated System (McColl et al., 2011)

5.3 Conclusion

This chapter has focused on the two complementary sides of robotic communication of emotions. In the first part, we discussed the way robots themselves can express their emotions and, in the second part, we took a look at how robots manage to detect human emotions.

Robots need to be able to display emotions, as they would otherwise never be accepted as a part of our daily lives (Beck et al., 2012). After all, without emotions, robots would be nothing more than another cold and dead kitchen tool. NAO, however, cannot express emotions through facial expressions (which is one of the main channels humans use) and therefore, body language is the best alternative. Several studies have been conducted to determine how NAO should use body language to display his feelings. These studies showed, for example, that the position of the robot’s head is an important factor in the expression of emotions (Beck et al., 2012). Furthermore, an affect space has been created for NAO, which allows for the generation of body language by blending his emotions on three different dimensions: arousal, stance and valance (Beck et al., 2012). This affect space was used to create

emotions for NAO which contain elements from several emotions (e.g. 30% Fear and 70% Anger), which were then presented to participants for evaluation (Beck et al., 2012). A similar experiment was also conducted with children, because their perception of emotional body language has been found to be slightly different to the one of adults (Beck et al., 2011). The results of these experiments indicate that blended emotions can indeed be used to express body language by robots, but that the age of the user is an important factor to keep in mind (children were less good at recognising some emotions) (Beck et al., 2011).

Other libraries of emotions were created for NAO by various researchers (sections 5.1.3 and 5.1.4) to enhance his expressiveness. Monceaux et al. pointed out the importance of consistency: if NAO expresses emotions through sounds for example, he should use sounds that are expected of a creature of NAO's size (Monceaux et al., 2009). Häring et al. made a similar library and the experiments based thereof showed that eye colour could not be considered a reliable emotional medium (Häring et al., 2011). After all, the lights in the environment alter the perception of NAO's eye colours, which might lead to undesired and unexpected emotional perceptions, which should of course be avoided (Häring et al., 2011).

The final four sections of part 5.1 dealt with other robots whose emotional expressions were compared to NAO's: iCat, Kismet, Brian and Kobian. Cohen et al. compared the expressiveness of iCat (a feline research robot by Philips with facial expressions) to NAO's and found that there were no significant differences between the perception rates of both robots (except for the perception of sadness) (Cohen et al., 2011). Breazeal created an experiment in which Kismet's emotions were used to teach her certain behaviour (she would try to feel good and thus avoid unwanted behaviour because that would make her feel bad) (Breazeal, 1999). This research uses the affect space which was the basis on which Beck et al. have built their affect space for NAO (Beck et al., 2012), which is why this was included here even though there was no explicit comparison with NAO's expressiveness. Furthermore, similar experiments with NAO could prove to be very interesting to study robotic drives and motivations. Next, we discussed Brian, a robot designed to support the elderly (McCull & Nejat, 2014). His recognition rates were surprisingly similar to the ones found in the experiment described above with NAO (Beck et al., 2012), except for the perception of fear (McCull & Nejat, 2014). While the participants in the NAO experiment scored the best rates for fear, those in the Brian experiment often failed to recognise this emotion, which might either be due to the different embodiment of the robot or to the motions chosen to express that particular emotion. Finally, research with KOBIAN has shown that – in his case – emotions are perceived more accurately when expressed through body language and facial expressions at the same time (Zecca et al., 2009). However, when compared to the results of the NAO experiment, KOBIAN scored lower on all four shared emotions, even though NAO does not have facial expressions.

As mentioned earlier, robots should not only be able to express their own emotions; they should be capable of recognising human emotions as well. In section 5.2, we have discussed several experiments which examined NAO's perceptual capabilities. Zhang et al. have conducted a study in which NAO could use both facial and semantic cues to determine the emotional state of the user (Zhang et al., 2013). NAO's face detection API allowed him to detect the first kind of cues; semantic cues, on the other hand, were discovered through the use of Latent Semantic Analysis (Zhang et al., 2013). The combination of the perceived emotion and the detected topic then allowed NAO to formulate adequate responses to the user (Zhang et al., 2013). The results showed that both techniques worked well in the context of human emotion detection (Zhang et al., 2013).

Furthermore, it is important to keep in mind that robots will need to deal with all sorts of people, ranging from young children to the elderly. Therefore, experiments should be conducted to see to which extent emotion detection models could be applied to different age-groups. After all, a part of emotion detection is achieved through detection in audio signals, which differ greatly according to the age of the speaker. Tahon et al. have conducted cross-corpora experiments to examine this, and their results indicate that multiple models should be provided to avoid recognition issues (Tahon et al., 2011).

Next to the experiments with NAO, we have also include one with Brian, as this provided us with a second robot of which we could examine both the expression as the detection of emotions. In this experiment, Brian detected the accessibility of a human user to determine how to proceed with the interaction (McColl et al., 2011). This study shows that the DNNS system used by Brian functioned almost as well as a human DNNS expert to determine a human's accessibility, which means that DNNS is indeed one possible way to detect human emotions (McColl et al., 2011).

6. Influence of Personality Traits

6.0 Introduction

In the previous chapter, we discussed the expression and perception of emotions. Emotions are one of the factors that influence the topic of this chapter, namely personality. Studies have shown that humans would be more willing to interact with robots if these had their own compelling personalities, just as humans themselves have (Breazeal, 2002). However, creating personalities for robots is not straightforward. It is still unclear whether implemented or learned personalities would be most beneficial for HRI, neither is it known if robots should mimic specific human personalities or not (Fong et al., 2003).

Either way, robotic personality has proven to be an important component of HRI. After all, personality can influence the affection felt by the user towards the robot in a positive or negative way (Fong et al., 2003). This is of course vital when creating social robots, as users should first and foremost be comfortable around their companions. Yet, research has also indicated that personality might have a complex influence on the efficiency of a robot, as charming robot personalities do not always inspire users to cooperate better with their robots (Goetz & Kiesler, 2002).

Robotic personalities can be created in different ways. Many studies indicate which verbal and non-verbal behaviours match particular personality types. One of the most visible traits is extraversion, and therefore, it is the trait that is most often discussed in research on robotic personalities. Certain speech characteristics, associated with introverted or extraverted humans, are transferred to robots to imitate introversion or extraversion.

In this chapter, we will first discuss five different robotic personality types. In the second part, we will take a closer look at the influence of personality matching on HRI. This, however, is not an easy topic. Psychology has taught us that humans can apply two different social attraction rules: one based on similarity and one based on complementarity. In other words, interhuman relationships are not at all based on uniform principles. This thus leads to the hypothesis that this would also hold true for human-robot relationships. It needs to be examined, therefore, if HRI would improve if robots resembled their users or complemented them.

Many studies suggest that HRI would benefit if the personality of the robot matched the personality of the user. Therefore, in the second section, we will discuss an experiment in which introverted and extraverted robots interacted with introverted and extraverted humans.

However, there are likewise plenty of studies that indicate that humans would prefer robots with complementary personalities. As literature does not seem to find any consensus on this topic, we have also included three other visions. Sections three, four and five of this chapter are dedicated to the views of three students of the University of Twente who participated in the 2012 Student Conference on IT organised by their university. They proposed the hypothesis that similarity or complementarity do not determine the perception of a robot. In the fourth section, we will present Windhouwer's argument that it is in fact the task context which is important for the preference of certain robotic personalities (Windhouwer, 2012). Next, we will examine Leuwerink's view (Leuwerink, 2012). He proposes that people prefer different robots when they have to interact with them in the context of a group. In the fifth part, we will discuss Waalewijn's idea that perception and preference is influenced by cultural background (Waalewijn, 2012). More specifically, he examines the differences between neighbouring cultures (Germany and the Netherlands). While the results of these three works were not exactly as positive as the researchers hoped, they did lead to some insights into the perception of robotic personalities.

6.1 Personality Types for Social robots

In general, there are five different personality types that are typically used when designing social robots (Fong et al., 2003).⁵⁸ Please see Table 25 for an overview of these types. All of these personalities can be conveyed in various ways: through emotion, embodiment, motion and communication style. Furthermore, the perception of a robot's personality is likewise influenced by the tasks the robot performs.

Personality Types		
	Type of Robot	Characteristics
Tool-like	Smart appliances	Dependability & Reliability
Pet or creature	Toy & Entertainment	Domestic animal traits
Cartoon	Robots for interaction with non-specialists	Exaggerated traits
Artificial being	NA	Science fiction traits
Human-like	NA	Depending on function

Table 25 Personality Types for Social Robots (Based on: Fong et al., 2003)⁵⁹

6.2 Personality Matching

In order for robots to be truly integrated into society as human-like companions, they need a certain insight into the personalities of the users with which they are interacting (Aly & Tapus, 2013)⁶⁰. Furthermore, according to some psychological theories, humans would be more attracted to those who share the same personality traits than to those who are completely different. This theory has been empirically proven in the context of human-machine interaction by various studies (i.a. Tapus & Matarić, 2008).⁶¹ This means that HRI could be improved if robots could match the personality of their users. Aly & Tapus designed an experiment with a NAO robot in which this hypothesis was tested. Furthermore, they also investigated whether or not robotic behaviour would benefit if gestures were added to speech-only behaviour. For an overview of the architecture used in this experiment, please see Figure 35.

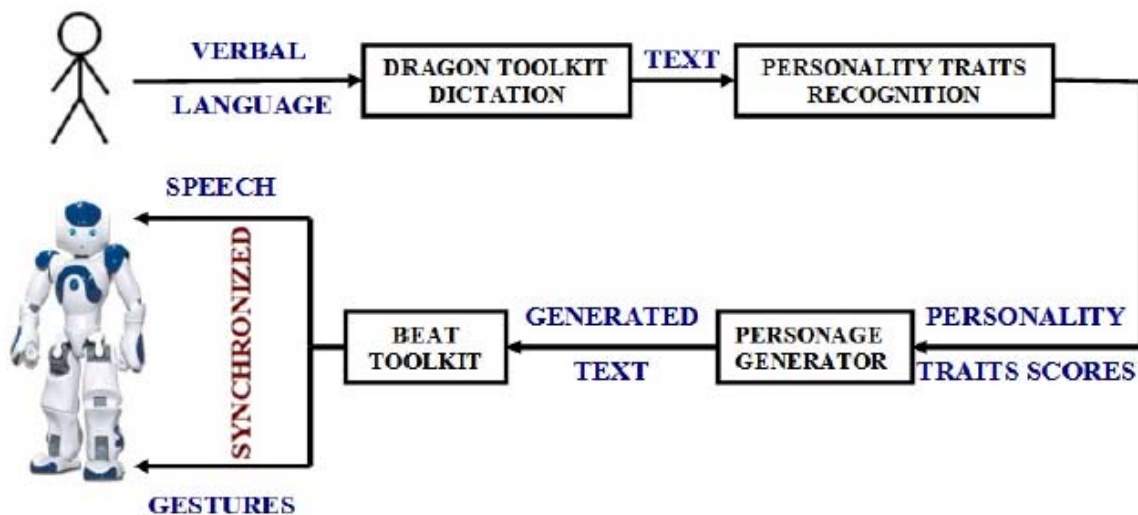


Figure 35 System Architecture (Aly & Tapus, 2013)

⁵⁸ The section *Personality Types for Social Robots* is based on (Fong et al., 2003), unless otherwise indicated.

⁵⁹ Fong et al. do not mention which kinds of robots are usually associated with the artificial and human-like personality types.

⁶⁰ The section *Personality Matching* is based on (Aly & Tapus, 2013), unless otherwise indicated.

⁶¹ Yet, as discussed in chapter 6.3, other studies have empirically proven exactly the opposite.

As shown in Figure 35, the user's speech is transformed into text by the Dragon Naturally Speaking Toolkit. Based on this generated text, a personality analysis of the human is performed by the Personality Recogniser module. The discovered personality traits are then passed on to the PERSONAGE generator, which is a module that generates natural language adapted to the personality traits of the user. The generated text is then passed on to the BEAT toolkit, which transforms it into robotic behaviour using the linguistic and contextual information in the text. Figure 36 shows the architecture of the BEAT pipeline.

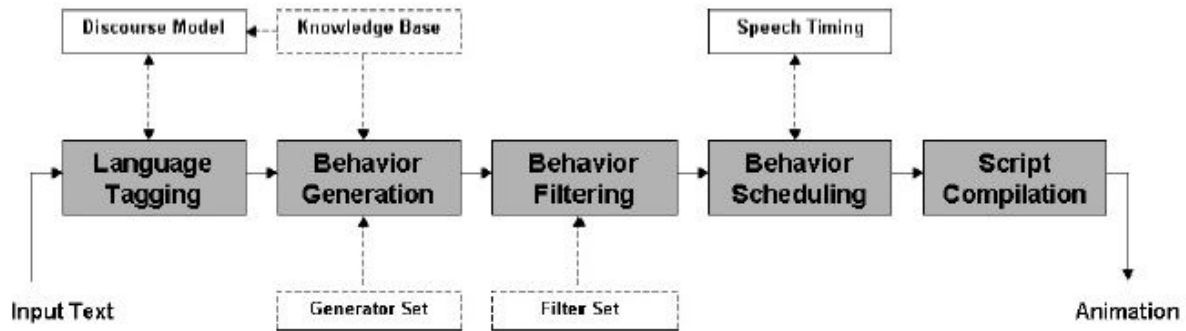


Figure 36 BEAT Architecture (Aly & Tapus, 2013)

BEAT consists of five XML-based modules which are ordered as a pipeline. PERSONAGE creates an XML tagged text and sends it to BEAT's language tagging module, where it is transformed into a parse tree with discourse annotations. The behaviour generation module then suggests all possible gestures based on the output of the previous module. Next, the behaviour filtering module selects the most appropriate set of gestures, based on user-definable data structures (Generator set and filter set). This way, a new XML tree is generated and delivered to the behaviour scheduling module which converts it into a script of synchronised speech and gestures. Finally, the script compilation module compiles the output of the previous module into executive instructions which can be performed by the NAO robot.

In this experiment, Aly & Tapus wanted to explore whether or not users would prefer interacting with a robot with similar personality traits. Therefore, participants were confronted with two different robotic personalities: one extraverted, one introverted. Furthermore, as the research team also wanted to know whether a personality expressed through combined behaviour (speech and gestures) would be perceived as more expressive, participants were confronted with the robot in two different conditions: once when NAO only used speech and once when he combined speech and gestures. Before meeting the robot, participants were asked to fill in a Big 5 Inventory Test to determine their personality traits. The study focuses solely on the extraversion-introversion dimension of the human's personality.

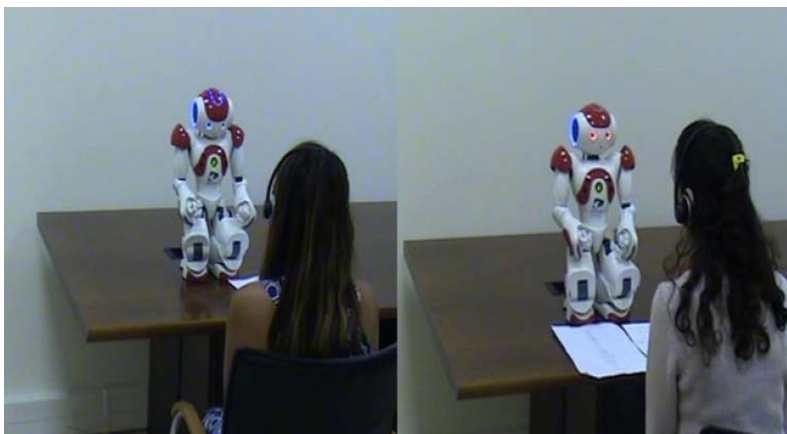


Figure 37 Introverted and Extraverted Conditions (Aly & Tapus, 2013)

Figure 37 shows the difference between the introverted condition of the robot (left) and the extraverted condition (right). When the robot has an introverted personality, his head is down and he uses less gestures. When he has an extroverted personality, however, his head is held up and he gestures much more. Furthermore, in the extraverted mode, the robot will talk more than in the introverted mode.

At the beginning of the experiment, NAO introduced himself and asked the participant to tell him something about New York City. This is an important part of the interaction because this is when NAO's Personality Recogniser uses linguistic cues to analyse the personality of the participant. The robot then presented the participant with a list of restaurants in New York City. When the participant had chosen a particular restaurant, NAO provided information about the food quality, service and price of the restaurant. The PERSONAGE generator and the BEAT toolkit manage the way this information is delivered to the participant by selecting appropriate speech and gestures for the personality type of the participant. The experiment ended when the participant did not want any other information about other restaurants anymore.

Results indicate that both the extraverted as the introverted participants felt closer to the robot when it matched their personality traits. Furthermore, both types of participants agreed that the robot was more engaging, natural and appropriate when using both speech and gestures.

6.3 The Effects of the Task on the Perceived Personality of a Robot

To fulfil social tasks accurately, robots will need clearly defined personalities (Windhouwer, 2012)⁶². Literature on HRI provides inconclusive – and even contradictory – evidence of how robotic personalities should be matched to human personalities. After all, some people feel attracted to others with a similar personality, while others want complementary personalities. It is thus important to know which social attraction rule applies for HRI. But this seems not an easy question to answer. Some experiments indicated that HRI should be based on similar social attraction rules (Reeves & Nass, 1996), while others claim the contrary (Lee et al., 2006). Windhouwer, however, thought both visions might be incorrect. He believed that it was the task context which determined whether people preferred a certain robotic personality over another, and not their own personality.

He conducted an experiment in which the role of introversion and extraversion in personality matching was examined. If we want humans to believe that robots have their own personality, they too should score either extravertly or introvertly on the Big Five personality test (Digman, 1990), just like humans would. By manipulating verbal and non-verbal cues, a human's perception of the robot's extraversion can be altered. For example, extraversion is linked to fast and loud speech in a higher tone of voice, while introversion is associated with the opposite features (Nass & Lee, 2001). The perception of extraversion is also altered by non-verbal cues: extroverts move wider and faster compared to introverts and they are more likely to come closer to other people (Isbister & Nass, 2000). In his experiment, Windhouwer used Aldebaran's program Choregraphe to create these behaviours for NAO.

Windhouwer's experiment consisted of showing 32 Dutch and German students a video in which NAO interacted with a human. The human conversational partner was the same in each video, to avoid unwanted side-effects on the results. The goal of this experiment was to determine if a human's preference for a certain type of robot might really depend on the task the robot needs to perform. Three research questions were chosen to examine this hypothesis: 1) Does a task strengthen the perceived robotic personality; 2) Does a robotic personality strengthen the perception of a task; 3) Does consistency between the personality of a robot and the characteristics of a task improve the human's perception of the robot?

⁶² The section *The Effects of the Task on the Perceived Personality of a Robot* is based on (Windhouwer, 2012), unless otherwise indicated.

The characteristics mentioned in the third research question are based on a study by Barrick & Mount in which they stated that people associate certain personality types to professions (Barrick & Mount, 1991). Based on this study, Windhouwer made videos of two NAO robots with opposite personality traits (one introverted, one extraverted) that performed three different professions: teacher (introversion), manager (extraversion) and apothecary (ambivalent). This resulted in six different videos. As a teacher, NAO gave a private lesson to a student. When playing the role of a CEO, NAO held a meeting with one of his managers. When behaving like a nurse in an apothecary, NAO gave advice and sold medicines to a customer. To avoid influence on the results, NAO was made to look human-size in these videos.

Participants were asked to fulfil a personality test which would indicate where they were situated on the introversion-extraversion scale. They were then asked to assess the introversion or extraversion of the NAO robot in one single video. Participants likewise had to do this for the task performed by the robot. Furthermore, NAO's intelligence was also examined. They also had to rate how fun and enjoying the robot seemed to them. The results can be found in Table 26.

Perception of both robots						
	Introverted NAO			Extraverted NAO		
	CEO	Pharmacist	Teacher	CEO	Pharmacist	Teacher
Introversion	6.2071	4.2940	5.4375	5.7000	6.1429	5.6250
Extraversion	5.4750	6.2679	5.9087	5.4750	5.0000	5.5250
Intelligence	3,0000	4,3000	4.8333	3.8000	3.9167	2.0000
Fun	5.9667	6.2500	6.4167	4.6000	4.5833	6.2500
Enjoyable	6.2667	6.4667	6.8333	4.5333	5.2778	5.4667

Table 26 Perception of Introverted and Extraverted NAO (Based on: Windhouwer, 2012)

The results shown in Table 26 might seem counterintuitive at first. These are numbers of a nine-point scale in which 1 equals 'Describes very well' and 9 equals 'Describes very poorly'. This means that if a characteristic scores very low, it is actually very present. Thus, the robot that was perceived as the most introverted one is the introverted NAO during the ambivalent pharmacy task. The robot that was perceived as the most extraverted one, on the other hand, was the extraverted Nao during the ambivalent task.

The answer to the first research question is that a task does not strengthen the personality perception of the robot if both are of the same type. The difference between the results of the extraverted NAO is negligible, however, the introverted NAO does show significant differences in perception when introversion was measured. For example, when the introverted NAO played a pharmacist, it was perceived as a lot more introverted than during the other tasks. This means that the task does have an effect on the perception of the introverted robot, as opposed to the extraverted robot. However, introverted NAO was not considered more introverted when performing the role of a teacher. We can thus conclude that a task does not strengthen the personality perception of a robot, nor weaken it.

The second question is likewise negative: consistency does not strengthen the perception of a task either. As the results show, the perception of a robot's introversion is not influenced by the tasks he performs (although the CEO tasks is slightly different from the two others). The perception of extraversion is only influenced during the CEO task (the task is perceived as more extraverted when performed by the introverted NAO). We can thus conclude that consistency does not strengthen the perception of a task.

The third question, again, should be answered negatively: consistency does not improve the perception of the robot's intelligence, fun or enjoyability. As shown in Table 26, the extraverted NAO was always considered to be more fun and more enjoyable than the introverted NAO, except as a teacher. The task does not seem to influence the perception of a robot's enjoyability at all. The task does not seem to influence the perception of the fun-side of the introverted NAO, however, it does influence the extraverted NAO. This robot was perceived as considerably less fun in the role of a teacher. However,

we can conclude that consistency does not have any influence on how fun or enjoyable a robot is perceived to be.

NAO's perceived intelligence, however, does seem to be influenced by the task. When the introverted NAO performed the role of the CEO, he was considered to be more intelligent than during his other roles. The extraverted NAO, on the other hand, was considered smartest as a teacher. This is completely the opposite than the expected result, as the CEO is the extraverted task and the teacher the introverted task. This could be explained by the fact that a calm and collected CEO would be seen as more intelligent than a loud and active one. A teacher, on the other hand, would be seen as more intelligent when acting energetically. Consistency does thus not seem to influence a human's perception of a robot's intelligence. We can therefore conclude that the task had indeed an influence on the perception of a robot's intelligence, but not on its "fun" or "enjoyable" features.

If we return once again to Windhouwer's hypothesis, we must thus conclude that the preference of humans for a certain type of robot is only determined by the context of the task on the level of intelligence. The task does not influence the perception of the robot as either "fun" or "enjoyable", however, introversion or extraversion on the other hand did influence these perceptions. So even though most research questions were answered negatively, the experiment does teach designers that if a robot needs to look intelligent, the tasks to be performed should be chosen carefully.

6.4 The Effects of Group Interactions on the Perceived Personality of a Robot

The experiment discussed in 6.3 examined the effect of the task on the perceived personality of a robot. This experiment was conducted in the context of the Student Conference on IT of the University of Twente. One of Windhouwer's fellow students, Leuwerink, conducted a parallel experiment for this conference in which he examined the effect of group interactions on the perceived personality of the robot (Leuwerink, 2012)⁶³. After all, studies on HRI in real-life environments have indicated that robots are more or just as likely to be confronted by a group than by an individual (Sabanovic et al., 2006). The chance of a robot being approached by a group depends of course on the particular setting in which the robot has to operate, but in general, it can be considered common for robots and groups of humans to interact with each other.

If we return to the hypothesis that the way robots are perceived depends on the personality of the human, we encounter a problem when dealing with group interactions. Social attraction rules (whether based on similarity or complementarity) cannot be simply applied to a group, because not everyone in the group has the same type of personality.

Like Windhouwer, Leuwerink focussed on the extraversion-introversion personality trait of the Big Five personality test (Digman, 1990). In his experiment, he examined the influence of this trait of the robot's personality on the way the group perceives him. He used studies on intra-group interactions between humans as a starting point: these are usually considered more positive if there are higher levels of extraversion in the group (Barrick et al., 1998). This led Leuwerink to the hypothesis that extraverted robots would be more suitable for group interactions. To determine whether a particular robotic personality is suited for a particular task, three factors need to be examined: likeability, usefulness and fun (Isbister & Nass, 2000).

The experiment proposed by Leuwerink consisted of 21 German and Dutch students watching one out of four possible videos. Each of these videos showed a NAO robot (either introverted or extraverted) performing the role of a teacher (either teaching an individual or a group of students). According to previous research, the profession of a teacher is usually associated with introversion (Barrick & Mount, 1991). Like in the experiment discussed in chapter 6.3, NAO's verbal and non-verbal characteristics

⁶³ The section *The Effects of Group Interactions on the Perceived Personality of a Robot* is based on (Leuwerink, 2012), unless otherwise indicated.

were adapted using Choregraphe to match an introverted or extroverted personality (Lee et al., 2006). Furthermore, NAO was once again transformed in these videos to be as tall as a human being.

Participants were asked to fill in an online questionnaire to determine their personality, after which they were shown a single video. They then had to answer questions about their perception of the extraversion/introversion of the task and of the robot.

Before continuing to the results of the experiment, it needs to be noted that the programming of NAO to match an introverted or extraverted personality was not entirely successful: both robots were perceived as extraverted. This needs to be kept in mind when examining the results, as they can be affected by this error.

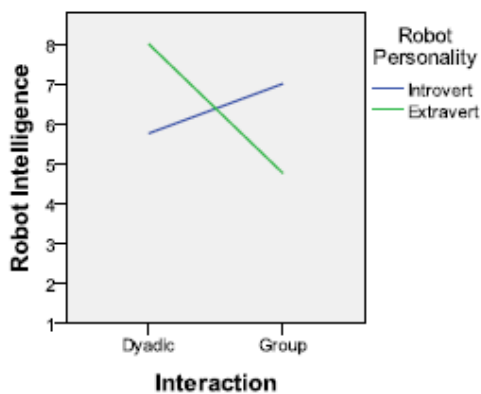


Figure 38 Perceived Intelligence of both Robots (Leuwerink, 2012)

Figure 38 shows the perceived intelligence of the introverted and extraverted robots. As can be seen, the extraverted robot is considered to be more intelligent when interacting with individuals than with groups. The introverted NAO, on the other hand, is considered more intelligent when dealing with groups than with individual humans. This result is contrary to the one expected by Leuwerink based on the studies with humans (Barrick et al., 1998). To explain these results, we must keep in mind the fact that the participants perceived both robots as extraverted. This means that these results are not accurate to form any conclusions on the perceived intelligence of introverted or extraverted robots. However, they can be used to draw conclusions about particular sets of verbal and non-verbal cues. Even though they failed to represent certain personality types, these cues can still be evaluated simply as cues. Designers should thus consider carefully which cues should be implemented in the context of group interactions or in the context of one-on-one interactions, as they influence the perceived intelligence of the robot.

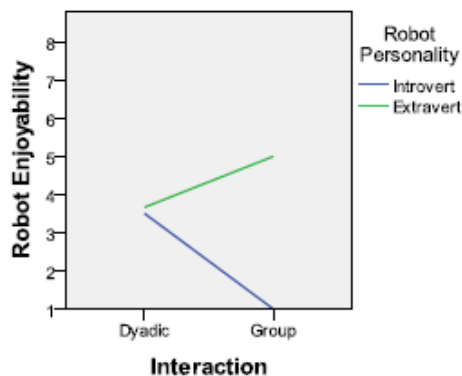


Figure 39 Perceived Enjoyability of both Robots (Leuwerink, 2012)

As shown in Figure 39, both robots are perceived as almost equally enjoyable in one-on-one interactions. However, in group interactions, the extraverted NAO is considered much more enjoyable than the introverted one. Similar results are found for the feature “fun”. Figure 40 shows that the introverted robot was perceived as slightly more fun than the extraverted robot when interacting with individuals, although there is not a lot of difference between the two. In the context of group interactions, however, the extraverted robot is considered to be more fun than the introverted robot. These results all confirm the older study on human-human interactions (Barrick et al., 1998).

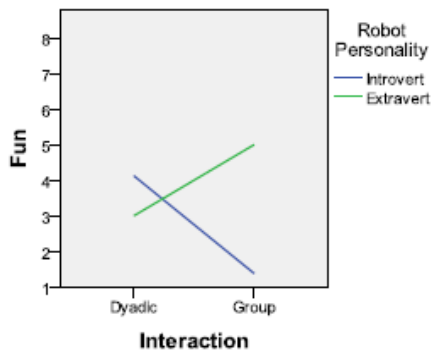


Figure 40 Perceived Fun of both Robots (Leuwerink, 2012)

In conclusion, this experiment has indicated that a set of introverted verbal and non-verbal cues are most suitable for interaction with a group of humans in a teaching environment when it comes to intelligence. This might create a difficult situation for robot designers when creating a robotic teacher. After all, teachers should be intelligent to be taken seriously, but they should also be fun and enjoyable to motivate their students. Therefore, Leuwerink suggests to make robotic teachers neither explicitly extraverted nor introverted in general. To suit particular purposes, designers could then slightly tip the scales into the required direction to create an expert teacher or a motivational teacher.

6.5 The Effects of Neighbouring Cultures on the Perceived Personality of a Robot

Yet another student participated in the joint experiment of Windhouwer and Leuwerink (see chapters 6.3 and 6.4) for the 2012 Twente Student Conference on IT. Waalewijn examined the effect of neighbouring cultures on the perception of a robot (Waalewijn, 2012)⁶⁴. In previous studies it had been shown that there were significant differences between Western and Eastern perceptions of robots (Li et al., 2010). Waalewijn wanted to determine whether or not the cultural background of a human could influence which type of robot he or she preferred. More specifically, he examined the effect of neighbouring cultures on the perception of a robot’s personality. Based on earlier psychological research, five characteristics were used to classify countries (Hofstede, 2001)⁶⁵. The results for the two countries examined in this experiment can be found in Table 27.

Cultural dimensions of Germany and the Netherlands		
	Germany	The Netherlands
Power-distance	35	38
Individualism	67	80
Masculinity	66	14
Uncertainty avoidance	65	53
Long-term orientation	31	44

Table 27 Cultural Dimension Ratings of Germany and the Netherlands (Based on: Hofstede, 2001)

⁶⁴ The section *The Effects of Neighbouring Cultures on the Perceived Personality of a Robot* is based on (Waalewijn, 2012), unless otherwise indicated.

⁶⁵ In the meanwhile, a sixth dimension was added by Hofstede in 2010: Indulgence (Hofstede Centre, 2015). Germany scores 40 on Indulgence, while the Netherlands score 68 (Hofstede Centre, 2015).

These dimension ratings (ranging from 0 to 100) indicate that Germany and the Netherlands are less alike than one would assume at first sight. They score significantly different on Individualism and Masculinity. Furthermore, studies on the association between personality traits and cultural dimensions show that extraversion is linked to individualism and that introversion is linked to masculinity (Hofstede & McCrae, 2004). This would mean that, in general, people from Germany are more introverted than people from the Netherlands.

The experiment by Waalewijn consisted of 13 German students and 28 Dutch students who were shown a video in which a NAO robot (either introverted or extraverted) performed one out of three possible tasks: teaching, managing or nursing. These professions are associated with respectively introversion, extraversion or neither (Barrick & Mount, 1991). Participants were first asked to complete a questionnaire to determine their personality. Then, they watched a video for which they had to evaluate the task and the robot. Furthermore, they had to indicate how much time they would be willing to spend with that particular robot on a weekly basis. As in the previous two experiments (chapters 6.3 and 6.4), Choregraphe was used to manipulate the verbal and non-verbal behaviours of the robot to suit the desired introverted or extraverted personality (Isbister & Nass, 2000).

Like in the experiment conducted by Leuwerink, the matching of verbal and non-verbal behaviours to certain personality types failed. Both robots were perceived as either introverted in almost 50% of the cases.

First of all, Waalewijn examined if Germans were indeed more introverted than people from the Netherlands. His results indicated that Germans, in general, score higher on the extraverted scale, which is exactly the opposite of the studied hypothesis. However, he did not consider these results significant because there were not enough German participants to generalise his findings. Furthermore, these German participants all lived abroad and therefore, they might not be the best representatives of their nation.

When considering the amount of time people want to spend with a robot, it was found that, in general, the Dutch are less willing to spend time with robots than Germans. This is shown in Table 28⁶⁶. Yet, there does not seem to be any link between cultural background and robot preference.

Average time wanted to spend with the robot				
	Germans		Dutch	
	Introversion	Extraversion	Introversion	Extraversion
Robot programmed personality				
Introversion	X	0.75	2.20	1.60
Extraversion	6.67	6.30	3.00	4.50
Robot perceived personality				
Introversion	7.50	1.16	4.00	2.17
Extraversion	5.00	8.50	2.14	3.70

Table 28 Average Time Wanted to Spend with the Robot (Waalewijn, 2012)

The results did thus not prove anything about cultural differences, and therefore, Waalewijn decided to consider the results without taking culture into account. This way, he discovered that, in general, people like to spend more time with robots if they match their own personality. This is shown in Figure 41 and Figure 42. This would mean that the social attraction rule based on similarity is more important to HRI

⁶⁶ The X in this table is not explained by Waalewijn. It might be linked to the fact that most Germans were considered extraverted, but this is not sure, as *most* does not equal *all*.

than the one based on complementarity. Designers should thus consider that the longer a person needs to spend time with a robot, the more their personalities should be matched.

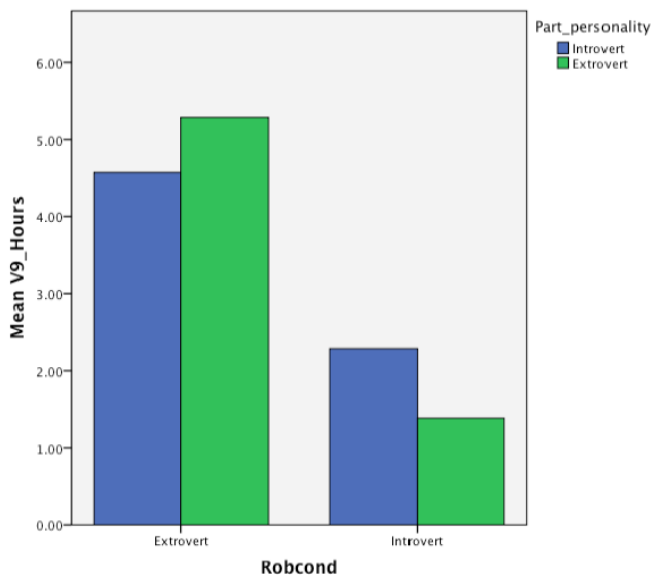


Figure 41 Average Time Wanted to Spend with Programmed Personality (Waalewijn, 2012)

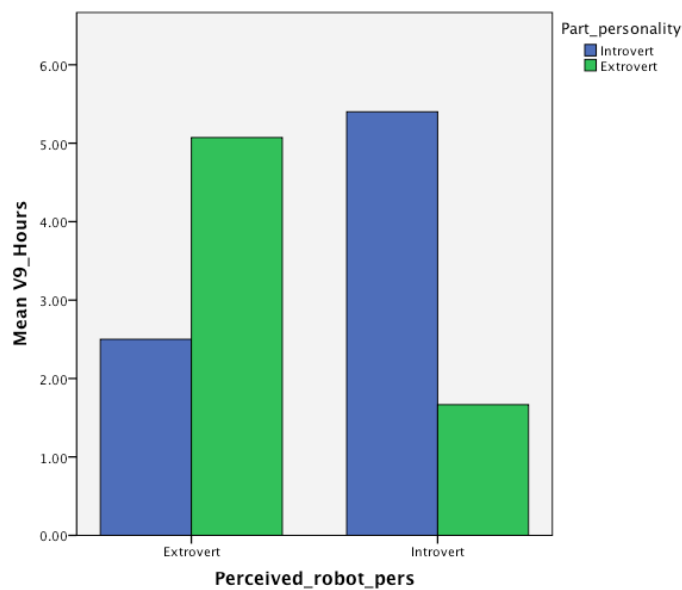


Figure 42 Average Time Wanted to Spend with Perceived Personality (Waalewijn, 2012)

On the level of perception, however, the results did indicate cultural differences. People from the Netherlands rated the perceived introverted NAO as more intelligent than the perceived extraverted NAO. Germans, on the other hand, thought the perceived extraverted NAO was the most intelligent. This could be linked to the cultural dimensions found in Table 27: Germans score higher on the dimension of masculinity which could explain why they prefer an extraverted robot. The Dutch, on the other hand, are considered more feminine and thus prefer an introverted robot.⁶⁷

⁶⁷ This conclusion seems to be in conflict with the earlier statement that masculinity was connected to introversion.

6.6 Conclusion

Robotic personalities remains a field which requires much future research, as there are many opposing views and none of them seem to get the upper hand in the debate. However, one fact is questioned by no one: robotic personalities are vital to create natural and comfortable HRI. Therefore, more research needs to be done into this still cloudy field of NLP.

In section 6.1, we have taken a brief look at the five personality types which are generally assumed to be the most important types of robotic personalities. Next, in section 6.2, we presented one of the two main views on robotic personalities, namely the necessity of matching a robot's personality to the user's. As mentioned earlier, opinions in this field vary widely. Some believe that robots should match the personality of the human (such as Aly & Tapus, whose research can be found in section 6.2), while others believe that HRI would benefit from robotic personalities that complement the personality of the human. These views are based on the two social attraction rules that psychology has determined for inter-human relationships: matching and complementing. However, there are other views as well, which seem to contradict both opinions described above. Therefore, we have included three papers by Dutch students that presented different perspectives. The first, Windhouwer, examined whether the context of the task that was to be performed by the robot influenced a human's preference for certain personality types (Windhouwer, 2012). This did not turn out to be the case, although the task context did have an influence on the perceived intelligence of the robot (Windhouwer, 2012). Participants' perception of NAO as more enjoyable or less enjoyable did not depend on the task context, however, even though this was expected by Windhouwer. They attributed a certain enjoyability to NAO based on his extraversion or introversion instead (Windhouwer, 2012).

The second student, Leuwerink, examined whether humans would prefer another type of robotic personality when engaging in one-on-one interactions or when participating in group interactions (Leuwerink, 2012). In order to do this, Leuwerink conducted an experiment in which NAO performed the role of a teacher (Leuwerink, 2012). It was found that the group of students perceived NAO as more intelligent when he acted introvertly and as more enjoyable when he acted extravertly (Leuwerink, 2012). This might lead to important design issues, as teacher should be both if they want to transfer knowledge efficiently and capture the attention of the group (Leuwerink, 2012). Leuwerink therefore suggests that teaching robots should neither be expressively extraverted or introverted generally, although these traits could be added to suit particular situations (Leuwerink, 2012).

Waalewijn, the third and final student, studied the influence of neighbouring cultures on the perception and preference of a robot's personality (Waalewijn, 2012). His experiment failed in finding an explicit link between a person's cultural background and his or her preference of a certain robotic personality type (Waalewijn, 2012). Neighbouring cultures only seemed to influence the perception people had of each robot (Waalewijn, 2012). He did find, however, that in general, people like to spend more time with robots that match their own personality type (Waalewijn, 2012), which seems to confirm the experiment discussed in section 6.5. However, many things went wrong during this research and therefore its results should not be considered conclusive evidence.

7. Case Studies

7.0 Introduction

In the previous chapters, we discussed different aspects of NLP with NAO robots. We examined HRI through natural language, the communication of emotions and the influence of personality traits. Now, we will take a closer look at some case studies in which NLP and the various aspects discussed earlier play an important role. In all of these cases, NAO is used in the context of healthcare. This kind of interaction is complex and it needs to be proven reliable and safe before it can be considered as part of an approved therapy (Shamsuddin et al., 2012a). Therefore, experiments are needed to confirm the usability of robots in such contexts.

First, we will examine the influence of a robot's embodiment on its successfulness in a medical context. Different embodiments are suited to fulfil different tasks. Cute, pet-like robots perform well as companions as they encourage their users to behave affectionately towards them. Humanoids, however, are more suited as coaches because of their authorial and motivational capacities. In order to investigate this, an experiment in which children had to choose between different robots such as the humanoid NAO and the toy-like dinosaur PLEO was conducted. These children were asked to explain why they preferred a particular robot. Furthermore, they needed to describe what they expected of the robot and if their expectations were met after having played with the robot of their choice. They were also asked to think about possible improvements, which should be considered important if designers truly want to develop suitable companions for young children.

Second, we will focus on NAO's possibilities in autism therapy. Many studies have shown positive results when NAO is introduced in this context and therefore, multiple initiatives have been created. We will first discuss ASK NAO, Aldebaran's own initiative. Then, we will take a closer look at some of the studies conducted in Malaysia. Despite all the benefits and positive results of multiple studies, it should be noted that not all autistic children react to robots in the same way (Tapus et al., 2012). Some children show significant progress when participating in robot assisted therapies, but others advance a lot slower and some even show negative reactions to the humanoid (Tapus et al., 2012). It remains thus important to investigate further the possible benefits and drawbacks of using humanoids in therapies for autistic children. This connects to the issue of personalisation. As each child is unique, each therapy should be unique as well. To meet this requirement, we will discuss a customizable, easy-to-use platform in the third part of the section on autism. In the fourth part, we will examine the use of a platform as described in part three in Pivotal Response Trainings, an important methodology used in autism therapies.

Thirdly, we will discuss NAO's performances in the context of diabetes therapies, more specifically in the context of the European ALIZ-E Project that ran from 2010 till 2014. First, we will list all the requirements needed for a robotic companion for diabetic children. Then, we will take a look at the two NAO companions that were developed by the ALIZ-E project team, ROBIN and NAO. These two robotic brothers fulfil complementary roles, as ROBIN represents an infant-companion and NAO a skilful monitor. These two robots are intended to interact with hospitalized children and thus need to be able to establish meaningful long-term relationships with these children. Therefore, we will examine the way in which these children adapt their interactions style to the robots in the course of time.

In the final section of this chapter, we will discuss NAO as a teaching assistant for sign language. As sign language is very dependent on the individual style of the teacher, this often causes problems when the teacher needs to be replaced. Therefore, sign language therapies would benefit from the introduction of a robotic assistant, as this would make the lessons more tutor-independent. NAO's physical limitations, however, likewise cause problems for the teaching of sign language. Therefore, his performances are compared to a robot with greater degrees of freedom, namely Robovie R3 robot.

7.1 Influence of Embodiment

When children are hospitalised for an extended period of time, they are inevitably confronted with serious stress, fear, and usually physical pain (Diaz et al., 2011)⁶⁸. Robots can be used to make this period in their lives easier, by supporting them and taking on the role of a faithful and friendly companion. This means robots can be used to fulfil at least two different roles in a healthcare context: rehabilitation monitor and companion. As a rehabilitation monitor, robots can be used in a mainly goal-oriented way. They can assist the children in their exercises to get well again. As a companion, however, they are used in a need-oriented way. They provide stress-relief, friendship and entertainment to brighten the children's time in the hospital. Both roles require the maintenance of long-term relationships. It is known that children are often attracted to robots, but if this effect is mainly created by the novelty of being allowed to play with a mechanical friend, it might be problematic to create long-lasting relationships between children and robots. Therefore, it is important to choose the right robot for the right tasks. Each robot has its own features (e.g. some look like pets, others like humans) and interaction capabilities (e.g. some can talk, others only use body language). It is thus vital that well-considered decisions are made to ensure that each task is accomplished by a well-suited robot.

In order to examine this, Diaz et al. have conducted two experiments with non-hospitalized children. The first experiment took place at a school as part of a science class for 49 11 to 12-year-olds. The second one took place at the laboratory: four children of the previous group were selected to participate in this. The research team assumed that a robot should have specific skills to match particular situations to be effective as a companion. Furthermore, they predicted that the embodiment of the robot would significantly influence the children's perception of it.

During the first experiment, it was examined which factors influence the emergence of a bond between the children and the robot. Children were presented with four switched-off robots among which they could choose the one they preferred to play with during the experiment. These robots can be seen in Figure 43.



Figure 43 Robots used in the Experiment. A. NAO; B. AIBO; C. PLEO; D. Spykee (Diaz et al., 2011)

Diaz et al. chose to publish only the results of the experiment with NAO and PLEO⁶⁹ as these two are the models which are most often employed as companions and rehabilitation monitors. NAO, as a humanoid, is often chosen to fulfil the role of rehabilitation monitor. His anthropomorphic embodiment allows him a certain authority over the child and accords him a certain degree of expertise in the matter. This research team also chose NAO because of the fact that he has been used successfully in the past in medical experiments due to his interaction and motor skills. Examples of this include the ALIZ-E project, which is discussed in chapter 7.3.

⁶⁸ The section *Influence of Embodiment* is based on (Diaz et al., 2011), unless otherwise indicated.

⁶⁹ PLEO is a commercially available dinosaur robot developed by UGOBE. It has multiple tactile sensors, speakers and microphones. PLEO presents different creature-like behaviours and moods. More information can be found in (Innvo Labs, 2015).

PLEO, on the other hand, has been chosen to fulfil the role of the companion. It is known that children greatly benefit from therapies with pets, as they provide entertainment, support, happiness and relief. However, it is not always possible to allow animals in medical environments. Therefore, experiments with robotic pets have been conducted to evaluate whether or not they could perform the same functions as real animal companions. While a humanoid has a certain authority over the children, robotic pets enter master-pet relationships with their companions. Because the children feel responsible for their pets, they will train important social skills such as showing affection and taking care of other creatures. PLEO was chosen to be the companion in this experiment because it meets all the embodiment requirements: it is expressive, baby-like and can perform different behaviours. Multiple studies have shown that PLEO can form long-term relationships with humans.

Once the children had chosen their preferred robot, the robots were switched on and carried out non-interactive behaviour. NAO started out by introducing himself in a loud voice, accompanied by arm and head motions and changing eye colours. He then danced to a song he played, after which he demonstrated a Tai Chi routine. PLEO woke up after being petted by the children. They could then make him go back to sleep by rubbing its back. He would get angry when the children lifted him by the tail. He demonstrated his head movements when being hugged and his leg movements while being walked across a desk.

Out of the 49 children, 33 chose to play with NAO or PLEO. PLEO turned out to be the most popular robot (18 children chose to play with him, all of which – surprisingly – were girls). NAO was the second most popular robot as 15 children (4 girls and 11 boys) chose him. Their perception and expectations can be found in Table 29. These indicate which features of a robot’s appearance are important to make children willing to engage with them. This might lead to conclusions about the possible roles each of these robots could play in the healthcare sector.

Perception and expectations		
	NAO	PLEO
Reasons for preference	Seems/is a person Seems an ape Seems more articulated Is the biggest robot	Nice aspect <i>So cute!</i> Animal likeness Baby likeness
Expectations (before the performance)	To walk To grasp things To speak To move hands To dance To do <i>Matrix</i> To follow instructions To sing	Love and affect responsiveness Baby-like behaviour Emotional expressiveness To make sounds
Liked most after self-presentation	Tai Chi routines Dancing	Seems a baby How it moves
Interactive behaviour	Spontaneous imitation Admiration Spontaneous applause Amazement / <i>Wow!</i> Curiosity about technical issues - <i>What’s X for?</i> - <i>Is that an USB plug?</i> Exploring physical, cognitive and social capabilities and constraints - <i>Is he hearing me now?</i> - <i>Does he see me?</i>	Baby talk Affection giving Taking care activities

Wish it could do / have / be	Hold a conversation Capability to communication in natural language Non-verbal communication skills: - Gaze - Intonation - <i>When looking at people, it should look in the face</i> Talk about itself - <i>Say what he is thinking</i> Improve motor competences - <i>Play football or hockey</i> - <i>Perform moonwalk</i> Assist - <i>Help with my homework</i>	More life-likeness - <i>Talk</i> - <i>Eat</i> - <i>Grow up</i> Responsiveness - <i>Not so sleepy</i>
-------------------------------------	--	---

Table 29 Perceptions and Expectations of the Children Concerning NAO and PLEO (Diaz et al., 2011)

After this first meeting with the robots, four children who chose PLEO were selected to have a second meeting with him at the laboratory. Two PLEO robots with different behaviours were presented to each child in this experiment: one purring while moving about slowly and one growling and agitated.

Children were asked to select a robot, after which they were allowed to play with it, both with and without the supervision of an adult. Afterwards, they were encouraged to discuss the robot with one of the other children. All of them chose the calm robot over the agitated one and they wanted to know which of the four robots was the one they had met in school. The interactive behaviour displayed by the four girls can be seen in Table 30.

Interactive behaviour	
Selecting a PLEO	All the participants chose the “nice” one, picked it up and took it in their arms
In the lab with the facilitator	Petting Hugging Feeding
Alone with PLEO in the lab	New activities appeared - Putting it into the doghouse - Grabbing it by the tail - Insisting on feeding
With a classmate and PLEO in the lab	The presence of a peer helped the girls to express their feelings and reinforce their role of owner.
	Similar to real pets Owner feelings Differences & similarities with the one in the school - More active - More fun

Table 30 Interactive Behaviour of the Children during Experiment II (Diaz et al., 2011)

When a classmate was called into the laboratory, each child naturally adopted the role of owner. They explained PLEO’s behaviour to the other girl as if PLEO was truly their own pet. For example, they told the other child that ‘*it was difficult for him to fall asleep*’ or that ‘*he was not hungry at the moment*’.

The results of these experiments show that children indeed behave differently towards these two different robots. Their salient features are completely different, which leads to other relationships being established. NAO caused the children to imitate him spontaneously and feel admiration and amazement towards him. It is therefore a logical choice to assign him the role of rehabilitation monitor, as children

are encouraged to be curious and to explore. PLEO, on the other hand, makes children behave in an affective and caring way. Its animal-like characteristics encourage children to take care of him as if they had a real pet. Furthermore, during the second experiment, children asked for “their baby dinosaur”, which indicates that it is indeed possible to form long-term relationships with robotic companions. PLEO can thus be considered to be a good option as a companion for hospitalized children. Judging from the comments made by the children, PLEO could be improved by showing more life-like behaviours, while NAO should act more like a human being.

7.2 NAO in Autism Therapy

Autism Spectrum Disorder (ASD) is a lifelong neurodevelopmental disability without a cure characterised by impaired social skills and repetitive behaviour (Shamsuddin et al., 2012a)⁷⁰. Appropriate therapy is thus vital for young children diagnosed with autism. Recent research indicates that humanoid robots could be an important asset to ASD therapies, as HRI is usually perceived as less demanding for autistic children. Furthermore, autistic children tend to be fascinated by robots, which greatly improves the success of such therapies as their interest is automatically triggered.

Although all these advantages are known, these kinds of programmes are not yet widely used as ASD therapies. This is due to the rising costs and to the fact that there are not enough skilled specialists to work with humanoids. Yet, it cannot be denied that humanoids show great potential to teach autistic children how to deal with emotions and social behaviours. Moreover, robots could be used at home as well, to help parents communicate with their children.

NAO seems an ideal platform to provide all of these advantages and reduce the problems concerning the use of robots in ASD therapies. After all, multiple experiments have indicated that NAO is both effective in this domain and cost-efficient (as it is a relatively affordable humanoid). Because of this, NAO is now a standard in robotic ASD research and Aldebaran has even created its own programme for ASD therapy, called ASK NAO.

7.2.1 ASK NAO

In 2013, Autism Solution for Kids (ASK NAO) was launched by Aldebaran. The purpose of this initiative was to ‘offer a new teaching approach to teachers and [to] children with autism thanks to robots’ (Aldebaran Robotics, 2015)⁷¹. This programme was created after research proved that autistic children were indeed highly attracted to technology (ASK NAO Information Kit, 2015)⁷². NAO was thus customized with several behavioural applications to support autistic children in developing social and learning skills.

Aldebaran believes that NAO is a perfect humanoid for HRI with autistic children because of the fact that NAO’s behaviour is very predictable. This reduces the natural anxiety which is typical of autistic children. Furthermore, NAO has tireless features and a judgment-free demeanour, which is also beneficial to this kind of robot supported ASD therapies. Moreover, the clean design of the robot allows these children to interact with NAO in a comfortable way. After all, such a design reduces the amount of sensory data that needs to be processed by the children.

Because each child is unique, their education assisted by NAO needs to be unique as well. Therefore, sessions with NAO are personalised to meet the needs of individual children accurately. Many applications have been developed to improve the social and learning skills through customizable programmes. These applications are based on encouragement and rewards, which will motivate the children and improve their confidence and independence. To make learning with NAO fun, all of these are interactive games. While enjoying a game with a non-threatening artificial friend, these children

⁷⁰ The information on ASD in this section is based on (Shamsuddin et al., 2012a), unless otherwise indicated.

⁷¹ ASK NAO now has its own website, please see (ASK NAO, 2015) for more information.

⁷² The section ASK NAO is based on (ASK NAO Information Kit, 2015), unless otherwise indicated.

learn valuable communicative skills such as turn taking, emotion detection and instruction following. All of these games are personalised to the individual child, taking his or her own motivators, internal states, personality and learning goals in consideration.

When schools or hospitals enter the ASK NAO programme, they also get access to a web portal through which caretakers can easily interact with parents. A personalised programme can be created with this portal through means of custom profiles that manage individual needs and progress.

7.2.2 NAO and the National Autism Society of Malaysia

Next to Aldebaran’s ASK NAO initiative, other ASD therapies supported by NAO have been developed all over the world. An example of this is the research conducted by the University of Technology Malaysia in collaboration with NASOM, the National Autism Society of Malaysia. Some of the reasons why this initiative has chosen to work with the NAO platform are affordability, availability and ease of programming (Shamsuddin et al., 2012b). They believe that the most important role for NAO is to create a stable environment (to avoid frightening autistic children) and to attract the attention of the participating children (Miskam et al., 2014)⁷³. Robots can help these children to develop social skills and to improve the communication with other children. As children with ASD are known to lack imitation skills, NAO can also be used to improve their imitation abilities. Furthermore, autistic children lack the ability to develop joint attention skills in the same way as other children (Shamsuddin et al., 2012b). This skill, likewise, can be improved through interaction with robots such as NAO.

Miskam et al. have designed an experiment in which NAO assisted autistic children in learning numbers and emotional gestures. The presence of a robotic tutor instead of a human one significantly reduces the stress experienced by these children. Furthermore, robots are ideal as they can repeat gestures endlessly, without becoming tired or losing their patience. This is important because children with ASD usually like repetition.

In this experiment, NAO will play three different games with mildly autistic children between 7 and 12 years old.⁷⁴ During each of these games, eye contact between the child and the robot will be monitored. At the beginning of each game, NAO will greet the child and ask some basic questions to establish a bond of familiarity with the child. This is important to create the safe and comfortable environment that is absolutely vital for ASD therapies. Please see Table 31 for an overview of the objectives of each component of the experiment and the associated target skills.

Game objectives and target skills				
	Methods	Objectives	Social Skills	Communicative Skills
Introduction	Two-way communication with voice recognition	To create a friendly environment for the child	- Eye contact - Emotions	- Word repetition - Speech tone - Answering
Game I	Object & vision recognition	To observe the child in a writing numbers activity.	- Behaviour - Emotions	- Answering

⁷³ The section *NAO and the National Autism Society of Malaysia* is based on (Miskam et al., 2014), unless otherwise indicated.

⁷⁴ Although the set-up of this experiment has already been presented to the public, it has not yet been conducted as the research team is still waiting for ethics approval from the government. It is still included in this bachelor’s thesis though, as its goals and research methods are considered important and interesting for the discussion of robotic assistants in ASD therapies.

Game II	Two-way communication with voice recognition	To observe how the child answers questions	<ul style="list-style-type: none"> - Eye contact - Emotions - Behaviours 	<ul style="list-style-type: none"> - Word repetition - Speech tone - Answering
Game III	Imitation	To observe the imitation abilities	<ul style="list-style-type: none"> - Eye contact - Behaviours - Imitation 	/

Table 31 Game Objectives and Target Skills (Miskam et al., 2014)

During the first game, children will have to draw numbers (1 to 10) that are dictated by NAO. They will then present the paper to the robot for evaluation.

In the second game, NAO will present the child with five simple mathematical problems. The robot will perform different movements associated with correct and wrong answers.

During the final game, NAO will display six different emotions or emotional behaviours (hunger, happiness, anger, fear, hug and kiss) which the children will need to imitate. These displays of emotional body language can be seen in Figure 44.

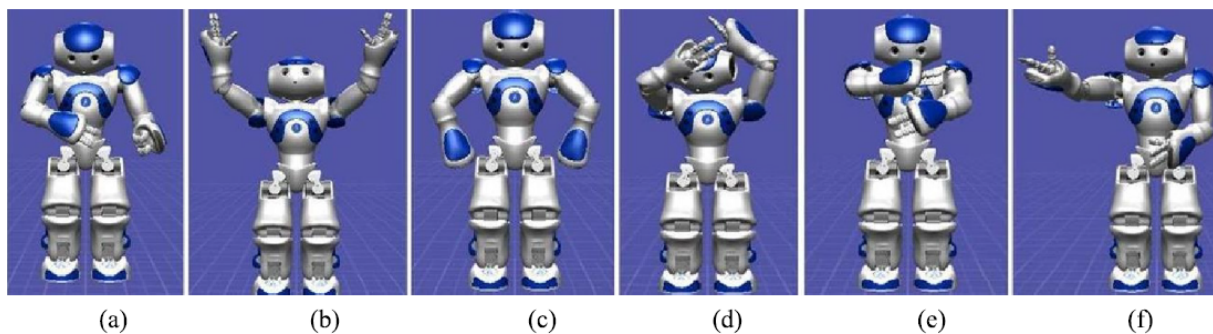


Figure 44 NAO during Game III: a. Hunger; b. Happiness; c. Anger; d. Fear; e. Hug; f. Kiss (Miskam et al., 2014)

These games will be used to examine the benefits humanoid robots such as NAO in ASD therapy sessions.

Earlier research within the context of the NASOM has tried to establish a connection between the reaction of an autistic child towards NAO and his or her IQ (Shamsuddin et al., 2012b)⁷⁵. In order to do this, an experiment consisting of seven modules has been conducted, testing the reactions of five young children with IQs ranging between 40 and 54. The description of each of these modules can be found in Table 32.

70% of children with ASD have an IQ of less than 70 (Charman et al., 2011), yet other autistic children have extremely high IQs. Therefore, it might be interesting to examine whether or not children's IQs influence their perception of NAO and the progress they make through robot assisted ASD therapy. In their experiment, Shamsuddin et al. studied the influence of a moderately impaired or delayed IQ. Furthermore, the capacity of NAO to reduce a child's autistic behaviour was likewise examined.

The experiment consisted of two parts. Each child was observed during the interaction with the robot and during the normal daily classroom routine. This allowed a comparison of the children's autistic behaviours in both situations. The research team expected that the children's behaviour would be less autistic during their interaction with NAO. The results can be seen in Figure 45 and Figure 46.

⁷⁵ The information on the connection between IQ and perception of NAO in this section is based on (Shamsuddin et al., 2012b), unless otherwise indicated.

Modules		
		Description
1	Static Interaction	No music, speech or movement. NAO sits on the table, facing the child.
2	Head turning	NAO turns head to left, right and back to the child.
3	Eye blinking	NAO's eyes blink continuously and randomly red, green and blue.
4	Talking	NAO greets the child (until the child answers). NAO asks some simple questions (one repetition for each question if the child does not answer) <ul style="list-style-type: none"> - 'How are you?' - 'What is your name?' - 'Where do you live?' - 'How old are you?'
5	Song playing	NAO plays the nursery rhyme <i>Twinkle, Twinkle, Little Star</i> . If the child does not respond, the song is repeated. Then, the song <i>Humpty Dumpty</i> is played. If the child does not respond, the song is once again repeated.
6	Hand moving	NAO waves to the child with his right hand, twice. NAO waves to the child with his left hand, twice.
7	Song playing & hand moving	NAO repeats module 6 while playing the song <i>ABC</i> . NAO repeats module 6 while playing the song <i>Itsy Bitsy Spider</i> .

Table 32 Modules of the Experiment (Based on: Shamsuddin et al., 2012a)

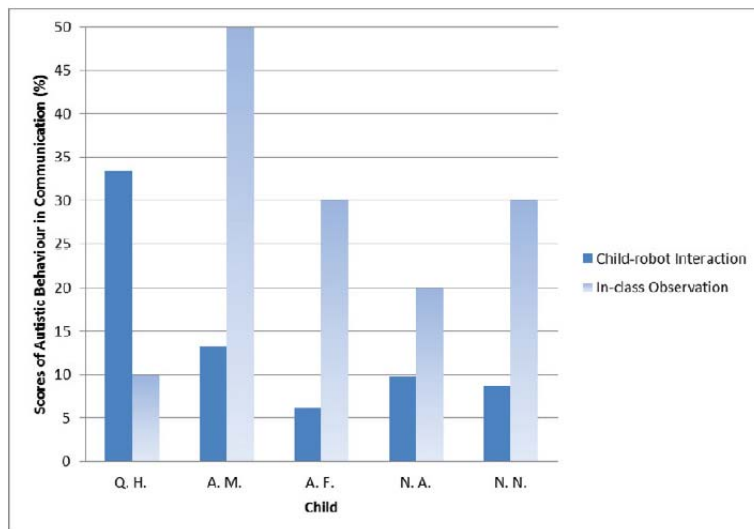


Figure 45 Comparison of Autistic Behaviour in the Classroom and during HRI (Shamsuddin et al., 2012b)

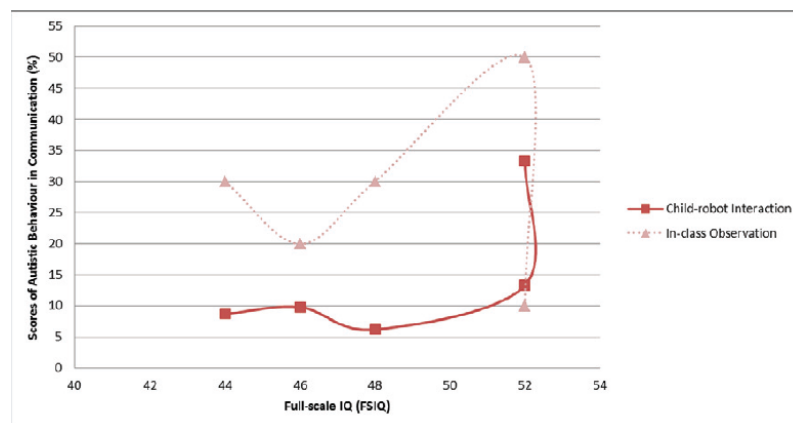


Figure 46 Connection between Autistic Behaviour and IQ (Shamsuddin et al., 2012b)

Figure 45 shows that four out of five children indeed displayed less autistic behaviour when interacting with NAO. One child, however, displayed more autistic traits. This might be due to personal differences and the research team suspects that repetitive interaction with NAO might change the child's perception of the robot and thus lessen the autistic behaviour. These results indicate that, in general, humanoids can indeed be used to successfully capture the attention and interest of an autistic child.

Figure 46 shows that children with a moderately impaired IQ are indeed receptive to humanoid robots. This is once again proven by the fact that four out of five children behaved less autistic during their interaction with NAO, which proves that robot assisted ASD therapy would be beneficent to these children.

It is important to keep in mind, however, that this study was only conducted with five participants. As this number is not large enough to prove their hypothesis, the research team plans to conduct larger experiments on the same topic in the future.

7.2.3 A Customizable Platform for Robot Assisted ASD Therapy

As most studies mentioned above point out, unique children need unique therapies. Therefore, platforms have been developed to support personalised ASD therapies with robotic assistants. One of these platforms is developed in the Netherlands by Gillesen et al. and is specifically designed to provide an intuitive interface (Gillesen et al., 2011)⁷⁶.

Gillesen et al. realised the need for personalisation when observing the heterogeneity of the autism spectrum. This heterogeneity is so great that therapy needs for different individuals might even be contradictory. Therefore, they designed an environment for training that is easily adaptable and reusable. Furthermore, they recognised the issue of non-technical people needing to work with robots. Their platform thus had to be user-friendly. To make this possible, they chose to work with a visual programming environment in which scenarios could be created using building blocks. Furthermore, medical experts usually do not have the time to learn new training methods. Therefore, robots such as NAO should be included into established training methods if they ever want to become the standard in ASD therapies.

In their experiments, Gillesen et al. chose to work with a NAO robot, because of its affordability and availability. Yet, as they did not want to be bound to a particular robotic platform forever, they made sure their platform would be compatible with other robots as well.

The visual programming environment chosen for these experiments is TiViPE, which is a box-wire based model. Behavioural components are represented by boxes in this model and can be connected to create a network. Such networks can then be transformed into components themselves, allowing for very complex and intelligent robotic behaviour. The process flow in Figure 47 shows the different steps that are needed to transform a scenario into a TiViPE network.

There are two disciplines involved in the creation of TiViPE networks: both trainers (therapists) and clients (autistic children) need to work together with robotic experts. Scenario builders are both technical experts as therapists in the first stages of this experiment. In the future, this role should be taken up solely by trainers.

Furthermore, two paths are indicated on the process flow: one top-down which splits a scenario in different components and one bottom-up which combines these components into networks that represent scenarios.

⁷⁶ The section *A Customizable Platform for Robot Assisted ASD Therapy* is based on (Gillesen et al., 2011), unless otherwise indicated.

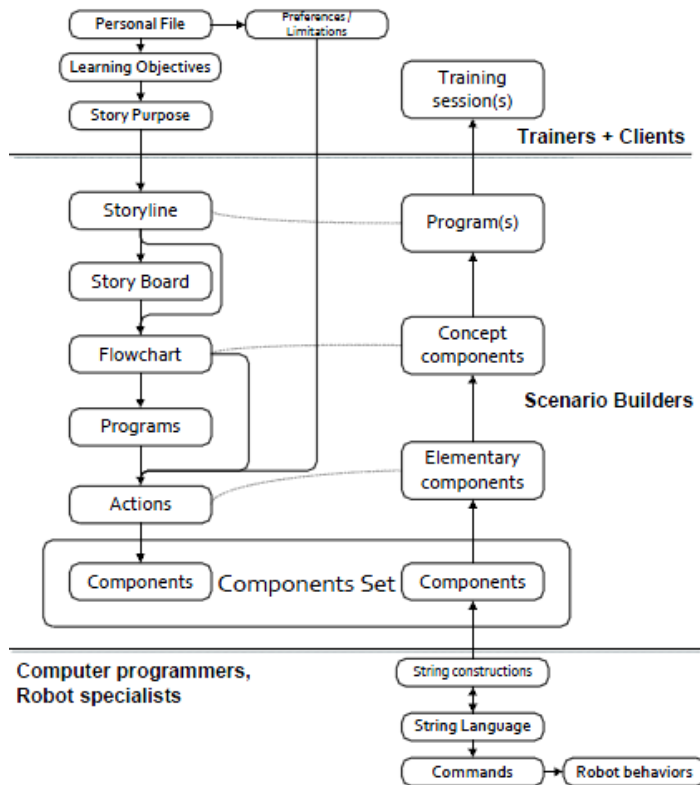


Figure 47 Process Flow to Transform Scenarios into TiViPE Networks (Gillesen et al., 2011)

The process flow starts with personalised learning objectives found in the medical files of a particular child. These are created by therapists, based on the individual needs of the child. These objectives determine which story purposes should be developed, which are embedded on their turn in one or more storylines. These storylines are then made visible through story boards: visual representations of the story. As the second arrow indicates, this step is optional. Sometimes, in case the story is clear, flowcharts can be created immediately based on the selected storylines.

A flowchart shows the boxes of the storyline and the network in which the boxes are structured. Each of these boxes correspond to certain actions and feelings of the robot. All the actions that are associated to a single box are called “programs”. These programs are then implemented as re-usable components. As mentioned before, personalisation is an important factor. Therefore, each of these components need parameters that can be adjusted to fit the needs of each child individually.

As an example, an introduction scenario can be discussed. This is a very important scenario, as the child’s first impression of NAO will often be a lasting one and can significantly influence the success or failure of the therapy. In this scenario, NAO will greet the child and ask for his or her name. This leads to the child asking the name of the robot, after which NAO will suddenly become sad. The robot will then explain to the child that he does not yet have a name and ask if the child can perhaps think of one.

During this scenario, a component called “TalkNod” is used. This is a complex network of three other components. The first one makes NAO talk using the text-to-speech engine. The second one makes NAO’s eye light up to accompany his speech activity. The third component enables NAO to show non-verbal behaviours such as, for example during the greeting, nodding his head. The parameters of the component can be used to make NAO say ‘*What is your name?*’ during the first step of the introduction scenario or to adjust the duration of the light effects. When NAO says to the child: ‘*I am sad because I don’t have a name*’, the non-verbal behaviour can be changed to looking away.

This platform meets the requirement of an easy-to-use and flexible environment that allows therapists to develop scenarios based on the needs of individual children. Before it can be used in clinics, however, it needs to become more stable as to ensure that the children feel safe when interacting with NAO and that the trainers can use it in exactly the right way.

7.2.4 Robot Assisted Pivotal Response Training

In the previous section, a flexible platform was discussed that would meet the needs of individual children and that would be easy-to-use for non-technical people. Such a system can also be used to design Pivotal Response Training (PRT) scenarios in which robots play an assisting role (Kim et al., 2014)⁷⁷. PRT is a behaviour intervention methodology used in ASD therapies which aims to teach children “pivotal” behaviour skills. These skills include motivation, empathy and interaction with peers, for example. PRT is called a “naturalistic approach”, which means that it is important that children are allowed to learn these skills in a natural way. This can be done by making them interact with robots during game activity scenarios. In this experiment, the robot NAO played three different games with an autistic boy (aged 7) during three different sessions. Two of these were collaborative games (*Boomgaardje*⁷⁸ and an adapted version of *Memory*) and the third one was competitive (a traditional version of *Memory*).

Kim et al. have designed a process to create PRT scenarios which uses the TiViPE programming environment. This allows a robot engineer and a PRT therapist to work closely together to create and evaluate robot based PRT. This process consisted of three stages: 1) The creation of game activity scenarios; 2) The insertion of learning opportunities in the scenarios by the therapist; 3) The development of a hierarchical model of the robot assisted PRT scenario⁷⁹.

Prompting children into behaving in a certain way is an important part of PRT. In order for children with ASD to learn the pivotal skills, NAO can provide prompts to help them (or refrain from doing so). Based on the amount of help required from the robot, there are four different levels of prompts in the proposed system. These determine which of the following four actions NAO will undertake: 1) wait for a reaction; 2) give an indirect hint; 3) ask an open-ended question or 4) provide a model of the expected behaviour.

In this hierarchical structure of prompts, it is important that NAO makes sure that the child is paying attention to him. Furthermore, the child must get an appropriate amount of time to react to each prompt (for example, seven seconds).

If the child answers correctly, it is rewarded by the robot. If the child is still unable to answer correctly after the final prompt level, NAO will end the activity by saying that they will repeat the activity next time.

The results of the experiment indicate that children are inclined to pay attention to the robot during the interaction. However, during the first session, it was clear that there should have been a way to stop the introduction phase of a game if the child understood the rules. This was adapted before the start of the second session to avoid the child losing interest.

Furthermore, the robot’s non-verbal behaviour were successful in creating an atmosphere of joint-attention (something which is quite difficult for children with ASD). However, NAO’s verbal expressions were not enough to motivate the child to participate in the games. For example, his prompting could not tempt the child into choosing a game to play during the first session. This means that the scenario scripts should be more flexible, to allow the robot to respond to hesitant or stubborn

⁷⁷ The section *Robot Assisted Pivotal Response Training* is based on (Kim et al., 2014), unless otherwise indicated.

⁷⁸ *Boomgaardje* is a game in which players have to work together to harvest cherries before the ravens eat them.

⁷⁹ Please see Appendix VIII for an example of the scenarios developed in steps 1 and 2 of the process for the game *Boomgaardje* and for a diagram representing the four levels of prompts.

behaviour of particular children. To this end, more variations of the same prompt should be included to avoid boring the child.

During the third session, the child seemed to have learned one of the desired skills, namely self-initiation. They were playing a memory game and NAO said that the cards he had turned over did not match. The child, realising the robot was wrong, protested.

Moreover, the child seemed to be more willing to pay attention to the robot during the sessions when he was alone with him (the second and third session). During the first session, he was still too dependent on the therapist to truly enjoying playing with NAO.

7.2.5 Conclusion

In this section, we have discussed several robot assisted ASD therapies. It has been shown that autistic children tend to react positively towards robots (Shamsuddin et al., 2012a), and therefore, therapies might be greatly enhanced if robots were included. However, this is not yet commonly done, as clinics often lack the budget and the expertise necessary to make such programmes work. The promising results of the studies discussed in this chapter, however, indicate that this should be changed as soon as possible. Autistic children would benefit from these kinds of therapy as they can help them learn a broad range of social skills.

One of the programmes developed in this context is ASK NAO, an initiative by Aldebaran itself. The reasons Aldebaran gives as to why NAO would be the perfect companion for autistic children are rather general and could in fact apply to many robots (e.g. predictable behaviour, clean design, not judging). However, the applications developed within the scope of the ASK NAO programme are valuable however, as they offer great possibilities, such as the creation of a network between children, teachers, therapists and parents, allowing for information to be communicated rapidly and effectively (ASK NAO Information Kit, 2015).

The advantages of the NAO platform are further confirmed by Miskam et al., who conducted studies for the National Autism Society of Malaysia. NAO has proven to be a good teacher for social and learning abilities and therefore, they have designed an experiment in which NAO will help children to learn numbers and emotional gestures (Miskam et al., 2014). Shamsuddin et al. have also conducted experiments for NASOM. Their research has indicated that 80% of the observed children showed less autistic behaviour when accompanied by NAO than in a normal classroom setting (Shamsuddin et al., 2012b). Moreover, as these children all had low IQs (the IQ's of autistic children range from very high to very low), this experiment also indicates that robot assisted ASD therapies would be a good option for autistic children with an impaired or delayed IQ (Shamsuddin et al., 2012b).

As mentioned before, all children are unique and therefore, they require unique therapies. This indicates the need of customizable platforms that can be adapted to suit the needs of each individual child. One of these platforms is being developed in the Netherlands, by Gillesen et al. Such platforms are absolutely necessary as the needs of different autistic children might be contradictory to each other (Gillesen et al., 2011). This platform allows technical and non-technical people to build scenarios together, based on the needs of individual children (Gillesen et al., 2011). Before it will be able to be used in clinics however, the platform should become more stable first (Gillesen et al., 2011).

In the final part of this section, an experiment by Kim et al. showed that NAO could also be integrated in PRT training, an established method to teach children motivation, empathy and interaction with peers in a natural way (Kim et al., 2014).

7.3 NAO in Diabetes Therapy: The ALIZ-E Project



Figure 48 NAO in the ALIZ-E Project (ALIZ-E, 2014c)

The ALIZ-E Project was an international project between 2010 and 2014 that was supported by the European 7th framework programme⁸⁰. Its aim was to ‘contribute to the development of integrated cognitive systems capable of naturally interacting with young users in real-world situations, with a specific goal of supporting children engaged in a residential diabetes management course’ (ALIZ-E, 2014a)⁸¹.

In order for robots to become true companions for children with diabetes, they need to be able to establish long-term relationships with them. This is thus the main research topic for the ALIZ-E project. Their target group are hospitalized children aged 8 to 11 who have metabolic disorders (mainly diabetes but also obesity). The experiments are conducted in two hospitals: one in Italy and one in the Netherlands. This project emphasises the need of conducting experiments in real-life settings, as they want to observe child-robot interaction “in the wild”. They believe that it is first and foremost important to keep in mind that children are not to be considered “mini-adults”. They have far more imagination than most adults, which should be remembered when designing robotic companions. Furthermore, many children nowadays are used to highly sophisticated toys. Therefore, they often lose interest in the robots if their responsiveness is too limited. Experiments have indicated that less complex but flexible robotic behaviour achieves better results than complex but repetitive behaviour.

The research team expects that using robots as companions for diabetic children will lead to reduced stress and anxiety levels. Furthermore, they believe that children are more likely to respond in a positive way towards the proposed treatment if they are supported and motivated by a robotic friend.

Several of the studies mentioned earlier in this bachelor’s thesis were conducted in the larger context of the ALIZ-E project. Among these were studies on frameworks needed for child-robot interaction “in the wild” (Kruijff-Korbayová et al., 2011 & 2012; Ros Espinoza et al., 2011; see chapter 4.2.2), studies on children’s turn-taking behaviour (Kruijff-Korbayová et al., 2013; see chapter 4.3.1), studies on the use of flexible sub-dialogues (Cuayáhuitl & Kruijff-Korbayová, 2012; see chapter 4.6), and studies on children’s interpretation of emotional body language (Beck et al., 2011 & 2013; see chapter 5.1.2).

In order for a robot to be eligible as a companion in diabetes therapies, it should meet several requirements listed by the ALIZ-E project team.

⁸⁰ The European 7th framework programme was an initiative of the European Commission that funded European research and technological development between 2007 and 2013 (European Commission, 2015). Its ICT section was responsible for funding the ALIZ-E project, as part of the Cognitive Systems and Robotics Objective (European Commission, 2015). It received 8.29 million Euros to complete its goals (

⁸¹ The introductory information on the ALIZ-E project in this section is based on (ALIZ-E, 2014a), unless otherwise indicated.

7.3.1 Requirements for Robots in Diabetes Therapy

Children diagnosed with diabetes are confronted with serious challenges on different levels: physically, mentally, socially and emotionally (Nalin et al., 2012a)⁸². Furthermore, hospitalization in itself is quite challenging as well to these children: they have to leave their familiar environment behind, together with their friends and families. To make this period in their lives more bearable, the ALIZ-E project team wants to introduce robotic companions. These robots could provide the young patients with an interesting combination of entertainment and education.

In the past, Animal Assisted Therapy (AAT) has proven to be very successful. However, as mentioned earlier, not all hospitals are willing to welcome animals onto their grounds. Furthermore, animals can be quite unpredictable, which is not ideal in the context of medical therapies. The ALIZ-E project team believes that robotic companions could fulfil the same roles as these therapeutic animals, but without the hygiene issues. To accomplish this, they have established four goals for their robotic development: 1) Reducing the stress and anxiety levels of the children; 2) Improving their response to the treatments; 3) Improving their self-efficacy; and 4) Motivating them to do physical activity. Sets of requirements have been created to meet each of these goals. Please see Table 33 for these requirements.

Requirements for robotic companions	
Reducing stress and anxiety levels	
1	Allow physical contact
2	Provide a tactile feeling which is realistic and pleasant
3	Provide a feedback of enjoyment of the physical contact
4	Provide non-verbal cuddles and affection expressions
5	Behave like an animal: not contradicting or judging the child
6	Accept the child immediately
7	Adapt to the child
8	Have a long term memory
9	Express recognition and familiarity
10	Propose games with inversion of roles <ul style="list-style-type: none"> - For example: the child plays the role of the doctor and the robot plays the patient - Goal: teach the child what will happen to him or her
11	Distract the child from his or her own problems and condition
12	Amuse and entertain the child
Improving response to treatments	
13	Become an educational companion
14	Motivate the child (overcome fear & follow healthier lifestyles)
Improving self-efficacy	
15	Be non-independent and care needy
16	A set of action should be defined to take care of the robot
17	Have specific needs which force the child to use an appropriate behaviour with it
Motivating physical activity	
18	Teach the child to do physical exercises while having fun

Table 33 Requirements for Robotic Companions (Nalin et al., 2012a)

Although NAO does not meet all of these requirements (NAO is not furry and pet-able for example), he was still chosen by the ALIZ-E research team to be their robotic platform. His selection is due to his friendly, cartoon-like appearance, his ability to express emotions through body language and speech and his advanced movements.

⁸² The section *Requirements for Robots in Diabetes Therapy* is based on (Nalin et al., 2012a), unless otherwise indicated.

In order to create a credible robotic companion, it is important that a robot manages to balance on the line between being a toy and being a living creature. This is influenced by the imagination typical of children: they tend to humanize NAO, even while being aware that he is a robot and not a real human being. Furthermore, it is important that the robotic companion should not be flawless. Experiments showed that children lose interest in the robot if they perceive them as perfect. For example, Kruijff-Korbayová et al. conducted an experiment in which children had to compete with NAO in a quiz game (see chapter 4.3.1). Results indicated that children lose interest in the interaction if the robot never makes a mistake (Kruijff-Korbayová et al., 2013). However, if the robot makes a few mistakes in a row, the child will be encouraged to be kind to the robot and try and help him get the next question right (Kruijff-Korbayová et al., 2013). This could be an important method to reach the third goal of the ALIZ-E project. Moreover, robots should introduce contextual comments into their interaction with the children. Children do so as well when they interact with other children, and therefore, it is needed to make the HRI feel more natural to them. For example, when playing a quiz, NAO could say sentences such as *'I really didn't know this answer, did you?'* or *'That was a hard question'*. The creation of an affect space, as suggested by Beck et al. (see chapter 5.1.2), could significantly contribute to the creation of a believable companion. This, in turn, would be an important step towards the accomplishment of goal 1 of the ALIZ-E project. Finally, NAO should never appear static to the child when switched on, because it would give the children the impression that their companion was more a statue than a real friend. Therefore, “breathing” and other continuous behaviour should be implemented.

7.3.2 Two Robotic Companions

As mentioned in chapter 7.1, different roles require different types of companions. The ALIZ-E project team, however, has chosen to use one single robot model, NAO, to perform the function of a companion and of a monitor. They developed two complementary robot prototypes which were introduced to the children as brothers: NAO and ROBIN (ALIZ-E, 2014b)⁸³.

NAO, the elder brother, fulfilled the function of a teaching companion, helping the children to understand their medical conditions and encouraging them to live healthy life styles. Whenever possible, it was NAO who introduced his brother to the children after they had gotten used to interacting with a robotic coach.

ROBIN (which stands for Robotic Infant), the younger brother, played the role of a teachable and affective toddler companion (comparable to the pet companion described in chapter 7.1). ROBIN, like the children with whom he interacts, suffers from diabetes and needs to be taken care of. The children play with ROBIN in a toddler's playroom, filled with objects that can help them cater to ROBIN's needs. Figure 49 shows the playroom and a child interacting with ROBIN. The toddler companion needs feeding, drinking, playing, sleeping, socializing and correcting his glucose levels. To satisfy these needs, ROBIN depends on the children, especially for the last one.

ROBIN is an autonomous robot, which means that he can make his own decisions based on the situation he finds himself in. This decision making behaviour is called “action selection”. Due to his affective action selection architecture (as seen in Figure 50), ROBIN is motivationally autonomous as well. This means that he will act based on his own motivations. When he interacts with his environment, his motivation changes, which leads to action. The intensity of his motivations determine which particular action he will undertake.

⁸³ The section *Two Robotic Companions* is based on (ALIZ-E, 2014b), unless otherwise indicated.



Figure 49 ROBIN's Playroom (ALIZ-E, 2014b)

Homeostatic variables influence the affection felt and expressed by the robot. For example, when the pleasure hormone is increased, ROBIN will produce non-linguistic affective vocalisations. In addition to these variables, ROBIN is also influenced by his blood-glucose levels. When these levels are too high or too low, ROBIN's tiredness essential variable will increase, leading to him being motivated to rest. The children are responsible for checking his levels with a similar device to the one they use to test their own blood-glucose levels. This is an important experience for the children, as the interaction with their companion makes them more comfortable with their own medical condition. When they test the robot's levels, they can use their own knowledge of diabetes (which can be enforced by their teaching sessions with NAO) to choose the necessary treatment.

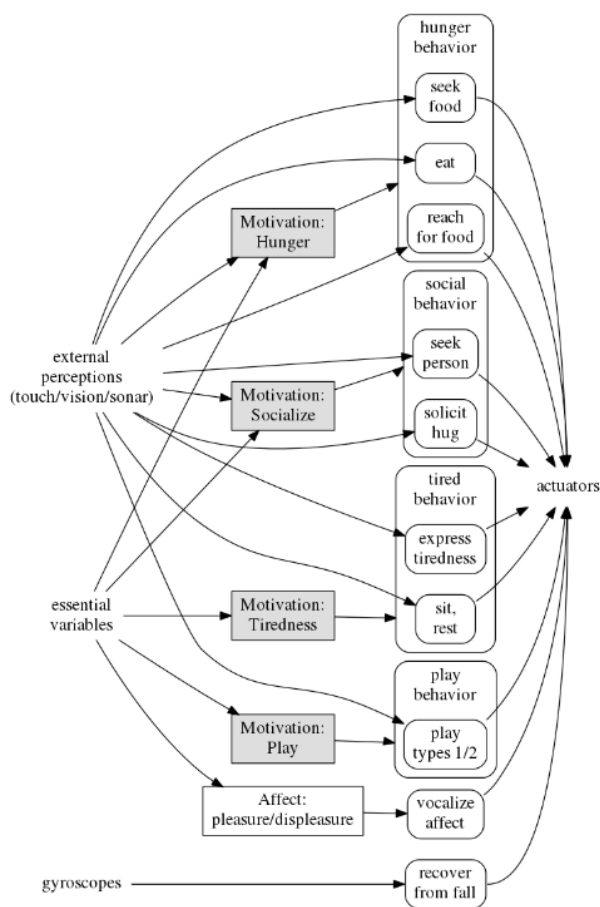


Figure 50 ROBIN's Architecture (ALIZ-E, 2014b)

7.3.3 Children's Adaptation in Multiple Interactions with NAO

The ALIZ-E Project also tried to solve one of the most important bottlenecks in robotic research: achieving sustainable, long-term HRI (Nalin et al., 2012b)⁸⁴. Because of the fact that HRI is still often perceived as unnatural and repetitive, robots are not yet completely suitable to be employed as daily human companions. Human users often lose interest in their robot after the novelty period expires. There are two possible ways to solve this bottleneck: robots could either be designed taking into account the perception of users, or they could (in the case of autonomous robots) learn from interactions. The ALIZ-E project used a combination of both methods to improve their robotic companion NAO.

Autonomous robots should learn both verbal and non-verbal skills from their human conversational partners. This is even more important when they interact with children, as children are more sensitive to non-verbal behaviour than adults (Davies, 2003).

It is, however, very important to remember that social adaptation is bi-directional. Robots must adapt to the humans during interactions, but humans likewise adapt to the robot, even though this happens mostly unconsciously. Experiments indicate that NAO can learn arm-gestures through imitation because humans automatically adopt helping behaviour, even though they believed that NAO's success depended on the verbal feedback they provided (Hiolle et al., 2010). In this experiment, the motivational and caring behaviour of the human participant was thus vital for the achievement of the interaction goal, however, this is not always the case. Cañamero et al. designed an experiment in which humans had to recognise the emotions expressed by the LEGO robot Felix (Cañamero & Fredslund, 2001). Humans did also match their non-verbal behaviour to the robot's in this experiment, but this was not due to motivation or caring (Cañamero & Fredslund, 2011). They simply did it because they were focussed on the task they were asked to perform (Cañamero & Fredslund, 2011). This proves that humans automatically adapt their non-verbal behaviour to robots in the case of successful HRI.

Furthermore, as stated above, humans do not only adapt their non-verbal behaviour to the context of HRI but also their verbal behaviour. They do this as well in normal human-to-human interactions. For example, humans match their expressions to their conversational partner, which is called "convergence". This accommodation strategy is used to reduce social differences between the conversational partners (Burgoon et al., 1995). Furthermore, they also match specific features of their speech signals, such as pitch, phonological features and amplitude. In general, it is believed that humans respond in similar ways to robots during interaction. In chapter 4.3.1, a study was discussed which showed that children adapted their turn-taking behaviour to the robot (Kruijff-Korbayová et al., 2013).

However, during normal conversations both partners adapt themselves to each other. In this experiment, NAO did not match his own behaviour to the children's. For example, he did not re-use vocabulary of syntax used by his conversational partners.

To examine the adaptation of the children's verbal and non-verbal behaviour, an experiment with 13 children was conducted in a hospital in Milan. Each of these children participated in three separate interactions with NAO. During each interaction, the child could choose to play one out of three possible games: a quiz, a dance session or an imitation game. NAO introduced himself to the child at the beginning of the first session, asking for the child's name. This created an atmosphere of familiarity which was considered important to evaluate correctly the establishment of a long-term relationship between the child and the robot. This familiarity was enhanced by the fact that NAO remembered whether or not he had already played a particular game with a child: he only explained the rules at the beginning of the first time they played that particular game. The architecture of the system is very similar to the one described in chapter 4.6.

⁸⁴ The section *Children's Adaptation in Multiple Interactions with NAO* is based on (Nalin et al., 2012b), unless otherwise indicated.

The results indicate that children do indeed adapt their verbal and non-verbal behaviour to NAO. Two communication problems were analysed: overlaps and sentences ignored by the robot. Speech timing (and its influence on turn-taking patterns) was clearly adjusted by the children over the course of the three sessions to match the timing of the robot. This leads to a decrease in the number of both types of communication problems. This resembles the results of the experiment described in chapter 4.3.1 (Kruijff-Korbayová et al., 2013). Figure 51 shows the average number of communication problems per minute per session⁸⁵.

The children likewise adapted their verbal behaviour to NAO: they started to re-use specific structures and words. For example, during session one, children were more inclined to provide feedback to the robot in the way they normally do: ‘Yes! Right’. In the third session, however, they had often adapted their expressions to match the more formal language of the robot: ‘The provided answer is correct’.

Furthermore, the children also match the features of their speech to the robot. They use a more uniform tone which resembles NAO’s speech and speech slower in general.

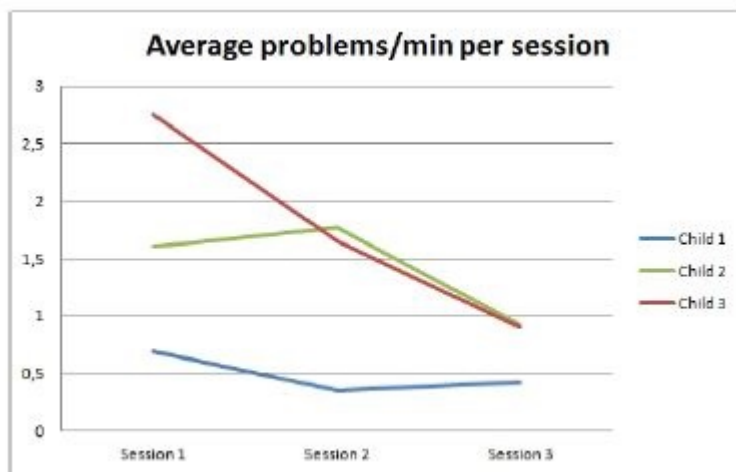


Figure 51 Average Problems/Minute per Session (Nalin et al., 2012b)

It is also interesting to notice that most children – even though they had very different non-verbal behaviours – matched their body language to NAO’s. For example, some of them adopted a stance close to NAO’s normal “rest” position. Others imitated the arm movements the robot made while asking questions or providing feedback.

It is important to keep in mind that these children adapted their verbal and non-verbal behaviour even though there was no practical need nor any visible results. The interpretation of the children’s speech, after all, was always correct as a human Wizard controlled the robot. This means that their adaptation did not help NAO to understand more of the interaction. Furthermore, the children’s adaptation was also not caused by repetitive verbal patterns as the robot used several variations of each sentence.

Even though earlier research indicated that humanoids such as NAO are usually perceived as creatures with authority (see chapter 7.1), the participants of this experiment thought NAO resembled a younger child than they were. This might partly be caused by the fact that they were explained that they would participate in an experiment which would improve the robot. They seemed to have naturally taken on the role of the teacher, wanting to help their robotic friend (which is closer to the role of the pet-like robot described in chapter 7.1). This also resembles the result of the experiment by Hiolle et al. that

⁸⁵ The graphs used in the report by Nalin et al. are based on the data collected from only 3 out of the 13 children.

indicated that people automatically adopt their behaviour when they care about the robot they are interacting with (Hiolle et al., 2010).

However, while many of the adaptations lasted throughout the three sessions, some also disappeared before the end of the experiment. This is due to the fact that even though the children are enthusiastic about teaching and helping the robot, they lose interest after a while because NAO does not adapt his own behaviour to theirs. It is thus vital that both communicative partners adapt their behaviour to develop a credible and natural robotic companion. Many children thus lose interest in the robot, but all of them reported having enjoyed playing with the robot (even though they are usually used to playing far more entertaining video games than the games played with NAO).

7.3.4 Conclusion

In this second set of case studies, experiments with NAO as a companion for children with diabetes were discussed. These were conducted in the larger context of the ALIZ-E Project, an international initiative supported by Europe. The goal of this project was to develop robots that would be suitable companions for long-term relationships with children that suffer from metabolic issues (ALIZ-E, 2014a). In order to accomplish this goal, the ALIZ-E project team has established a set of requirements for robots to be used in diabetes therapies, which were discussed in section 7.3.1 (Nalin et al., 2012a).

As seen in section 7.1, different functions require different robots. Yet, the ALIZ-E project has chosen to use one model (NAO) to fulfil both the function of a robotic teacher as the function of a robotic companion (ALIZ-E, 2014b). These two robots were presented to the children as brothers: the elder brother NAO who performed the role of a teacher and ROBIN, the younger brother who presented himself as a companion to the children (ALIZ-E, 2014b). These two robots fulfilled both the children's needs for education as for entertainment, which leads to the conclusion that the NAO robot can indeed be used for both functions (ALIZ-E, 2014b), even though other research suggested more pet-like robots for the companion role (section 7.1).

Moreover, the ALIZ-E Project also focussed on solving one of the main problems of robotic research: humans get bored when HRI is no longer considered new and exciting. This can be solved in two different ways (both of which were used in the ALIZ-E Project): either by keeping this in mind when designing robots or by teaching robots to learn from previous interactions (Nalin et al., 2012b). In normal inter-human conversations, both partners will unconsciously adapt their communicative behaviour to each other. Research has shown that children do this as well when interacting with a robot (see section 4.3.1). To examine the influence of this adaptation, Nalin et al. have conducted an experiment in which NAO would not match his behaviour to that of the child (Nalin et al., 2012b). As expected, the children did start to mirror NAO's verbal and non-verbal communicative behaviour after a while; however, as NAO did not react the way they unconsciously expected (by adapting himself to them as well) the children started to lose interest near the end of the experiment (Nalin et al., 2012b). This indicates that robots should be programmed to adapt their communicative behaviour as well to match their conversational partner to establish long-term HRI (Nalin et al., 2012b).

7.4 NAO as a Teaching Assistant for Sign Language

Another domain in which NAO can be used as an assistive agent is the teaching of sign language. Computers and videos are used increasingly to help hearing-impaired people learn sign language (Kose & Yorganci, 2011)⁸⁶. Recent studies have been conducted to determine the benefits of robots assisting sign language teachers. No matter the experience of the human teacher, each individual teacher has his or her own style, which leads to difficulties when the teacher needs to be replaced for some reason.

⁸⁶ The section *NAO as a Teaching Assistant for Sign Language* is based on (Kose & Yorganci, 2011), unless otherwise indicated.

Therefore, a robotic teaching assistant could be useful, as a robot could provide tutor-independent teaching.

In 2011, Kose & Yorganci have conducted an experiment in which a NAO H25 taught some elementary Turkish Sign Language (TSL) vocabulary to preschool children. The physical limitations of NAO, however, led to quite some problems, which resulted in the research team opting for another robotic platform in subsequent experiments. For example, most TSL gestures are based on five fingers signs, while NAO only has three fingers on each hand. Furthermore, NAO's degrees of freedom were not considered sufficient to teach sign language accurately.

In the experiment, 106 6-year-olds listened to NAO telling a simple yet interesting story⁸⁷. During his interactive story, NAO supported five basic vocabulary words with TSL (*car, friend, dad, three* and *table*). Afterwards, the children were given a test to examine whether or not they learned sign language through the interaction with the robot. This test consisted of the robot showing the signs one by one and the children putting stickers of the relevant words on story cards. The research team expected that there would be a positive effect on the performance and interaction compared to video-based studies because of the toy-like embodiment of the NAO robot.

During the experiment, children were expected to provide feedback to the robot. They were given flash cards with the TSL signs on them and each time NAO used a sign, the children showed the robot which sign they thought he performed. If they were right, NAO said the name of the sign. If they were wrong, NAO's eyes flashed from green to red and to green again.

Even though NAO comes with a text-to-speech function, this was not used in this experiment. The research team was unable to find a natural sounding Turkish TTS programme, and therefore, they chose to implement the voice of a 6-year-old into NAO.

The results indicate that 90% of the children were able to complete the final test without making any mistakes. This recognition rate is a lot higher than those of previous experiments conducted with adults and teenagers, even though the same set of words was used (Kose et al., 2012). These words were chosen because NAO would not be able to perform them perfectly due to his physical limitations. Yet, in this experiment, the children were presented these words in the context of a story, which resulted in the high recognition rates.

In later experiments, NAO's sign language teaching abilities were compared to those of the five-fingered Robovie R3 robot (see Figure 52) (Kose et al., 2014). In this study, ten TSL words were taught by both robots to children and adults (Kose et al., 2014). The results of this experiment are shown in Table 34.



Figure 52 Robovie R3 Robot (Kose et al., 2014)

⁸⁷ The story is included as Appendix IX.

Recognition rates		
	NAO (10 participants)	Robovie R3 (21 participants)
Spring	60%	100%
To throw	40%	98%
I / Me	40%	98%
Big	100%	98%
Mountain	60%	98%
Table	^{/88}	98%
Black	60%	98%
To come	60%	92%
Baby	80%	90%
Mother	40%	88%

Table 34 Recognition Rates of NAO and Robovie R3 (Kose et al., 2014)

As shown in Table 34, participants were significantly better at recognising signs when performed by the Robovie R3 Robot, most likely because of his greater degrees of freedom and because of the fact that this robot has five fingers (Kose et al., 2014). Based on these results, the research team decided that further research into sign language teaching would be conducted with the Robovie R3 Robot instead of with NAO (Kose et al., 2014).

⁸⁸ The word *table* was replaced by *car* in the NAO experiment, but this word does not count for the final results.

8. Conclusion

In this bachelor's thesis, I have tried to present the state-of-the-art of NLP with NAO robots. In chapter 2, the difference between artificial and natural languages was presented. Because of these differences, NLP is not a simple task. The most problematic characteristic of natural language utterances is the fact that they are almost always ambiguous (Beardon et al., 1991). This makes it very hard for computers to identify their meaning correctly, which complicates natural language based HRI.

Another problem with which HRI is faced is the uncanny valley theory by Mori (Mori, 2012). Although it has typically been considered as a limiting factor in the design of robots, new research suggests that this might not be the case – or at least not in the way Mori described. Personal factors influence the position humans assign to a particular robot on the graph, and these factors are not included in the traditional vision on the uncanny valley. Therefore, designers of robots should not necessarily feel constricted by the limitations imposed through the uncanny valley theory.

In chapter 3, NAO and the other two robots by Aldebaran were introduced. These three humanoids are all being developed to interact with humans in social situations. Pepper is even created especially to become a daily companion at home (Aldebaran Robotics, 2015). For robots such as these, NLP is crucial. Without a sufficient command of natural language, they will not be able to fulfil their roles, nor would humans be inclined to allow them into their lives.

From chapter 4 onwards, NLP with NAO robots was discussed in many different applications and domains. Chapter 4 itself dealt with the most straightforward application, namely HRI through natural language. We have first examined a study in which NAO was taught a “native language” through joint attention and multi-instance learning (Dindo & Zambuto, 2010). This proved to be a successful method, but only words that are grounded in perceptual data can be learned in this way. It is therefore most likely not a method that can be used on its own to allow a robot to acquire a complete native language.

Next, to illustrate different approaches to allowing robots to use natural language, three frameworks for NLP were presented. First, the research into frame-based dialogue frameworks stressed the importance of domain-adaptivity and language-adaptivity (Barabás et al., 2012). Domain-adaptivity does no longer truly pose problems anymore, but systems that are completely language-independent do not yet exist. Therefore, a system is currently considered language-adaptable if it can deal with a limited – but extendable – list of languages. The second framework discussed was an event-based dialogue system which improved greatly when the developers decided to replace their finite-state dialogue manager by one that used probabilistic methods and optimisation of dialogue policies based on reinforcement learning (Kruijff-Korbayová et al., 2012). This was necessary because the finite-state machine was deemed unsuitable as it was not flexible enough to deal efficiently with the uniqueness and unpredictability of children, which were the target audience in this experiment (Kruijff-Korbayová et al., 2012). The third framework discussed was a reward-based meta-cognitive framework, based on dialogue rewards that were given to a robot to encourage him to keep the human user interested (Pipitone et al., 2014). The moment the human loses interest, the reward levels will go down, which incites the robot to either change the topic or to provide more detailed information to recapture the interest of the human (Pipitone et al., 2014).

Section 4.3 deals with turn-taking, an important aspect of human communication. Humans automatically expect HRI to be regulated by turn-taking patterns as well (Baxter et al., 2013), which means that robot designers should meet this expectation. We have therefore compared the turn-taking abilities of NAO to those of Kismet, a sociable robot developed at MIT. Experiments with NAO showed that children adapt their turn-taking behaviour to the robot between their first and second interaction with him (Kruijff-Korbayová et al., 2013). Further research has also shown NAO's turn-taking abilities can be improved by implementing non-verbal behaviour to regulate turn-taking (Meena et al., 2012). Kismet, likewise, used these paralinguistic cues to convince humans to lower their conversational pace to allow

her to interact at her own rhythm (Breazeal, 2003). Experiments showed that after a while, humans and Kismet manage to find a balance in which the human's expectations of a conversation are met and Kismet's abilities are not perceived as insufficient (Breazeal, 2003). This is an ideal situation in which Kismet can learn from her interaction with the human, which is her ultimate goal (Breazeal, 2003).

In sections 4.4, 4.5 and 4.6, issues concerning dialogues were discussed (and some possible solutions), as these problems do not limit themselves to the uncanny valley problem as described in sections 2.5 and 2.6. First of all, it should be avoided that robots confuse humans by giving them the wrong impression about their functionalities and capabilities (Fong et al., 2003). Secondly, to avoid repetitiveness, manually created templates could be enhanced by templates developed by the crowd (Mitchell et al., 2014). However, Mitchell et al. also stress that crowdsourcing could not yet entirely replace manual labour, as the new templates should still be checked afterwards to make sure that they fit the desired purpose (Mitchell et al., 2014). Thirdly, in section 4.5, open-domain and closed-domain dialogues were compared, which stresses once again the importance of domain-adaptivity. WikiTalk, an open-domain dialogue application that allowed NAO to use Wikipedia to manage conversations and topic changes, has been discussed as an example of such domain-adaptable systems (Wilcock, 2012). Fourthly (section 4.6), the above described system might score high on the level of domain-adaptivity, it still manages an entire conversation as one single dialogue – a method that has proven to be less efficient than the use of multiple sub-dialogues (Cuayáhuitl & Kruijff-Korbayová, 2012). Such systems are called hierarchical dialogue control and they allow for more flexibility and a better performance in general (Cuayáhuitl & Kruijff-Korbayová, 2012).

Next, a crucial function of natural language is discussed in section 4.7, namely coordination. NAO has proven to be able to perform cooperative tasks with a human being, which is an important accomplishment as this is one of the main functions for which humans use natural language (Petit et al., 2013). Natural language has a double function in this context: it can control the cooperation and it can be used to teach the robot specific actions (Petit et al., 2013). However, as NLP is not yet able to process complex sentences in this kind of systems, the user is limited to the use of relatively simple grammatical structures (Petit et al., 2013). This does only impose problems for the input, however, and not for the output as complex tasks can be coordinated efficiently by simple language (Petit et al., 2013).

In section 4.8, we took a short look at semantic gestures, which are an important aspect of human interaction (as human communication is almost never exclusively verbal). As Zheng & Meng pointed out, semantic gestures cannot simply be transferred to robots without testing them on the target audience first (Zheng & Meng, 2012). These gestures are culturally defined and, more importantly, are perceived differently when performed by a robot than by a human (Zheng & Meng, 2012). This could lead to miscommunication, which is an important reason to test each gesture before it is being implemented (Zheng & Meng, 2012).

And finally, in section 4.9, we examined one of the problems that prevent a faster rise of robots in our society: the fact that controlling and programming them still requires technical expertise. Robots are meant to be used by non-technical people, and therefore, the dream of NLP would be to develop robots that can be controlled in natural language. This, however, is not yet entirely possible (some researchers even believe that it will never be possible) and therefore, alternatives and intermediary solutions are being examined. One of these alternatives is a user-friendly programming environment which requires only a basic knowledge of programming (Lourens & Barakova, 2012). Another alternative – or as more pessimistic researchers believe, the only option – would be to use artificial languages rather than natural languages, such as RIOLA. This language is easy-to-learn for humans and easy-to-understand for robots, which would allow it to increase the performance of HRI (RIOLA, 2015). Less sceptic researchers, however, continue their experiments with intermediary solutions. For example, motion description languages, such as Cybele, allow end-users to define robotic motions through natural language (Shukla & Choi, 2013). This domain is thus clearly divided over two opposing opinions, and whoever succeeds best will strongly influence the way we will interact with robots in the future.

In chapter 5, another application of natural language in HRI was discussed: the communication of emotions. This chapter consisted of two parts: section 5.1 dealt with the expression of emotions by robots and section 5.2 with the detection of human emotions. As robots need to be accepted into the daily lives of humans as warm companions, they need to be able to display emotions much in the same way as humans do (Beck et al., 2012). To allow NAO to express his emotions through body language, an affect space has been created to blend emotions along three axes (arousal, stance and valence) as to generate hybrid emotions (such as an emotion of 30% happiness and 70% excitement) (Beck et al., 2012). The results of experiments conducted with this affect space indicate that body language is a suitable emotional medium, but that age and cultural background might have significant influence on the perception of emotions expressed in this way (Beck et al., 2012). Other libraries of emotions were discussed as well, which indicate the importance of consistency between different emotional media (Monceaux et al., 2009) and showed the unreliability of eye colour as a medium (Häring et al., 2011).

In the final part of section 5.1, the expressiveness of four other robots was compared to NAO's. First, a comparison with iCat indicated that there was no significant difference between their expressiveness, even though NAO did not possess facial expressions (Cohen et al., 2011). Secondly, an experiment with Kismet showed how a robot's drives, emotions and motivations could lead to learning social behaviour (Breazeal, 1999). Thirdly, Brian's recognition rates (McColl & Nejat, 2014) were compared to those of NAO, which showed that they were largely similar except for some unexpected differences connected to the perception of fear. This might either be due to the different embodiment or to the motions chosen to accompany fear. Fourthly, an experiment with KOBIAN showed that facial expression enhanced the perception of emotions (Zecca et al., 2009). However, when compared to NAO's recognition rates, KOBIAN scored less good even though NAO does not have facial expressions.

Section 5.2 dealt with two main problems: 1) how to detect human emotions in general and 2) how to detect emotions of different age groups. In the context of problem 1, we have discussed an experiment in which NAO successfully determined a user's emotions based on a combination of facial and semantic cues (Zhang et al., 2013). The results indicate that a combination of LSA and NAO's face detection API are sufficient to allow the robot to formulate adequate responses given the user's emotion and the topic discussed (Zhang et al., 2013). Problem 2 proved to be a more difficult hurdle to cross. Human emotions can be detected in speech, but adult speech is highly different from child speech. Therefore, robots such as ROMEO that need to interact with different age groups are faced with a problem when trying to detect emotions. Cross-corpora experiments have indicated that separate models should be provided for each age group to assure high performances (Tahon et al., 2011).

We have concluded this section with an experiment with Brian, as it seemed enriching to have a second robot of which both the expression (section 5.1.7) and the detection of emotions (section 5.2.3) was examined. One of the main emotions which should be detected is accessibility, as it has an enormous influence on how a user will react to a robot at a particular time (McColl et al., 2011). Experiments have shown that DNNS systems are as efficient as trained human DNNS to detect this particular emotion (McColl et al., 2012).

In chapter 6, we have discussed yet another application of NLP in HRI: robotic personalities. This is a domain which will require much more research as a consensus on the best type of robotic personality has not been reached. Many people believe that the personalities of robots should either match or complement their user's personality (as these are the two social attraction rules that govern inter-human relationships), yet a few voices express radically different opinions. Whoever turns out to be right, one fact remains: robotic personalities are crucial for the development of natural HRI, as humans automatically expect robots to have compelling personalities, as they do themselves.

After discussing the five basic robotic personality types (Fong et al., 2003), we have presented an experiment by Aly & Tapus which belongs in the personality-matching camp. Ideally, experiments illustrating the views of the other camp (personality-complementation) with NAO robots should have

been included as well, but these could not be found. However, some of the independent voices were included. In this case, these voices belong to three Dutch students who participated in the University of Twente 2012 Student Conference on IT. Although the results of these experiments were not always conclusive or even reliable (many things went wrong), they were included because they were considered to be enriching perspectives in the robotic personality debate. Windhouwer examined the effect of the task context on a human's preference for certain robotic personalities in section 6.3. Next, in section 6.4, the effect of group interactions on the preference of a user was presented by Leuwerink. Lastly, in section 6.5, Waalewijn discussed the effect of neighbouring cultures on this preference.

In the final chapter of this bachelor's thesis, chapter 7, we have discussed several case studies in which NAO was used in medical contexts (namely autism therapy, diabetes therapy and sign language teaching). First, we have taken a look at the influence of the embodiment of a robot on the tasks that it can perform in such a context. Research has shown that robots can fulfil two main functions in a medical setting, which require other types of robots: a rehabilitation monitor or a companion (Diaz et al., 2011). Humanoids, such as NAO are ideal to play the part of a monitor, as they inspire children to be curious and to explore their own abilities (Diaz et al., 2011). Cute, pet-like robots, such as PLEO, are more suitable to be companions, as they inspire children to be caring and affective (Diaz et al., 2011). The embodiment of a robot is thus an important factor to keep in mind when selecting a particular model for a particular task.

Second, we have studied several experiments on NAO's performance in the context of autism therapy. ASD therapies can be greatly improved by incorporating robots, as many autistic children react in a positive way to these artificial companions (Shamsuddin et al., 2012a). This is not yet commonplace, however, as these kinds of therapies demand a considerable amount of money and expertise. To facilitate this, many initiatives are established internationally, among which the ASK NAO programme by Aldebaran. This programme presents many applications that are valuable in the context of ASD therapies, such as the creation of a network that would enhance communication between parents, children and therapists (ASK NAO Information Kit, 2015). Another series of experiments were conducted in the context of the National Autism Society of Malaysia. In these, it has been shown that robots positively influence the behaviour of autistic children in 80% of the cases (Miskam et al., 2014) and that children with impaired or delayed IQs would benefit from robot assisted ADS therapies (Shamsuddin et al., 2012b). Furthermore, as all children are unique, they require unique therapies. Therefore, customizable platforms are needed that can meet the needs of each particular child. The platform discussed in section 7.2.3 enables technical and non-technical people to work together to create scenarios adapted to individual needs (Gillesen et al., 2011). This is important, because autism is a very broad spectrum and the needs of one child might be contradictory to the ones of another child. Finally, we have also seen how robots can be integrated in the established PRT therapy (Kim et al., 2014).

Third, we have studied some experiments conducted in the context of the ALIZ-E project, an internationally funded initiative to develop robotic companions for children with diabetes. In order to create such a companion, a set of requirements were established, after which NAO was chosen as the model to be used in this project (Nalin et al., 2012a). Two versions were used, which were presented as brothers: the elder NAO who acted like a teacher, and the younger ROBIN who acted like a companion (ALIZ-E, 2014b). This shows that even though normally, pet-like robots are preferred for the role of a companion (see section 7.1), NAO could perform this function efficiently as well (ALIZ-E, 2014b). The ALIZ-E project team also tried to find a solution for one of the main problems of robotic research: how to avoid that humans get bored when the novelty effect wears off. When humans interact with other humans, they tend to adapt their communicative behaviour to their conversational partner. As this is done automatically, humans unconsciously apply this to HRI as well (see, for example, turn-taking behaviour in section 4.3.1). When a robot does not adapt its own behaviour as well, children will lose interest in the interaction (Nalin et al., 2012b). Enabling NAO to adapt his communicational behaviour to his interaction partners would thus prevent the novelty effect from wearing off and would thus keep

HRI interesting longer (Nalin et al., 2012b), which is exactly what is needed to develop long-term robotic companions for medical contexts.

Finally, the benefits of a robotic teaching assistant to teach sign language to the hearing-impaired were discussed. Sign language is taught slightly different by each human teacher, which means that problems can arise when the teacher needs to be replaced (Kose & Yorganci, 2011). Therefore, robotic teaching assistants could prove useful, as they can repeat a specific gesture a hundred times in exactly the same way (Kose & Yorganci, 2011). Research in this area with NAO has not been a complete success, however. Kose & Yorganci used NAO in their earlier experiments, but soon decided to replace him with another model because of NAO's physical limitations (e.g. insufficient degrees of freedom and too few fingers) (Kose et al., 2014).

We can thus conclude from chapter 7 that even though NAO seems to be a highly efficient and suitable choice in some medical contexts (such as ASD therapies and diabetes therapies), he would not be the best candidate to become a sign language teaching assistant. It is important to keep in mind that no robot would ever be able to fulfil all possible roles, much in the same way that each human also has his or her own individual talents and weaknesses. Several factors need to be considered when choosing a robot for a particular task, both internal (i.a. embodiment, physical limitations, HRI possibilities and software applications) and external (i.a. target audience, function and context).

9. Further Research

In this chapter, we will take a brief look at some suggestions for future research, based on gaps in the literature as presented by the authors of the above-described studies and based on unsatisfying results of some of the experiments.

In chapter 2, new research on the uncanny valley has led to new insights into this traditionally accepted limitation on robotic design. However, although it is becoming clear that personal factors should be included in the graph, many uncertainties remain. For example, further research should point out which personal factors are necessary and sufficient to be considered in a new uncanny valley theory. Studies focussing on the influence of a robot's appearance on the uncanny effect have suggested that EQ and experience with video games might be important factors (Beck et al., 2012). It might thus be interesting to conduct further experiments to see to which measure these two also influence the perception of an eerie feeling inspired by the use of natural language.

Chapter 4 presented several studies on HRI through natural language which greatly enhanced the knowledge of NLP with NAO robots. Yet, some of these works might benefit from new experiments based on their results. Section 4.1 discussed a learning-technique in which joint attention and multi-instance learning were used (Dindo & Zambuto, 2010). The results indicate that NAO was indeed able to learn several words by using this technique. However, this was only possible because the to-be-learned nouns were grounded. Furthermore, verbs were not learned but hardcoded. It might be enriching to widen this research to include a larger number of lexical items (such as verbs) and to see to which extent joint attention and/or multi-instance learning could be usable in the teaching thereof.

In section 4.2, several frameworks for HRI with NLP were discussed. The frame-based system proposed by Barabás et al. showed promising results (Barabás et al., 2012). However, as the authors mention themselves at the end of their paper, NAO was only capable of performing 18 actions in this study. The results of the experiment were good, but it might be useful to see how many functions could be performed by this kind of system without significantly lowering its response time and effectiveness. Furthermore, Barabás et al. focus on the importance of language-adaptivity. Current NLP systems are not truly language-independent yet, although they can often work with an extendable set of languages. This is an area of NLP in which there remains a lot of research to be done. Furthermore, the developers of the framework discussed in section 4.2.3 do not provide any of the results of their experiments, yet they claim that their system was efficient and versatile. It might thus be interesting to conduct further research into this type of frameworks and to make the result public as to prove its efficiency and suitability.

The crowdsourcing technique discussed in section 4.4 seems a promising method to reduce the amount of manually created dialogue templates and to avoid repetitive HRI. However, as Mitchell et al. indicate, it is not yet a possibility to replace entirely manually created templates by those developed by the Crowd (Mitchell et al., 2014). This could be an interesting method to be developed in the future, as HRI should absolutely not feel repetitive and unnatural to humans. This would interfere with the desired function of robotic companions, as humans would become less willing to engage in interaction with a boring robot.

In section 4.5, WikiTalk, an open-domain dialogue system was presented, which allowed NAO to use Wikipedia as a source for conversational topics. Each time he mentioned a word that was marked as a hyperlink in the original article, he would make a beat gesture to indicate that the human could repeat the word to get more detailed information (Wilcock, 2012). This has proven to be an efficient system, but as Wilcock mentions, some people believe that Wikipedia will one day only contain hyperlinked words (Wilcock, 2012). This would mean that the speech recogniser would have to deal with an unlimited set of words, which is not yet possible today (Wilcock, 2012). Therefore, new ways to manage topic changes should be examined. Furthermore, some words are only once marked as a hyperlink, which results in problems when humans react to the second occurrence of such a word (Wilcock, 2012).

The author mentions that drawing up lists of hyperlinks beforehand might be a possible solution which should be carefully examined in the future (Wilcock, 2012).

We have concluded the fourth chapter by a section on one of the most crucial problems in HRI: end-user programming. Three alternatives have been suggested to solve this issue: (partial) natural language based HRI, end-user friendly programming and artificial languages. As robots are first and foremost meant to interact naturally with humans, each of these alternatives should be extensively tested on different end-users (especially on the elderly and on the very young) to see which methods would work for them.

In chapter 5, Beck et al. suggest that adding motion would improve NAO's expressiveness (which was only based on static postures in their experiments) (Beck et al., 2012). It might be interesting to see which effect motion would have on the perception rates of NAO's emotions.

Various authors also mention the need to examine further the effect of age and cultural background on the perception of a robot's emotions (e.g. Beck et al., 2012; Zhang et al., 2013; Monceaux et al., 2009). This is important to avoid miscommunication and to make HRI smoother and less likely to offend anyone.

In chapter 6, multiple opinions on robotic personalities were discussed. The field is mainly divided in two camps: those who believe that the personality of a robot should match the personality of the user and those who believe in complementary personalities. While research on personality-matching with NAO robots is sparse, studies on personality-complementation with NAO could not even be found. It might thus be interesting to conduct more experiments with NAO to see which side of the argument could be proven empirically. Furthermore, sections 6.3 to 6.5 contained interesting suggestions by students of the University of Twente. These brought enriching new perspectives to the debate, although their methods and results proved to be rather unsatisfactory. Re-conducting these experiments in a more reliable manner might lead to important new information on the perception of robotic personalities.

Chapter 7 discussed several case studies with NAO in a medical context. Among these case studies, several experiments with robot-assisted ASD therapies have been presented. An experiment in Malaysia showed that children with low IQs reacted positively towards NAO (Shamsuddin et al., 2012b). A similar experiment with autistic children with high IQs would provide interesting complementary information to get a full overview of the possibilities of robot-assisted ASD with NAO.

Finally, the ALIZ-E project, as discussed in section 7.3, might have come to an end, but this does not mean that research into robotic companions for hospitalized children with diabetes should end as well. Humanoids, of which NAO has proven to be an exemplary representative, are highly useful either to encourage children as an experienced monitor or to make their lives brighter by being their playful companion. Therefore, the event-based dialogue framework (Kruijff-Korbayová et al., 2012), as discussed in section 4.2.2, might be further developed in the future to support a wider range of dialogues.

10. Bibliography

- Aldebaran Robotics. 2015. *Aldebaran*. www.aldebaran.com. (1 April, 2015.)
- ALIZ-E. 2014a. The ALIZ-E Project: Adaptive strategies for sustainable long-term social interaction. <http://www.dfki.de/KI2012/PosterDemoTrack/ki2012pd09.pdf> (4 March, 2015.)
- ALIZ-E. 2014b. The ALIZ-E Project: Deliverable 2.4 specialized adaptive non-verbal behaviour. <http://www.tech.plym.ac.uk/socce/ALIZ-E/deliverables/D2.4%20-%20Specialised%20adaptive%20non-verbal%20behaviour.pdf> (6 March, 2015.)
- ALIZ-E. 2014c. *The ALIZ-E Project Facebook*. <https://www.facebook.com/pages/Aliz-e-Project/110021579070918?fref=ts>. (10 April, 2015.)
- Aly, Amir & Adriana Tapus. 2013. A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction. In *Proceedings of the 8th ACM/IEEE International Conference on HRI 2013*. 325-332.
- Andry, Pierre, Philippe Gaussier, Sorin Moga, Jean Paul Banquet and Jacqueline Nadel. 2001. Learning and communication via imitation: An autonomous robot perspective. In *Transactions on Systems, Man, and Cybernetics* 32(5). 431-442.
- ASB Lab. 2015. *Assistive Robotics*. <http://asblab.mie.utoronto.ca/research-areas/assistive-robotics>. (4 March, 2015).
- ASK NAO. 2015. *Aldebaran*. <https://asknao.aldebaran.com/>. (8 April, 2015.)
- ASK NAO Information Kit. 2015. *Aldebaran*. https://asknao.aldebaran.com/sites/default/files/material/aldebaranrobotics_asknao_information.pdf (8 April, 2015).
- Barabás, Péter, László Kovács & Maria Vircikova. 2012. Robot controlling in natural language. In *Proceedings of the 3rd IEEE International Conference on Cognitive Infocommunications [CogInfoCom]*. 181-186.
- Barakova, Emilia, Jan C.C. Gillesen, Bibi E.B.M. Huskens & Tino Lourens. 2013. End-user programming architecture facilitates the uptake of robots in social therapies. In *Robotics and Autonomous Systems* 61(7). 704-713.
- Barrick, Murray R., Michael K. Mount. 1991. The big five personality dimensions and job performance: a meta-analysis. In *Personnel Psychology* 44(1). 1-26.
- Barrick, Murray R., Greg L. Steward, Mitchell Neubert & Michael K. Mount. 1998. Relating member ability and personality to work-team processes and team effectiveness. In *Journal of Applied Psychology* 83(3). 377-391.
- Bartneck, Christoph. 2002. *EMuu: An emotional embodied character for the ambient intelligent home*. Eindhoven: Technische Universiteit Eindhoven.
- Bates, Joseph. 1994. The role of emotion in believable agents. In *Communications of the ACM* 37(7). 122-125.
- Baxter, Paul, James Kennedy, Tony Belpaeme, Rachel Wood, Ilaria Baroni & Marco Nalin. 2013. Emergence of turn-taking in unstructured child-robot social interactions. In *Proceedings of the 8th ACM/IEEE International Conference on HRI 2013*. 77-78.
- Beardon, Colin, David Lumsden & Geoff Holmes. 1991. *Natural language and computational linguistics*. Melksham: Ellis Horwood.

- Beck, Aryel, Lola Cañamero & Kim A. Bard. 2010a. Towards an affect space for robots to display body language. In *Ro-Man 2010*. 464-469.
- Beck, Aryel, Antoine Hiole, Alexandre Mazel & Lola Cañamero. 2010b. Interpretation of emotional body language displayed by robots. In *Affine [Affective Interaction in Natural Environments] 2010*. 37-42.
- Beck, Aryel, Lola Cañamero, Luisa Damiano, Giacomo Sommavilla, Fabio Tesser & Piero Cosi. 2011. Children interpretation of emotional body language displayed by a robot. In *Social Robotics 2011*. 62-70.
- Beck, Aryel, Brett Stevens, Kim A. Bard & Lola Cañamero. 2012. Emotional body language displayed by artificial agents. *ACM Transactions on Interactive Intelligent Systems* 2(1). 1-29.
- Beck, Aryel, Lola Cañamero, Antoine Hiole, Luisa Damiano, Piero Cosi, Fabio Tesser & Giacomo Sommavilla. 2013. Interpretation of emotional body language displayed by a humanoid robot: A case study with children. In *International Journal of Social Robotics* 5(3). 325-334.
- Becker-Asano, Christian, Kohei Ogawa, Shuichi Nishio & Hiroshi Ishiguro. 2010. Exploring the uncanny valley with Geminoid HI-1 in a real-world application. In *Proceedings of IADIS [International Association for Development of the Information Society] International Conference Interfaces and Human Computer Interaction*. 121-128.
- Boone, Thomas R. & Joseph G. Cunningham. 1998. Children's decoding of emotion in expressive body movement: The development of cue attunement. In *Developmental Psychology* 34(5). 1007-1016.
- Breazeal, Cynthia. 1999. Robot in society: Friend or appliance? In *Proceedings of the 1999 Autonomous Agents Workshop on Emotion-Based Agent Architectures*. 18-26.
- Breazeal, Cynthia. 2002. *Designing sociable robots*. Cambridge: MIT Press.
- Breazeal, Cynthia. 2003. Toward sociable robots. In *Robotics and Autonomous Systems* 42. 167-175.
- Breazeal, Cynthia. 2015. *Kismet*. <http://www.ai.mit.edu/projects/sociable/overview.html>. (8 February, 2015.)
- Burgoon, Judee K., Lesa A. Stern & Leesa Dillman. 1995. *Interpersonal adaptation: Dyadic interaction patterns*. Cambridge: Cambridge University Press.
- Cañamero, Lola & Jakob Fredslund. 2001. I show you how I like you: Can you read it in my face? In *IEEE Transactions on Systems, Man and Cybernetics. Part A: Systems and Humans* 31(5). 454-459.
- Cap Digital. 2015. *Organisation*. <http://www.capdigital.com/en/capdigital/organization/>. (12 April, 2015.)
- Cassell, Justine. 1999. Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents. In Cassell, Justine, Joseph Sullivan, Scott Prevost & Elizabeth Churchill (eds.), *Embodied Conversational Agents*, 1-27. Cambridge: MIT Press.
- Charman, Tony, Andrew Pickles, Emily Simonoff, Susie Chandler, Tom Loucas & Gillian Baird. 2011. IQ in children with autism spectrum disorders: Data from the Special Needs and Autism Project (SNAP). In *Psychological Medicine* 41(3). 619-627.
- Cohen, Iris, Rosemarijn Looije, Mark A. Neerinx. 2011. Child's recognition of emotions in robot's face and body. In *Proceedings of the 6th International Conference on Human Robot Interaction [HRI]*. 123-124.

- Csapo, Adam, Emer Gilmartin, Jonathan Grizou, JingGuang Han, Raveesh Meena, Dimitra Anastasiou, Kristiina Jokinen & Graham Wilcock. 2012. Multimodal conversational interaction with a humanoid robot. In *Proceedings of the 3rd International Conference on Cognitive Infocommunications [CogInfoCom]*. 667-672.
- Cuayáhuítl, Heriberto & Ivana Kruijff-Korbayová. 2012. An interactive humanoid robot exhibiting flexible sub-dialogues. In *Proceedings of the NAACL-HLT [North American Chapter of the Association for Computational Linguistics] [Human Language Technologies] 2012*. 17-20.
- Daelemans, Walter. 2013. *Syllabus computerlinguïstiek*. Antwerpen: CLiPS Universiteit Antwerpen.
- Davies, Martha. 1991. *Guide to movement analysis methods*. Unpublished manuscript.
- Davies, Mollie. 2003. *Movement and dance in early childhood*. Cambridge: SAGE Publications Ltd.
- Delaborde, Agnes & Laurence Devillers. 2010. Use of nonverbal speech cues in social interaction between human and robot: Emotional and interactional markers. In *Proceedings of the 3rd International Workshop on Affective Interaction in Natural Environments*. 75-80.
- Diaz, Marta, Neus Nuño, Joan Sàez-Pons, Diego Pardo, Cecilio Angula & Amara Andrés. 2011. Building up child-robot relationship: From initial attraction towards social engagement. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition and Workshops [FG 2011]*. 927-932.
- Digman, John M. 1990. Personality structure: Emergence of the five-factor model. In *Annual Review of Psychology* 41(1). 417-440.
- Dindo, Haris & Daniele Zambuto. 2010. A probabilistic approach to learning a visually grounded language model through human-robot interaction. In *The 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 790-796.
- Ekman, Paul & Wallace V. Friesen. 1969. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. In *Semiotica* 1(1). 49-98.
- European Commission. 2015. *CORDIS*. http://cordis.europa.eu/fp7/home_en.html. (9 April, 2015.)
- Fong, Terrence, Illah Nourbakhsh & Kerstin Dautenhahn. 2003. A survey of socially interactive robots. In *Robotics and Autonomous Systems* 42. 143-166.
- Gillesen, Jan C.C., Emilia Barakova, Bibi E.B.M Huskens & Loe M.G. Feijs. 2011. From training to robot behavior: Towards custom scenarios for robotics in training programs for ASD. In *IEEE International Conference on Rehabilitation Robotics [ICORR] 2011*. 387-393.
- Gillis, Steven, Walter Daelemans & Koenraad De Smedt. 1995. Artificial intelligence. In Östman, Jan-Ola & Jef Verschueren (eds.), *Handbook of pragmatics*. Amsterdam: John Benjamins.
- Goetz, Jennifer & Sara Kiesler. 2002. Cooperation with a robotic assistant. In *Proceedings of CHI 2002*. 578-579.
- Hall, Patrick A.V. & Geoff R. Dowling. 1980. Approximate string matching. In *ACM Computing Surveys* 12(4). 381-402.
- Hanson, David, Andrew Olney, Ismar A. Pereira & Marge Zielke. 2005. Upending the Uncanny Valley. In *Proceedings of the national conference on artificial intelligence* 20(4). 1728-1736.
- Häring, Markus, Nikolaus Bee & Elisabeth André. 2011. Creation and evaluation of emotion expression with body movement, sound and eye color for humanoid robots. In *Ro-Man 2011*. 204-209.

- Hiolle, Antoine, Lola Cañamero, Pierre Andry, Arnaud Blanchard & Philippe Gaussier. 2010. Using the interaction rhythm as a natural reinforcement signal for social robots: A matter of belief. In *Proceedings of the 2nd International Conference on Social Robotics [ICSR]*. 81-89.
- Hofstede, Geert. 2001. *Culture's consequences: Comparing values, behaviors, institutions, and organisations across nations*. Thousand Oaks: Sage.
- Hofstede, Geert & Robert R. McCrae. 2004. Personality and culture revisited: Linking traits and dimensions of culture. In *Cross-Cultural Research* 38(1). 52-88.
- Hofstede Centre. 2015. *Cultural dimensions*. <http://geert-hofstede.com/cultural-dimensions.html>. (11 April, 2015.)
- Invo Labs. 2015. *Pleoworld*. http://www.pleoworld.com/pleo_rb/eng/index.php. (9 April, 2015.)
- Isbister, Katherine & Clifford Nass. 2000. Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics. In *International Journal of Human Computer Studies* 53(2). 251-267.
- Jokinen, Kristiina & Graham Wilcock. 2012. Constructive interaction for talking about interesting topics. In *Proceedings of English International Conference on Language Resources and Evaluation [LREC] 2010*.
- Jurafsky, Daniel & James H. Martin. 2008. *Speech and language processing*, 2nd edition. New Jersey: Pearson.
- Kim, Min-Gyu, Iris Oosterling, Tino Lourens, Wouter Staal, Jan Buitelaar, Jeffrey Glennon, Iris Smeekens & Emilia Barakova. 2014. Designing robot-assisted pivotal response training in game activity for children with autism. In *2014 IEEE International Conference on Systems, Man and Cybernetics*. 1101-1106.
- Kose, Hatice & Rabia Yorganci. 2011. Tale of a robot: Humanoid robot assisted sign language tutoring. In *Proceedings of the 11th IEEE-RAS International Conference on Humanoid Robots*. 105-111.
- Kose, Hatice, Rabia Yorganci, Esra H. Algan & Dag S. Syrdal. 2012. Evaluation of the robot assisted sign language tutoring using video-based studies. In *International Journal of Social Robotics* 4(3). 273-283.
- Kose, Hatice, Nezih Akalin & Pinar Uluer. 2014. Socially interactive robotic platforms as sign language tutors. In *International Journal of Humanoid Robotics* 11(1).
- Kraf, Rogier & Jantine Trapman. 2006. *Syntactic ambiguity*. http://languagelink.let.uu.nl/~lion/?s=Grammar_exercises/grammar_4. (4 February, 2015.)
- Kruijff-Korbayová, Ivana, Georgios Athanasopoulos, Aryel Beck, Piero Cosi, Heriberto Cuayáhuitl, Tomas Dekens, Valentin Enescu, Antoine Hiolle, Bernd Kiefer, Hichem Sahli, Marc Schröder, Giacomo Sommavilla, Fabio Tesser & Werner Verhelst. 2011. An event-based conversational system for the Nao robot. In *Proceedings of the Paralinguistic Information and its Integration in Spoken Dialogue Systems Workshop*. 125-132.
- Kruijff-Korbayová, Ivana, Heriberto Cuayáhuitl, Bernd Kiefer, Marc Schröder, Piero Cosi, Giulio Paci, Giacomo Sommavilla, Fabio Tesser, Hichem Sahli, Georgios Athanasopoulos, Weiyi Wang, Valentin Enescu & Werner Verhelst. 2012. Spoken language processing in a conversational system for child-robot interaction. In *3rd Workshop on Child Computer Interaction [WOCCI]*. 32-39.

- Kruijff-Korbayová, Ivana, Heriberto Cuayáhuitl, Bernd Kiefer, Ilaria Baroni, Alberto Sanna & Marco Nalin. 2013. Children's turn-taking behavior adaptation in multi-session interactions with a humanoid robot. In *Special Issue of the International Journal of Humanoid Robotics*. 1-27.
- Landauer, Thomas K., Peter W. Foltz & Darrell Laham. 1998. Introduction to Latent Semantic Analysis. In *Discourse Processes* 25. 259-284.
- Lee, Kwan Min, Wei Peng, Seung-A Jin & Chang Yan. 2006. Can robots manifest personality? An empirical test of personality recognition, social responses, and social presence in human-robot interaction. In *Journal of Communication* 56(4). 754-772.
- Leuwerink, Kevin. 2012. A robot with personality: Interacting with a group of humans. In *Proceedings of the 16th Twente Student Conference on IT*. 1-10.
- Li, Dingjun, Patrick P.L. Rau & Ye Li. 2010. A cross-cultural study: effect of robot appearance and task. In *International Journal of Social Robotics* 2(2). 175-186.
- Lourens Tino & Emilia Barakova. 2011. User-friendly robot environment for creation of social scenarios. In Ferrández, José Manuel, José Ramón Álvarez Sánchez, Félix De la Paz, F. Javier Toledo (eds.), *Foundations on natural and artificial computation*, 212-221. Berlin Heidelberg: Springer.
- McColl, Derek, Zhe Zhang & Goldie Nejat. 2011. Human body pose interpretation and classification for social human-robot interaction. In *International Journal of Social Robotics* 3(3). 313-332.
- McColl, Derek & Goldie Nejat. 2014. Recognizing emotional body language displayed by a human-like social robot. In *International Journal of Social Robotics* 6(2). 261-280.
- McColl, Derek, Wing-Yue Geoffrey Louie & Goldie Nejat. 2013. Brian 2.1: A socially assistive robot for the elderly and cognitively impaired. In *IEEE Robotics & Automation Magazine* 20(1). 74-83.
- Meena, Raveesh, Kristiina Jokinen & Graham Wilcock. 2012. Integration of gestures and speech in human-robot interaction. In *Proceedings of the 3rd IEEE International Conference on Cognitive Infocommunications [CogInfoCom]*. 673-678.
- Mehrabian, Albert & James A. Russell. 1974. *An approach to environmental psychology*. Cambridge: MIT Press.
- Miskam, Mohd Azfar, Nur Fareeza S. Masnin, Mohd Hazwan Jamhuri, Syamimi Shamsuddin, Abdul Rahman Omar & Hanafiah Yussof. 2014. Encouraging children with autism to improve social and communication skills through the game-based approach. In *International Conference on Robot Pride 2013-2014 [ConfPRIDE] – Medical and Rehabilitation Robotics and Instrumentation*. 93-98.
- Mitchell, Margaret, Dan Bohus & Ece Kamar. 2014. Crowdsourcing language generation templates for dialogue systems. In *Proceedings of the INLG [International Natural Language Generation] and SIGDIAL [Special Interest Group on Discourse and Dialogue] 2014 Joint Session*. 16-24.
- Monceaux, Jérôme, Joffrey Becker, Céline Boudier & Alexandre Mazel. 2009. Demonstration: First steps in emotional expression of the humanoid robot Nao. In *International Conference of Multimodal Interfaces [ICMI] 2009*. 235-236.
- Mori, Masahiro. 2012. The uncanny valley. *Robotics and Automation Magazine* 19(2). 98-100.
- Murray, Ian & John Arnott. 1993. Towards the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. In *Journal of Acoustic Society of America* 93(2). 1097-1108.

- Nalin, Marco, Ilaria Baroni, Alberto Sanna & Clara Pozzi. 2012a. Robotic companion for diabetic children : Emotional and educational support to diabetic children, through an interactive robot. In *Proceedings of the 11th International Conference on Interaction Design and Children [IDC]*. 260-263.
- Nalin, Marco, Ilaria Baroni, Ivana Kruijff-Korbayová, Lola Cañamero, Matthew Lewis, Aryel Beck, Heriberto Cuayáhuitl & Alberto Sanna. 2012b. Children's adaptation in multi-session interaction with a humanoid robot. In *Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication [RO-MAN]*. 351-357.
- Nass, Clifford & Kwan Min Lee. 2001. Does computer-generated speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistence-attraction. In *Journal of Experimental Psychology: Applied* 7. 171-181.
- Petit, Maxime, Stéphane Lallée, Jean-David Boucher, Grégoire Pointeau, Pierrick Cheminade, Dimitri Ognibene, Eris Chinellato, Ugo Pattacini, Ilaria Gori, Uriel Martinez-Hernandez, Hector Barron-Gonzalez, Martin Inderbitzin, Andre Luvizotto, Vicky Vouloutsi, Yannis Demiris, Giorgio Metta & Peter Ford Dominey. 2013. The coordinating role of language in real-time multimodal learning of cooperative tasks. In *Transactions on Autonomous Mental Development* 5(1). 3-17.
- Reeves, Byron & Clifford Nass. 1996. *The Media Equation: How people treat computers, television, and new media like real people and places*. Cambridge: Cambridge University Press.
- Philips. 2014. *Philip's Research – Technologies*.
<http://www.research.philips.com/technologies/projects/robotics/>. (22 April, 2015.)
- Pipitone, Arianna, Vincenzo Cannella, Roberto Pirrone & Antonio Chella. 2014. A cognitive architecture for understanding and producing natural language in support of robotic creativity. In *Humanoid Robots and Creativity 2014: Workshop Humans and Robots Face-to-Face*. 1-4.
- Plasticpals. 2015. *MIT Kismet-15*. http://www.plasticpals.com/?attachment_id=30236. (11 April, 2015.)
- Puterman, Martin L. 2005. *Markov Decision Processes: Discrete stochastic dynamic programming*. Hoboken: Wiley.
- RIOLA. 2015. *Robot Interaction Language*. <http://roila.org/about/>. (29 April, 2015.)
- RoboCup. 2014. *RoboCup*. www.robocup.org. (13 November, 2014.)
- Ros Espinoza, Raquel, Marco Nalin, Rachel Wood, Paul Baxter, Rosemarijn Looije & Yiannis Demiris. 2011. Child-robot interaction in the wild: Advice to the aspiring experimenter. In *Proceedings of the 13th International Conference on Multimodal Interfaces*. 335-342.
- Sabanovic, Selma, Marek P. Michalowski & Reid Simmons. 2006. Robots in the wild: Observing human-robot social interaction outside the lab. In *9th IEEE International Workshop on Advanced Motion Control*. 596-601.
- Shamsuddin, Syamimi, Hanafiah Yussof, Luthffi Ismail, Fazah Akhtar Hanapiah, Salina Mohamed, Hanizah Ali Piah & Nur Ismarrubie Zahari. 2012a. Initial response of autistic children in human-robot interaction therapy with humanoid robot NAO. In *Proceedings of the 2012 IEEE 8th International Colloquium on Signal Processing and its Applications [CSPA]*. 188-193.
- Shamsuddin, Syamimi, Hanafiah Yussof, Luthffi Ismail, Saline Mohamed, Fazah Akhtar Hanapiah & Nur Ismarrubie Zahari. 2012b. Humanoid robot NAO interacting with autistic children of moderately impaired intelligence to augment communication skills. In *International Symposium on Robotics and Intelligent Sensors [IRIS] 2012*. 1533-1538.

- Shukla Kankana & Ben Choi. 2013. Humanoid robot interactions by motion description language and dialogue. In *International Journal of Electrical Energy* 1(2). 86-91.
- Tahon, Marie, Agnes Delaborde & Laurence Devillers. 2011. Real-life emotion detection from speech in human-robot interaction: Experiments across diverse corpora with child and adult voices. In *Interspeech Conference 2011*. 3121-3124.
- Takanishi Laboratory. 2011. *Emotion expression biped humanoid robot Kobian-RII*. <http://www.takanishi.mech.waseda.ac.jp/top/research/kobian/KOBIAN-R/index.htm>. (11 April, 2015.)
- Tapus, Adriana & Maja J. Matarić. 2008. Socially assistive robots: The link between personality, empathy, physiological signals, and task performance. In *Proceedings of the AAAI Spring Symposium on Emotion, Personality and Social Behavior 2008*. 133-140.
- Tapus, Adriana, Andreea Peca, Amir Aly, Christina Pop, Lavinia Jisa, Sebastian Pintea, Alina S. Rusu & Daniel O. David. 2012. Children with autism social engagement in interaction with Nao, an imitative robot. In *Interaction Studies* 13(3). 315-347.
- Tonks, James, Huw W. Williams, Ian Frampton, Phil Yates & Alan Slater. 2007. Assessing emotion recognition in 9-15-years olds: Preliminary analysis of abilities in reading emotion from faces, voices and eyes. In *Brain Injury* 21(6). 623-629.
- Trovato, Gabriele, Tatsuhiro Kishi, Nobutsuna Endo, Kenji Hashimoto & Atsuo Takanishi. 2012. Evaluation study on asymmetrical facial expressions generation for humanoid robots. In *1st International Conference on Innovative Engineering Systems [ICIES]*. 129-134.
- Waalewijn, Dennis. 2012. Neighboring cultures, what difference do they make in the perception of socially interactive robots? In *Proceedings of the 16th Twente Student Conference on IT*. 1-6.
- Wilcock, Graham. 2012. WikiTalk: A spoken Wikipedia-based open-domain knowledge access system. In *Proceedings of the Workshop on Question Answering for Complex Domains, Coling [Computational Linguistics] 2012*. 57-70.
- Windhouwer, Dennis. 2012. The effects of the task context on the perceived personality of a Nao robot. In *Proceedings of the 16th Twente Student Conference on IT*. 1-7.
- Woods, Sarah, Kerstin Dautenhahn & Joerg Schultz. 2005. Child and adults' perspectives on robot appearance. In *AISB [Artificial Intelligence and Simulation of Behaviour] Symposium on robot companion 2005*. 126-132.
- Zecca, Massimiliano, Yu Mizoguchi, Keita Endo, Fumiya Lida, Yousuke Kawabata, Nobutsuna Endo, Kazuko Itoh & Atsuo Takanishi. 2009. Whole body emotion expressions for KOBIAN Humanoid robot: Preliminary experiments with different emotional patterns. In *Ro-Man 2009*. 381-386.
- Zeng, Zhiping, Anthony K.H. Tung, Jianyong Wang, Jianhua Feng & Lizhu Zhou. 2009. Comparing stars: On approximating graph edit distance. In *Proceedings of the VLDB [Very Large Data Base] Endowment* 2(1). 25-36.
- Zhang, Li, Ming Jiang, Dewan Farid & Alamgir Hossain. 2013. Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot. In *Expert Systems with Applications* 40(13). 5160-5168.
- Zheng, Minhua & Max Q.H. Meng. 2012. Designing gestures with semantic meanings for humanoid robot. In *Proceedings of the 2012 IEEE International Conference on Robotics and Biomimetics*. 287-292.

11. List of Figures

FIGURE 1 PHYSICAL STRUCTURES AND PROCESSES REPRESENT MENTAL FUNCTIONS (GILLIS, DAELEMANS & DE SMEDT, 1995).....	13
FIGURE 2 THE UNCANNY VALLEY (MORI, 2012).....	15
FIGURE 3 THE ALDEBARAN ROBOTIC FAMILY (ALDEBARAN ROBOTICS, 2015)	18
FIGURE 4 NAO EVOLUTION (ALDEBARAN ROBOTICS, 2015).....	19
FIGURE 5 DIAGRAM OF SYSTEM (DINDO & ZAMBUTO, 2010).....	23
FIGURE 6 NAO FOLLOWING INSTRUCTIONS (DINDO & ZAMBUTO, 2010)	23
FIGURE 7 COGNITIVE AND META-COGNITIVE MDPs (PIPITONE ET AL., 2014)	26
FIGURE 8 A META-COGNITIVE ARCHITECTURE BASED ON TWO MDPs (PIPITONE ET AL., 2014)	27
FIGURE 9 CHILDREN PLAYING WITH NAO. LEFT TO RIGHT: QUIZ, DANCING, IMITATION GAME (KRUIJFF-KORBAYOVÁ ET AL., 2013).	29
FIGURE 10 FORCED WAITING DURING TURN-TAKING OVER 3 SESSIONS (KRUIJFF-KORBAYOVÁ ET AL., 2013).....	30
FIGURE 11 IGNORED SPEECH ACTS OVER 3 SESSIONS (KRUIJFF-KORBAYOVÁ ET AL., 2013).....	30
FIGURE 12 NAO'S TURN-TAKING GESTURES. A. HEAD NOD UP; B. LISTENING KEY POSE; C. SPEAKING KEY POSE (MEENA ET AL., 2012)	30
FIGURE 13 KISMET (PLASTICPALS, 2015)	31
FIGURE 14 PIPELINE FOR CROWD-BASED DEVELOPMENT OF NATURAL LANGUAGE GENERATION TEMPLATES (MITCHELL ET AL., 2014)	33
FIGURE 15 CLOSED-DOMAIN DIALOGUE FLIGHT RESERVATION SYSTEM (WILCOCK, 2012)	34
FIGURE 16 OPEN-DIALOGUE WIKITALK SYSTEM (WILCOCK, 2012).....	35
FIGURE 17 OPEN-DIALOGUE WIKITALK SYSTEM WITH SMOOTH TOPIC-SHIFTS (WILCOCK, 2012).....	36
FIGURE 18 STRICT (A) VERSUS FLEXIBLE (B) HDC (CUAYÁHUITL & KRUIJFF-KORBAYOVÁ, 2012)	37
FIGURE 19 OPERATIONS ON STACK-BASED HDC (CUAYÁHUITL & KRUIJFF-KORBAYOVÁ, 2012)	37
FIGURE 20 BASSIS ARCHITECTURE (PETIT ET AL., 2013)	39
FIGURE 21 KEY POSES FOR NAO; A. WAVE WITH FOREARM; B. BECKON WITH PALM UP; C. CLAP (ZHENG & MENG, 2012)	42
FIGURE 22 GRAPHICAL REPRESENTATION OF THE EXECUTION ORDER (LOURENS & BARAKOVA, 2011)	45
FIGURE 23 NAO INTERACTIONS. A. GREETING; B. SITTING DOWN; C. APPLE RECOGNISING; D. PETTING AIBO; E. PLAYING WITH PLEO DINOSAUR; F. REACHING (SHUKLA & CHOI, 2013)	46
FIGURE 24 STATIC POSES EXPRESSING EMOTIONS. A: ANGER. B: SADNESS. C: FEAR. D: PRIDE. E: HAPPINESS. F: EXCITEMENT. (BECK ET AL., 2010A).....	52
FIGURE 25 RESULTING SYSTEM BASED ON 2 DIMENSIONS: AROUSAL & VALENCE (BECK ET AL., 2010B).....	53
FIGURE 26 FIVE GENERATED KEY POSES: A: 100% SADNESS; B: 70% SADNESS & 30% FEAR; C: 50% SADNESS & 50% FEAR; D: 30% SADNESS & 70% FEAR; E: 100% FEAR (BECK ET AL., 2010B)	54
FIGURE 27 COMBINED EMOTIONAL EXPRESSIONS: ANGER 1, ANGER 2, FEAR 1, FEAR 2, JOY 1, JOY 2, SADNESS 1, SADNESS 2 (HÄRING ET AL., 2011).....	56
FIGURE 28 RESULTS FOR THE PLEASURE – AROUSAL – DOMINANCE (PAD) MEASUREMENTS (HÄRING ET AL., 2011)	58
FIGURE 29 KISMET'S ARCHITECTURE (BREAZEL, 1999)	60
FIGURE 30 BRIAN 2.0 (ASB LAB, CONSULTED: 4/03/2015)	61
FIGURE 31 KOBIAN (TAKANISHI LABORATORY, 2011).....	64
FIGURE 32 SYSTEM ARCHITECTURE (ZHANG ET AL., 2013)	67
FIGURE 33 MULTI-LEVEL DETECTION OF EMOTIONAL AND INTERACTIONAL CUES (DELABORDE & DEVILLERS, 2010)	69
FIGURE 34 COMPARISON BETWEEN RESULTS OF DNSS SPECIALIST AND THE AUTOMATED SYSTEM (MCCOLL ET AL., 2011)	71
FIGURE 35 SYSTEM ARCHITECTURE (ALY & TAPUS, 2013)	75
FIGURE 36 BEAT ARCHITECTURE (ALY & TAPUS, 2013).....	76
FIGURE 37 INTROVERTED AND EXTRAVERTED CONDITIONS (ALY & TAPUS, 2013).....	76
FIGURE 38 PERCEIVED INTELLIGENCE OF BOTH ROBOTS (LEUWERINK, 2012)	80
FIGURE 39 PERCEIVED ENJOYABILITY OF BOTH ROBOTS (LEUWERINK, 2012)	80
FIGURE 40 PERCEIVED FUN OF BOTH ROBOTS (LEUWERINK, 2012).....	81
FIGURE 41 AVERAGE TIME WANTED TO SPEND WITH PROGRAMMED PERSONALITY (WAALEWIJN, 2012)	83
FIGURE 42 AVERAGE TIME WANTED TO SPEND WITH PERCEIVED PERSONALITY (WAALEWIJN, 2012).....	83
FIGURE 43 ROBOTS USED IN THE EXPERIMENT. A. NAO; B. AIBO; C. PLEO; D. SPYKEE (DIAZ ET AL., 2011)	86
FIGURE 44 NAO DURING GAME III: A. HUNGER; B. HAPPINESS; C. ANGER; D. FEAR; E. HUG; F. KISS (MISKAM ET AL., 2014)	91

FIGURE 45 COMPARISON OF AUTISTIC BEHAVIOUR IN THE CLASSROOM AND DURING HRI (SHAMSUDDIN ET AL., 2012B)..... 92

FIGURE 46 CONNECTION BETWEEN AUTISTIC BEHAVIOUR AND IQ (SHAMSUDDIN ET AL., 2012B) 92

FIGURE 47 PROCESS FLOW TO TRANSFORM SCENARIOS INTO TiVIPE NETWORKS (GILLESEN ET AL., 2011) 94

FIGURE 48 NAO IN THE ALIZ-E PROJECT (ALIZ-E, 2014C)..... 97

FIGURE 49 ROBIN'S PLAYROOM (ALIZ-E, 2014B) 100

FIGURE 50 ROBIN'S ARCHITECTURE (ALIZ-E, 2014B)..... 100

FIGURE 51 AVERAGE PROBLEMS/MINUTE PER SESSION (NALIN ET AL., 2012B) 102

FIGURE 52 ROBOVIE R3 ROBOT (KOSE ET AL., 2014)..... 104

12. List of Tables

TABLE 1 DIFFERENCES THAT LEAD TO DIFFICULTIES FOR NLP (BASED ON: BEARDON ET AL., 1991)	14
TABLE 2 DIFFERENT KINDS OF KNOWLEDGE NEEDED FOR CONVERSATIONAL AGENTS (BASED ON: JURAFSKY & MARTIN, 2008)	14
TABLE 3 NAO EVOLUTION SPECIFICATIONS (BASED ON: ALDEBARAN ROBOTICS, 2015)	20
TABLE 4 GESTURES FOR TURN-TAKING (MEENA ET AL., 2012)	30
TABLE 5 KISMET'S ENVELOPE DISPLAYS (BREAZEAL, 2003)	32
TABLE 6 NON-VERBAL CATEGORIES (EKMAN & FRIESEN, 1969).....	42
TABLE 7 PERCEPTION OF GESTURES PERFORMED BY HUMAN (ZHENG & MENG, 2012)	43
TABLE 8 PERCEPTION OF GESTURES PERFORMED BY NAO (ZHENG & MENG, 2012)	44
TABLE 9 MOTION DESCRIPTION LAYERS (SHUKLA & CHOI, 2013).....	45
TABLE 10 RECOGNITION RATES (BASED ON: BECK ET AL., 2010A).....	52
TABLE 11 COMBINED EMOTIONAL EXPRESSIONS (BASED ON: HÄRING ET AL., 2011).....	57
TABLE 12 RECOGNITION RATES (BASED ON: HÄRING ET AL., 2011).....	57
TABLE 13 RECOGNITION RATES OF NAO AND ICAT (BASED ON: COHEN ET AL., 2011)	59
TABLE 14 BODY LANGUAGE DESCRIPTORS FOR POSITIVE AND NEUTRAL EMOTIONS (BASED ON: MCCOLL & NEJAT, 2014)	62
TABLE 15 BODY LANGUAGE DESCRIPTORS FOR NEGATIVE EMOTIONS (BASED ON: MCCOLL & NEJAT, 2014)	62
TABLE 16 RECOGNITION RATES FOR EMOTIONS DISPLAYED BY BRIAN 2.0 (BASED ON: MCCOLL & NEJAT, 2014)	62
TABLE 17 RECOGNITION RATES FOR EMOTIONS DISPLAYED BY THE ACTOR (BASED ON: MCCOLL & NEJAT, 2014).....	63
TABLE 18 RECOGNITION RATES FIRST EXPERIMENT (BASED ON: ZECCA ET AL., 2009)	64
TABLE 19 RECOGNITION RATES 2ND EXPERIMENT (BASED ON: ZECCA ET AL., 2009).....	65
TABLE 20 SYMMETRICAL - ASYMMETRICAL PREFERENCE (BASED ON: TROVATO ET AL., 2012).....	66
TABLE 21 RECOGNITION RATES (BASED ON: ZHANG ET AL., 2013)	68
TABLE 22 PERFORMANCE IN CROSS-CORPUS CLASSIFICATION (BASED ON: TAHON ET AL., 2011)	69
TABLE 23 ACCESSIBILITY LEVELS. T: TOWARD; A: AWAY; N: NEUTRAL (MCCOLL ET AL., 2011).....	70
TABLE 24 ASSOCIATION OF THE RESULTS OF THE SELF-STUDY AND THE RESULTS OF THE DNSS SYSTEM (MCCOLL ET AL., 2011)	71
TABLE 25 PERSONALITY TYPES FOR SOCIAL ROBOTS (BASED ON: FONG ET AL., 2003)	75
TABLE 26 PERCEPTION OF INTROVERTED AND EXTRAVERTED NAO (BASED ON: WINDHOUWER, 2012)	78
TABLE 27 CULTURAL DIMENSION RATINGS OF GERMANY AND THE NETHERLANDS (BASED ON: HOFSTEDE, 2001)	81
TABLE 28 AVERAGE TIME WANTED TO SPEND WITH THE ROBOT (WAALEWIJN, 2012)	82
TABLE 29 PERCEPTIONS AND EXPECTATIONS OF THE CHILDREN CONCERNING NAO AND PLEO (DIAZ ET AL., 2011)	88
TABLE 30 INTERACTIVE BEHAVIOUR OF THE CHILDREN DURING EXPERIMENT II (DIAZ ET AL., 2011)	88
TABLE 31 GAME OBJECTIVES AND TARGET SKILLS (MISKAM ET AL., 2014)	91
TABLE 32 MODULES OF THE EXPERIMENT (BASED ON: SHAMSUDDIN ET AL., 2012A)	92
TABLE 33 REQUIREMENTS FOR ROBOTIC COMPANIONS (NALIN ET AL., 2012A).....	98
TABLE 34 RECOGNITION RATES OF NAO AND ROBOVIE R3 (KOSE ET AL., 2014).....	105

Appendix I: NAO versions

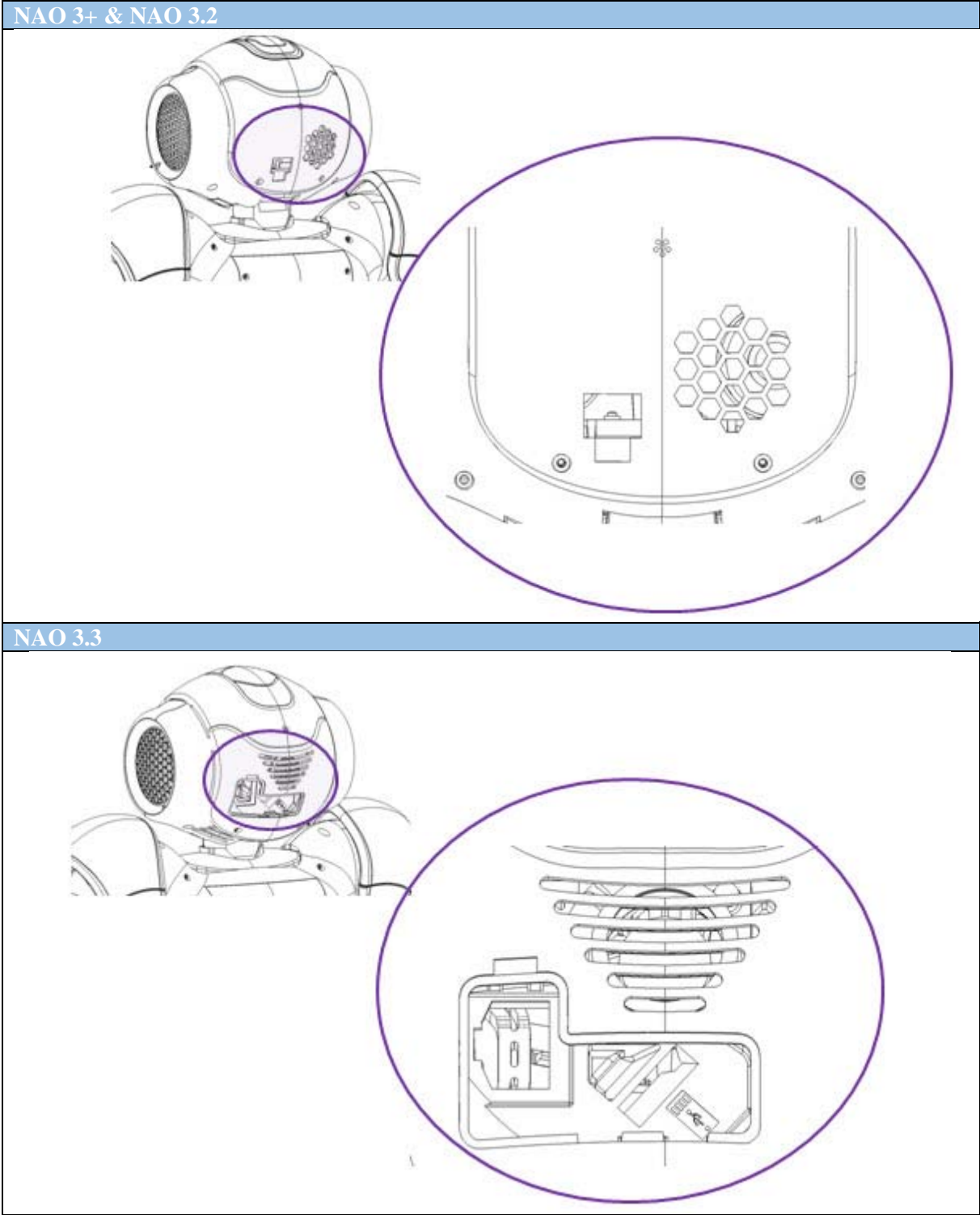
Prototypes	
January 2005 – March 2006	AL-01, AL-02, AL-03
September 2005 – July 2006	AL-04
June 2006 – June 2007	AL-05a
May 2007 – December 2007	AL-06b
Releases	
2008	NAO RoboCup Edition (V2)
2009	NAO V3
2009	NAO V3+
2010	NAO V3.2
2010	NAO V3.3
2011	NAO Next Gen (V4)
2014	NAO Evolution (V5)

Based on: www.aldebaran.com (consulted 05/02/2015)

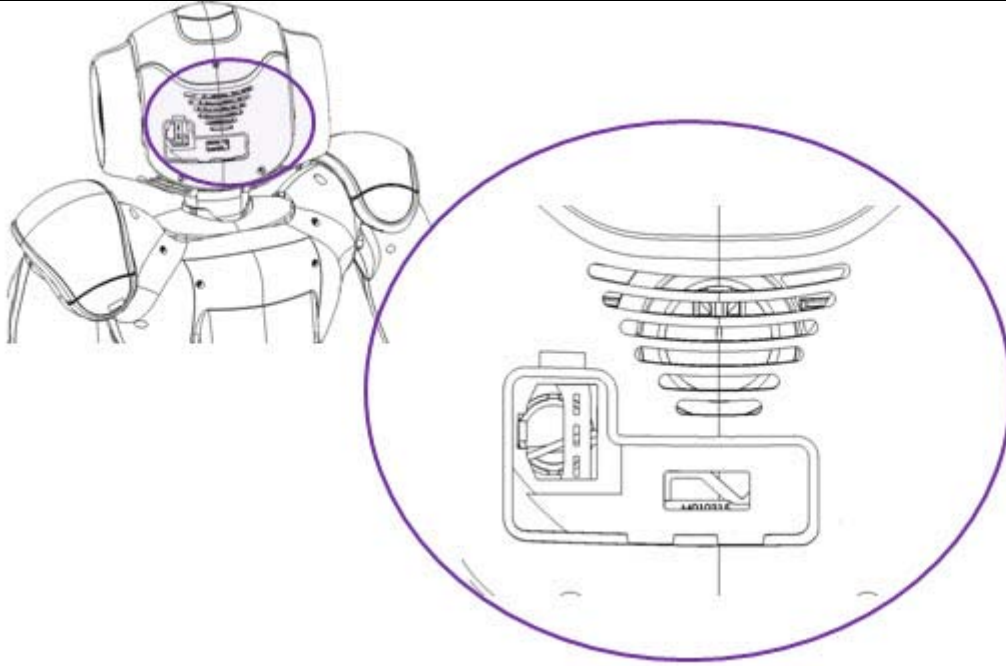
Appendix II: NAO version and body types diagrams

1. Versions

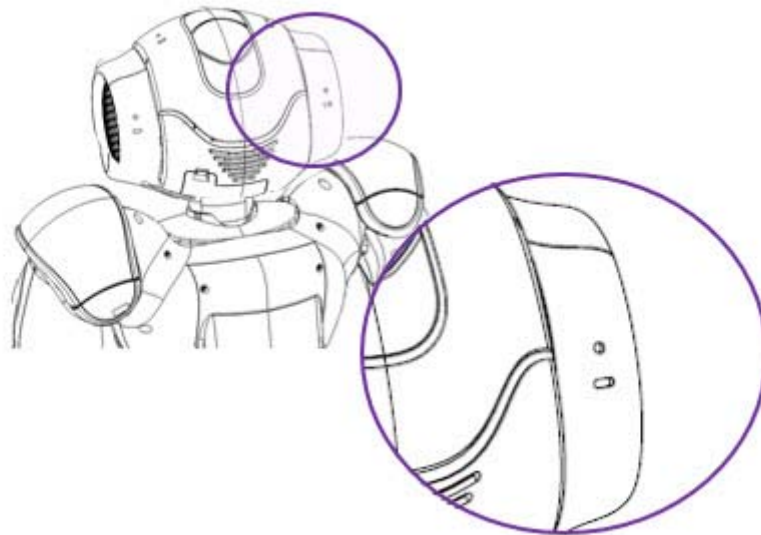
The design of the back of the head is used to differentiate between the NAO releases (Aldebaran Robotics, 2015).



NAO Next Gen (V4)



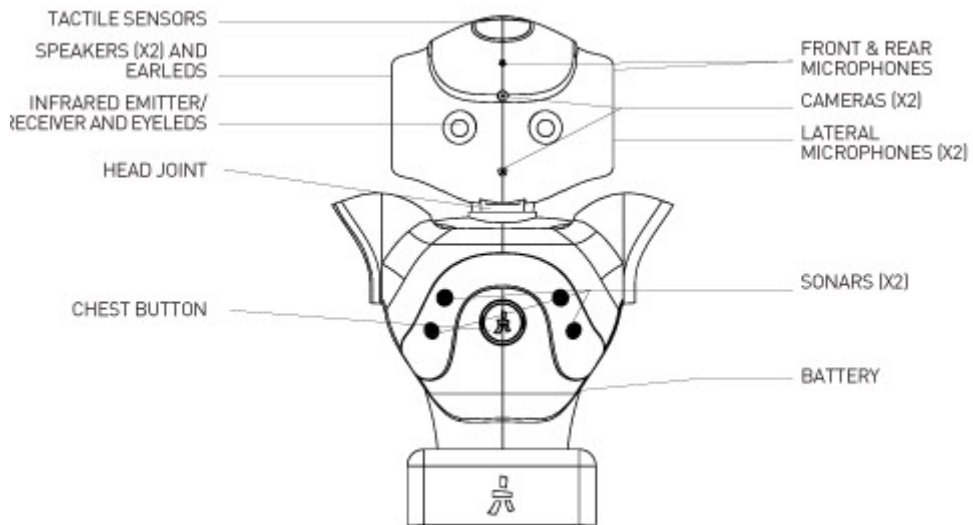
NAO Evolution (V5)



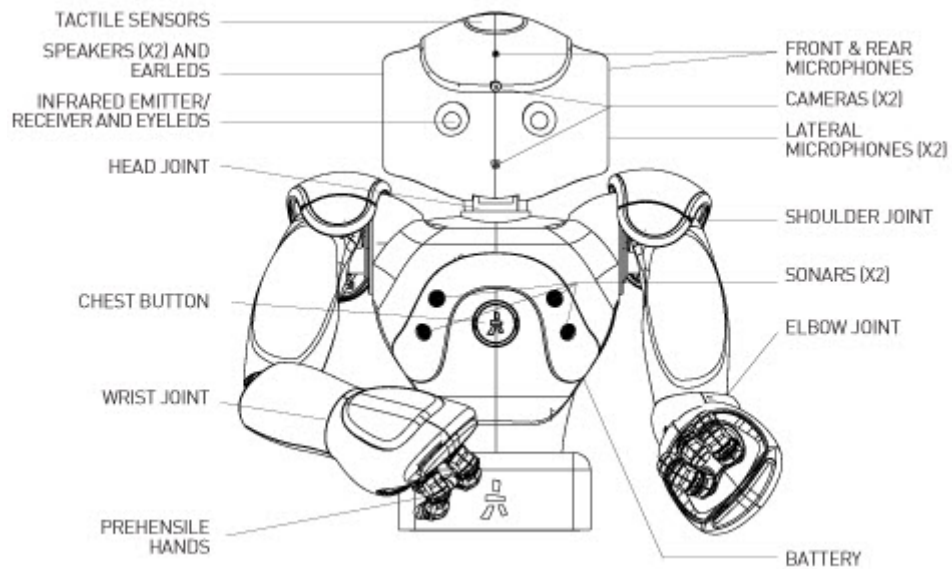
Based on: http://doc.aldebaran.com/2-1/family/body_type.html (Consulted 05/02/2015)

2. Body Types

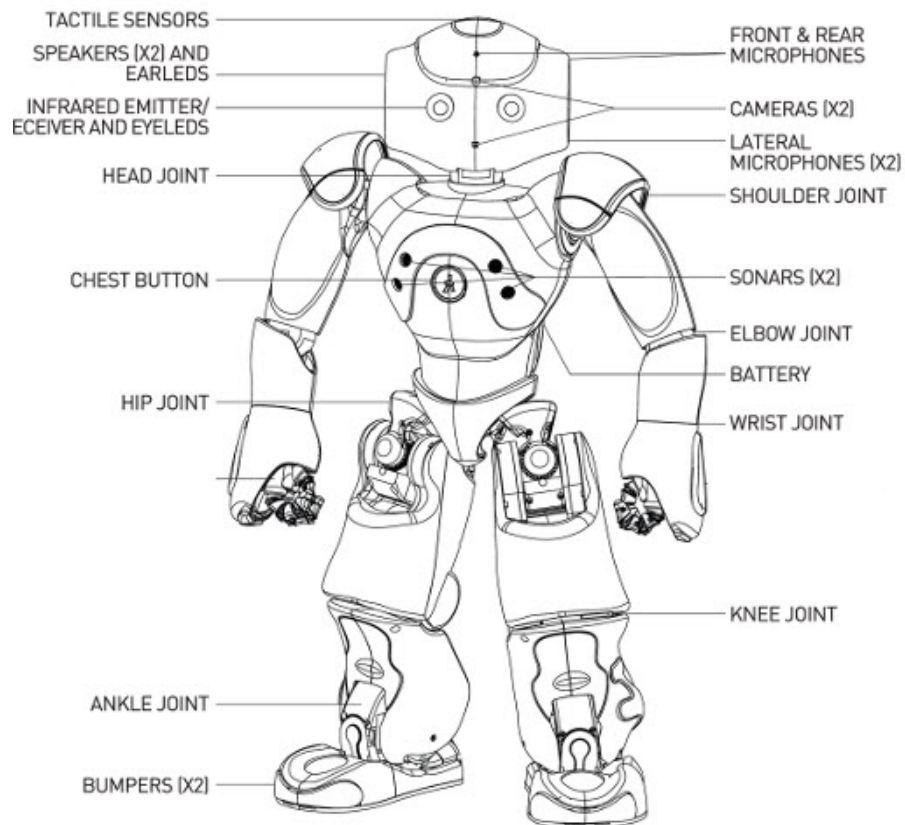
NAO T2



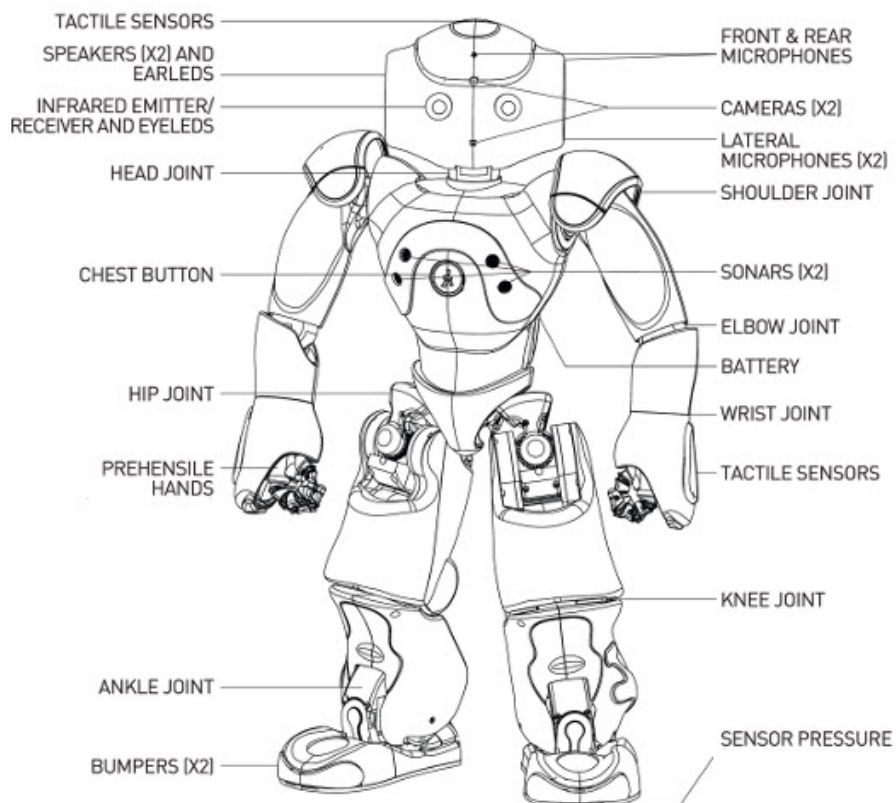
NAO T14



NAO H21



NAO H25



Based on: http://doc.aldebaran.com/2-1/family/body_type.html (Consulted 05/02/2015)

Appendix III: NAO Evolution Datasheet



ELECTRICAL

Input	100 to 240 Vac – 50/60Hz – Max 1.2A	
Output	Max 1.2A 25.2 Vdc – 2A	
Battery	Type	Lithium-Ion
	Nominal voltage/capacity	21.6V / 2.25
	Max charge voltage	25.2V
	Recommended charge current	1.8A
	Max charge / discharge current	2.3A / 2.0A
	Energy	48.6Wh
	Charging duration	5h
	Autonomy	60min (Active use) 90 min (Normal use)

MOTHER BOARD

CPU	CPU processor	ATOM Z530
	Cache memory	512KB
	Clock speed	1.6GHZ
	FSB speed	533mHz
RAM	1GB	
Flash memory	2GB	
Micro SDHC	8GB	

CONNECTION

Ethernet	1xRJ45 – 10/100/1000 BASE T
WIFI	IEEE 802.11b/g/n

AUDIO

Loud Speakers	x2 lateral	
	Diameter	36mm
	Impedance	8ohms
	Sp level	87dB/w +/- 3dB
	Freq range	up to ~20kHz
	Input	2W
Microphone	x4 on the head	
	Sensitivity	43 db +/- 3dB
	Frequency range	150Hz-12kHz
	Signal/noise ratio	58dBA

CONSTRUCTION

Dimension (HxDxW)	574x275x311 mm / 22.6x10.8x12.2 inch
Weight	5.4kg / 11.9 lb
Construction material	ABS-PC/PA-66/XCF-30

LANGUAGES

Text to speech & Automatic speech Recognition	Czech, Danish, Dutch, English, Finnish, French, German, Italian, Japanese, Korean, Polish, Portuguese, Spanish, Swedish, Russian, Turkish
---	---

VISION

Cameras	x2 on front	
Sensor model	MT9M114	
Sensor type	SOC Image Sensor	
Imaging array	Resolution	1.22MP
	Optical format	1/6inch
	Active Pixels (HxV)	1288x968
	Pixel size	1.9µm
Sensitivity	Dynamic range	70dB
	Signal/Noise ratio (max)	37dB
	Responsivity	2.24 V/lux-sec (960p)
		8.96 V/lux-sec (VGA)
Output	Camera output	960p@30fps
	Data Format	YUV422
	Shutter type	ERS (Electronic Rolling Shutter)
View	Field of view	72.6° DFOV
		(60.9° HFOV, 47.6° VFOV)
	Focus range	30cm – infinity
	Focus type	Fixed focus

Framerate

Resolution	Embedded	Gigabit Ethernet	100Mb Ethernet	Wifi g
160x120px	30fps	30fps	30fps	30fps
320x240px	30fps	30fps	30fps	11fps
640x480px	30fps	30fps	12fps	2.5fps
1280x960px	29fps	10fps	3fps	0.5fps

Note: using the video stream in remote highly depends on the network and the video resolution chosen. All frame rates depend on the CPU usage. Values are calculated with a CPU fully dedicated to images gathering.

Make interactive robots for the well being of humans





IR

Number	×2 on front
Wavelength	940nm
Emission Angle	+/-60°
Power	8mW/sr

FRS (FORCE SENSITIVE RESISTORS)

Range	0 to 110N
	×4 per feet

POSITION SENSORS

MRE (Magnetic Rotary Encoder)	×36 Using hall effect sensor technology Precision: 12bits / 0.1°
-------------------------------	--

SOFTWARE

Open Nao	Embedded GNU/Linux Distribution based on Gentoo
Architecture	×86
Programming	Embedded: C++ / Python Remote: C++ / Python / .NET / Java / MatLab

CONTACT SENSOR

Chest Button	✓
Foot Bumper	✓
Tactile Head	✓
Tactile Hand	✓

SONAR

Emitters	×2 on front
Receivers	×2 on front
Frequency	40kHz
Sensitivity	-86dB
Resolution	1cm
Detection Range	0.05m to 3m
Effective Cone	60°

INERTIAL UNIT

Gyrometer	×1	
	Axis	3
	Precision	5%
	Angular speed	-500°/s
Accelerometer	×1	
	Axis	3
	Precision	1%
	Acceleration	-2g

LEDS

Placement	Quantity	Description
Tactile Head	×12	16 Blue levels
Eyes	2×8	RGB FullColor
Ears	2×10	16 Blue levels
Chest button	×1	RGB FullColor
Feet	2×1	RGB FullColor

CONTACT SENSOR

Head	×2 dof
Arm (in each)	×5 dof
Pelvis	×1 dof
Leg (in each)	×5 dof
Hand (in each)	×1 dof

Make interactive robots for the well being of humans



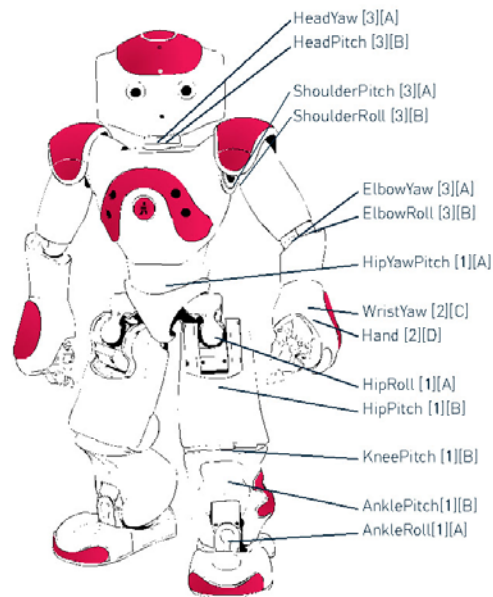


MOTOR SPECIFICATIONS

Motor type Brush DC Coreless

POSITION OF MOTORS

		Motor	Reduction Ratio
Head joints	HeadYaw	Type 3	Type A
	HeadPitch	Type 3	Type B
Arm joints	ShoulderPitch	Type 3	Type A
	ShoulderRoll	Type 3	Type B
	ElbowYaw	Type 3	Type A
	ElbowRoll	Type 3	Type B
	WristYaw	Type 2	Type C
Leg joints	Hand	Type 2	Type D
	HipYawPitch	Type 1	Type A
	HipRoll	Type 1	Type A
	HipPitch	Type 1	Type B
	KneePitch	Type 1	Type B
	AnklePitch	Type 1	Type B
	AnkleRoll	Type 1	Type A



Legend:
Joint Name[Motor Type][Reductor Type]

DESCRIPTION OF THE MOTORS

	Motor type 1	Motor type 2	Motor type 3
Model	22NT82213P	17N88208E	16GT83210E
No load speed	8300rpm ±10%	8400rpm ±12%	10700rpm ±10%
Stall torque	68mNm ±8%	9.4mNm ±8%	14.3mNm ±8%
Continuous torque	16.1mNm max	4.9mNm max	6.2mNm max

Speed Reduction Ratio TYPE A

	Motor type 1	Motor type 3
Reduction ratio	201.3	150.27

Speed Reduction Ratio TYPE C

	Motor type 2
Reduction ratio	50.61

Speed Reduction Ratio TYPE B

	Motor type 1	Motor type 3
Reduction ratio	130.85	173.22

Speed Reduction Ratio TYPE D

	Motor type 2
Reduction ratio	36.24

CERTIFICATIONS & APPROVALS

Region	Classification	Electromagnetic compatibility	EN 301-1 / EN 301 489-17 / EN 300 328 EN 62311 : 2008 / FCC PART15, Class B
Europe	CE (Attestation of conformity)		
USA	FCC	Safety	IEC 60950-1 : 2005 (2nd edition)

www.RobotsLAB.com

75 Broadway St. STE 202 San Francisco, CA 94111 USA Tel. +1-415-702-3033



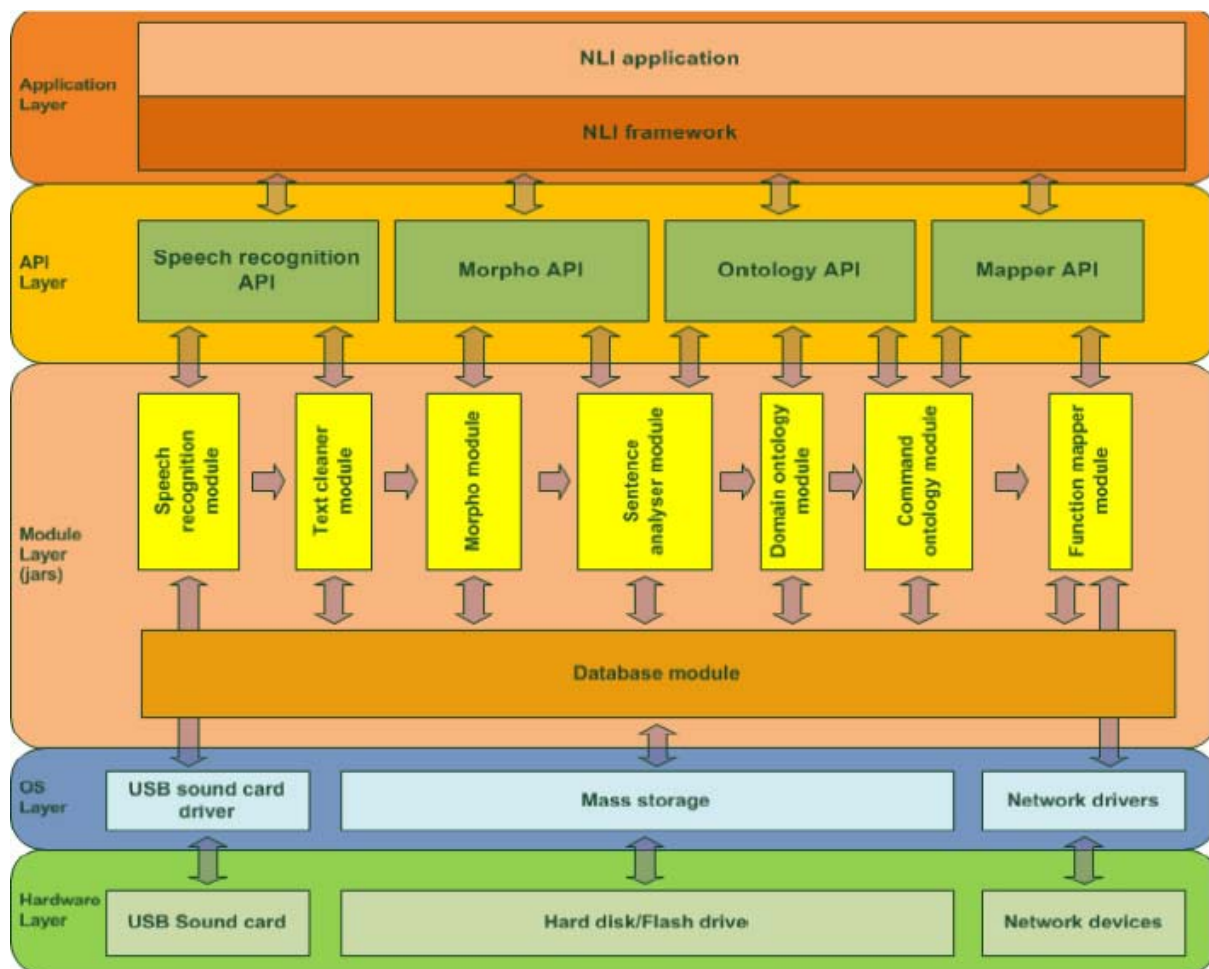
Make interactive robots for the well being of humans



Source: <http://cdn2.hubspot.net/hub/314265/file-1072832042-pdf/PDFs/NAO/NAO-Evolution-Datasheet.pdf?t=1422465793478> (Consulted 05/02/2015)

Appendix IV: A frame-based Dialogue System

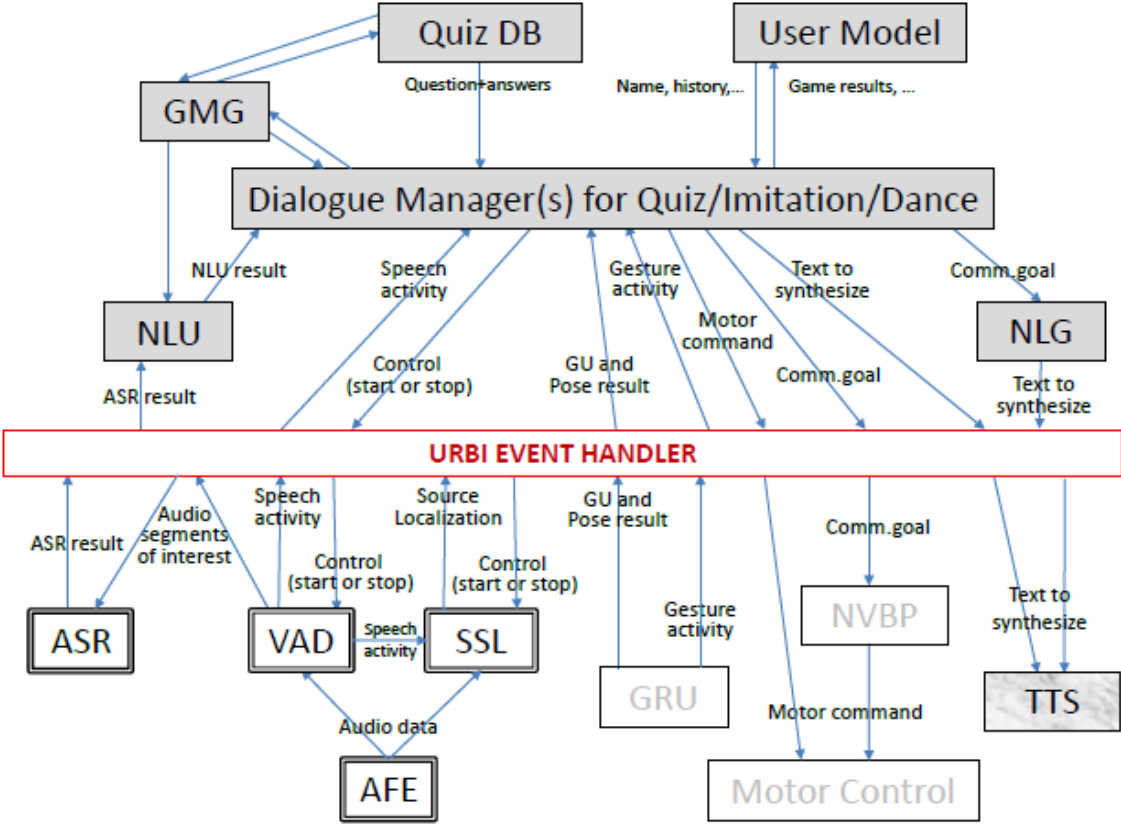
In chapter 4.2.1, a frame-based dialogue system was discussed, which used the following architecture for its NLP engine:



Source: Barabás et al., 2012.

Appendix V: An Event-based Dialogue System

In chapter 4.2.2, an event-based dialogue system was presented, based on the following architecture:



Source: Kruijff-Korbyová, 2012.

Appendix VI: WikiTalk-based Open-dialogue with NAO

State: Hello

- **NAO:** [Stands up]

State: Select New Topic

- **NAO:** [Tells the user that his favourite topics include Shakespeare, Metz and Paul Verlaine]
- **Human:** [Asks the robot to talk about Metz]

State: Start New Topic

- **NAO:** *Metz is a city in the northeast of France located at the confluence of the Moselle and the Seille rivers. Metz is the capital and the prefecture of both the Lorraine region and the Moselle department. Located near the tripoint along the junction of France, Germany, and Luxembourg, Metz forms a central place of the European Greater Region and the SaarLorLux euroregion.*

New Info Links:

[France, Germany, Luxembourg, Moselle (river), Seille, prefecture, Lorraine (region), Moselle, department, tripoint, Greater Region, SaarLorLux, Euroregion]

State: Continue Topic

- **Human:** *Continue.*
- **NAO:** *A Celtic oppidum, an important Gallo-Roman city, the Merovingian capital of the Austrasia kingdom, the birthplace of the Carolingian dynasty, a cradle of the Gregorian chant, and one of the oldest republics of the common era in Europe, Metz has a rich 3.000-year-history. The city has been steeped in Romance culture, but has been strongly influenced by Germanic culture due to its location and history.*

New Info Links:

[France, Germany, Luxembourg, Moselle (river), Seille, Prefecture, Lorraine (region), Moselle, Department, Tripoint, Greater Region, SaarLorLux, Euroregion, Celts, Oppidum, Gallo-Roman, Merovingian, Austrasia, Carolingian dynasty, Gregorian chant, Common era, Romance languages, Culture of Germany]

State: Start New Topic (Smooth Topic-shift)

- **Human:** *Celtic?*
- **NAO:** *The Celts or Kelts were an ethno-linguistic group of tribal societies in Iron Age and Medieval Europe who spoke Celtic language and had a similar culture.*

New Info Links:

[Ethno-linguistics, Tribe, Iron Age, Middle Ages, Europe, Celtic languages]

State: Start New Topic (Smooth Topic-shift)

- **Human:** *Celtic languages?*
- **NAO:** *The Celtic or Keltic languages are descended from proto-Celtic, or “Common Celtic”, a branch of the greater Indo-European language family. The term “Celtic” was first used to describe this language group by Edward Lhuyd in 1707.*

New Info Links:

[Proto-Celtic, Indo-European languages, Language family, Edward Lhuyd]

State: Start New Topic (Smooth Topic-shift)

- **Human:** *Indo-European?*
- **NAO:** *The Indo-European languages are a family of several hundred related languages and dialects. It has about 449 languages and dialects, according to the 2005 Ethnologue estimate, about half belonging to the Indo-Aryan sub-branch. It includes most major current languages of Europe, the Iranian plateau, and South Asia, and was also predominant in ancient Anatolia. With*

written attestations appearing since the Bronze Age, in the form of the Anatolian languages and Mycenaean Greek, the Indo-European family is significant to the field of historical linguistics as possessing the longest recording history after the Afroasiatic family.

- **Human:** *Thank you.*

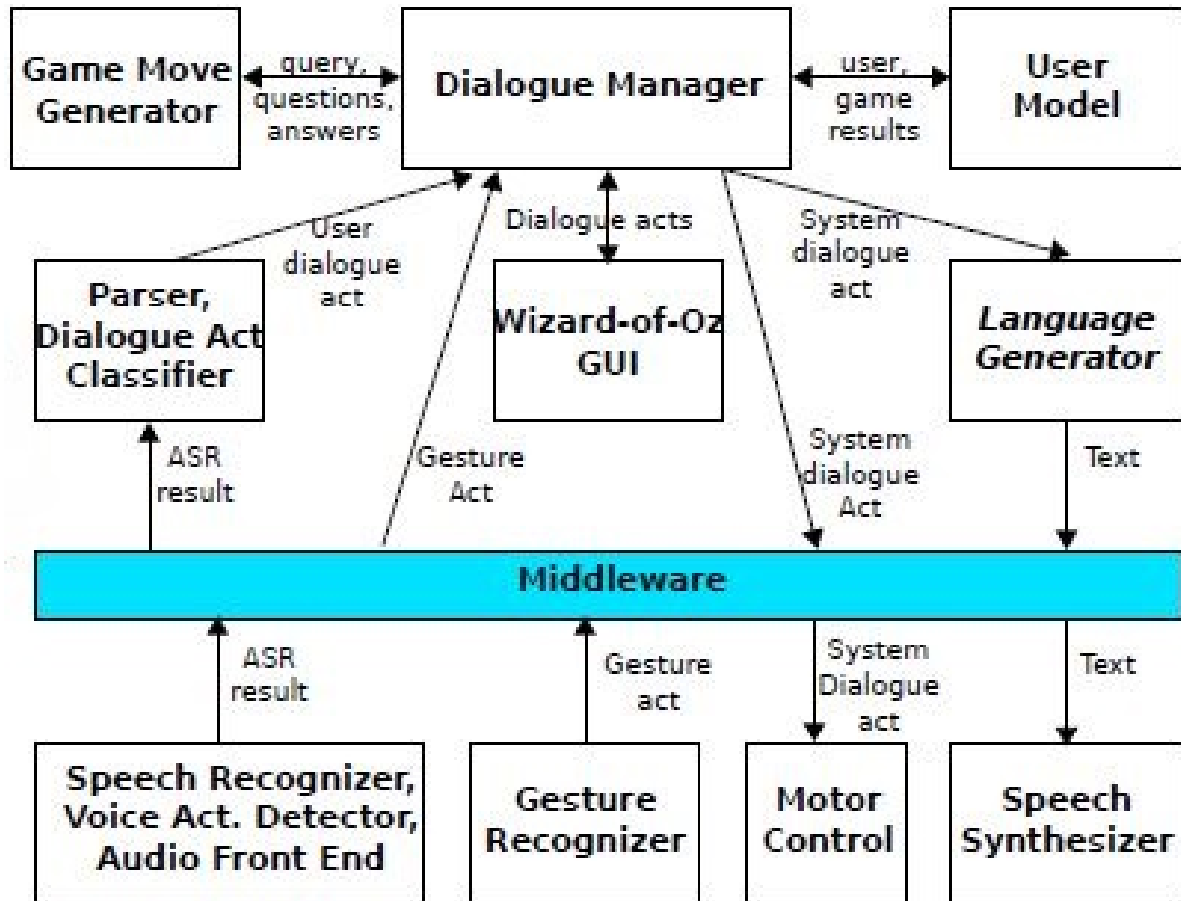
State: Goodbye

- **NAO:** *Thank you. Goodbye. [Sits down]*

Based on: Jokinen & Wilcock, 2012 (Quoted in: Mitchell et al., 2012).

Appendix VII: High-level Architecture of a HDC System

In chapter 4.6, Hierarchical Dialogue Control was discussed. The following is the high-level architecture of a flexible HDC system for NAO:

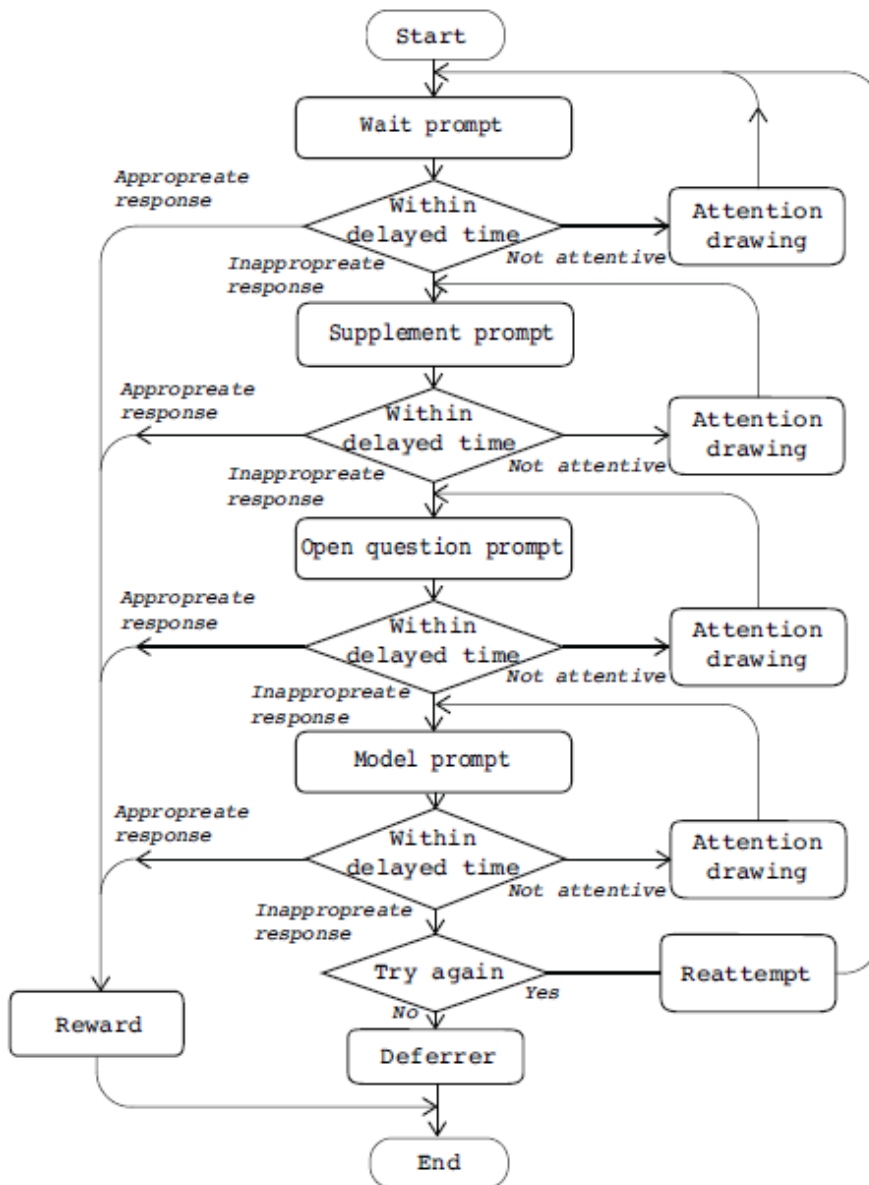


Source: Cuayáhuatl & Kruijff-Korbayová, 2012

Appendix VIII: PRT Scenario for *Boomgaardje* and prompting diagram

PRT Scenarios
Setting, agent and agent's goal
<p>Setting: The toys are placed at the initial positions.</p> <p>Agents: The robot and child are involved in the game.</p> <p>Goal: The robot and the child can win by pinching all cherries from the tree before the raven.</p>
Step 1: Designing a scenario
<p>General: A turn taking between a robot and a child occurs until the <i>Boomgaardje</i> game ends.</p> <p>Robot: When the robot's turn comes, it asks the child to give it the dice. After the robot throws the dice, it asks the child to turn over the flower card that corresponds to the flower symbol on the top of the dice. Next, the robot performs the appropriate action according to the figure on the back. If the card with the raven is picked, the robot asks the child to move the raven to the next footprint. If the cherry is on the card, the robot asks the child to pinch one of the cherries from the tree and put it into the basket. If sleeping animals appear on the card, the robot passes its turn to the child without doing anything.</p> <p>Child: In the child's turn, the child first throws the dice and turns over the corresponding card. The robot often says which flower appeared on the top of the dice or which card the child turned over: <ul style="list-style-type: none"> - 'You got a white flower on the dice.' - 'Wow, it's a cherry!' If the figure on the back is a raven, the child moves it to the next footprint. If the card has a cherry on it, the child picks up a cherry from the tree and puts it into the basket. If sleeping animals appear on the card, the child passes his turn to the robot. The robot sometimes verbalizes which toy the child moved: <ul style="list-style-type: none"> - 'You picked a cherry!' </p>
Step 2: Adding learning opportunities to the scenario
<p>General: see above</p> <p>Robot: see above</p> <p>Child: In the child's turn, the child first throws the dice and at this time the robot says it is his turn. <ul style="list-style-type: none"> - Learning opportunity: protesting If the child does not know what to do, the robot teaches the child to ask for help. <ul style="list-style-type: none"> - Learning opportunity: asking for help After throwing the dice, the child turns over the corresponding card. If the child does not know what to do, the robot teaches the child to ask for help. <ul style="list-style-type: none"> - Learning opportunity: asking for help The robot often says which flower appeared on the top of the dice or which card the child turned over. If the figure on the back is a raven, the child moves it to the next footprint. If the card with the cherry is picked, the child picks up a cherry from the tree and puts it into the basket. If sleeping animals appear on the card, then the child passes his turn to the robot. If the child is hesitating to do something for a long time, the robot teaches the child to ask for help. <ul style="list-style-type: none"> - Learning opportunity: asking for help The robot sometimes verbalizes which toy the child moved.</p>

Based on: Kim et al., 2014



Source: Kim et al., 2014

Appendix IX: Story with TSL for NAO

The words supported by TSL are indicated in blue and bold. The story is an English translation of the Turkish original, created by the research team. The story is constructed in such a way that each relevant word appeared twice, mostly at the beginning of the sentence. Furthermore, no more than two signs were used per sentence.

Three close **friends** decided to go to picnic to the forest.

Dad drove them to picnic with his red **car**.

When they arrived they put their food on the **table**.

They had very delicious food and cakes.

Three little rabbits suddenly jumped to the **table**.

Everything falls on the ground.

Dad said “Don’t worry; we can continue our picnic at home.”

Friends arranged their belongings and got on the **car**, and drove to their home to continue the picnic.

Little rabbits continue to play in the forest, too.

Based on: Kose & Yorganci, 2011.