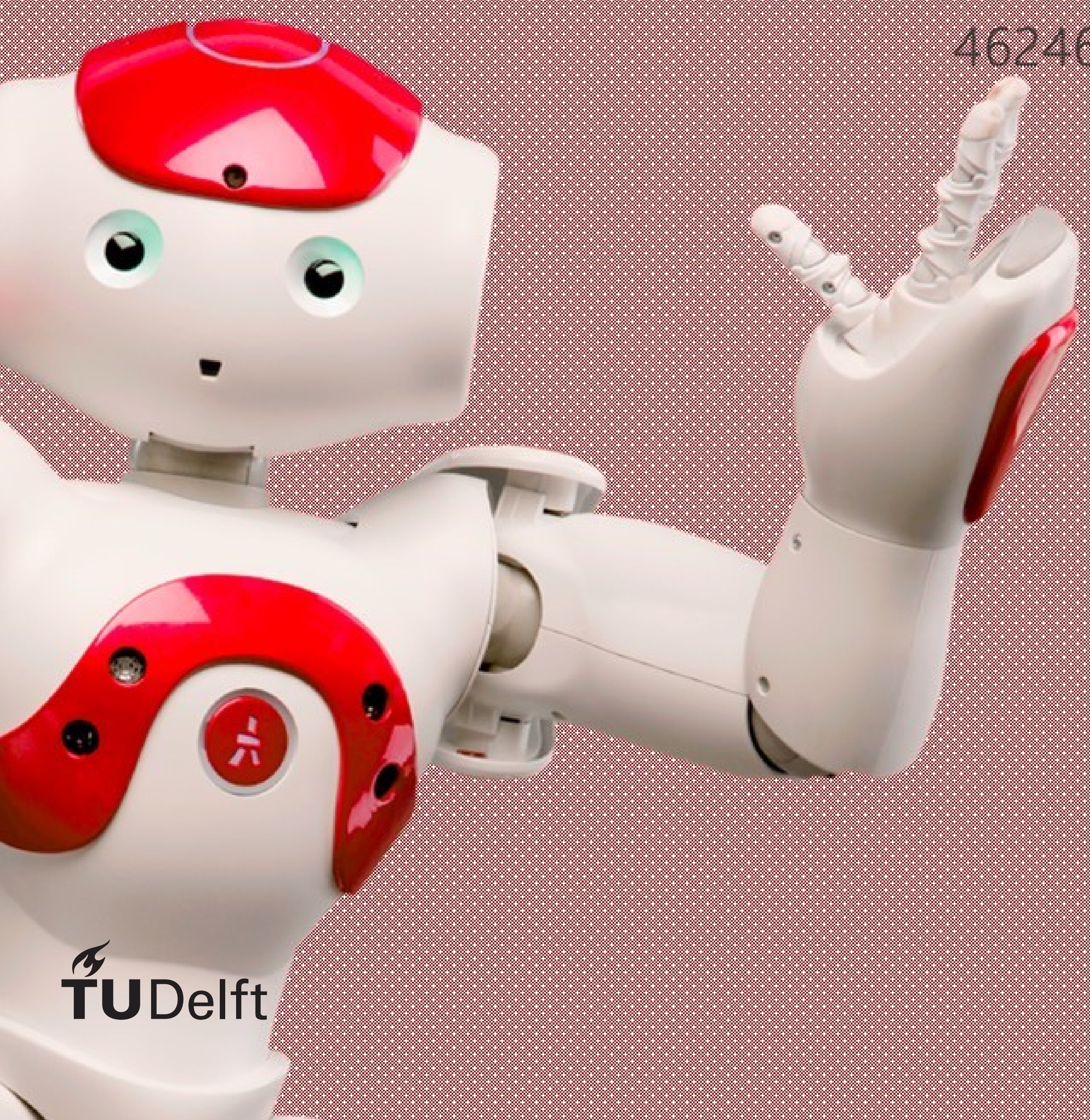
User Perception of a Humanoid Robot's Developing Behavior

Romi Kharisnawan
4624645



Technische Universiteit Delft

TU Delft

User Perception of a Humanoid Robot's Developing Behavior

by

.Romi Kharisnawan

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday August 30, 2018 at 09:30 AM.

Student number:	4624645	
Project duration:	November 25, 2017 – August 30, 2018	
Thesis committee:	Dr. ir. D.J. Broekens,	TU Delft, supervisor
	Dr. Ir. Koen V. Hindriks,	TU Delft
	Dr. Hayley S. Hung ,	TU Delft

This thesis is confidential and cannot be made public until August 31, 2018.

TU Delft

Preface

This thesis was prepared as a master degree completion in computer science department at TU Delft. The work has been done in 9 months in Interactive Intelligence group with the topic in human-robot interaction (HRI). This topic has attracted me since taking an Artificial Intelligence course in the first quarter. Moreover, I gained more interest in HRI field, especially in developing robots that can have a better engagement with humans. This thesis covers a research on growth and adaptive behavior in a human-robot interaction from the user perceptions. During the literature survey, both academic research paper and commercialized toys/games reviews were studied to give different angles of an intelligent agent development, especially robots. By conducting this study, I hope it can give a different perspective and a valuable insight on the implementation of adaptation and growth mechanism in human and robot interaction design which is able to obtain a better interaction experience.

I am grateful to work with all people who have been helping me during the thesis project. My highest gratitude is articulated to Joost Broekens who has guided and supervised me as my thesis supervisor. Big thank you note is also expressed to Koen Hindriks and Hayley Hung as thesis committee members. Your critical thought and evaluation are highly appreciated to improve this study and contribution to HRI in general. I also would like to thank Ruud de Jong who helped me on managing resources for the experiment. All gratitude is also uttered to all participants who were willing to join the experiment. Your time and feedback are really appreciated and useful to make the study happened. Furthermore, I would like to say thank you to Nuffic Neso for providing financial funding through Studeren in Nederland (StuNed) scholarship. Lastly, I would like to say thank you to all friends who always gave support in up-and-down during the thesis work.

.
Delft, August 2018

Abstract

An important challenge in developing a social robot is making the interaction between human and robot to be more pleasant and convenient. It could be obtained by making the robot to develop, i.e. change its behavior over the course of time. Here we study two aspects of development: behavioral adaptation and behavioral complexity to which we refer as growth. In this study, the adaptive behavior was implemented as a finite state machine with probability state transitions shaped by human feedback, while the behavioral growth was implemented as an unlocking behavior stages approach which was inspired by the development capability theory. The goal of this study is to examine if there is a significant effect of the adaptation and growth mechanism on human perceptions of aliveness, learning ability, and the behavior shaping control; and moreover how these perceptions influence interaction experience. We used a NAO robot for our studies. There were four conditions experimented from combination of adaptive and growing behavior. Twenty four (24) participants joined to interact with the robot in a within-subject experiment design where each participant interacted in two different conditions. As a result, we did not find a significant effect of the behavior manipulation in the experiment towards the measured perceptions. However, there is a significant positive correlation between the perception of learning ability and interaction experience.

Contents

1

Introduction

> I do think, in time, people will have, sort of, relationships with certain kinds of robots - not every robot, but certain kinds of robots - where they might feel that it is a sort of friendship, but it's going to be of a robot-human kind
>
> *Cynthia Breazeal*

What excites you the most about a robot which is featured in a movie? Let's take an example of a well-known BB-8 robot in Star Wars or TARS in Interstellar. Do the advanced anthropomorphism, smartness, friendliness, or all of them that make you impressed with the robot? For me, the way a robot engages with human becomes the most attractive aspect. Despite of the incapability of BB-8 to talk, it can still convey a message that humans might understand. Another impressive example is the ability of TARS to make a joke. There is a scene when Cooper, the main character in the movie, needs to reset its configuration, including its humor and honesty level, through speech. Although both of the robots are only fictional characters, they certainly can engage well with humans.

Human-Robot Interaction (HRI) is a field in the robotics domain which has drawn a lot of attention. Its complexity comes from different fields of study, such as computer science, mechanical study, psychology, and many others. It is not only related to the advanced robot development that can do various tasks, but also taking human as an important integral part of the robot development. Because of its complexity, there are still many aspects to explore in this study domain. Many researches have been working on making the interaction between human and robot more pleasant and convenient by embedding social aspects to the robot. It has been growing from the traditional HRI which only includes simple tasks to a very complex and smarter interactions, such as the Sophia robot by Hanson Robotics which is able to understand context of conversation, displaying numerous emotions, and embedded more human-like behaviors, such as making jokes and pauses in the speech.

The social aspect of the robot has been discussed in early 2000 where the paradigm of robot as sociable partner was introduced [31]. It raises the question of how to properly interface untrained humans with robots in intuitive, efficient, and enjoyable way. Endowing robots with social skills and capabilities could help humans to feel more natural experience and make robots know better on their intention. Despite its benefit, there are several design issues of the robot development which include its morphology, aesthetics appearance, physical skillfulness, perceptual capabilities, communicative expression, and its intelligence. These issues can be addressed from several point of views, such as naturalness of the interaction, user expectation, personality, acceptance, and other views.

Growth and adaptive behavior are some of the social aspects that are important to make a social robot different from a non-social robot. Growth is related to a development of cognitive and movement from simple to more advanced tasks. This reflects what humans experience in their real life where the knowledge and the capabilities are developed from an early age until the end of life. It has been implemented not only in a robot but also in virtual agents in several computer games. Some games, such as Black and White [13] and Creatures [18], successfully attracted people to play with their implementation of an artificial intelligence to grow its agent's character. Another social aspect that will be discussed in this document is an adaptive behavior. Adaptive behavior is defined as the capability of robot to understand humans and react based on their preferences or behaviors. Adaptiveness can arguably give some benefits in achieving the goal and create a better interaction experience. For example, robot which was implemented with adaptive behavior gave better learning progress compares to non-adaptive robot [48]. Another example is Phillos, a low-cost social robot, which was designed as a personalized robot, can engage each user in a personalized interaction via uniquely defined behavioral responses tailored for each user [47]. Moreover, the experiment by Hemminghaus and Koop [41] displayed that robot with ability to learn user behavior gives better result in helping human solving memory games faster.

Adaptive robot needs to implement sensory data to detect feedbacks that displayed by humans and react to it [27]. Verbal and nonverbal modalities could be used to detect and show the feedback. Verbal communication is related to the language, such as diction, while nonverbal communication pertains to a process of generating meaning using behavior other than words, such as body gesture and facial expression [26]. As an example for feedback detection, speech could be a cue to detect if a user favors the robot behavior by processing their speech and transcript it to certain dictionary for classifying whether it is a positive or a negative feedback. In addition, an emotion could also be a cue to understand humans behavior by processing images and analyze it using an emotion recognition algorithm. The similar approach using images is also used to detect gazing behavior of user as cue for the engagement or excitement. However, in this experiment, only speech will be used as social cue to determine a positive or a negative feedback. Another essential item is how robot portrays the behavior so user will be able to understand that it shapes its behavior based on their feedback. This transparent approach by making adaptiveness and growth feasible to human, for instance personality changes in Furby, makes the toy perceived more alive and interactive which arguably gives positive value to the bond [20].

The main problem of an adaptive behavior implementation in the human-robot interaction is no single guideline to indicate the adaptiveness due to various natures of the interaction which created many different approaches. One have used simple logic by identifying errors that human made and give the hint based on it [48] while other implemented a reinforcement learning algorithm to study human behavior then giving an explicit response [40]. Both analyzing perception and giving response are important in the adaptation design. Many of researches examined how to improve the perceptions of robots while not many studies discussed the way to give a good response. High accuracies of perception systems is important, however, a reasonable response by robots has a significant impact [54]. Human will not care if its perception system can achieve high accuracy as long as the robot might serve a reasonable response during the interaction.

On the growth aspect, there are limited studies in a physical embodiment robot. This might be caused by the fact that growth is heavily related to a physical growth, from small to bigger physical appearances. Physical embodiment robot has limitation to display the growth because the body is rather static to the possibility to be a taller or bigger shape. It is different with virtual agents where most of the computer games implement the growth aspect in the agent, such as in Black and White and Creatures. However, there is a study related to cognitive growth has been conducted to study the influence of growth in the physical embodiment robot towards interaction experience using an AIBO robot [44]. It used a developmental capability method where the development divided into stages and each stage contains certain tasks. The higher the stage is, the more complex the task becomes.

In this study, we implemented growth and adaptive behavior into NAO as a humanoid robot. We argue that growth and adaptive behavior implementation in the robot can positively influence the human-robot interaction which is illustrated in Figure 1.1. The main purpose of the study is to understand whether or not the this robot's behavior development can improve the interaction experience, which furthermore can lead to create and maintain bond between human and robot. A finite state machine was implemented to model the interaction by modeling each behavior as a state. State transition probability was implemented to formulate adaptiveness by changing its probability upon the user feedbacks, while unlocking behavior stages was implemented to show growth aspect.

This thesis also tries to bridge the gap on how humans perceive the interaction with a robot. There are a lot of researches that studied an interaction experience in a direct correlation to the behaviors produced by the robot. They usually asked the participants to rate how enjoyable the interaction in a response to the robot's actions. In this study, user perceptions towards a robot is used as mediating variables to measure the interaction experience. There are three types of perception that are interesting to explore: the perception of aliveness, the perception of learning ability, and the perception of behavior shaping control. These perceptions are measured to breakdown in detail what make humans perceived the interaction as a pleasant experience. It hopefully can help to understand more about which factor has more significant effect and need to be focused in the robot's behavior design.

Figure 1.2 shows a work flow model to answer the main purpose of the study. Growth and adaptive behavior arguably give a positive effect to these measured perceptions. Then, the positive perceptions will improve the interaction experience which lead to a stronger bond between human and robot. Bonding is an important aspect in establishing and maintaining a long term relationship. In the practice, most people are excited to interact with robot in the first meeting, or usually called by novelty effect. However, it might decrease along with the interaction goes if robot only performs statically. Human might lose interest and the bond get lessen gradually. Thus, a long term oriented design is important in the robot development [43]. The design was implemented to minimize novelty effect in the interaction, which was formalized by implementing growth and adaptive behavior to the robot.

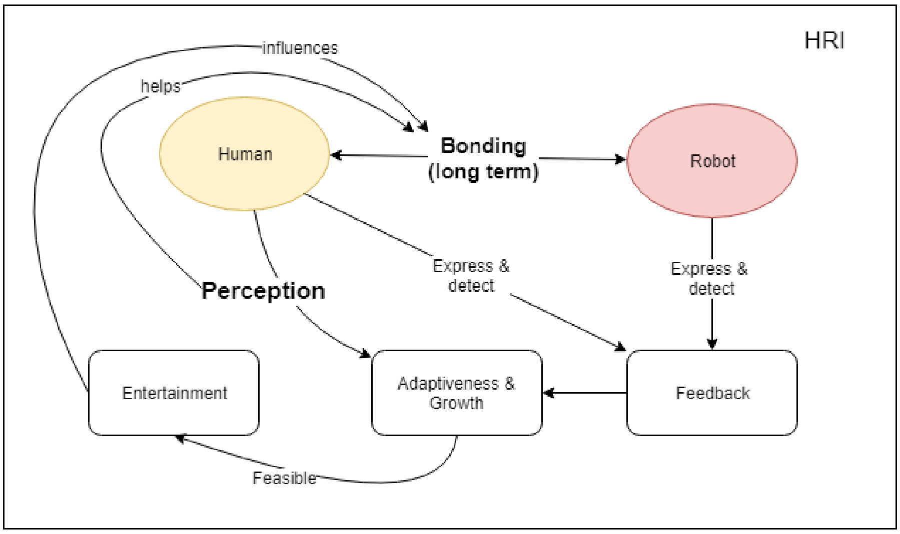


Figure 1.1: Human-Robot Interaction Adaptive Model

The main purpose of this study is to understand the effect of adaptation and growth on the interaction through the perception of aliveness, learning ability, and behavior shaping control. It raised two main research interests for this exploratory experiment, as mentioned below:

1. What is the effect of growth and adaptation on the perception of the robot being alive, the perception of the learning ability of the robot, and the perception of power/control?

2. Is there a correlation between the perception of the robot being alive, the perception of the learning ability of the robot, and the perception of power/control with the interaction experience?

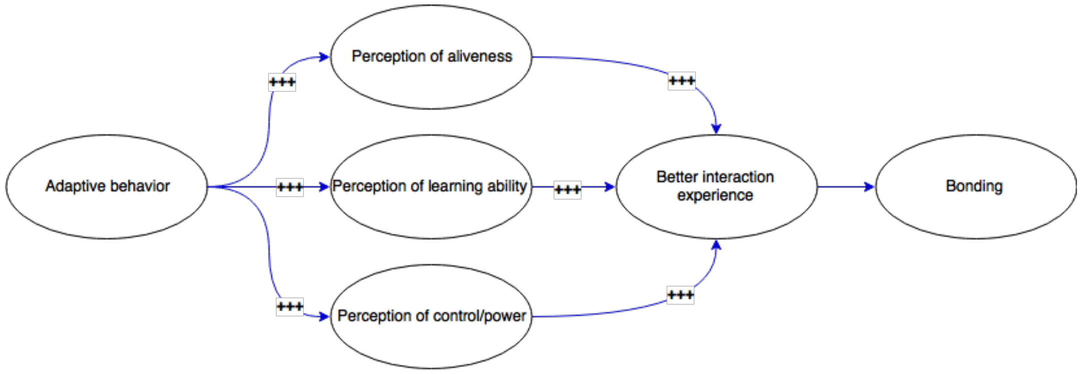


Figure 1.2: Work Flow Adaptive Behavior Model

This thesis document is divided into 8 chapters. It starts with the introduction to give brief explanation about the study and its motivation. Then, it is followed by literature reviews on supporting theories and related works in chapter 2. Chapter 3 will discuss the method which was used in the experiment, while in chapter 4 will elaborate more on the implementation of the experiment. Chapter 5 will discuss the result from the conducted experiment and the analysis of gathered data. Then, chapter 6 will discuss reflection of the result based on the hypothesis and recommendations for further research. This document will be closed by a conclusion in chapter 7.

2 Literature Review

This section describes the supporting theories and literature survey related to growth and adaptive behavior. The literature survey did not only cover research studies but also included reviews of commercial products which implement growth and adaptive behavior in their agents. Moreover, it also consists of previous studies on the user perceptions and the interaction experience. Furthermore, the behavior modeling approach was also examined in the algorithm and model section.

2.1. Adaptation

Adaptation as a definition refers to a subject changing and becoming better suited, or fit, to an environment. The initial concept of an adaptation starts with a biological adaptation. It relates to human anatomy and genetics where the long run changes are considered as evolutionary changes. The same concept was used in understanding human behaviors, which is usually called by evolutionary psychology. This study often uses social constructs that are not directly observable, such as love, extraversion, conservatism, and depression, which need to be validated [35]. Cronbach and Meehl discussed about three important steps to establish validity of these constructs: articulating the concepts of the constructs and their expected interrelations, develop ways to measure the proposed concepts, empirically test the hypothesized relations among the concepts.

Adaptive behavior was successfully implemented in three experiment examples [41, 47, 48]. They have different approaches to study an adaptive behavior of robot, including method, setup, and evaluation metrics. Further explanation of the experiments are described below:

- Simple adaptive behavior was implemented by Ramachandran, et.al. [48], where the experiment was intended to reduce sub-optimal help seeking behavior. Two condition groups, with and without adaptive behavior, were studied. Adaptive behavior follows simple strategies: next hint will be automatically provided after two consecutive incorrect attempts and the participant can not request three consecutive hints without any attempt. The interaction was conducted in 4 sessions with the same treatment of both groups in the first session to see the learning ability difference between pre and post test, then a different treatment in the following weeks. The evaluation metrics used for this experiment are a number of triggers and score changes. The number of triggers evaluates optimality of help-seeking, while difference between pre and post test score represent a learning gain.
- Hemminghaus and Koop [41] studied an adaptive social behavior generation using a reinforcement learning in a memory game setup. The experiment is using Furhat robot which has the ability to show affective behavior cues, such as gaze, speech, facial expression, and head gestures. There are two experimental groups: a robot which helped with a random condition and a robot which helped with a learning condition. The robot with random condition gave help with probability 50%, whether to help or not, and

then randomly chose an action during the game with an interval time 2 seconds. On the other hand, the robot with learning condition determines if the participant needed a help based on the user states and then randomly chose either an action with maximal Q-value or explore action combinations with probability 60% and 40%. Temporal difference method, specifically with Q-learning, was chosen for this setup as it is model free and work online using step-by-step computation. User states were determined by Wizard-of-Oz, which include user gaze and speech to the robot, with an additional information of the game state (number of the remaining pairs that are still not matched). The evaluation metric used is a completion time and q-values metric of each behavior combination. An additional qualitative evaluation is also provided based on the feedback from the participants.

- In Phillos experiment by Puehn [47], five predefined personality types were used to generate internal characteristic of the robot. Each of personality type is represented by values of 5 personality dimensions: extraversion, agreeableness, self-control, emotional stability, and independence. These values determine behavior parameters of the robot, such as number of behaviors the robot may exhibit in the idle time and the level of positive behavioral response. These values are initially set by an operator and will be adaptively changed based on an external input by user along the interaction. There are several kind of external inputs by the user: touch, face tracking, face recognition, motion detection. Speech recognition was omitted in this version but the writer mentioned that this is an important areas of the exploration. The evaluation method used is a percentage of positive response enacted by Phillos based on the user behavior input.

2.2. Growth

Growth is one of the common aspects which implemented in *artificial intelligence* agent, both virtual or physical agent. The concept of a growth has been discussed as a development aspect in the essence of intelligence by Brooks, et.al. [33]. It was discussed as a humanoid design principle which referring to human intelligence. The development aspect explains the growth of the human reasoning, motor, and sensory system as a gradual process from an infant to an adult. The similar concept was used in *cognitive development robotics* (CDR) by Asada, et.al. [29]. This concept was used as a paradigm to design humanoid robots which covers a design of an embedded structure which can learn and develop; and creation of a social environment which capable to support development of cognitive processes. This paradigm uses social interaction and communication developmental approaches, such as the transition from non-verbal to verbal communication which is grown from a baby until an adult.

Growth can be seen from two different perspectives: mental and physical. In the physical embodiment agent, such as a robot, it is hard to show a physical growth in the interaction because the body design is rather static, while it is possible in the virtual agent by changing the physical appearances along the interaction. Therefore, a mental growth is the growth that can be implemented in the physical agent. Mental growth is also related to an intellectual growth which can be broke down in a cognitive and an affective point, for example, showing more complex movements, understand the situation better, or showing more emotions. One of growth implementation in physical agent is developmental capability by Lee et.al. [44].

2.2.1. Developmental Capability

Developmental capability is an approach which has been used to model development stages of the AIBO robot. There are 4 stages introduced: Baby (D1), Kid (D2), Adolescent (D3), and Adult (D4). The starting stage is a baby which grows into a kid, then an adolescent, and an adult as a final stage. It reflects the development process of a human. The growth transition from one stage to the next stage was shown by a cognitive development, which are translated to four factors, namely number of tasks, the sophistication of tasks, the speed of learning, and the sophistication of random and spontaneous behaviors. The tasks implemented were simple yet showing the growth transition. To move from one stage to the next stage, the participants needed to train AIBO to do certain tasks which triggered by verbal commands, patting on its head, and patting on its chin.

Each development stage occurs in a week, which resulted 4 weeks in a total for the development from a baby to an adult stage. Figure G.1 and Figure G.2 display the detail information of all stages. As one of example explanation, AIBO was only able to perform two tasks, namely make AIBO understands its name and learns how to say goodbye, in week 1. After giving certain amount of training through verbal commands, patting on head, and patting on chin, the AIBO can unlock the next stage. For example, the AIBO will unlock the kid stage after giving trainings in verbal, patting on head, and patting on chin for six times on each task in baby stage: understand its name and learn to say goodbye. The similar approach was also used in our experiment. After unlocking the new stage, the previous tasks will be still stored and remembered.

This approach has been successfully immersed into the AIBO which was recognized by the participants that the robot in the growing condition is indeed grew its behavior step-by-step. Also, it brought a positive impact by obtaining a better social presence and more positive social response. Thus, this approach was being used as an inspiration to build interaction design for our experiment.

2.2.2. Growth and Adaptation as a Coupling Factor

Adaptation and growth are often coupling one to another. Jean Piaget viewed an intellectual growth as a process of adaptation (adjustment) to the world [16]. Process of adaptation is occurred through assimilation, accommodation, and equilibrium process. Assimilation is defined as the usage of an existing schema to deal with a new object or situation. On the other hand, the accommodation process occurs when the existing schema (knowledge) does not work in a certain situation and needs to be modified to deal with a new object or a situation. The equilibrium process is a development force from applying an existing schema in the assimilation to change it based on a new situation. It drives the learning process to adjust a human frustration in a new situation and will seek to restore balance by mastering the new challenge (accommodation).

A mental model schema might be used in building a mental model for the robot as well. The idea of learning something new and an adaptation to a new situation is well-suited to improve cognitive and affective aspects of the robot. In the learning adaptive behavior, a behavior might be perceived as a building block which is linked one to another. In the case of a robot adapts with human behavior, it can apply its existing schema (initial pre-programmed mental schema), as mentioned in the assimilation process, and modify it when it does not fit to the situation. After many interactions to shape the robot's mental schema, hopefully the robot can achieve a converged mental schema to fit with human preferences or behaviors.

2.3. Commercial Toy and Game Review

Besides several previous studies in the research domain, reviews on commercial toys and games are also beneficial in the human-robot interaction design. These reviews show best practices or empirical insights of what work well in the existing market. The complexity of adaptation and growth implementation vary from one to another but yet can provide an enjoyment in the interaction with the agent, both physically and virtually. There were 2 toys and 2 computer games were reviewed: Furby, AIBO, Black and White, and lastly Creatures.

2.3.1. Furby

Furby is a toy which represents an animal-like embodiment, which similar to an owl or a hamster. It was initially created by Dave Hampton and Caleb Chung which later on Richard C. Levy joined the effort to sell the toy. They sold it to Tiger Electronics and made the debut appearance in 1998 [23]. Then in 2005, a new release of Furby was released and later on the next major change of Furby was released in 2012. Further details on the Furby changes will be described on the following subsections. The current Furby in market has several sensors to perceive the user behaviors and actuators to perform actions. Some sensors are placed in a Furby's body, such as light, motion, gesture, and sound sensor. To convey message to the user, Furby is equipped with motors, LCD eyes, and a speaker.

Another interesting interaction is the Furby's speech. It speaks Furbish language which has been already predefined before in the system. Furby will change the language skill gradually from Furbish to English. As a matter of fact, it does not learn the language but it is simply pre-programmed to incorporate basic English words after certain period of the interactions [21]. It gives the intuition that Furby can learn from the interaction. In fact, the response by Furby is predefined based on the action given by the user and its *learning state*.

Furby 1998

The initial release of Furby offered a physicality touch which differentiated it from other popular toys at that time, such as Tamagochi and Digimon. It has several sensors and motors in the body which enable Furby to [22]:

1. Respond to strokes on the back.
2. Respond to touches on the stomach.
3. Respond to 'fed' when something is put in its mouth.
4. Identify being turned upside-down.
5. Respond to a change in light levels e.g. a light being turned on.
6. Respond to sounds over a certain volume.
7. Communicate with another Furby via IR (infra red) connectivity.

It creates a perception of an intelligent robot pet where it likes to be petted and dislikes to be turned upside down or shaken off. Another intelligent perception created is on ability to grow up. It has maturity process where it begins to talk in Furbish (Furby language) and learn words of English, then will start responding to people in human language. It has rumor that Furby can understand language from the environment around or repeating words, but in fact, it is only simple pre-programmed feature to introduce English words into its vocabulary after a set amount of time has passed [22], as per mentioned before. It also has a basic software capability where it could only respond to stimulus and 'learn' language in a linear fashion way without any personality development along the interaction.

Furby 2005

After line of Furby products was discontinued in 2003, later in 2005, a new version of Furby, Emoto-Tronic Furby, was introduced by Hasbro with voice-recognition and more complex facial movements. The prominent changes in this versions were a bigger physical design and a voice recognition. While the sensors in the stomach and the back are pretty much same with a 1998 Furby.

This version of Furby is more focused on the ability to understand speech and respond accordingly [19]. It improved the previous version where it only recognized loud sounds. Although it embedded new speech recognition technology, it only detects limited phrases and commands. For example, user has to say "Hey Furby!" to get its attention and continue to give a command. Some commands that you can give includes "Sing me a song", "Tell me a story", "I love you", etc. There is no explicit evidence of an improvement on the learning ability. This improvement shows that the speech recognition technology is a big part to improve the user experience.

Furby 2012

In 2012, a new line of Furby was created with more expressive LCD eyes, more motion range, its own iOS and Android app, and the ability to change personality in regards to the user behaviors [20]. Physical eyes to digital eyes and personality development were the major changes in this version. The Furby will respond to the user's actions promptly, such as a change of LCD eyes to lovely eyes upon rubbing on back of the body. It helps the user to understand whether or not the Furby likes the action given to him.

The collection of user's actions will shape the Furby's personality. There are several personalities available on the Furby: chatterbox, evil, sweet, crazy, sassy. As an example, a chatterbox personality could be obtained by by talking a lot to the Furby. She will often say ”Blah” and ”Like me like, to say like” while chatting. Her eyes are oblongs with the eyelashes. As a note, the personality is not perpetual and can change along with the user behaviors. Whenever the personality changes, the Furby will shout *”Changeeee!”* and react based on the next personality. The stand out respond that a user can see is on the eyes, where Furby will change type of gazing on its LCD eyes. It triggers the perception of more intelligent robot by learning on the interactions between human and the robot. On the language learning part, there is no improvement made and thus the maturity of spoken language by Furby is developed throughout time and not based on how advance the user trains the Furby. The development items display that the explicit feedback to the user is important to make an engagement.

2.3.2. AIBO

AIBO, abbreviation of Artificial Intelligence (AI) Robot, is a robot dog developed by Sony. It was first released in 1999 and discontinued in 2006 [2]. Later in 2017, Sony just announced a new version of AIBO, which has been launched in January 2018 [3]. The new version of AIBO is equipped with the latest AI technology which allowing it to learn tricks, be aware of its surroundings, and develop a bond with its owner through facial and voice recognition [5]. From the video, AIBO is able to simulate how a dog learn tricks by teaching it using a punishment and reward mechanism. Also, it can learn how the map of a house look like, so it might run smoothly without too many collisions to objects and go to a designated place.

AIBO in 1999

The original AIBO was launched in 1999 with main purpose as a companion to human as a pet. It was carried out with programmed actions such as barking but its movements were clunky and its learning capability was limited due to a dependency to internal memory sticks inside the robot. This version of AIBO was equipped with 20 motorized joints for conveying motoric movements, touch sensor (on his head, chin and back), hearing sensor (stereo microphones), sight (a camera in his head) and balance sensor, infrared distance sensor, an acceleration sensor, and a temperature sensor. Nevertheless its limitation, it is still one of the most advanced personal robots in the market at that time.

In this version of AIBO, Sony was focused on maximize its lifelike appearance. Due to a lack of good evaluation method for *lifelike appearances*, Sony introduces following factors as solutions to assess its liveliness [40]:

1. a configuration with high degree of freedom
2. multiple motivations for movement
3. a non repeated behavior exhibition

Liveliness of the robot was resulted by implementing an artificial intelligence. It creates emotions which become the motivations for the movements. Lifelike aspect was also shown by the developmental stages of an infant, child, teen and adult [8]. Maturity was affected by daily communication and attention given to an AIBO, which makes it learns to do certain actions and recognize vocabularies. In addition, AIBO is also able to learn its name which was given to him through the voice recognition.

AIBO used three different motivations for movement: time, internal/external, and body parts of robot [40]. Related to the time, it will respond quickly to some stimuli, such as a loud sound, while also have to behave slowly with deliberation to show more real effects. In second motivation, AIBO has four instincts: affection, investigation, exercise, and appetite where each of the instinct related to current status of robot. For example, an appetite instinct will appear when the battery discharges and it will perform a motion and sound to notify user. In addition, six artificial emotions are implemented as well from the Ekman's six basic emotions: joy, sadness, anger, surprise, fear, and disgust [38]. The third motivation is body parts of robot which means the availability of the behavior exhibited is based on the combination of body parts, consist of head, tail, and legs.

In order to avoid repeated behaviors, there are behavior control architectures implemented: artificial emotions and instincts, probabilistic of state machine for behavior generation, reinforcement learning for probabilistic state machine, and development through interaction [40]. Artificial emotions and instincts are quick responses of stimuli given to robot, for example giving a paw when human hand is nearby. In addition, the reinforcement learning was implemented to modify probabilities of the state machine. It enables AIBO to learn based on the rewards given by humans. Final method is the development where it enables the robot to develop a different persona for each user based on its state machine.

AIBO in 2018

Latest version of AIBO, which is also a revival version from 2006, has been released in January 2018 with price around $1,800. The dog robot understands a handful of English-language directions, including hand-shaking and commands to sit [4]. From hardware perspective, the revived AIBO has glassy OLED eyes and a camera inside its nose, which can also be used as a webcam that you can access real-time.

It has more advanced technology implemented to the robot, such as 64-bit quad-core CPU, built-in LTE and WiFi, motors and gyroscopes to augment the 22 different articulated parts. It also has a speaker for robotic yips and yaps and four microphones to pick up voice commands even in a noisy environment. The use of WiFi enables data transmission to the cloud which benefits in performing more data processing and more complex actions compare to memory sticks on the previous version.

In this version, it has more power in processing the input data and make decision on what action should be performed. This can be performed by a data collection on the cloud which results ability of the robot to learn more, such as an ability to identify its owner and remember behaviors that make its owner happy [6]. In addition, AIBO will be able to differentiate the interacted users and learn which user gives the best snuggles, or at least whoever pets it the most [4].

This reborn AIBO has more capabilities to emote and convey messages to its user by body language, perking its ears, making eyes, wagging its tail towards people’s action. The upgraded eyes allow the robot to display diverse and nuanced expressions [7]. It also recognizes its owners voices and feeds off interactions with family members. Another improvement made is Simultaneous Localization and Mapping (SLAM) system which allows the dog to map out your house, avoid obstacles, and figure out the shortest distance from one location to another. In addition, more tricks are available to be learned by the dog.

2.3.3. Creatures

Creatures is a simulation game of artificial life created by Steve Grand in mid 1990s. It features the creature, called Norn, as a main subject which has to be raised by the player. The player has to teach them to survive, helping them to explore their world, defending them against other species, and breeding them [18]. The agents, known as ”creatures”, have an artificial neural networks for a sensory-motor control and learning, an artificial biochemistries for energy metabolism and hormonal regulation of behavior, and both the network and the biochemistry are ”genetically” specified to allow for the possibility of an evolutionary adaptation through sexual reproduction [50].

The creature’s life represents life stages, where it starts as an infant, grow older, and finally dies. In the growing period, the creature can learn to do things and language. The creature is able to be trained by giving a reinforcement signal: stroking to generate positive signal/reward or slapping to generate negative signal/punishment [50]. Furthermore, the language might be learned by typing word(s) when the creature see object or command it to study from the computer.

Creatures 1

The first release has already included two and half dimension where the area is in two dimensional plane and additional view on what Creature sees. The player needs to act as Norns who live like living creature in real life. Player needs to teach Norns to do some actions and vocabulary. Norns has unique genome, which resembled from its heritage, and functional brain and body.

Creatures 2

There is no significant difference on brain structure of Creatures 1 and Creatures 2 [17]. There are additional features added to the brain lobe functionality that could give advantage of in the future genomes for more new brain modifications potential that has not been done in Creatures 1. The main differences are the additional of new lobe, called regulator lobe, and the additional new state variable rules.

Regulator lobe is used for supplying functionality similar to the receptor and emitter genes but provides a whole lot more flexibility [17]. It provides correlation between receptor and emitter where relates to feedback loops. For example, neuron 1 has couplings of water as a receptor and thirst as an emitter where it gives loop feedback signal. Additional state variable rules function to have more operations of states in the brain.

2.3.4. Black and White

This game implemented the advanced use of an artificial intelligence compare to the other games during its release period. There are two types of intelligent agents in the game: community of villagers and a creature [55]. The desire tendency was developed by an artificial intelligence using a punishment and reward training. In the training, the player can choose an action, either pet or slap the creature to show a reward or a punishment. Each action has spectrum of a punishment to a reward, so it is not binary 0 or 1. Based on Cass' review, the creature uses emphatic learning technique by watching the player's actions and attempts to divine the intent behind them [36]. For example, a rage action towards the foreign tribes will make the creature to think that the player does not like them and help to defeat them. In addition, the player can also train the creature to do certain actions rather only do its free will (autonomous actions). These actions create the creature's belief represented by symbolic attribute-value pairs which is used to give a basic intelligence about objects to the creature with a rule-based AI [53].

There are numerous interactions and indicators of a current status of the creature. The creature has two main indicators on toolbar: free will and good/evil. This indicator will change along with the interactions by the player, such as adding buildings, attacking other tribes, or assigning citizens as disciples. In addition, the creature also shows what it feels at the moment through a chat bubble box, such as *"I'm hungry", "I'm tired", "I want to eat these people"*. Moreover, the creature also shows a feedback after the player rub as a positive reward or slap as a punishment in the training process. These kind of interactions make the player ables to notice easier on the implication of the given actions, especially a direct action to the creature.

Black and White 1

Black and White 1 is an initial release of the game and it was published in 2001 with artificial intelligence as important key of the game. The main concept of the game is being good or evil, represented by how the player plays the game. Player acts as a God who control the game play, including train the creature to be good or evil. In the first version, it has implemented quite complex artificial intelligence on the creature. It was designed by blending traditional frameworks of AI in game, which are mostly hard-coded behavior or multiple behavior sates, with decision trees, which create a branching map of the AI's beliefs and enables it to make choices, and the perceptron networks [10]. It enables the creature to make certain assumptions about its behavior and responds those assumptions depending on the external responses. There are three ways of learning: looking from what God does, assign certain task to the Creature, and pet or slap the creature on his action.

Black and White 2

Black and White 2 was released as a real-time strategy desktop game by EA (Electronic Arts) in 2005. It was the updated version of the first release of Black and White in 2001. The game play is still same where a Greek tribe plays against the others: Aztecs, Norse, and Japanese in order to rebuild people's livelihood. The player still plays a role as a God who is called to help the people. In addition, the player should choose which of the creatures that they want to play with, from gorilla to wolf. Moreover, it keeps the ability of the player to develop the environment and creature's good or evil desires like in first version [11, 13]. Some reviews on second version of the game said this version has a better building game while it lost the superior god power, especially in teaching the creature [12]. The creature is able to learn quickly by a spoken instruction without an intense learning by example process [15]. In contrast, the first version has a difficulty to train the creature to do exactly what you wanted, and it inevitably messed up in the training period, but that made the the Black and White creature seems has a real personality and able to learn [14] although it made the perception of artificial intelligence becomes powerless. On the other hand, the second version has a better city-building portion and a simplified real-time strategy game. It also has more options in choosing a starting creature, although it has no significant affect on which creature has been chosen [9]. The creature has an initial personality and ability which might affect on learning process, for example a lion has a good combat skill while a cow is good at looking after the villagers.

2.4. Adaptive and Growing Behavior Requirements

From the previous experiments and commercialized products discussed above, we can conclude there are couple requirements which make growth and adaptive behavior can be noticeable and well-perceived by human. Some of them are mentioned below:

2.4.1. Perception towards learning

Human perception towards learning ability of the robot is one of few aspects which make an interaction more realistic and smooth, especially related to the adaptation experience. Adaptation process takes time and usually comes through trial-and-error but it has to be perceived to human as a learn-able agent. In Black and White game, the player has to develop the environment and the creature's good or evil desires by training them [13, 36]. The desire is developed through an adaptation mechanism. One said the first version of the game is more exciting because of the sense of learning ability of the creature. It is hard to train the creature to do exactly what you wanted and it is inevitably messed up in the training period, which makes the Black and White creature seems as a real personality and ables to learn [14]. Empirically, the implementation of an adaptive behavior in the agent enables humans to perceive that the agent can learn.

2.4.2. Ability to shape behavior transparently

It is important for a robot to communicate its comprehension of human behavior and the intention on an action transparently, including its behavior changes. The transparency to shape the behaviors might be shown implicitly or explicitly. Robot expressions is one way to display implicitly whether it understands what humans want and shape its behavior towards the input. Expression can be shown in a form of facial or body expression. It helps human to understand better its behavior, affects, and intent which might lead to better interaction between human and robot. As an example, Breazeal found that people are willing to help the robot to address difficult tasks or get future actions when the robot shows transparency [32]. By giving guidance, feedback, and motivational intents in various way it helped user to understand that the robot is learning to adapt with what human teach [52]. On another hand, a robot can also ask the question regarding its behavior favors human preferences as an explicit approach. This approach will give more clarity to user on which and how the behavior should be shaped.

2.4.3. Reasonable response is more important than high accuracies of perception system

In survey paper presented by Yan et.al., the evaluation criteria for perception in social robot has a slight difference from pure research. High accuracies of perception systems is important, however, reasonable response by robot has significant impact [54]. Human will not care if its perception system can achieve high accuracy as long as robot might serve a reasonable response during the interaction. Some of commercial toys and games in the market implemented this concept. For instance, Furby as a toy gives the perception of learning by changing its personality or expression, although it is scripted and not actually listen to the feedback from humans.

2.4.4. Use non-verbal cues as response from and to human

Implicit channel is related to act and emotion of human or an agent. Research by Mehrabian and Friar [45] shows that about 93% of emotional meaning of a message is communicated implicitly through non-verbal channels. Thus, understanding and displaying correct meaning from the non-verbal channels are essential for a robot to be an emotionally intelligent. In addition, Bruce et.al. studied about a role of expressiveness and attention in a human-robot interaction which yielded that facial expressions and tracking behavior had statistically significant effects on attracting people to talk to the robot [34].

2.5. Perception

This section describes the perception of humans towards robot that interacts with them as the measured constructs. There are three main perceptions which will be focused on, namely perception of aliveness, learning ability, and control/power. To be able to add personalization and obtain those three perceptions, verbal and nonverbal cues were implemented, such as speech, motion, affective.

2.5.1. Aliveness

Perception of aliveness pertains to a perception of human to robot as a living entity. Being alive is one of distinctive characteristics that distinguish human beings from machines. In robot which performs movements and intentional behaviors, it is not obvious how the aliveness perception of human to the robot. In research by Bartneck et.al., aliveness is related closely to the animacy factor of a robot and has some relation to the anthropomorphism [30]. Animacy is commonly referred by a classic perception of life which is centered on ”moving of one’s own accord” based on Piaget’s framework [16]. It covers movements and intentional behaviors of the nonliving creature, in this case is a robot. Moreover, anthropomorphism refers to the attribution of a human form, human characteristics, or human behaviors to nonhuman things, such as robots, computers, and animals. Those two aspects contribute to make robot is able to be perceived alive.

Piaget found that children considered every objects that moved as alive at first, but later, only things that moved without external push or pull. Then, the definition of alive entity move to only those things that breathed and grew. There were several approaches explored in Bartneck’s paper but Lee et.al.’s approach has been chosen as foundation with 4 items (lifelike. machinelike, interactive, and responsive) [44] in 10-point Likert scale. The final questionnaire, commonly called by Godspeed Questionnaire, has animacy aspect which is transformed into 5 items in 5-point Likert scale, namely: Dead/Alive, Stagnant/Lively, Mechanical/Organic, Artificial/Lifelike, Inert/Interactive, Apathetic,/Responsive [30]. Appendix B shows the full questionnaire. This questionnaire has been tested with Cronbach’s Alpha 0.702 which is sufficiently high to ensure the internal consistency reliability.

In the anthropomorphism study, highly anthropomorphism usually perceived as a better experience although it has a main challenge related to *uncanny valley*. This theory stated that if a robot is created to be more human-like in its appearance and movements, it will result more positive and emphatic emotional response from humans, until it reaches one point and the response will quickly becomes an intense repulsion. This aliveness attribute can be attached to a robot if humans think that a robot is not only as a machine. The Godspeed questionnaire, developed by Bartneck et.al., asks humans to score 5 items in 5-point Likert scale related to anthropomorphism, namely: Fake/Natural, Machinelike/Human-like, Unconscious/Conscious, Artificial./Lifelike, and Moving rigidly/Moving elegantly [30]. Appendix B shows the full questionnaire.

2.5.2. Learning Ability

Perception of learning ability will influence future actions given by human. When human perceived that robot can learn from his feedback, he will change his behavior to compensate the robot's behavior. Learning ability might be shown through the behavior changes, for example repeating the favorable actions or showing an expression which indicates that robot understand the human action. Giving guidance, feedback, and motivational intents in various way can help the user to understand that agent is learning to adapt with what human teach to them [52]. In the experiment, gaze is used as a feedback to humans to show a specific area where they can help the agent to learn better. In addition, an explicit feedback can also be used as an indication of learning by the robot.

This perception is highly related to the perception of intelligence. If robot has high learning ability, it will be perceived as an intelligent creature. One might use Wizard-of-Oz methods to give a learning ability flavor but it is very limited to the practical environment. Also, the perception of learning ability will take time because learning is not a result but more as a process. It will be hard to perceive if the robot is learning or not by a single and short interaction. But, the perception of learning ability might be yielded by an explicit feedback as a response to the human behavior. Thus, this perception will be seen from the intelligence perspective. Robot which shows a responsive and/or correct feedback, which is considerably defined as a competent robot, will give better perception of learning. Bartneck et.al. developed 5 items in 5-point Likert scale, namely Incompetent/Competent, Ignorant/Knowledgeable, Irresponsible/Responsible, Unintelligent/Intelligent, Foolish/Sensible. The questionnaire has been tested with a satisfactory internal consistency.

2.5.3. Power/control

The third construct is the perception of behavior shaping power/control. It represents human perception of his/her influence on a robot. For instance, the changes of robot behaviors in the future which are affected by human feedbacks. Anderson et.al. defined this perception as a personal sense of power, which has meaning as the perception of one's ability to influence another person or other people [28]. The more influence that a person might give results higher sense of power towards another person/other people. Moreover, the perception of being powerful is not only yielded by using more power but also an ability to influence the other's behavior.

In study by Anderson et.al., personal sense of power is measured by Sense of Power Scale Items in Appendix A. The measurement tool is a questionnaire form with 8 questions using 7-point Likert scale and detail instruction based on relationship context (specific interaction, relationship, group, or generalized). In the form, it also covers all aspects of the personal sense of power as psychological constructs, which is described as follows:

1. Coherence to specific manifestation of power.
   There are several examples of manifestation of power which have been studied before, such as the ability to control joint decision, influence other's behavior, shape other's internal states, and satisfy one's own desires even it's conflicted to others' desires. In this study, the personal sense of power is coherent within a social-relational context, where people see themselves as more or less powerful along specific dimensions of influence.

2. Consistency across relationship context.
   Personal sense of power is considered as a relationship specific construct where it differs from one person to another. For example, personal sense of power of person 'A' towards his parents is different compare to his teacher or his good friend. Theoretically, there is an independent relationship between power and relationship. However, there are some studies show that individuals has consistency in power regardless the relationship context, such as the perception of power of leader across multiple group situations [56]. In the study by Anderson et.al., personal sense of power is moderately consistent across relationship context but shows a substantial specificity [28].

3. Generalization to multiple levels of abstraction.
   This study shows that people are able to form semantic perceptions of power in general,

across their relationships, and in groups. It includes four distinct levels of abstractions: in specific momentary social setting (single interaction), in long-term dyadic relationship, in a long-term group, and in generalized form. It shows that individuals who attain higher power levels of power have higher generalized sense of power.

4. Determinants as personal antecedents. This study proved personality variables play important role in the personal sense of power but social-contextual factors, such as social economic status, are not considered as a determinant factor to the perception of power. It shows that individuals who are more dominant have a tendency to have greater ability to influence others than introverted or submissive people.

There were 5 studies conducted in the experiment by Anderson, et.al., namely:

1. Power in long-term dyadic relationships
   This study is aimed to examine the individual's personal sense of power in context of close, long-term, intimate relationship. Some of examples are a child-parents and same-sex friend relationship. Result of the study shows that the sense of power was largely relationship-specific and sense of power to friends is higher than parents. This setup is not suitable in the experiment because it is designed for short-term interaction.

2. Power across multiple dyadic relationship contexts
   This is a leverage study from first study where number of relationships examined were broader. The study showed that individuals possess coherent beliefs about their power across specific manifestation of power although it is largely affected by relationship-specific. We assumed that it will be applicable in human-robot interaction as well.

3. Sociometric study of status hierarchies in social-living groups.
   This study examines position of person in larger group context. Someone might has a higher power than some group members but a lower power than others. The result of this study shows that the personal sense of power was related to the sociometric status. We will not examine this aspect in the experiment.

4. Experimental study of resource control in single interaction.
   This study shows personal sense of power in a single interaction with someone who has no prior experience before. Also, in this study, Anderson et.al. measured trait dominance and see the relationship with the sense of power. The results shows that individuals who belief having a higher control will also have a higher personal sense of power. It is applied to the experiment where the human will experience first time meeting the robot.

5. Power at generalized level: socioeconomic status and other power-related constructs.
   This study looked into the highest level of abstraction, which is the generalized level. It explored individual's beliefs about their power across relationships and group contexts. Also, it took a socioeconomic status into account as a determinant factor of personal sense of power. We will not apply this study to the experiment because we will focus more into study number 4.

2.6. Interaction Experience

The main purpose of this study is to find whether an growth and adaptive behavior can improve the interaction experience, through three types of perception that have been explained in the previous section. There are already several studies related to improving interaction experience in human-robot interaction domain. The term of interaction experience is still wide and can be interpreted in various ways. In this study, likability and affective gain were assessed as the way to measure an interaction experience.

Bartneck et.al., discussed likability as a measurement instrument which is also used as the key aspects in HRI [30]. Likability is a good impression that someone has against another person or object. Positive first impression of a person leads to more positive evaluations of that person [49]. It can be assumed that likability is also applicable to robots as social actors. The measurement instrument of likability was designed as a questionnaire of 5 items in 5-point Likert scale, namely Dislike/Like, Unfriendly/Friendly, Unkind/Kind, Unpleasant/Pleasant, and Awful/Nice. This questionnaire has been tested in two studies where results Cronbach's Alpha well above 0.7 which indicate that this questionnaire has a sufficient internal consistency reliability.

Another perspective of the interaction experience can be seen from the affective gain. Affective gain is defined as an increasing internalization of positive attitudes toward the content or subject matter [42]. Larger affective gain indicates better interaction experience. Affective learning can be measured by Affective Learning Scale which has been develop by McCroskey [1]. The measurement is situated in a learning environment where includes students and instructors. It addresses a problem of a single measure of affect which makes affect for content and affect for instructor were not considered separately. Later on, this perspective was not considered as a fit to the setup of the experiment and taken out from the measurement. Thus, we limited the interaction experience measurement only by the likability factor.

2.7. Algorithm and Model

2.7.1. Finite State Machine

Finite State Machine (FSM) is a common mathematical model of computation where the system can be in only one of a finite number of states at any given time. A FSM can move between states in response to some events or triggers. The movement between states is called a state transition. All transitions rely on the state transition without any probability of actions involved. FSM has an intuitive structure which makes easier to understand and implement in a sequential programming. However, it will result a problem for complex cases where many states are involved. An example of FSM can be seen in a simple grab-and-throw ball task in Figure 2.1. Moreover, Foukarakis did experiment on building an adaptable robot behavior by combining FSM and decision-making tools [39]. It has a task to find item in the user's house. It incorporates an adaptive factor by considering user preferences, for example the favorite places in the house.

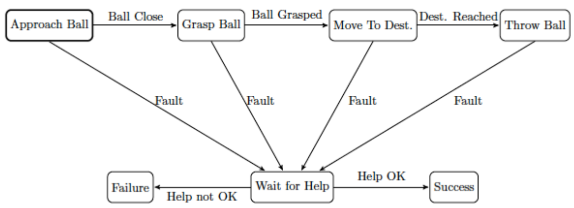


Figure 2.1: Finite State Machine example [37]

Using this model, all behaviors are modeled as states where each state has a trigger to move to another state. For example, sit and stand up are considered as states in the model, while user response by telling "please sit down" is modeled as a trigger to move to sit state. There is no probability of an action to move from one state to another state. This model only relies on a trigger action. For example, word "Dance" can be a trigger to move from any state to dance state. It is not possible to move to another state, such as idle state or talk state if this trigger has been started. The trigger is usually called as a state transition. If the state transitions are stochastic, the model is usually called as Markov Chain. Figure 2.2 illustrates example of two states with stochastic transition probability. For example, S1 has probability 0.2 to stay in S1 and 0.8 to move to S2. Considering its simplicity but yet powerful, this method was used in the implementation to model behavior of the robot.

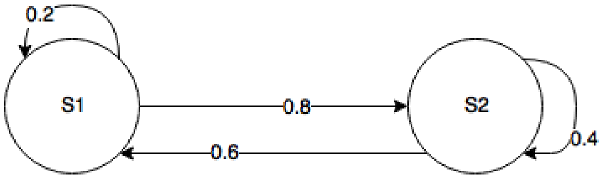


Figure 2.2: Markov Chain example

2.7.2. Markov Decision Process

Unlike Markov chain, Markov Decision Process (MDP) consists not only states but also actions and its reward to move from one state to another. The next state is determined only by the current state and current action, which holds Markov property. Each action has its own probability to move to another state and reward will be given upon the transition. Figure 2.3 shows the example of two states with two actions attached to each state. In this experiment case, the robot behaviors will be modeled as states, the user's feedback as actions, and a reward comes from the action state transition.

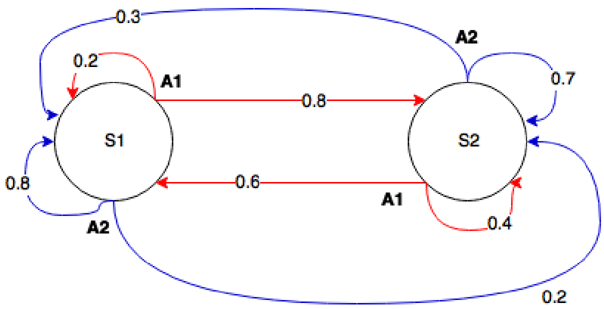


Figure 2.3: Markov Decision Process example

2.7.3. Partial Observable Markov Decision Process (POMDP)

POMDP is the extension of MDP with hidden states [51]. It happens where not all states (partially) are observable. It defines an optimal behavior for a given Markovian problem, taking into account uncertainty in observations as well as action effects over a potentially long time horizon [46]. This case commonly occurs in a real world where not all states are given and visible to an agent. For example, in a vacuum cleaner robot might not be able to see the condition of every corner of the house at one time. It only able to observe its surrounding and find the best way to find the dirt and clean it.

The difference between MDP and POMDP is on the state space, observation space, and belief. The state set in POMDP includes all possible states of the world, in which the agent is assumed to operate in. Agent is assumed not able to see all the states in one time. Observation space contains all observations made by the agent, when action a was executed and the world moved to the state *s'*. Also, POMDP has belief where represents state probability distribution. The goal of POMDP is to maximize the expected reward.

POMDP can be used in this experiment case to model the available observed states. It will give more efficient computation although it will not cover observations to all states. With the window, for example 3 states before and after, will make the computation more narrowed down and can give respond quickly. In this model, the state space is all possible behaviors performed by the robot and the observation space would be human nonverbal social behavior cues, such as user expressions. Moreover, the action space contains verbal responses from human, for example, ”yes”, ”no”, ”again”, ”dance”, ”sit”, etc.

3 Method

The research was an exploratory study which aimed to investigate how people perceive and respond to growth and adaptive behavior of a robot; and its relation to the interaction experience felt by human. This section will describe research questions, its hypothesis, the experiment setup, and the measurements used.

3.1. Research Questions and Hypothesis

RQ-1 - What is the effect of growth and adaptation on the perception of the robot being alive, the perception of the learning ability of the robot, and the perception of power/control?

Dependent variable(s): the perception of the robot being alive, the perception of the learning ability of the robot, and the perception of power/control.
Independent variable(s): condition groups.

Growth and adaptation manipulation is hypothesized can significantly improve the perception of aliveness, the perception of the learning ability of the robot, and the perception of power/control. In this study, the similar approach with the developmental capability [44] was implemented in a humanoid robot instead of animal-like robot which was used in the experiment by Lee et.al. We hypothesized that the improvement result of the lifelikeness perception and more enjoyable interaction could also be obtained even though in a different anthropomorphism. Moreover, the adaptation implementation successfully improved interaction experience [47] and also helped in achieving determined goals [41, 48]. We hypothesized that these successful evidences were influenced by the positive user perception towards the robot behaviors.

RQ-2 - Is there a correlation between the perception of the robot being alive, the perception of the learning ability of the robot, and the perception of power/control with the interaction experience?

Dependent variable(s): interaction experience
Independent variable(s): perception of the robot being alive, the perception of the learning ability of the robot, and the perception of power/control.

The perception of aliveness, the perception of learning ability, and the perception of control/power significantly and positively influence the interaction experience. Human will like more robot which is perceived as alive. In the study by Lee et.al. [44], a robot with development capabilities, which is perceived more alive, resulted better social presence and more positive social response. Moreover, a quality of the interaction will be better in a robot who gives the perception of high learning ability. By showing the ability to learn, it smoothen the interaction and create a better bonding by giving the feeling of being heard. Moreover, the perception of control/power to the robot is also arguably will increase a likability score because it tries to fit in to human preferences and give the impression of taking in charge compare to a condition where a robot does random or pre-scripted actions regardless the human feedbacks.

3.2. Experiment Setup

The experiment will be conducted in 2 (growth versus fully matured) x 2 (adaptive versus non-adaptive) conditions with a within-subject approach by coupling the condition to each participant. The four conditions are:

1. $NG-NA$: This is a control group where has condition of non-growth (fully matured) and non-adaptive.
2. $G-NA$: Growth and non-adaptive.
3. $NG-A$: Non-growth and adaptive
4. $G-A$: Growth and adaptive

Participants, who were assigned to the growth condition, interacted with the NAO who was programmed to gradually unlock its behaviors during the interaction. There are several options to determine an unlocking behavior: by number of states visited, number of interaction, length of interaction, and learning status of each state. In this study, the unlocking behavior is enabled by learning of all states in each stage in the interaction. For example, second stage of the interaction in Figure 4.2 will be unlocked after rolling and hand movement behavior have been learned. Moreover, an adaptation is represented by three conditions: explicit feedback by robot (happy or sad expression), temporary probability change, and long term probability change upon user's feedback. In the adaptive condition, the participants will experience the three conditions mentioned before, while the participants in the non-adaptive condition will not get any changes upon their feedback to the robot.

Each participant experienced two out of four conditions. Thus, there are 6 experiment pairs: $NG-NA$ & $G-NA$, $NG-NA$ & $NG-A$, $NG-NA$ & $G-A$, $G-NA$ & $NG-A$, $G-NA$ & $G-A$, $NG-A$ & $G-A$. The order of each pair will be swapped as well, which resulting a set of 12 different combinations. As an example, participant A will be assigned to $NG-NA$ & $G-NA$ and participant B will be assigned to a reversed order which is $G-NA$ & $NG-NA$. There will be 2 sets of these combinations. In a total, there will be 48 sessions with 12 sessions per condition.

Each participant will interact with the robot in about 40 to 45 minutes, consists of interacting with two different conditions, briefing, questionnaire, and open question. Every participant will enter the experiment room one-by-one. The experiment will be kicked off by signing a consent form and a short briefing on how the participant can interact with the robot, consists of an instruction and a short description what the robot can do, and also a notification about the questionnaire after the interaction. The interaction with the robot will be recorded for further analysis purposes. When the participant is ready, the administrator will start the interaction by choosing a condition in the WoZ control panel, fill in a participant name, then click 'Introduction' button. In the end of the experiment, each participant will be asked to grade some statements in the questionnaire and answer one open question related the interaction.

3.3. Participant

24 participants from TU Delft joined the experiment where each participant experienced two different conditions. There are 2 main reasons why target participants were chosen: (1) it is easier to predict an adult's response, which can be a good pilot study before testing to children; (2) time and resources constraint which makes visible to limit scope of the testing to TU Delft environment. Also, adult has abilities to comprehend language and their verbal memory. Thus, they can understand the follow-up questionnaire and give reasoning of what they perceived during the interaction. In addition, the participants were assigned randomly

to the condition. The random assignment also took a gender distribution into account to minimize the gender bias effect.

3.4. Measurement

There are 4 aspects in the measurements: the perception of aliveness, the perception of learning ability, the perception of power, and the interaction experience. Both quantitative and qualitative methods are used to measure these perceptions.

3.4.1. Quantitative

Quantitative method covers a questionnaire, a length of the interaction, and a personalization of state transition probability.

Proposed questionnaire can be seen in Appendix C. The questionnaire is developed from a Godspeed questionnaire by Bartneck et.al. [30] for the perception of aliveness, the perception of learning ability, and the interaction experience self-report; the questionnaire by Anderson et.al. [28] is used for the perception of power. Personal sense of power is added to Godspeed questionnaire because it does not cover the perception of power in the interaction. The measurement is conducted in the statements with 5-point Likert scale from agree to disagree. In a total, there are 27 statements, which is divided into 10 statements on the perception of aliveness, 4 statements on the perception of learning ability, 8 statements on the perception of power, and 5 of the perception of interaction experience. Some of the statements were made reversely, to make the statement easier to understand and validate user's understanding of the statements, such as ”robot moves rigidly”. The original statement was designed with score 1 if human finds that robot moves rigidly, but in this statement, it indicates the robot moves more elegantly/smoothly.

Besides the questionnaire as a self-report measurement, we will also look at overall duration of interaction. Longer duration of the interaction shows that the interaction works in a mutual way and each individual is interested to interact to each other. In addition, personalization of state transition probability shows the changes of robot behaviors based on user input.

3.4.2. Qualitative

Open question will be asked to each participant after the questionnaire session. The question is rather a general opinion about the interaction: ”Do you have any other thoughts on the interaction?” This question is aimed to capture any other thoughts which are not represented in the questionnaire.

4 Implementation

4.1. Tools and Software

The NAO robot is chosen as the embodiment to execute growth and adaptive behaviors. It was selected because it has functionalities that support adaptive and growing behaviors. Also, it has been successfully implemented in several studies to display adaptation [41, 47, 48]. NAO is a humanoid robot which was developed by Aldebaran Robotics and later on acquired by SoftBank. It has senses and actuators for natural interaction, namely moving, feeling, hearing, speaking, seeing, connecting, and thinking. By using these functionalities, it might give the expected implementation result. The NAO can be developed using Choreograph application, which is developed by SoftBank, or programmed using C++ or Python.

A computer was used to program the behavior design of the robot with a minimum specification needing a Python library and a NAOqi library. In this experiment, a DELL computer was used with the Windows operating system, Intel Core i7, and Python version 2.7.13 installed in the system. The implementation of this experiment used Python as programming language with capabilities to have API to the NAO robot using `naoqi` library. NAOqi is the name of a main software that runs and controls the robot. It also has programming framework used to program NAO [24]. To use it, `naoqi` SDK should be installed in the local machine and import it to the scripts. Moreover, to test whether the script is working well in NAO, a virtual robot in Choreograph is used for most of the functionalities, except the speech recognition and the face recognition. It needs input from NAO I/O device for these two functionalities. In addition, PyCharm IDE was chosen as programming tool because of its easiness to use for Python project with built-in Python interpreter. Moreover, a Pytransition library is used to implement a finite state machine in an easier way. It is an open source project which contains an object-oriented state machine implementation in Python [25].

In the experiment, a set of camera recorder was used to record the interaction and Typeform [1] was employed as a questionnaire platform. The video was used to validate the result when there is a doubt related to the result. Typeform was utilized to make the data gathering from the questionnaire became easier by the ability to generate the automated questionnaire result. Moreover, the questionnaire was designed to be private.

4.2. Architecture of The System

The main process of implementation was performing a set of behaviors. It was chosen based on the probability state transition. Speech was used as a cue to change the probability in the adaptation implementation and to indicate if next stage should be unlocked, as illustrated in Figure 4.1. The NAO behavior implementation used all basic APIs provided in naoqi module which are described below:

1. ALSpeechRecognition was used for the speech recognition. In the initiation, set of words

[1] https://www.typeform.com/

were defined as a dictionary. The threshold used to recognize words in predefined words was 0.4.

2. ALMotion was used to define all motions related.
3. ALMemory was used to store passed value from the program into robot memory, which is important to subscribe events, such as a function to process recognized words.
4. ALTextToSpeech was used to translate text to speech.
5. ALAnimatedSpeech was used to perform a small body movement while speaking.
6. ALRobotPosture was used to define posture of the robot in the existing state.
7. ALFaceDetection was used to track face of participant in walking towards state and when reacting to tactile touch.
8. ALAudioPlayer was used to play a music in dance state.
9. ALLeds was used to change color of LED eyes when the robot send message of positive or negative feedback.

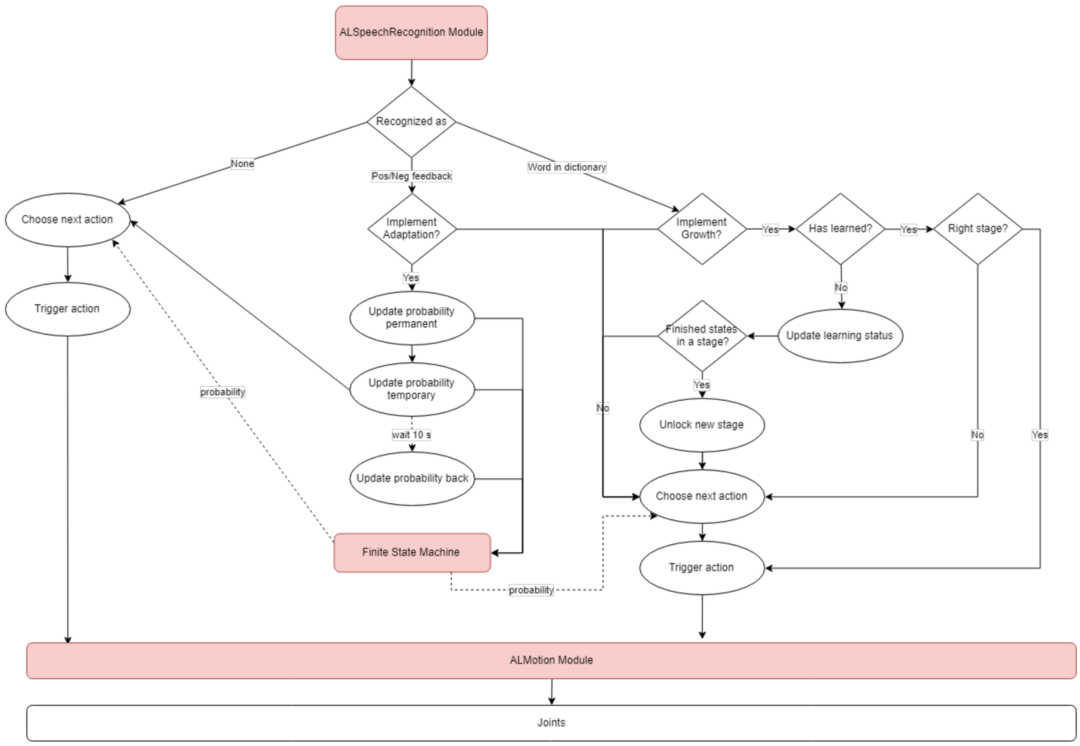


Figure 4.1: Architecture of the system

4.2.1. Adaptation

Adaptation was implemented based on the response received from a user in this set of words:

- Positive: yes, nice, good
- Negative: no, bad

If one of the words in the set list of positive/negative response is recognized, then the probability from the previous state to the current state (behavior that being praised/criticized) is changed permanently, using alpha (α) as a learning rate. In this study, a learning rate 0.6 was used to calculate a new weight. The greater α will make the learning faster which means probability to choose or not to choose to a certain behavior will change drastically. The learning rate α was also introduced to avoid a zero probability of the state transition as a reaction to the negative response. The transition probability from one state to another state in a stage will not be zero. It is designed to allow the participant to experience the least preferable behavior because he might not like the behavior at the moment but it might change in the future. The calculation follows a rule in Equation 4.1 for a positive response and Equation 4.2 for a negative response. The weight transition is stored in a configuration file which will be described more in subsection 4.5. Then, a temporary probability change will be performed to ensure that the next behavior chosen will/will not be the same by changing weight into 1000 for a positive response and 0 to a negative response which is shown in Equation 4.3 for 10 seconds. The probability matrix 4.5 is calculated whenever there is a change using the formula in Equation 4.4.

$$W(s,s') = \begin{cases} 1 & \text{if } W(s,s') = 0 \\ (1+\alpha) * W(s,s') & \text{if } W(s,s') > 0 \end{cases} \tag{4.1}$$

$$W(s,s') = (1-\alpha) * W(s,s') \tag{4.2}$$

$$W(s,s') = \begin{cases} 1000 & \text{if positive response} \\ 0 & \text{if negative response} \end{cases} \tag{4.3}$$

$$P(s,s') = \frac{W(s,s')}{\sum_{i=0}^{n} W(s,s_i)} \tag{4.4}$$

$$P_{m,n} = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m,1} & p_{m,2} & \cdots & p_{m,n} \end{pmatrix} \tag{4.5}$$

4.2.2. Growth

At the beginning, all states are tagged as *not learned yet*. It will change the tag as *learned* when a word associated to one of the behaviors is recognized. Then, it will process the recognized word as a command to the robot to perform the associated behavior by calling a function attached to the word (only if it has been learned and in the right state, please refer to 4.2). If all states in a stage have been learned, the robot will unlock the next stage by manually choosing a basic behavior in the next stage and add weight by 1 from the basic behavior from the previous stage to the next stage and vice versa. As an example, it will choose sit when all states in first stage: lie down, playful hand, and roll have been learned and add weight by 1 from lie down to sit and also sit to lie down. A value 1 was chosen as an initial weight to enable the transition between two stages, but later on, it can also be shaped through the adaptation mechanism whenever positive or negative response is received. Another reason to use an initial weight as 1 is to make a higher probability to explore all states within the next stage.

4.3. Interaction Model Using Finite State Machine

The interaction model was designed in the developmental stages. There are 4 main stages, which represents 4 developmental capability levels, with basic behavior states: lie down, sit, stand, and walk. These main basic behaviors indicate the growth step. Movement from one stage to another stage is only available through basic behavior states. In each stage, there are two other states with similar movement to its basic behavior. Each state has transition probability to every states within same stage. Between one stage to the next stage, there is also a transition probability which will be unlocked once all states are learned in previous stage which has been explained before. The *learning* process was indicated by the recognized user's word associated to the current behavior. For example, when the robot rocks the body, it will say *"rock"* or *"dance"* while performing the behavior. If the user repeats one of the word associated to the behavior, it will change the tag of the behavior as *learned*.

The implementation of choosing next behavior applies Finite State Machine approach where each behavior is treated as state. The set of behaviors is finite, which means already predefined before. There are in total 12 states consist of 3 states in each stage. The behavior design is illustrated in Figure 4.2. The behavior states were designed to illustrate growth of behavior complexity and showing progress from lying down to sit to stand to walk. All states are represented as hierarchical stages where the behaviors in the next stage can still access the behaviors in the previous state(s). Every state is connected with each other but the moving transition is based on the weight of transition, which has been explained in 4.4. For example, transition from sit to point is weighted as 2, sit to rock is 1, and other transitions from sit equal to zero will result probability from sit to point 66,67% and sit to rock as 33,33%.

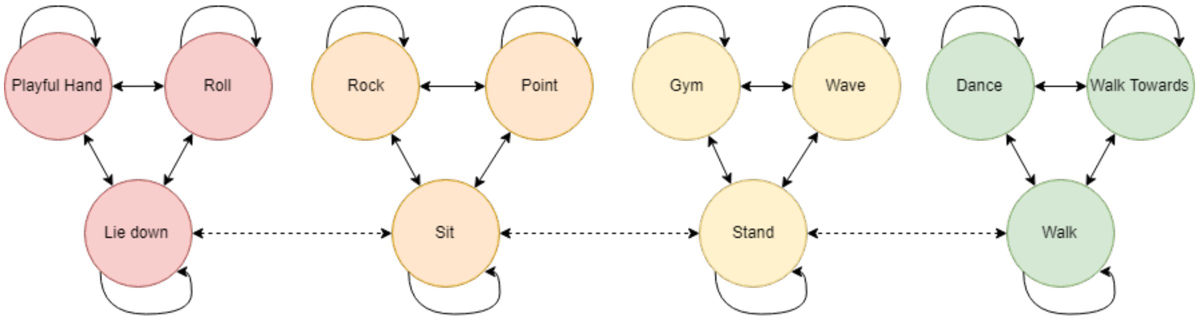


Figure 4.2: Interaction Behavior Model

4.4. Functionality modules

There are two main modules for the implementation: NAO related functionalities and control panel.

4.4.1. NAO Functionalities

All NAO functionalities are written in NAO class in Nao.py. It contains scripted behaviors, data loading, data transformation, data processing, and data writing. There are couple functionalities inside the class:

- Object creation/deletion

  `__init__`

  `__del__`

- Speech recognition

  `onLoad`

  `onUnload`

  `onInput_onStart`

  `wordRecognized`

- Touch recognition

  `subscribeAllTouch`

  `unsubscribeAllTouch`

  `onTouched`

- State transition related

  `update_journal`

  `goToNextState`

`getStateIndex`

`getFunctionName`

`calculate_prob`

`getProbabilityTransition`

`changeProbability`

`changeProbabilityOption`

`changeTemporaryProbability`

`changePermanentProbability`

- Configuration file related

  `loadConfigFile`

  `loadFunctionFile`

  `updateConfigJson`

- Behavior related

  `happyFace`

  `sadFace`

  `lieDown`

  `roll`

  `playfulHand`

  `sit`

  `rock`

  `point`

  `stand`

  `gym`

  `wave`

  `walk`

  `dance`

  `walkTowards`

4.4.2. Control Panel

Control panel is used as a Wizard-of-Oz (WoZ) panel to prepare the initial setup and manage the robot's behavior manually. It is important to have it to start the robot and also make an interference to the robot in the unexpected cases. The screen shot of the panel can be seen on Figure 4.3. There are 3 main areas:

1. Condition group
   There are 4 different radio buttons which represent the different conditions described in Section 3.2. The choice is mandatory to set up the experiment configuration based on a certain group. For example, in $G-A$ and $G-NA$ group, the initial configuration file loaded is based on growth condition (only first stage transition probability is enabled) while a configuration file with a complete and equal probability is loaded in $NG-NA$ and $NG-A$. Besides the initial transition state probability distribution, the selection will also activate the adaptive features, which are an explicit feedback (happy or sad face upon the feedback from the user), a temporary probability change, and a long term probability change.

2. Name
   This area is to define a participant name. Each participant will have his/her configuration file which yielded after the interaction for the participants who experienced the adaptive condition, which are $NG-A$ and $G-A$. Name will be an identifier of configuration file. The file contains weights of state transitions which shaped along the interaction.

3. Button
   This area contains *Introduction* button to start the interaction and other help buttons whenever robot performs unexpected cases. Buttons which are implemented are: introduction, restart, pause speech recognition, random state generator button to select manually next state, courage button to select manually next state with positive feedback from the participant, and discourage button to select manually next state with negative feedback from the participant.

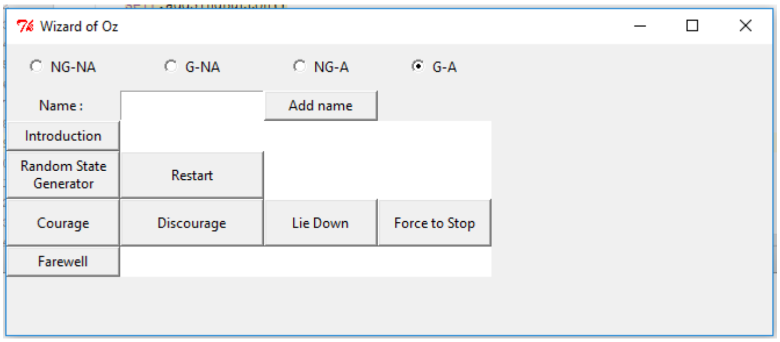


Figure 4.3: Wizard-of-Oz Control Panel

4.5. Configuration files

Configuration files are used to set the initial setting of the interaction. It is written in a json format. There are 2 types of configuration files: weight probability and function configuration file. The weight probability configuration file is used to determine an initial probability when the robot starts the interaction. In this implementation, there are two kinds of initial probability setup: growth and non growth, where only weight in first stage is available for the growth experiment group, while all transition weight with equally 1 value are assigned for the non-growth condition. The second configuration file is used to set a name of the function needs to be called in each state. User can define what kind of behavior will be shown in a state by calling the respective function. The structure of both configuration files can be seen in Figure 4.4, where left side shows weight configuration json file and right side shows function configuration json file. If there are additional states want to be added to the interaction design, the administrator can add the data to the both configuration files by providing transition weight and name of function to be called. If the function has not been implemented in the NAO program, user should also add the function to NAO.py file.

```json
[
  {
    "initial_id": 1,
    "initial_name": "lie_down",
    "class_id": 1,
    "target": [
      {
        "target_id": 1,
        "target_name": "lie_down",
        "transition_prob": 1
      },
      {
        "target_id": 2,
        "target_name": "roll",
        "transition_prob": 2
      },
      {
        "target_id": 3,
        "target_name": "react_movement",
        "transition_prob": 2
      },
```

```json
{ "function": [
  {
    "id": 1,
    "name": "lie_down",
    "level": 1,
    "function": "lieDown()"
  },
  {
    "id": 2,
    "name": "roll",
    "level": 1,
    "function": "roll()"
  },
  {
    "id": 3,
    "name": "react_movement",
    "level": 1,
    "function": "reactMovement()"
  },
  {
    "id": 4,
```

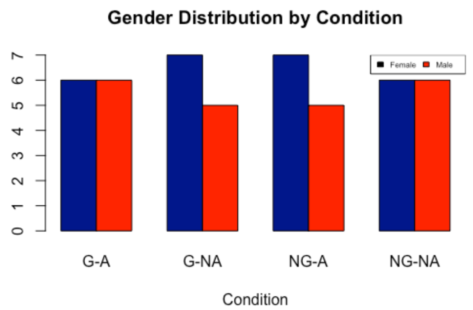

Figure 4.4: Initial Configuration File in Json Format

The structure of the transition weight json file consists of initial_id as an identifier for the initial state, initial_name as a name of the initial state, class_id as an identifier of the stage, target as a list of the target state, and there are target_id as an identifier of the target state, target_name as a name of the target, and transition_prob as a transition weight of the initial state to the target state inside the target parent. Similar structure is also implemented for the function configuration file. Function configuration file consists of a list of states to the functions mapping. In each mapping, there is an information about id as an identifier of the mapping, name as a name of the state, level as a level of the state, function as a name of the function to be called in the state.

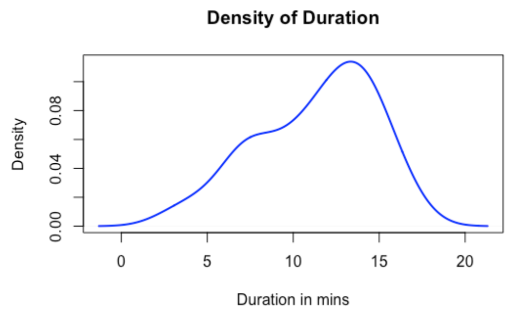
5

Result and Data Analysis

This section explains the gathered data from the experiment and various analysis conducted to answer the research questions. There are two kind of data gathered: quantitative and qualitative data. Quantitative data was yielded from the questionnaire result where it has 27 questions related to the perceptions and the interaction experience and also duration of the interaction. There were in total 25 participants in the experiment with 1 person was taken out from the data because of a battery issue during the experiment. So, the final gathered data consists of 48 (24 participants x 2 times interaction) set of questionnaire answers with the distribution of gender was 54% female and 46% male. The gender distribution per condition was set to be as even as possible which can be seen in Figure 5.1a. The age of participant varies within range of 21 to 28 years old. Related to the experience interacting with NAO, more than half (13 persons) have been interacted with NAO in another occasion before. The duration of interaction also varies a lot, from 3 minutes to 17 minutes, which can be seen in Figure 5.1b. These factors are also considered as contributing factors in the analysis. On the other hand, qualitative data was obtained through an open question after the interaction and also an observation during the interaction.



(a) Gender by Condition



(b) Density of duration

Figure 5.1: Gender and Duration

The data analysis starts by checking the coherence of the questionnaire. The questionnaire was built from the Godspeed and the Power Scale Item questionnaire. Both of the questionnaires have 5-Likert scale but have different purpose and approach. Godspeed is to study a human-robot interaction with a rating of two opposite statements, for example Machinelike to Humanlike. While, the Power Scale Item is intended to study the perception of power which was used in a human-human interaction before using a rating from strongly disagree to strongly agree. In this study, Cronbach's coefficient α was used to calculate the internal consistency coefficients of each construct (perception of aliveness, perception of learning ability, perception of power, and interaction experience) included in the questionnaire with 24 participants. Results of the reliability analysis showed that the items in the 5-Likert scales had a satisfactory discriminating power. The α values of the 4 constructs were: the perception of aliveness 0.85, the perception of learning ability 0.84, the perception of power 0.83, and the interaction experience 0.86. Therefore, the results indicated a satisfactory level of construct validity and an internal consistency of this modified questionnaire. Later on, the average value of each construct will be used as observed variables instead of individual question scores because we know that the questionnaire is consistent within each construct. The average score was calculated from all questionnaire answer score within one construct, as an example, the aliveness value was calculated as average of 10 statement scores in the aliveness section of the questionnaire.

5.1. RQ-1: Influence of Growth and Adaptive Behavior to The Perceptions

To analyze the growth and adaptive behavior effect, the dataset of 24 subjects was divided into two different datasets, namely adaptive and growth dataset. Adaptive dataset consists of all subjects who experienced working with two different adaptation condition, while growth dataset consists of all subjects who experienced working with two different growing behavior condition. Table 5.1 shows the data distribution to each dataset. Thus, each of the new dataset has 16 subjects records. There are two conditions, NG-NA G-A and NG-A G-NA, overlapped both in adaptive and growth dataset. Then, the analysis was performed separately for growth and adaptive behavior towards each construct.

Table 5.1: Dataset Division

Condition 1	Condition 2	Adaptive	Growth
G-NA	G-A	✓	
NG-A	G-A		✓
NG-NA	G-A	✓	✓
NG-A	G-NA	✓	✓
NG-NA	G-NA		✓
NG-NA	NG-A	✓	

To repeat from the previous section, there are 4 constructs used in the analysis: the perception of aliveness, the perception of learning ability, the perception of power/control, and the interaction experience. The interaction experience score was also included in the analysis to see if there is direct correlation of the manipulation. As a legend, row aliveness refers to the perception of aliveness, learning refers to the perception of learning ability, power refers to the perception of power, interaction_experience refers to favorableness score of interaction with the robot. Descriptive statistic was examined by seeing a marginal mean and a standard deviation using Bonferroni confidence interval adjustment between adaptive versus non-adaptive and also growth versus non-growth condition. This statistic result was covered in a pairwise comparison to see whether or not there is a statistically significant finding of the experiment modification on the adaptation and growth to the measured perceptions. The result is displayed in Table 5.2. By examining the marginal mean of the constructs individually, we can see that perception of power has large mean difference between the growth and the non-growth condition. On the other hand, there is only a small marginal mean difference in all constructs in adaptive versus non-adaptive condition. This information can be a good guess to see determinant variables for further analysis. In the interaction experience construct, there is no significant difference but the average score and its lower bound has the highest value compare to other constructs. The participants thought the interaction was enjoyable, regardless any condition that they experienced in general. It is shown by a relatively high score given by them compare to a low score on the perceptions that they got from the robot. The interesting insight was shown in Figure 5.2 and Figure 5.3 where the modification implementation has a positive effect to most of measured constructs, except the perception of learning ability and the perception of power in the growth condition. Figure 5.3 upper-right and bottom-left show that non-growth condition was perceived better in terms of learning ability and behavior shaping control.

Table 5.2: Pairwise Comparison

						95% Conf. Interval	
Measure	Cond 1	Cond 2	Mean diff	Std.Err	p-value	Lower B.	Upper B.
aliveness	A	NA	0.025	0.148	0.869	-0.291	0.341
	G	NG	0.144	0.181	0.439	-0.242	0.529
learning	A	NA	0.078	0.195	0.694	-0.337	0.493
	G	NG	-0.031	0.216	0.887	-0.492	0.430
power	A	NA	0.039	0.144	0.790	-0.268	0.346
	G	NG	-0.258	0.167	0.143	-0.613	0.098
interaction experience	A	NA	0.075	0.162	0.650	-0.271	0.421
	G	NG	0.150	0.196	0.456	-0.268	0.568

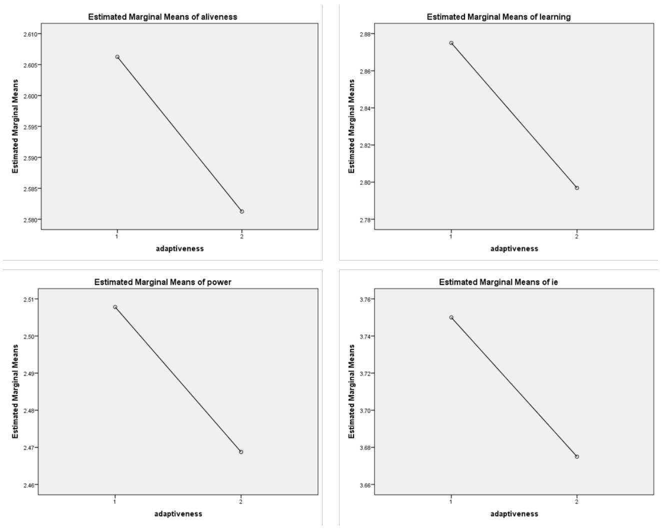


Figure 5.2: Marginal means of adaptive vs non-adaptive condition

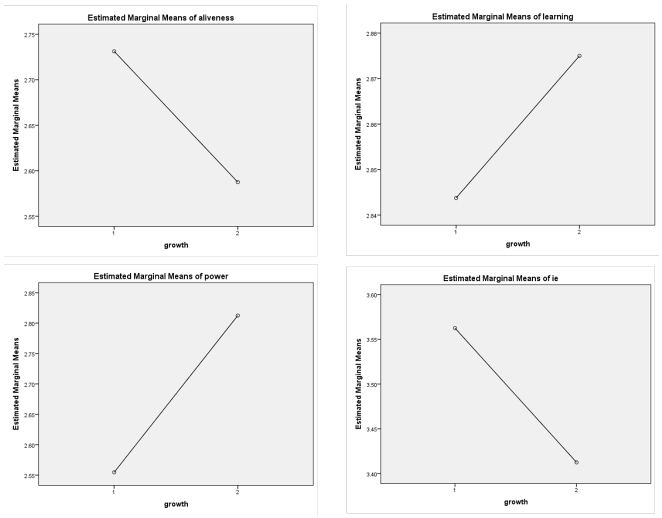


Figure 5.3: Marginal means of growth vs non-growth condition

Repeated measure analysis of variance (ANOVA) was conducted to test the difference means in multiple dependent variables using within subject approach. In the analysis, $\alpha = 0.05$ was adopted as the significance threshold. Multivariate within-subject effect result of both the adaptive dataset and the growth dataset showed non significant effect with p-value equal to 0.993 and 0.201 accordingly. Therefore, we can conclude that there is no significant effect of the growth and adaptation implementation in the experiment towards the perception of aliveness, the perception of learning ability, and the perception of behavior shaping control. Then, univariate test was performed to see if there is a specific contract that have significant effect individually. Table 5.3 shows the univariate test result of a within-subject contrast where no significant effect found neither in the adaptive nor the growth condition for all of the measured perceptions. The most prominent effect shown among the result is the effect of the growth implementation towards the perception of power with p-value = 0.143.

Table 5.3: Univariate Tests of Within-Subjects Contrast

Control Variable	Construct	Mean square	F	p-value
Adaptiveness	aliveness	0.005	0.028	0.869
	learning	0.049	0.161	0.694
	power	0.12	0.073	0.790
	ie	0.045	0.214	0.650
Growth	aliveness	0.165	0.631	0.439
	learning	0.008	0.021	0.887
	power	0.532	2.391	0.143
	ie	0.180	0.584	0.456

5.2. RQ-2: Correlation Between Measured Perceptions and Interaction Experience

To analyze the correlation among the perceptions and the interaction quality, linear models were developed to predict how enjoyable the interaction from the interaction experience score based on the measured perceptions. There are two different linear models developed using backward method:

1. Model 1: predicted variable: average interaction experience; predictors: average of perception of aliveness, perception of learning ability, and perception of power in the questionnaire.

2. Model 2: model 1 with interaction among these three perceptions) as the additional predictors.

Backward method was used to throw out one predictor at one time and calculate the significance of remaining predictors until it finds significance of the predictor(s) in the model. It is useful to see which of the construct(s) has more correlation to the interaction experience.

In model 1, it gave a result $F(3,20) = 3.709$, p-value = 0.029, residual standard error = 0.598, multiple R-squared = 0.357. As its p-value is less than 0.05, we can reject the null hypothesis and conclude that the perceptions gave a better fit than intercept-only model. The backward method used in the regression model was started by taking out perception of power and then perception of learning. So, there were three linear regressions run, namely:

(a) Predictors: perception of aliveness, perception of learning ability, and perception of power.

(b) predictors: perception of aliveness, perception of learning ability.

(c) predictors: perception of learning ability

Moreover, the corresponding coefficient variables of the model 1 are shown in Table 5.4 which display a significant effect of intercept and the perception of learning ability to the model with p-value ≤ 0.05 when the predictor is only perception learning ability left. It indicates that this variable is significant to determine the quality of interaction, while other variables, namely the perception of aliveness and the perception of power do not carry much weight. The estimated coefficient value of the perception of learning ability is positive which displays positive correlation between this perception and the interaction experience.

Table 5.4: Coefficient of Linear Model 1

Model	Variable	Est.	SE	tvalue	pvalue
a	(Intercept)	1.789	0.589	2.991	< 0.05
	aliveness	0.331	0.398	0.833	0.414
	learning	0.444	0.305	1.457	0.161
	power	-0.121	0.260	-0.465	0.647
b	(Intercept)	1.778	0.586	3.031	< 0.05
	aliveness	0.267	0.366	0.730	0.474
	learning	0.397	0.282	1.407	0.174
c	(Intercept)	2.013	0.485	4.154	< 0.05
	learning	0.561	0.169	3.322	< 0.05

In model 2, the interactions among variables are taken into account. The model resulted $F(7,16) = 1.992$, p-value = 0.12, residual standard error = 0.602, multiple R-squared = 0.108. As its p-value of overall significance test is greater than 0.05, we can conclude that the model does not provide a better result than the intercept-only model. Also, there is no significant effect of each perception score (p-value greater than 0.05) on the correlation to the interaction experience score, including the interaction term among the measured perceptions. So, we can conclude that the interactions among the perceptions are not statistically significant. It means that each perception is independent and not significantly correlate to each other. Thus, we can conclude that model 1 is the better fit compare to model 2 with interaction as predictor.

Furthermore, we would like to see if the interaction length can be used as a sign of interaction experience where the longer the interaction indicates the more enjoyable the interaction is. Repeated measure ANOVA was also performed both in the adaptive and the growth condition to see significance of the duration of interaction in within-subject. The result shows no significance of the duration with p-value in the adaptive versus non-adaptive condition is 0.509 and p-value in the growth versus non-growth condition is 0.203. Then, the length of interaction between first and second condition was averaged because there is no significant effect in either the adaptation or the growth condition. A linear regression is used to see if it

Table 5.5: Coefficient of Linear Model 2

Variable	Est.	SE	tvalue	pvalue
(Intercept)	3.384	0.169	20.077	< 0.05
aliveness	0.413	0.428	0.964	0.349
learning	0.426	0.399	1.068	0.302
power	-0.094	0.291	-0.324	0.750
aliveness:learning	-0.143	0.704	-0.202	0.842
aliveness:power	0.384	0.733	0.524	0.607
learning:power	0.380	0.545	0.697	0.496
aliveness:learning:power	-0.233	0.644	-0.361	0.723

has correlation with the interaction experience. A model was built by using the interaction experience as a dependent variable and the duration as a independent variable. From the regression, it displays there is no significant effect of the duration to the model with p-value = 0.313. Hence, we can conclude that the duration has no significant effect to how enjoyable the interaction perceived by the participants. Figure 5.4 supports the insignificance result by showing there is no linear correlation between the duration and the interaction experience score. Some participants felt the interaction was pleasant even they interacted only in a short time.

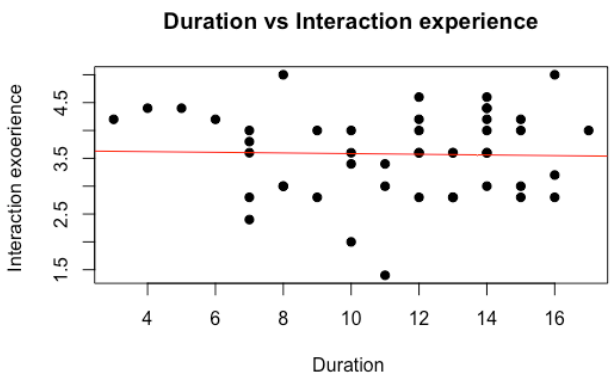


Figure 5.4: Correlation between duration and interaction experience

5.3. Probability Changes

Probability changes during the interaction display the adaptiveness implementation of the robot to the humans feedback. The changes only occurred in the adaptive condition and triggered by a positive or a negative feedback from the participant. The initial weight matrix is illustrated in Equation 5.1 where the row represents an initial behavior state and the column represents a target behavior state. *Yes, good, nice, again* are classified as positive feedbacks, while *no, bad* are considered as negative feedbacks. Once a positive or a negative feedback was received, the NAO will respond to it by saying an explicit feedback, *"Glad you like it"* for a positive feedback or *"I'm sorry, I will not do it again"* for a negative feedback. Then, probability of going to the same state again will be updated, both temporary and permanently. The probability changes in this section refers to the permanent probability change that is shaped by the participant along the interaction. In each interaction, the probability changes can be unique depends on what a participant acted towards the robot's behavior. The example of probability changes of one participant in the growth and the adaptive condition is drawn in Equation 5.2.

From the example below, we can see that there are changes in the weight of state transitions which were caused by participant's feedbacks. As an example, a participant did not like when the robot performs the same behavior of playful hand, which is shown in third row and third column. The weight was decreased from 2 to 0.512 which made the probability to repeat the playful hand behavior is less likely happen compare to performing roll behavior in third row and second column. Although a negative feedback was given to a state, it still has probability to move to this state again in the future because the weight was designed to be reduced but always greater than 0.

$$P_{m,n} = \begin{pmatrix} 1 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 2 \end{pmatrix} \tag{5.1}$$

$$P_{m,n} = \begin{pmatrix} 1 & 2 & 3.2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.16 & 2 & 0.512 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 2 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 3.2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 0.32 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0.8 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0.64 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.64 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 2 \end{pmatrix} \tag{5.2}$$

5.4. Qualitative

Qualitative result was obtained from an open question at the end of the experiment. The question covers another thought on the interaction which was asked to the participants to dig in other comments or feedbacks regarding both interactions. Some of the prominent findings from the qualitative results are described below:

1. Some participants did not know exactly what to do in the interaction. The guideline in the beginning about free interaction with the robot, including touching and talking to the robot, seems not so clear to them.

2. Some participants can distinguish between two conditions, especially when they experienced adaptive and non-adaptive condition. They hinted that the explicit feedback from the robot helped them to understand that robot reacts based on their feedback.

3. Most of participants thought that they can command the robot instantly. It raises frustration when the robot could not recognize their speech and did not do what they told the robot to do.

4. Most of participants did not know when to talk to the robot and just spoke to the robot based on their guts.

5. Some participants also do not understand about the unlocking behavior although they successfully unlocked to the next level.

6. Some behaviors are perceived as creepy or scary, especially when the robot produced some sounds, either a crying sound or a movement sound from joint. It also includes when robot fell down or almost fell down.

6
Discussion and Recommendation

The data analysis shows unexpected findings and an interesting insight in a response to the research questions. Due to the nature of this experiment as an exploratory experiment, this chapter will discuss how the growth and adaptive behavior implementation in a physical embodiment robot based on the result and analysis drawn from the previous chapter. Furthermore, this chapter will discuss the result in a broader context, including possible factors that influenced the experiment result, especially the factors which possibly led to the result that contradicts with the hypothesis. In the end of this chapter, there will be recommendations for the future research and the next development iteration for the developed system.

6.1. Insignificance Effect of Adaptive and Growing Behavior to The Measured Perceptions

First hypothesis postulates there is a significant effect of growth and adaptive behavior toward the perception of aliveness, the perception of learning ability, the perception of power. From the analysis of the experiment data, there is no significant effect of growth and adaptive behavior manipulation in the experiment toward how the participants perceived the measured perceptions mentioned before. But, the adaptive behavior has a positive impact to obtain better perceptions. There are many things that lead to the insignificance result, such as the brief explanation at the beginning, type of the research (within instead of between), the behaviors choice, robot malfunctions, and type of the setup.

The initial briefing was conducted at the beginning of the experiment which its manuscript is written in Appendix D. In the briefing, the main goal of the interaction was not mentioned and only the high level guideline was given to each participant. The reason is not to make participant feel biased during the interaction which might lead to a superficial perception. But, the goal of the interaction was explained in the end of the interaction while asking for improvement suggestions. The high-level briefing was perceived as an unclear instruction to interact with robot. This unclarity might affect the participants on focusing to main implementation manipulation which was designed to answer the research questions. Thus, it can explain why there is no significant effect of the implementation of growth and adaptive behavior condition to the perception of aliveness, the perception of learning ability, and the perception of behavior shaping control.

Another possible reason of the insignificance result is the research design which has been using within-subject approach. The within-subject approach was chosen to see the significance effect among the conditions namely adaptive versus non-adaptive and growth versus non-growth. In addition, it was also influenced by a limitation of time and resources to gather more subjects for conducting between-subject experiment. So, the experiment design applied *semi* within-subject where one subject only experienced two out of four conditions considering the boredom effect of doing all 4 conditions sequentially. Even though the distribution of conditions are equal and flipping order of condition has been implemented, this setup might also affect how participants perceived the interaction. In the common setting, the improved condition, in this case the adaptive and the growth condition, is usually put in the later order after the control condition. This order effect might also influence the result. Moreover, the opportunity to experience only two of the possible condition groups also limited the ability to compare throughout all conditions. It makes the analysis loses its power to analyze which variable, adaptation or growth, has a more prominent effect to the interaction experience and do a full comparison among conditions.

In the experiment, there were 12 implemented behaviors. For the short interaction, this number was also causing boredom effect. The participant interacted with the robot which performed the similar behaviors for several times. It might influence how they perceived the aliveness of the robot. But, a more advanced implementation, such as implementing chat bot, might cause a bias in the perception of learning ability because the participants can the impression of an intelligent robot regardless the condition that they experience. Thus, this feature was not implemented in the experiment.

Next thing that might influence the insignificance of the result is robot malfunctions. During the experiment, the movement of the robot is not fully smooth. It sometimes did stiff movements and fell down because of hot motor joints. It could affect the perception of aliveness of the robot because there are several movement-related statements in the questionnaire, for example machine-like versus human-like or move rigid versus move elegantly. It will make the participants perceived that the robot is not alive enough. Another aspect in the malfunctions that might highly influenced the perception is a misclassification in the speech recognition. Although the speech recognition works quite well, there were still some misinterpretation of the speech recognition to classify some words as a positive or a negative feedback. This problem leads to unnecessary probability changes. It also made participant frustrated when the robot recognized a positive feedback when user did nothing, which led to going back to the same behavior all over again. Moreover, it gave the impression that the robot is not smart and do not want to follow the participant's feedback. This problem could lead to a low score on the perception of learning ability and the perception of power.

The setup design can be distinguished into a subject choice related and the duration of the interaction. The choice of subjects might also give a prominent effect to the experiment. The initial study was intended to child-robot interaction. Due to time constraint, this study was shifted to adults as target participants as an exploratory study. Adults have more expectations when it comes to interacting with robot, especially if they already had some experiences interacting with NAO. This expectation set a standard of what the robot can or cannot do. In the case of a robot performs below their expectations, they will grade low to the perceptions and the interaction experience no matter the different condition given to them. Also, children usually have a more free will style compare to adults. Adults usually follow the instructions and react based on their experiences which limit the possible actions that they can give to the robot. Moreover, the interaction is rather short to understand what the robot can do. The longer interaction per condition might give a more opportunity for the participants to explore the robot's behaviors, especially related to the implicit changes in the adaptation condition.

Talking about the result, the perception of power in growth versus non-growth has the most significant effect compare to other perceptions in the same or different condition. It might be affected by the ability to process user commands of the basic behaviors, such as sit, stand, and walk. The participants tend to give a command to the robot to perform an action during the interaction. In the non-growth condition, the robot started in standing stage and has more probability to go to sit or walk, where participants usually gave command sit, stand, or walk. In contrast, the growth condition took a longer time because it needs a learning process by the participant. Therefore, the participants might perceive a more power to control the robot in the non-growth condition in a short interaction period.

6.2. Correlation of The Measured Perceptions in Respect to The Interaction Experience

Second hypothesis is related to significant correlation of the perceptions toward the interaction experience. From the linear regression performed, there is a significant effect of the perception of learning ability to achieve a better interaction experience. The participants enjoyed more interaction with the robot who is perceived able to learn, which implicitly means ”smarter”.

This analysis gave the insight that the perception of learning ability matters to engage nicely with the robot. Regardless the implementation strategy, humans can feel a better interaction experience as long as they perceive that the robot learns based on their feedback. The accuracy of learning and performing task is important but not becoming a distinctive factor. The implementation of an explicit response of adaptiveness by saying *”Glad you like it!”* or *”I am sorry, I will not do it again.”* might become the sign of a learning ability shown by the robot even though the next behavior choice was not preferable. The significant correlation between this perception with the interaction experience contradicts the insignificance result of the implementation towards this perception. Although the participants did not perceive the robot learning ability, they still appreciated the robot trials to learn which made the interaction was felt more enjoyable. Therefore, an approach to show a learning ability should be incorporated into an implementation strategy, so it could improve the perception of learning ability which arguably leads to a better interaction experience.

6.3. Further Research and Recommendation

The implementation of growth and adaptive behavior could be varied from one research to another. This study implemented general behaviors without one single focus topic. The main focus of the study was implementing a set of simple behaviors that can convey both an adaptation and a growth. Thus, basic behaviors were employed to the implementation, such as sit, stand, lie down, walk, roll, dance, etc., in general interaction setting. On the other hand, there are many other researches that are focused on particular topic, as an example is an adaptive behavior implementation in a help seeking of solving math problem [48]. This research resulted a significant positive effect difference between a robot with adaptive and non-adaptive behavior. Thus, a specific domain of the interaction can be chosen for the further research. It also helps to create a focus on the adaptation or growth style in the interaction design.

Another option to improve the implementation is by employing other modalities, such as emotion and gaze, as social cues. Currently, speech was the only modality used to recognize the humans feedback. The implementation is limited by only recognizing word that the participants gave and classified it as a positive or a negative feedback. This approach only can cover an explicit feedback given to the robot. But, there are more indirect feedbacks that human provided in the interaction from an emotion to a gesture. If the robot can improve its perception by incorporating also implicit human feedback, the adaptive behavior could also be improved better. Besides its modalities, the implementation of new behaviors also help to give more flavor in the interaction which make more interesting and exhilarating.

Related to the experiment procedure, a further research might use a full within-subject design to be able to see the comparison among the conditions, especially to see which factors (adaptation or growth) that have more significant effect to the interaction experience. Currently, there were only 24 participants gathered to experience 4 conditions with 10 to 15 minutes interaction for each of the condition. If the participants experienced all 4 of conditions sequentially, they might feel bored or less excited from the first condition to the last condition. The experiment might be split into two sessions to avoid this effect. The drawback of this implementation is more time resources to conduct the experiment. Another approach that can be applied to the study is using a between-subject design. It will enable the comparison between subjects independently. This approach will solve the effect of a longer experiment in a within-subject design but requires more participants to join the experiment.

The interaction with adults could be a good start for this exploratory study. The further research can choose children as target subjects. There are many differences on how children and adults interact with a robot. Children tend to have more surprising factor and excitement to interact with a robot. It might give a promising factor as they perceive the robot as alive, smart, and control-able. Children are also more curious on discovering a new toy or technology. It can create a better engagement which might influence the bonding between them.

6.3.1. New Research Questions

Due to nature of an exploratory study by this experiment, it is important to generate new research questions reflecting to both significant and insignificant results in the current study. The possible research questions that are interesting to examine are:

1. Is there any significant difference between child and adult in interaction with robot?
2. Is there any significant effect of learning ability perception without explicit feedback?
3. How does leveling of adaptation (explicit feedback, temporary probability change, and permanent probability change) influence perception of the participants?
4. Do previous experience interacting with NAO influence the perception? If so, why?
5. Do participants perceive stronger perceptions of aliveness, learning ability, and power in long-term interaction?
6. How many and what kind of behaviors should be assigned ideally to the robot?
7. How does the different modeling approach, besides FSM, affect the perceptions and the interaction experience?

6.3.2. Prototype Development Iteration

There are couple improvement points that could be implemented for the next prototype. One of the improvement items that has been mentioned in previous section is the speech recognition. In the existing implementation, the speech recognition used ALSpeechRecognition built-in module from naoqi. The classification is limited to certain words or phrases that needs to be defined in the initiation. Also, the classification accuracy is not perfect even though accuracy threshold has been introduced. Another approach has been investigated by implementing Python SpeechRecognition library and it performs well even with the longer sentences. This library and its dependencies have to be installed in the NAO system through SFTP protocol. This approach is worth to try for next iteration.

Other implementation on emotion recognition is also one of the aspects that can be implemented in the next iteration. Emotion can drive to better understanding of the feedback. But, this implementation should come with an algorithm to balance between explicit and implicit feedbacks. Moreover, an emotion recognition can fluctuate along the time. The next implementation should take into account when to use the emotion as a feedback.

Furthermore, there is some fix needs to be done related to the robot movements. There are couple movements that slightly differ among NAO robots. This movement should be generalized to all kind of NAOs and make it smoother. Moreover, there was also a problem in parallel task in changing a temporary probability where sometimes the process of choosing next behavior happened before the temporary probability change time frame. It caused the next chosen behavior was not considering the participant's feedback for repeating or avoiding the same behavior.

7 Conclusion

The exploratory study on a growth and an adaptive behavior using a NAO as a humanoid robot was conducted with participants from adult age range. A set of behaviors was implemented using a finite state machine. The behavior state transition probability and explicit feedback were chosen to convey the adaptation style, while unlocking behavior stages were employed to show the growth of the robot. The purposes of the study were to learn human perceptions towards growth and adaptive behavior and to find correlation between the perceptions with the interaction experience. This study used the perception of aliveness, the perception of learning ability, and the perception of behavior shaping control as measurement variables. The result showed there is no significant effect of growth and adaptive behavior manipulation in respect to these perceptions. But, there is a significant correlation of the perception of learning ability towards the interaction experience. A further research needs to investigate a different interaction design approach in growth and adaptation which can show whether there is any significant effect towards the measured perceptions and the interaction experience. Hence, we conclude that this current research is only an initial point but has a promising future in the contribution to the human-robot interaction field.

A

Sense of Power Scale Items

In rating each of items below, please use the following scale:

1. - Disagree strongly
2. - Disagree
3. - Disagree a little
4. - Neither agree nor disagree
5. - Agree a little
6. - Agree
7. - Agree strongly

Questions:

1. I can get him/her to listen to what I say.
2. My wishes do not carry much weight.
3. I can get him/her/them to do what I want.
4. Even if voice them, my views have little sway. (r)
5. I think I have a great deal of power.
6. My ideas and opinions are often ignored. (r)
7. Even when I try, I am not able to get my way. (r)
8. If I want to, I get to make the decisions.

have used the Sense of Power Scale with the following instructions:
Specific interaction:
In the negotiation . . .
(Note. All items were written in the past tense when assessing prior specific interactions.)

Relationship (multiple interactions):
In my relationship with my friend . . .
In my relationship with my mother . . .
In my interactions with my TA . . .
In my interaction with my date . . .
In my interactions with my supervisor . . .

Group (multiple relationships):
In my sorority . . .
In my dormitory floor . . .

Generalized (all relationships, groups):

B

Godspeed Questionnaire

Anthropomorphism
Please rate your impression of robot on these scales:

Fake	1	2	3	4	5	Natural
Machinelike	1	2	3	4	5	Human-like
Unconscious	1	2	3	4	5	Conscious
Artificial	1	2	3	4	5	Lifelike
Moving rigidly	1	2	3	4	5	Moving elegantly

Animacy
Please rate your impression of robot on these scales:

Dead	1	2	3	4	5	Alive
Stagnant	1	2	3	4	5	Lively
Mechanical	1	2	3	4	5	Organic
Artificial	1	2	3	4	5	Lifelike
Inert	1	2	3	4	5	Interactive
Apathetic	1	2	3	4	5	Responsive

Likability
Please rate your impression of robot on these scales:

Dislike	1	2	3	4	5	Like
Unfriendly	1	2	3	4	5	Friendly
Unkind	1	2	3	4	5	Kind
Unpleasant	1	2	3	4	5	Pleasant
Awful	1	2	3	4	5	Nice

Perceived Intelligence
Please rate your impression of robot on these scales:

Incompetent	1	2	3	4	5	Competent
Ignorant	1	2	3	4	5	Knowledgeable
Irresponsible	1	2	3	4	5	Responsible
Unintelligent	1	2	3	4	5	Intelligent
Foolish	1	2	3	4	5	Sensible

Perceived Safety Please rate your impression of robot on these scales:

Anxious	1	2	3	4	5	Relaxed
Agitated	1	2	3	4	5	Calm
Quiescent	1	2	3	4	5	Surprised

C

Proposed Questionnaire

Perception of aliveness
Please rate your impression of robot on these scales:

Fake	1	2	3	4	5	Natural
Machinelike	1	2	3	4	5	Human-like
Unconscious	1	2	3	4	5	Conscious
Artificial	1	2	3	4	5	Lifelike
Moving rigidly	1	2	3	4	5	Moving elegantly
Dead	1	2	3	4	5	Alive
Stagnant	1	2	3	4	5	Lively
Mechanical	1	2	3	4	5	Organic
Inert	1	2	3	4	5	Interactive
Apathetic	1	2	3	4	5	Responsive

Perception of learning ability
Please rate your impression of robot on these scales:

Incompetent	1	2	3	4	5	Competent
Ignorant	1	2	3	4	5	Knowledgeable
Unintelligent	1	2	3	4	5	Intelligent
Foolish	1	2	3	4	5	Sensible

Interaction Experience
Please rate your impression of robot on these scales:

Dislike	1	2	3	4	5	Like
Unfriendly	1	2	3	4	5	Friendly
Unkind	1	2	3	4	5	Kind
Unpleasant	1	2	3	4	5	Pleasant
Awful	1	2	3	4	5	Nice

Perception of Power
In rating each of items below, please use the following scale:

1. - Disagree strongly
2. - Disagree

3. - Neither agree nor disagree
4. - Agree
5. - Agree strongly

Questions: *In my interaction with the robot . . .*

1. I can get him/her to listen to what I say.
2. My wishes do not carry much weight.
3. I can get him/her/them to do what I want.
4. Even if voice them, my views have little sway. (r)
5. I think I have a great deal of power.
6. My ideas and opinions are often ignored. (r)
7. Even when I try, I am not able to get my way. (r)
8. If I want to, I get to make the decisions.

D

Experiment Guideline

1. You will interact with robot twice with total time about 30 minutes.
2. Before the interaction, you need to sign consent form for allowing us to record the video during the interaction.
3. During the experiment, the robot will perform set of behavior.
4. During the interaction you can touch, move, talk to robot but the robot can only understand few words, which will be used by robot as indication of your understanding of robot's behavior.
5. You can also give positive or negative feedback to the robot by saying yes, no, nice, etc.
6. You can talk to the robot when the eyes turning blue.
7. If robot falls down, you can help him to put in sit/lying down position.
8. Do not press the button in middle because it will turn off the robot.
9. You can end the interaction anytime if you feel bored or you think it is enough by saying stop to the robot. If the robot understand your command, it will ask confirmation whether or not you want to stop the interaction. Then, you can say yes or no as response to the question.
10. You can also stop the interaction anytime manually by talking to me if you feel uncomfortable.
11. At the end of interaction, you will be asked to fill in the questionnaire and answer one open question.
12. If you need some help, you can always ask me anytime.

E

Experiment Consent Form

In this study, we will research adaptive and growing behavior in human-robot interaction. You will be asked to interact with NAO (a commercially available research robot) twice. If you agree to participate, please know that you are free to withdraw at any point throughout the duration of the experiment. Your interaction will be recorded using camera recorder for interaction observation purposes. If you don't want to be recorded, please inform one of us, and we will terminate the experiment. All information you provide will remain confidential and will be used for this experiment only. The data will not be connected to your name or any other identifiable information, aside from your likeness captured on video. Only the researchers and their supervisors will have insight into the data. After the thesis done, the recordings will be deleted from personal PC and stored in TU Delft server for 5 years. If you wish, you may request the results of the research, after the research is done by contacting Romi Kharisnawan (romikharisnawan@student.tudelft.nl) or Joost Broekens (D.J.Broekens@tudelft.nl).

Please sign this form if you have understood the above and agree with your participation in the experiment.

Date: ______________________

Name: ______________________

Signature: ______________________

Researcher details
Romi Kharisnawan,
Student MSc Computer Science TU Delft
romikharisnawan@student.tudelft.nl

F Experiment Questionnaire and Open Question Result

F.1. Experiment Questionnaire

Legend of Table F.1, F.2, and F.3 is described below:

1. 1_* : questionnaire questions from section C part perception of aliveness
2. 2_* : questionnaire questions from section C part perception of learning ability
3. 3_* : questionnaire questions from section C part interaction experience
4. 4_* : questionnaire questions from section C part perception of power

F.2. Open Question Result

Participant 1

1. The participant thought second condition (G-NA) is more enjoyable and smarter than the first condition (NG-NA).

Participant 2

1. The participant felt the face tracking was nice.
2. The participant tries to mimic the robot.

Participant 3

1. The participant knows that the robot can do movement but do not know exactly how to make the robot does that.
2. It is useful to know what robot can do in the introduction.

Participant 4

1. The participant did not know how much robot can interact.
2. The robot did not answer the command even after repeating it.
3. The explicit feedback ("Sorry, I will not do it again") was perceived nice.

Participant 5

1. The participant was confused on what she should have done in the interaction because she did not know all possible actions.

Table F.1: Perception of aliveness questionnaire result

no	cond	gender	age	nao	1_1	1_2	1_3	1_4	1_5	1_6	1_7	1_8	1_9	1_10
1	NG-NA	Male	25	0	2	1	1	1	2	1	1	1	2	1
1	G-NA	Male	25	0	2	2	3	2	2	1	4	2	4	3
2	NG-A	Male	24	0	3	2	2	1	1	1	3	1	2	3
2	NG-NA	Male	24	0	2	1	1	2	1	3	4	1	3	3
3	G-A	Female	23	1	4	2	5	2	3	4	4	1	5	4
3	NG-NA	Female	23	1	4	2	4	2	4	4	2	2	4	4
4	NG-A	Female	24	1	4	3	3	3	4	4	4	3	4	4
4	G-NA	Female	24	1	2	2	3	3	4	2	3	2	2	2
5	G-A	Female	23	1	3	2	4	2	1	4	2	2	2	1
5	G-NA	Female	23	1	2	1	3	1	1	4	4	2	2	1
6	G-A	Female	23	1	1	2	1	3	2	3	4	2	1	3
6	NG-A	Female	23	1	2	1	1	2	1	1	2	1	2	2
7	NG-NA	Male	24	0	3	2	2	2	2	2	3	3	4	4
7	G-NA	Male	24	0	3	2	4	2	2	3	3	2	4	4
8	NG-NA	Female	28	1	2	2	3	3	3	3	3	3	3	3
8	NG-A	Female	28	1	3	3	2	3	3	3	3	3	4	4
9	NG-NA	Male	21	0	3	2	1	2	2	2	4	1	4	3
9	G-A	Male	21	0	2	1	1	1	2	1	2	2	2	1
10	G-NA	Female	24	1	3	4	1	4	3	4	5	4	4	2
10	NG-A	Female	24	1	4	5	2	2	1	3	1	2	1	2
11	G-NA	Female	26	1	3	3	4	2	2	4	3	3	4	3
11	G-A	Female	26	1	3	2	3	2	1	2	3	2	4	3
12	NG-A	Male	26	1	3	1	4	2	2	4	4	2	4	4
12	G-A	Male	26	1	3	2	4	2	2	4	4	1	2	3
13	G-NA	Female	25	0	2	2	3	2	3	3	3	2	3	2
13	NG-NA	Female	25	0	2	2	3	2	3	3	2	2	2	2
14	NG-A	Female	27	1	2	3	3	2	2	4	3	3	3	4
14	NG-NA	Female	27	1	2	2	3	2	1	2	2	1	3	2
15	G-A	Female	24	0	3	2	4	3	1	4	3	2	3	3
15	NG-NA	Female	24	0	3	1	4	3	1	4	3	2	3	4
16	NG-A	Female	25	0	3	2	4	3	2	4	4	3	4	4
16	G-NA	Female	25	0	3	4	3	2	2	4	4	3	4	4
17	G-A	Male	28	1	3	2	3	3	3	2	4	2	3	4
17	G-NA	Male	28	1	4	2	2	4	3	4	3	1	4	3
18	G-A	Male	25	0	4	4	3	4	3	4	4	3	4	4
18	NG-A	Male	25	0	2	3	2	2	1	2	3	1	2	3
19	G-NA	Male	24	0	2	2	2	1	3	3	3	2	2	2
19	NG-NA	Male	24	0	3	2	2	1	2	3	3	1	2	2
20	NG-NA	Female	27	1	3	2	4	3	2	3	3	2	4	3
20	NG-A	Female	27	1	2	2	3	3	2	2	3	2	3	3
21	NG-NA	Male	28	1	2	2	2	2	4	3	2	3	3	3
21	G-A	Male	28	1	3	3	3	1	2	2	3	3	3	3
22	G-NA	Male	24	0	2	1	4	1	1	2	2	1	2	2
22	NG-A	Male	24	0	2	2	4	1	1	3	2	4	2	1
23	G-NA	Female	27	1	1	1	1	1	3	1	2	1	1	1
23	G-A	Female	27	0	2	3	1	2	2	1	4	2	1	1
24	NG-A	Male	26	1	4	4	4	3	3	4	4	4	5	5
24	G-A	Male	26	1	3	4	4	3	2	5	4	4	4	4

Table F.2: Perception of learning ability and interaction experience questionnaire result

no	cond	2_1	2_2	2_3	2_4	3_1	3_2	3_3	3_4	3_5
1	NG-NA	1	2	1	2	1	1	1	2	2
1	G-NA	2	3	2	2	2	4	2	4	2
2	NG-A	1	1	1	1	5	3	3	4	3
2	NG-NA	2	3	1	2	5	4	3	3	3
3	G-A	4	4	4	4	4	5	5	4	4
3	NG-NA	4	5	4	4	4	4	5	5	4
4	NG-A	4	3	3	4	4	4	4	4	4
4	G-NA	1	2	2	3	3	3	3	2	3
5	G-A	4	4	2	3	4	3	3	3	2
5	G-NA	2	2	2	1	3	4	3	2	2
6	G-A	1	1	1	2	2	3	3	3	3
6	NG-A	1	2	3	3	3	3	3	3	3
7	NG-NA	2	3	2	2	3	3	3	3	3
7	G-NA	4	4	3	3	4	4	3	3	4
8	NG-NA	3	3	4	3	4	4	3	4	4
8	NG-A	3	4	3	3	4	4	4	5	4
9	NG-NA	2	3	2	2	3	3	3	4	3
9	G-A	1	1	2	1	2	1	3	2	2
10	G-NA	2	2	2	3	4	3	3	4	4
10	NG-A	2	1	4	3	4	3	4	4	5
11	G-NA	4	3	3	2	4	4	3	3	3
11	G-A	3	4	3	3	3	4	3	2	3
12	NG-A	4	4	3	4	4	3	3	4	4
12	G-A	4	4	4	3	4	3	3	3	4
13	G-NA	2	2	2	3	3	3	3	3	3
13	NG-NA	3	2	2	3	2	3	3	3	3
14	NG-A	3	3	3	3	4	5	5	5	4
14	NG-NA	2	2	3	1	4	5	5	4	4
15	G-A	4	3	4	3	5	4	3	4	5
15	NG-NA	3	4	4	3	5	3	3	4	5
16	NG-A	3	4	4	4	5	5	5	5	5
16	G-NA	4	4	4	4	5	5	5	5	5
17	G-A	3	4	3	3	4	4	5	3	4
17	G-NA	3	4	4	3	5	4	3	3	5
18	G-A	4	3	4	3	4	4	4	4	4
18	NG-A	2	2	1	2	2	4	3	2	3
19	G-NA	3	2	3	3	3	4	5	5	4
19	NG-NA	3	2	3	3	3	4	4	4	3
20	NG-NA	3	3	3	4	4	4	4	3	3
20	NG-A	3	2	3	3	3	4	5	4	4
21	NG-NA	2	2	3	3	2	3	3	1	3
21	G-A	3	4	3	3	2	3	4	2	3
22	G-NA	4	4	2	2	4	4	4	5	4
22	NG-A	4	4	2	2	4	3	3	2	2
23	G-NA	2	1	4	2	2	4	4	4	4
23	G-A	2	1	2	1	5	4	4	5	4
24	NG-A	4	3	4	5	4	5	5	5	4
24	G-A	4	3	3	2	4	5	5	4	3

Table F.3: Perception of power questionnaire result

no	cond	4_1	4_2	4_3	4_4	4_5	4_6	4_7	4_8	dur
1	NG-NA	1	2	1	2	1	4	4	4	11
1	G-NA	3	3	4	4	3	4	2	4	16
2	NG-A	2	4	2	4	2	1	5	1	13
2	NG-NA	3	4	3	5	2	5	5	2	12
3	G-A	4	2	5	4	5	2	1	4	4
3	NG-NA	5	4	4	4	3	2	2	4	5
4	NG-A	4	3	3	3	4	3	3	4	14
4	G-NA	1	4	3	3	2	4	4	4	15
5	G-A	1	5	1	5	1	5	5	4	8
5	G-NA	1	5	1	4	1	5	5	1	13
6	G-A	1	5	1	5	1	5	5	1	9
6	NG-A	3	2	2	4	1	3	3	2	11
7	NG-NA	3	3	3	3	3	4	4	4	8
7	G-NA	3	4	3	3	3	4	3	2	13
8	NG-NA	2	3	3	3	3	3	2	3	7
8	NG-A	2	3	4	4	4	4	3	3	6
9	NG-NA	4	4	2	4	2	4	4	1	16
9	G-A	1	5	1	4	1	3	4	1	10
10	G-NA	1	5	1	5	1	5	4	5	12
10	NG-A	1	4	1	4	1	5	4	5	10
11	G-NA	4	4	4	2	4	2	3	4	11
11	G-A	2	4	4	3	3	4	3	4	15
12	NG-A	4	3	4	4	3	2	2	2	10
12	G-A	2	3	4	3	3	3	3	4	10
13	G-NA	1	4	2	5	3	5	4	3	14
13	NG-NA	3	3	3	4	2	4	3	3	13
14	NG-A	3	4	2	3	3	4	4	2	14
14	NG-NA	1	5	1	5	2	4	4	2	14
15	G-A	3	3	2	4	2	4	3	2	12
15	NG-NA	4	3	3	3	2	4	3	4	12
16	NG-A	5	4	5	3	4	2	1	3	16
16	G-NA	5	4	4	4	3	3	2	4	8
17	G-A	2	2	2	4	3	2	3	2	15
17	G-NA	2	4	3	2	5	2	2	5	9
18	G-A	5	2	4	2	4	2	2	4	17
18	NG-A	4	3	4	2	4	3	2	3	7
19	G-NA	2	3	3	3	1	3	3	1	3
19	NG-NA	1	2	1	1	1	4	4	3	14
20	NG-NA	3	4	3	4	3	3	3	3	14
20	NG-A	3	3	3	4	2	4	3	4	7
21	NG-NA	1	5	2	4	2	4	4	2	7
21	G-A	2	4	3	4	1	4	3	2	13
22	G-NA	4	4	1	4	5	5	5	4	15
22	NG-A	1	4	1	2	2	5	4	4	12
23	G-NA	1	1	1	3	1	5	5	1	7
23	G-A	1	5	1	1	1	5	4	1	14
24	NG-A	3	2	4	2	5	3	1	2	12
24	G-A	2	2	3	5	2	3	4	4	14

2. The confusion was faded away along the interaction and start to understand after several behaviors.
3. It would be really nice if there is more interactive command.
4. The participant felt the face recognition is great because it can know where she sits or stands.
5. The participant felt the robot could nicely understand few words.
6. The malfunction of the robot felt pretty scary.

Participant 6

1. The participant felt it needs more guideline in the introduction about what she could say during the interaction.
2. The starting behavior (lying down) was perceived as creepy behavior.
3. The word that the robot said was not clear, for example *roll* was perceived as *run*.
4. The participant did not know what to say *"yes/no"* or *repeated word*

Participant 7

1. The first condition was felt better because of its responsiveness (G-NA compare to NG-NA).
2. The participant recognizes that the movement is only in the head.
3. The participant felt quite worried of the robot because its movement did not show stability.
4. The participant felt the hand movement was great on how the robot can grab user's hand.
5. The participant thought the robot has level of intelligence like ET.

Participant 8

1. The participant felt the voice recognition did not work well.
2. The participant was lost because she did not know in what extent that the robot can do.
3. The participant could not recall difference between two conditions (NG-A and NG-NA).
4. The confirmation question would be nice additional interaction, such as *"Do you like it?"*.

Participant 9

1. The participant felt the second condition (NG-NA) was better than the first one (G-A).
2. The participant thought the interaction was not so clear on what should he do and the possibilities.

Participant 10

1. The participant thought the robot follows what user said.
2. The word that the robot said was not clear, for example *gym* was perceived as *chill*.
3. The robot did different thing from what she commanded.
4. The second interaction (G-NA) was felt nicer and more human-like compare to first interaction (NG-A).

5. Overall, the participant felt the interaction was nice.

Participant 11

1. The participant felt the interaction was quite awkward and confused on what to do.
2. The participant shocked and worried that the robot will be broken after falling down.

Participant 12

1. The participant realized that the robot can only receiving reward, so he always gave positive or negative feedback towards its behavior.
2. The participant thought the robot has many different actions but the second condition (NG-A) only repeated some of these actions.

Participant 13

1. The participant thought the purpose of the study was not articulated well in the beginning.
2. The participant felt the interaction was more one-way and not inviting.

Participant 14

1. The participant did not know if she can give a command to robot.
2. The participant felt that the robot could understand positive and negative feedback (”yes/no”) and other words but sometimes it failed to recognize it.
3. The participant felt the robot always tried to comment on its action.
4. The participant felt the first condition (NG-NA) was more random and the second condition (NG-A) was more attentive by listening on what she said.
5. The participant realized that the robot has capability to track face and react to touch but not affecting other behaviors.

Participant 15

1. The participant felt strange that the robot said *”sit”* while sitting because she did not know if she has to repeat or do something else.
2. The participant felt the second condition (G-A) gave more response, such as ”Sorry, I will not do it again” as a response to ”No”, compare to first condition (NG-NA).

Participant 16

1. The participant felt the robot was responsive but needs time.
2. The robot was perceived has a struggle on coordinating what to do even though it knows what to do.
3. The robot was perceived as bit quiet.

Participant 17

1. The participant felt the turn-taking was confusing by looking at the robot’s eyes.
2. The participant felt the second condition was more responsive and kinder (G-A) than the first one (G-NA).
3. The command of positive response could be changed from ”yes” to ”ok”.

Participant 18

1. The participant felt the unlocking behavior was unclear because the robot suddenly said ”Horay, I successfully unlock new movements”.
2. The participant though the second condition (G-A) implemented learning condition even though it is not explicitly shown.
3. Overall, the experience was perceived good.

Participant 19

1. The participant has negative perception of robot in general.
2. The participant thought the robot was not obedient.
3. The participant was confused because he did not do anything but the robot kept doing some movements.

Participant 20

1. The participant thought some of the words could not be responded by the robot, for example ignoring command ”shake”.
2. The participant felt the face tracking was nice.

Participant 21

1. The participant felt the robot was not interactive and not understood what he said.
2. The participant though the robot repeated the same behaviors for many times.
3. The participant liked smoothness of the stand up movement.
4. The participant felt the second interaction (NG-NA) was better because he knew already what to do.

Participant 22

1. The participant felt the interaction was confusing at first but later could see how it works.
2. The participant felt the robot did not respond to ”yes/no” in every single time.
3. The word that the robot said was not clear, for example *gym* was perceived as *chew*.
4. The participant worried on the joint sound and found its laugh/sleeping sound was creepy.

Participant 23

1. The participant felt the robot did not get any emotion, such as happy or sad.
2. The participant felt the robot did not listen to user commands.
3. The participant did not know how the unlocking behavior works although she successfully unlocked the new stage.

Participant 24

1. The participant felt the first interaction was more engaging by its unlocking behavior (G-A), while the second interaction (NG-A) was smarter by understanding what user said better.
2. The participant found that the repeating behavior after saying ”no” was a mistake.

G

Literature Review Extras

Week	Task	Verbal Commands	AIBO Behaviors
Baby (D1: Week 1)	Make AIBO understand its name	AIBO	Answers with a simple melody
	Make AIBO learn how to say good-bye	See you	Waves its front paw left and right
Kid (D2: Week 2)	Make AIBO sit down	Sit down	Sits down
	Make AIBO stand up	Stand	Stands up
	Make AIBO lie down	Lie down	Lies down
Adolescent (D3: Week 3)	Make AIBO walk forward	Go forward	Walks forward
	Make AIBO walk backward	Go backward	Walks backward
	Make AIBO turn right	Turn right	Turns right
	Make AIBO turn left	Turn left	Turns left
Adult (D4: Week 4)	Make AIBO cheer you up	Cheer up	Shakes its shoulders and tries to cheer you up
	Make AIBO sing a song	Sing a song	Sings a song
	Make AIBO dance	Dance	Dances by lifting and shaking its two front legs
	Make AIBO proud of itself	Show time	Opens its mouth and makes the sound of clapping
	Make AIBO wish you a happy day	Happy day	Dances with happy music

NOTE: Tasks that AIBO can do were cumulative. That is, AIBO in Week 2 can do all tasks in Week 1. Thus, AIBO in Week 4 can do all tasks listed above.

Figure G.1: Lee et.al.'s weekly tasks of the developmental condition [44]

The Number of Verbal and Touch Commands Required in Each Developmental Stage

Baby (D1: Week 1)		*Kid (D2: Week 2)*		*Adolescent (D3: Week 3)*		*Adult (D4: Week 4)*	
Verbal commands	*Speed of learning*	*Verbal commands*	*Speed of learning*	*Verbal commands*	*Speed of learning*	*Verbal commands*	*Speed of learning*
AIBO	V=6 H=6 C=6	AIBO	V=0 H=0 C=0	AIBO	V=0 H=0 C=0	AIBO	V=0 H=0 C=0
See you	V=6 H=6 C=6	See you	V=0 H=0 C=0	See you	V=0 H=0 C=0	See you	V=0 H=0 C=0
		Sit down	V=5 H=5 C=5	Sit down	V=0 H=0 C=0	Sit down	V=0 H=0 C=0
		Stand	V=5 H=5 C=5	Stand	V=0 H=0 C=0	Stand	V=0 H=0 C=0
		Lie down	V=5 H=5 C=5	Lie down	V=0 H=0 C=0	Lie down	V=0 H=0 C=0
				Go forward	V=3 H=4 C=4	Go forward	V=0 H=0 C=0
				Go backward	V=3 H=4 C=4	Go backward	V=0 H=0 C=0
				Turn right	V=3 H=4 C=4	Turn right	V=0 H=0 C=0
				Turn left	V=3 H=4 C=4	Turn left	V=0 H=0 C=0
						Cheer up	V=0 H=0 C=0
						Sing a song	V=0 H=0 C=0
						Dance	V=0 H=0 C=0
						Show time	V=0 H=0 C=0
						Happy day	V=0 H=0 C=0

NOTE: V= the number of verbal commands needed to accomplish the task; H= the number of patting touches on head needed to accomplish the task; C= the number of patting touches on chin needed to accomplish the task.

Figure G.2: Lee et.al.'s requirements of stage transition [44]

Bibliography

[1] Measuring affective learning and teacher evaluation. http://www.jamescmccroskey.com/measures/affect.pdf. Accessed: 2017-03-14.

[2] Sony puts aibo to sleep. https://www.cnet.com/news/sony-puts-aibo-to-sleep/, . Accessed: 2017-12-11.

[3] Sony just announced a new aibo robot dog. https://www.theverge.com/circuitbreaker/2017/10/31/16588878/sony-aibo-2017-announced-price-release-date, . Accessed: 2017-12-11.

[4] Sony aibo first impressions: old robot dog, new tricks. https://www.engadget.com/2018/01/08/new-sony-aibo-first-impressions/, . Accessed: 2017-12-16.

[5] Sony uses artificial intelligence to bring aibo robot dog back to life. https://www.dezeen.com/2017/11/02/sony-artificial-intelligence-aibo-robot-dog-revived/, . Accessed: 2017-12-11.

[6] Sony looks to reboot 1990s glory days with relaunch of robot dog. https://www.ft.com/content/ec2ca8de-f6a2-11e7-88f7-5465a6ce1a00, . Accessed: 2017-12-16.

[7] Sony's aibo robot is back, ready to learn new tricks. https://asia.nikkei.com/Business/Companies/Sony-s-aibo-robot-is-back-ready-to-learn-new-tricks?page=1, . Accessed: 2017-12-16.

[8] Sony aibo robot dog. http://www.robotbooks.com/sony_aibo.htm, . Accessed: 2017-12-16.

[9] Differences between creatures. https://www.gamefaqs.com/boards/535410-black-and-white-2/41422166, . Accessed: 2017-12-14.

[10] Black and white combined the sublime with the stupid. http://www.eurogamer.net/articles/2015-07-26-black-and-white-combined-the-sublime-with-the-stupid, . Accessed: 2017-12-15.

[11] Review: Black and white 2. https://www.macworld.com/article/1138954/black_and_white_2.html, . Accessed: 2017-12-10.

[12] Black and white 1 vs 2. https://www.gamespot.com/forums/pc-mac-linux-society-1000004/black-and-white-1-vs-2-28615924/, . Accessed: 2017-12-14.

[13] Black & white 2. https://en.wikipedia.org/wiki/Black_%26_White_2, . Accessed: 2017-12-10.

[14] Gamespy: Black and white 2. http://pc.gamespy.com/pc/black-white-2/656318p1.html, . Accessed: 2017-12-14.

[15] Black and white 2 review trusted review. http://www.trustedreviews.com/reviews/black-white-2-black-white-2-page-2#rRB1cIau35P12mLE.99, . Accessed: 2017-12-14.

[16] Jean piaget cognitive theory simply psychology. https://www.simplypsychology.org/piaget.html. Accessed: 2017-25-01.

[17] Cdr brain lobes. http://double.nz/creatures/brainlobes/differences.htm, . Accessed: 2018-01-07.

[18] Creatures (video game series). https://en.wikipedia.org/wiki/Creatures_(video_game_series), . Accessed: 2018-01-07.

[19] Furby 2005. http://official-furby.wikia.com/wiki/Furby_2005, . Accessed: 2017-01-05.

[20] The new furby review: Absolute horror. https://gizmodo.com/5937402/the-new-furby-review-absolute-horror, . Accessed: 2017-12-08.

[21] No, you can't teach your furby to swear: How furbies "learn" language. https://www.listenandlearn.org/blog/no-you-cant-teach-your-furby-to-swear-how-furbies-learn-language/, . Accessed: 2017-12-10.

[22] Case study: Furby. http://tombattey.com/design/case-study-furby/, . Accessed: 2017-01-05.

[23] Furby. https://en.wikipedia.org/wiki/Furby, . Accessed: 2017-12-08.

[24] Naoqi framework. http://doc.aldebaran.com/1-14/dev/naoqi/index.html. Accessed: 2018-04-06.

[25] Pytransition. https://github.com/pytransitions/transitions. Accessed: 2018-04-10.

[26] *Communication in the Real World: An Introduction to Communication Studies*. University of Minnesota Libraries Publishing Edition, Minnesota, 2016.

[27] M.I. Ahmad, O. Mubin, and Orlando J. A systematic review of adaptivity in human-robot interaction. *Multimodal Technologies and Interaction*, 2017.

[28] C. Anderson, O.P. John, and D. Keltner. The personal sense of power. *Journal of Personality*, 2012.

[29] M. Asada, K.F. MacDorman, H. Ishiguro, and Y. Kuniyoshi. Cognitive development robotics as a new paradigm for the design of humanoid robots. *Robotics and Autonomous System*, 2001.

[30] C. Bartneck, D. Kulic, E. Croft, and S. Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int J Soc Robot*, 2009.

[31] C. Breazeal. Social interactions in hri: The robot view. *Trans on Systems, Man, and Cybernetics*, pages 181–186, 2004.

[32] C. Breazeal. Role of expressive behaviour for robots that learn from people. *Phil Trans of the Royal Soc B*, pages 3527–3538, 2009.

[33] R.A. Brooks, C. Breazeal, R. Irie, C.C. Kemp, M. Marjanovic, B. Scassellati, and M.M. Williamson. Alternative essences of intelligence. pages 961–968. AAAI, 1998.

[34] A. Bruce, I. Nourbakhsh, and R. Simmons. The role of expresiveness and attention in human-robot interaction. pages 4138–4142. IEEE, 2002.

[35] D.T. Campbell and D.W. Fiske. Convergent and discriminant validation by the multitrait multimethod matrix. *Psychological Bulletin*, 1959.

[36] S. Cass. Mind games: To beat the competition, video games are getting smarter. 2002.

[37] M. Colledanchise. *Behavior Trees in Robotics*. PhD dissertation, KTH Royal Institute of Technology, 2017.

[38] P. Ekman and R.J. Davidson. *The Nature of Emotion*. Oxford Univ. Press, Oxford, U.K., 1994.

[39] M. Foukarakis, A. Leonidis, M. Antona, and C. Stephanidis. Combining finite state machine and decision-making tools for adaptable robot behavior. *UAHCI/HCII 2014, Part III*, pages 625–635, 2014.

[40] M. Fujita. On activating human communications with pet-type robot aibo. 2004.

[41] J. Hemminghaus and S. Kopp. Towards adaptive social behavior generation assistive robots using reinforcement learning. *Human Robot Interaction*, pages 332–340, 2017.

[42] P. Kearney. *Affective learning scale. In R. B. RubinP. Palmgreen& H. E. Sypher(Eds.). Communication research measures: A sourcebook*. The Guilford Press, New York, 1994.

[43] J. Kennedy, P. Baxter, and T. Belpaeme. Comparing robot embodiments in a guided discovery learning interaction with children. *International Journal of Social Robotics*, pages 293–308, 2015.

[44] K.M. Lee, N. Park, and H. Song. Can a robot be perceived as a developing creature? *Hum Commun Res*, 2005.

[45] A. Mehrabian and J.T. Friar. Encoding of attitude by a seated communicator via posture and position cues. *Journal of Consulting and Clinical Psychology*, pages 330–336, 1969.

[46] J. Pajarinen and V. Kyrki. Robotic manipulation of multiple objects as a pomdp. *Artificial Intelligence*, pages 213–228, 2015.

[47] C.G. Puehn, T. Liu, Y. Feng, K. Hornfeck, and K. Lee. Design of a low-cost social robot: Towards personalized human-robot interaction. *UAHCI/HCII 2014*, pages 704–713, 2014.

[48] A. Ramachandran, A. Litoiu, and B. Scassellati. Shaping productive help-seeking behavior during robot-child tutoring interactions. pages 247–254, 2016.

[49] T. Robbins and A. DeNisi. A closer look at interpersonal affect as a distinct influence on cognitive processing in performance evaluations. *J Appl Psychol*, 1994.

[50] A. Malhotra S. Grand, D. Cliff. Creatures: Artificial life autonomous software agents for home entertainment. *Autonomous Agents 97*, 1997.

[51] D. Silver. Lecture 2: Markov decision process. 2015.

[52] A.L. Thomaz and C. Breazeal. Teachable robots: Understanding human teaching behavior to build more effective robot learners. *Artificial Intelligence*, pages 716–737, 2008.

[53] J. Wexler. Artificial intelligence in games: A look at the smarts behind lionhead studio's ”black and white” and where it can go and will go in future. 2002.

[54] H. Yan, M.H. Ang Jr., and A.N. Poo. A survey on perception methods for human-robot interaction in social robots. *Int J Soc Robot*, pages 85–199, 2014.

[55] S. Yildirim and S.B. Stene. A survey on the need and use of ai in game agents. 2008.

[56] S.J. Zaccaro, R.J. Foti, and D.A. Kenny. Self-monitoring and trait-based variance in leadership: An investigation of leader flexibility across multiple group situation. *Journal of Applied Psychology*, pages 308–315, 1991.