LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
Department "Institut für Informatik"
Lehr- und Forschungseinheit Medieninformatik
Prof. Dr. Andreas Butz

Bachelorarbeit

Automated Robotic Concierge: User Experience Study

Phoenix [redacted]

Bearbeitungszeitraum:	15. 9. 2018 bis 02. 02. 2019
Betreuer:	Dr. Daniel Ullrich
Verantw. Hochschullehrer:	Prof. Dr. Andreas Butz

Abstract

With the litany of new automated robotic systems coming to the market, focus is shifted from the physical limitations of robots to software implementation and execution. In this paper the acceptance of robots as an alternative to, or as a low cost supporter of, reception staff shall be evaluated using user studies. The impression of autonomy and a fluid interaction will be the focus.

Aufgabenstellung

UX und Akzeptanz einer robotischen Empfangsdame or Acceptance of a robotic concierge with a focus on user experience(UX)

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbstständig angefertigt, alle Zitate als solche kenntlich gemacht sowie alle benutzten Quellen und Hilfsmittel angegeben habe.

München, January 29, 2019

. . . .

Contents

1 Introduction

The media has been fascinated by the idea of robots since the inception of the computer. Today we stand at a point where this fantasy may turn into reality. The future is robotic, hopefully allowing humans to be freed from menial labour. Traditionally human jobs, such as restaurant greeter or delivery man, are already being aided, or to a certain degree, being replaced by robots. More traditional consumer robots such as Pepper and Jibo shall be assessed, in order to understand the the current state of the art. As well as scrutinizing commercially functioning robots. Further, we shall illustrate the various mechanisms employed in order to create some semblance of autonomy.

The hardware that shall be used is one Segway Robotics' Loomo. This robot is built similarly to a traditional Segway, with a built in Android device. This device is the means of creating the robotic personality and assistant that is the ultimate aim of this bachelor thesis. The specific use case being built and implemented is a robotic concierge, it should wait in the reception area of iteratec GmbH. Ultimately it should be able to greet guests and personnel alike, hopefully guiding guests to their desired meeting points with a minimum of human intervention. Taking consumer robots into account, Loomo should also provide a welcoming attitude to all humans in its vicinity. Personnel should be assisted in their basic daily activities such as making meetings and communicating with other employees, all the way to more arbitrary services such as the weather forecast or brightening the office environment with a joke.

In order to determine the effectiveness and legitimacy of such robotic personnel, a user study has been be conducted and contrasted with the effectivness of a real human receptionist. As this paper represents an early work of this type, an explorative analysis has also been made in order to assess the viability and the potential of this type of technology.

The ultimate aim of this project is to provide a proof of concept and prove the viability of robots as sources of menial labour in undesired positions. However before this can be attempted, we must first look to the past and the present.

2 State of the Art

Robotics is an industry which is growing at an incredible pace, some projections expect this industry to hit $14.91 Billion by 2023[25]. Before beginning to evaluate the effectiveness, practicality and necessity of using robots in traditionally human positions, what is already possible has to be assessed. In order to do that we shall be looking at commercially available robots on the market. The current market leaders in consumer robotics, in no particular order are: iRobot, Jibo, 3D Robotics, Honda, Bossa Nova Robotics, DJI and Softbank[25]. iRobot will be ignored with it's Roomba, as it is dissimilar to this papers purpose. The drones from 3D Robotics and DJI will also be ignored for similar reasons. This leaves us with the predominantly entertainment robots: Asimo from Honda, Jibo from Jibo and Pepper from Softbank. We shall also touch on two robots that are purely used in a commercial setting, namely the grocery store scanning robots from Bossa Nova Robotics and the MK service robots from a collaboration between Bangkok University and MK Restaurants Group Company Limited. Additionally we shall assess rescue robots as an measure for intuitive usability. Loomo was chosen for its mobility and modest price in comparison with the majority of the following robots.

2.1 Entertainment Robots

To a certain extent Humans affinity for empathy is storied and well documented, humanoid robots need only be passably intelligent in order for humans to stop thinking of them as mere appliances and begin thinking of them as emotional constructs. Children often do not know or understand the difference between the emotions they feel and robotic programming; thinking the robot will get bored or feel pain. Whether or not this empathy is good or appropriate, is however the subject of another paper. In that vein, we will begin with the robots whose primary focus is or was to entertain the user, with no concrete use case beyond that.

2.1.1 Asimo

ASIMO or Asimo as it colloquially known is an acronym which stands for "Advanced Step in Innovative Mobility"[6]. The aim was to create a robot which could walk like a biped. Research started in 1986 with the first version, the prototype P2, being released in 1996. The final iteration was finished in 2011 and has since been completely retired. Each robot cost just under $1 million to produce.[4]

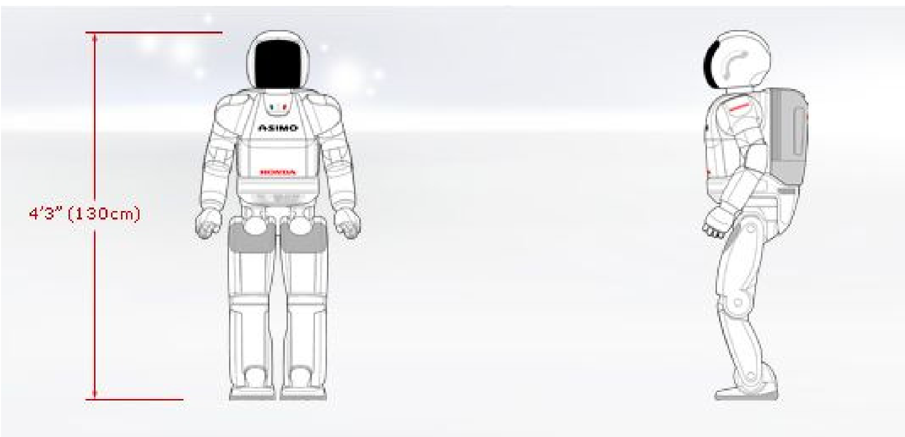

Figure 2.1: Honda's Asimo robot[5]

This kind of technology was very much a flagship, showing the community at large what was possible with robotics. In its last iteration from 2011 Asimo was able to independently avoid obstacles and people, as well as a few other minor functionalities. It never found much of a function beyond showing walking, reacting robots were a reality. Honda used what they learned, perhaps spurred by the natural disasters in Japan, to make rescue robots that would have a more practical function.[3]

2.1.2 Jibo

Figure 2.2: Jibo[13]

The Jibo robot made more sense in a pre Alexa and Google Home world. Jibo was always intended to be a companion, not assistant. Jibo can learn up to 16 different faces and the voices of each person[17]. It dances and sings and generally tries to provide entertainment[23]. Jibo was marketed as "The World's First Social Robot for the Home"[10] and as such it delivers. Unfortunately the large price tag, the multitude of low cost alternatives and lack of commercial applications mean that Jibo is soon to be defunct. Jibo is more limited than other virtual assistants although as stated, this was not the aim of the robot.

2.1.3 Pepper

As one of the most prevalent and widely available robots on the market, Pepper is an example of a thriving robot, with a multitude of uses. It combines the best of Asimo's mobility with the personality of Jibo, although Pepper lacks the ability to walk up stairs. It has small wheels which are not intended to be used for extended periods of movement, and even trips over small objects[1]. The creators of Pepper, SoftBank Robotics, have created an 87 page document with the dos and don'ts of the creation of the personality that is Pepper, with gems such as “Pepper is a character” or “Pepper tries to be smart”.[7](p.14, p.23) The aim being to create a seamless user experience, where all newly created applications made for Pepper seamlessly blend with and build onto this personality. Pepper uses different colored eyes to indicate the different modes it is in. For instance the color to indicate listening is navy blue and the color for processing requests is green. This increases usability, giving the user clear feedback as to the status of the robot. The number one source of frustration when using similar technology is a lack of visual and audio feedback. In order to know the basic and start commands there is an instruction manual with a section titled Basic Channel.[20] The most basic functionalities and the necessary command words are listed here.

Figure 2.3: Pepper[20]

Today Pepper has already found some commercial use, with 14 robots being used in the Smithsonian in Washington D.C.[18] as well as a trial in the Japanese sushi restaurant chain Hamazushi.
Although some dissenters have suggested that a move towards automation may not necesarily be the best decision, as it lacks the social and bonding experience of a real human connection.[19]

2.2 Commercially Used Robots

The following robots have been created exclusively for mid to large businesses in a commercial environment.

2.2.1 Grocery Store Robots

The grocery store robot from Bossa Nova roams aisles, scanning the shelves. These robots check the amount of a product still in stock or check that a product hasn't been put back in the wrong spot in the supermarket by a distracted shopper. According to Bossa Nova co-founder and chief

Figure 2.4: Grocery store robot[14]

business officer Martin Hitch, "the[se] robots work three times faster and can be up to twice as accurate as humans".[15] The robots move extremely slowly, at a rate of 20 cm/s and use sensors to detect and avoid shoppers or other miscellaneous objects that may be in its path.[14] One of the main aims of this robot is decreasing human labour, especially labour considered tedious by humans.[12] Bossa Nova is still in its infancy and is currently trying to scale up. The company hopes to increase the amount of useful statistics that the robot garners from driving around the store through the use of deep learning of all data.[8]

2.2.2 Service Robots

The following robots were developed as part of a collaboration between MK Restaurants Group Public Company Limited(MK) and Bangkok University. The aim was "[t]o develop a robot for restaurant service tasks and customer entertainment"[11](p.2). Basically the use case was to create two robots, one appropriate for ordering and interacting with customers and one to bring meals to the tables.

Figure 2.5: MK Service Robots, from left to right: ServeTwo(2012), OrderOne(2009), Slim(2012), OrderTwo(2009) and ServeOne(2010)[11]

Instead of creating a robust collision avoidance system or making the system dynamic, the creators decided to use line tracing sensors. This means that large black lines were taped or drawn on the ground. If any of the sensors mounted to the front and rear of the robot sense something on its intended path, it would wait on its line until the person or thing has moved out of the way. Customers for the most part found the robotic operation fascinating.

2.3 Rescue Robots as Usability Benchmark

Usability is always hotly contested - what may be intuitive to one person may not be intuitive to another. A paper from the year 2004 tried to use objective measures in order to test human-robot interaction(HRI). Four teams agreed to take part in the study in 2002 at the American Association for Artificial Intelligence(AAAI) Robot Rescue Competition. A simulated urban search and rescue situation was created and each robot was expected to rescue as many people as possible, whilst avoiding collisions. The collisions, percentage of traversed arena area and people rescued were tallied and given minus and plus points respectively. This was used as an objective measure of usability. These tasks were performed by two different groups, the systems developer followed by a domain expert. The system developers were the creators of the software and robots, whereas the domain expert was a special operations fire chief who had participated in training sessions with robots for search and rescue. The domain expert was given five minutes to familiarise himself with the robots interface, which would be similar to a real life disaster given the primary user was, for some reason, unavailable. They were then allowed to attempt the rescue situation, the scored points during the event were tallied under the assumption that a highly usable, intuitive robot design would lead to more points in the event.[2]

Figure 2.6: All robots from the 2002 AAAI robot competition[26]

3 An Automated Concierge in Practice

Programming with Loomo was centered around Android programming, as an Android device was the interface to the robot.

3.1 The Hardware

We shall start with the hardware capabilities given. We have one Loomo Robot from Segway:

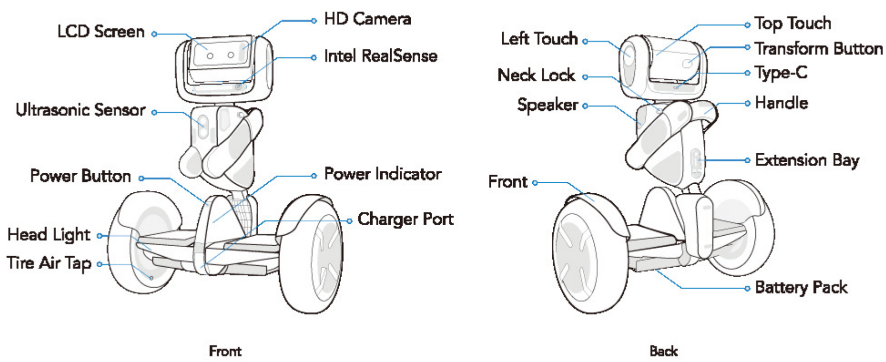

Figure 3.1: Loomo

The robot is autobalancing and has two base states: a robot state, where the Android counterpart can be interacted with and a driving mode, wherein Loomo is passively driven as with a real segway vehicle. In order to change states, one must press the button located on the top of the robots head, whilst in driving mode or the button on the back of the Android device. The Android component has 4GB of RAM and 64 GB of hard disk space. The device is outfitted with a standard touchscreen and forward facing camera. In addition to the Androids pieces there are the following sensors:

Sensor Specs	Descriptions
RealSense	**Real time 30Hz RGB-Depth streaming for both indoor and outdoor uses**
HD Camera	High resolution with 104° FOV (Pending fine turning. Image quality is not optimal)
Mic Array	**Beamforming and voice localization, powered by 5 microphones**
Ultrasonic Sensors	Obstacle distance calculation
Touch Sensors	**Located at robot head left, right & back**
Wheel Odometry	Approximately 4 degrees precision
Base IMU	**6-axis IMU**

Figure 3.2: Loomo sensor details

3.2 The Program

The program was created using a variety of open source projects, as well as using some of the functions provided by the Segway SDK. There are many different moving pieces that go into creating a certain sense of sentience.

3.2.1 Program Structure

The program structure is split into three distinct parts:

Common: Logic for all fundamental operations such as Networking, SpeechToText, TextToSpeech as well as the State Machine.

Functionalities: This is where new features can be built and tested. There is the 'Repeat' functionality, which was used to test and integrate the State Machine, but has no real use now. The second and last functionality is the ConversationManager, this uses intelligent parsing to ascertain what a user wants and responds appropriately.

View: Anything shown to a user or anything involving displays is processed here.

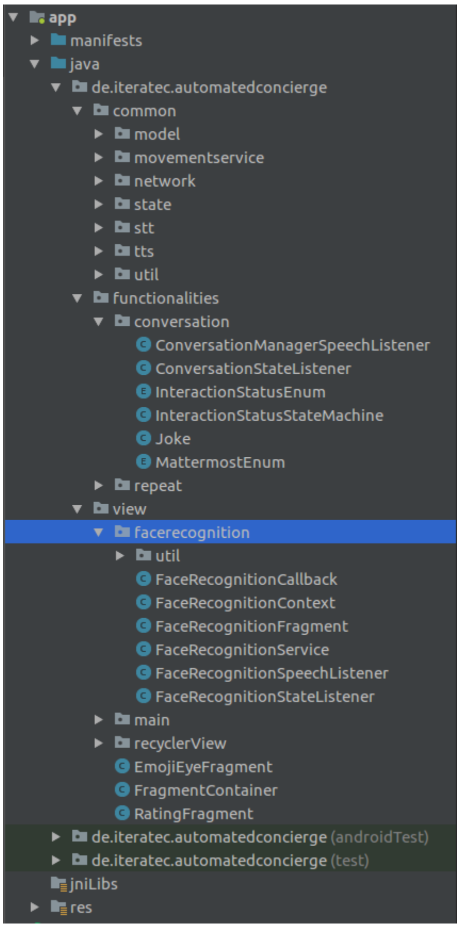

Figure 3.3: A snapshot of packages in program

As these represent the bulk of this thesis they shall in following be illustrated.

3.2.2 Common

The State Machine

In order to keep the system logically differentiated in a clear and structured manner, a state machine was implemented. Although the robot is theoretically able to accomodate many more functionalities, only a start up state and three further states were implemented. The program begins in State.MAIN; after booting necessary processes, the program then is switched into the state State.RECOGNIZE.

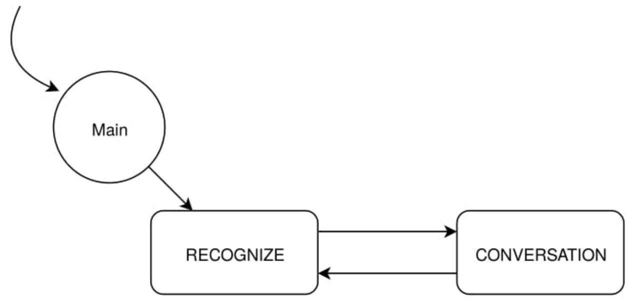

Figure 3.4: Base states

Following recognize are two sides of the same coin: State.CONVERSATION_UNKNOWN and State.CONVERSATION_EMPLOYEE.

A StateMachine allows legal changes and disallows illegal states. All possible states begin by registering themselves using so called state listeners. All subscriebers are informed in the case of a state change. If a state notices it is the relevant state, it initialises all necessary modules and begins. The prior active state also notices the state change and deactivates any unnecesary modules.

Networking
The networking folder is used to send messages to employees, book meetings and meeting rooms and perform face recognition calls as well as more innocuous tasks such as checking the weather.

Speech-To-Text
The speech to text (STT) used was an opensource project called DroidSpeech. DroidSpeech uses a standard android text to speech program, this program explicitly states that it is not meant for continuous listening. Continuous listening is the constant input of audio and consequent translation in strings, which can then be further parsed and interpreted as necessary. The lack of continuous listening then presents a hurdle for an autonomous bot. Droidspeech merely cheats the standard android speech to text by turning on the functionality as soon as it has received the interpreted text data, thereby approximating continuous listening. The interpreted text data is then given to the so called Interpreter, which informs all listeners of the text data.

Text-To-Speech
The text to speech module also uses android library. In the case that Loomo has to talk, a trigger turns off continuous listening and deactivating the STT module. Once Loomo has finished talking a second trigger restarts listening. This avoids Loomo hearing what he says and getting caught in loops or trying to converse with himself.

Util
Various functions were also implemented in order to minimise some developmental problems. As the android STT was not intended to listen contiuously, a scheduled threads restart service was created. This checks how long ago the program had last heard someone talk and if necessary restarts the speech to text module. In additionto that, a person map service is held here. In order for

Loomo to give employees special attention and feel known, a databank with all names and distinct internally used abbreviations was built. This is used, for instance, when recognizing someone with face recognition or booking a meeting or writing a personal message.

3.2.3 Functionalities

Repeat
As with any program and especially with so many moving pieces, a simple functionality was needed in order to test basic functionalities. Such as the interplay of speech to text and text to speech, or the correct functioning of the state machine. The repeat functionality showed the basic functions were performing as was expected.

ConversationManagerSpeechListener
Once switched into State.CONVERSATION_UNKNOWN or State.CONVERSATION_EMPLOYEE the ConversationManagerSpeechListener(CMSL) is started. The CMSL is responsible for understanding the majority of commands that are dictated by communication and interaction, distinctive from commands that are used purely to start and stop processes. The CMSL began simply, but became increasingly complex in order to contain all the different functionalities required.

Hierarchy of Understanding
Once the audio files have been converted to text, this module is then informed. This is the central brain, if you will, and interprets and tries to respond to all commands. Many were created puely for the purposes of this work, whereas others were programmed to give a semblance of a personality. When asked if Loomo knows where the bathroom is, he responds “Do I look like I use toilets?”.

A hierarchy is used to give Loomo the ability to react to previous stimuli. A set of booleans is run through to see if the user is currently awaiting a response. Is for instance a joke punchline expected? Or are they in the process of sending an employee a message? The second level of the hierarchy attempts to understand natural language. It uses different ways of saying the same thing and keywords to accurately interpret what is said.

Local states
In addition to the global state, there are two further minor states: one responsible for Mattermost[1] communication and one responsible for the so called interaction status. The Mattermost state guides users through the process of sending a message:

[1]Messaging application used at iteratec GmbH for internal communications

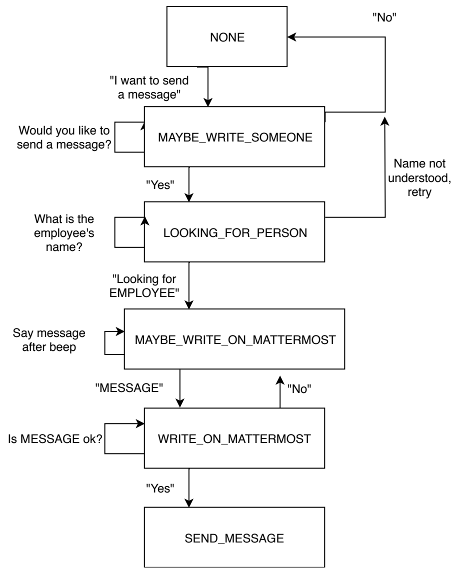

Figure 3.5: Mattermost state

All messages enclosed with "" represent user feedback, bold non state text represent the variable written.

The second minor state is the interaction status. After some beta testing a lack of visual feedback was noticed. The interaction status was therefore introduced. It is intended to tell the user that their needs have been understood, or give them the ability to use the touch screen interface instead of relying purely on audio commands. The interaction status provided a problem to implement, as the relevant functions provided a cross section of the entire application. The user needed to be able to send a message or get a map to their specified location. It needed to dynamically react in relation to sending messages or whatever was being done.

3.2.4 View

View holds all user interface related code. From the main class, we are immediately switched into the FragmentContainer. All activity screens were modeled as an Android fragment, this meant that pieces could be switched in and out at runtime without losing any context information. The state machine was responsible for starting and stopping the relevant fragments. The recycler view allows the use of lists which is used as part of the interaction status, for instance to select a room from a list of rooms which is then shown as a map.

The EmojiEyeFragent provided by Segway was always active and provided a sense of aliveness. It is an eye that moves and sometimes makes what can be described as cute noises.

The face recognition fragment became quite complicated, as the code necessary to take photos of users then convert these images into ByteString format and send to the face recognition backend is quite complicated. When a user has identified themselves as an employee, a camera view is shown in place of the stock emoji eye. This was to give some feedback to the user and allow them to position themselves in a way more conducive to face recognition.

3.2.5 Happy Flow

We shall take the reader through one standard program run of Loomo.

During this phase all necesary modules are booted in order to avoid startup time overhead needed to start modules when the application is running. This allows for a more fluid experience for the user. Listeners are registered by the StateEventService as well as beginning the TextToSpeech and SpeechToText modules.

We begin, as with the program, in the main class. This is used as a means to initialise all the components and give them an Android context in order to function. The components in order of their code debut are: PersonMapService, initializeStateListeners, TextToSpeech, SpeechToText, Interpreter and the start of the next activity.

A singleton programming strategy has been used whenever possible to reduce memory overhead in the relatively constrained device that an Android phone represents. The TextToSpeech component requires a context in order to function, it may after this point be used by any part of the program. The program then starts the base Activity: the FragmentContainer holds all the other Fragments, the skilled switching of which hopefully creates the impression of autonomy and intelligence.

First State: RECOGNIZE

The first State is RECOGNIZE which is used to tell if someone has run into the camera's viewing area. This works by using Android's native CaptureCallback which analyses the image, in this case for faces. A seen face is used as a trigger, which prompts Loomo to ask if the person works there. After hearing a 'yes' the robot will expose the camera's view, so that the employee may position their face in a way that may be more easily recognized. People can be more easily recognized if their face is level and perpendicular to the camera, which should hopefully become apparent for the users. The camera takes a multitude of photos and sends them to the the backend provided by iteratec GmbH. The face-recognising backend has the faces of the majority of workers within the company. Loomo receives a multitude of responses with the likeliness and abbreviation of the person in the photo. As the backend has not been optimized to work with Loomo, the likeliness can vary. Once a person with a likeliness above a certain threshhold has been seen, Loomo switches to the state CONVERSATION_EMPLOYEE. In the case that the person is not an employee, if they answered 'no' to the question: 'are you an employee', then they are taken to the state CONVERSATION_UNKNOWN.

Second State: CONVERSATION

The next state is conversation. This has been encapsulated in one class, this class is responsible for trying to interpret and understand what a guest or employee wants. The functionalities are currently constrained to the use-case. If a guest arrives and wants to meet someone or go to a meeting room, he would ideally say as much and Loomo goes through the process of finding the right person. After the correct person has been ascertained, we find ourselves in the message crafting phase. Finally the message is sent to the intended member of staff over the internal chatting application. The staff member then has the option of writing a message to the waiting person, which would be said by Loomo, or they can write "quickguide" followed by a destination, this will give verbal directions to the room or kitchen and show a small map on Loomo's screen. If users are unsure of how to proceed, Loomo is adorned with a large question mark which says a helpful message. The aim is hereby to say key or trigger words in order to lead the user to the expected way to interact with Loomo. This was attempted in a subtle way, which should lead to a relatively natural experience.

3.3 Discussion

There were a wide range of challenges when programming with Loomo. Despite the amount of sensors, primarily the microphones, the robot does not hear particularly well. Loud noises in the background make use difficult. Precise enunciation and leaning down to the robot's microphones are required in order for him to understand and accurately parse what is said. When recognizing a face, if the camera was directed up towards a light, he could no longer accurately see the person.

This was unfortunate as due to Loomo's small stature Loomo always has to look up in order to see someone.

Another limiting factor in many regards is the API version of the Android device. The Android version came out in November 2014, the speed of development of technology meant that many newer software solutions could not be implemented: originally Loomo would have video called the desired employee using Webrtc or Skype. Despite this limitation, the wealth of information and opensource projects available to use, led to a legitimate concierge.

4 Acceptance of an Automated Concierge

In order to test the acceptance of a robotic concierge with a focus on user experience, a variety of hypotheses were tested using the research method usability benchmarking. According to the Unified theory of acceptance and use of technology(UTAUT) there are four key criteria for usability: performance expectancy, effort expectancy, social influence and facilitating conditions.[24] UTAUT has been been empirically validated and attempts to unify popular acceptance measurement tools.

4.1 Hypotheses

As the majority of literature was dissimilar to this line of questioning; a certain amount of conceptualisation was required in order to create hypotheses that were appropriate. Common sense advantages of robots were taken into consideration as well as the current status of the technology vis-à-vis natural language recognition and the state of artificial intelligence.

Hypothesis 1: Affective state before interaction strongly influences perception of the reception[2] after interaction.

Hypothesis 2: The younger a user, the more likely they are to be able to intuitively interact with the robot.

Hypothesis 3: A robotic concierge is on average faster than a human.

Hypothesis 4: Human - human interaction is still preferential to human robot interaction(HRI)

Hypothesis 1 attempts to nullify potential skewed data in order to realistically judge the potential of this technology. Hypothesis 2 would be primarily judged as the effects of social influence as shaped by the prevalence of complex technology during the participants youth. Hypothesis 3 would fall primarily under performance expectancy. Lastly hypothesis 4 can be described as the interplay of all factors combined with human nature and the state of the technology.

4.2 User Study Details

This user study took the form of usability benchmarking, a script was given (can be found in the attached disc) and employees were asked to put themselves in the role of a guest. Usability benchmarking has been proven to produce reliable usability metrics.[21] Before attempting the assigned task, each person was asked to fill out a short survey with the following questions: age, gender, attitude towards receptions in general and the character trait extraverted/introverted as well as the self-assessment manikin (SAM) a metric for affective state which will be further detailed below.

The tasks given to participants, employees from iteratec GmbH, were conceived to accurately model a real world use of such a technology. There were two available tasks which were randomly distributed: a client that needed to know where a meeting room is and a spouse trying to find their spouse, the examiner in this case. These task groups were then further split into two groups, one interacting with Loomo and one interacting with the real receptionist. Normally visitors would be buzzed in from the receptionist but due to Loomo's inability to open the door and in order to keep test conditions constant, participants were let into the foyer by the examiner. The examiner then hastily retired to his desk in the office in the case that a participant was looking for their spouse. Alternatively the examiner retired out of view of participants in order to assess when the participant had been acceptably guided to their assigned meeting room.

[2]Robotic and human

Two standard metrics were used, SAM or the self assessment manikin and AttrakDiff mini. SAM was used to accurately measure the affective state of the participants prior to and after their interaction with man or machine, this follows a 9 point scale.

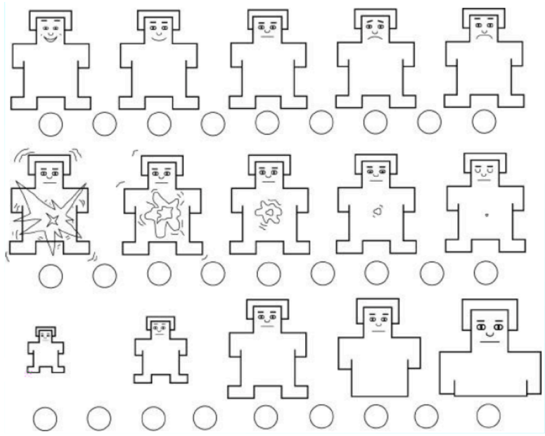

Figure 4.1: The self-assessment manikin(SAM)

From top to bottom the rows determine pleasure, arousal and dominance; pleasure ranges from happy to sad, arousal ranges from excited to calm and dominance ranges from submissive to powerful.

AttrakDiff was exclusively used to measure the usability of Loomo, with metrics such as pragmatic quality, attractiveness of the product and hedonic qualities: stimulation and identity. As well as a mass of other questions to attempt to gauge how positive of an experience was had and if the participant felt all their needs were satisfied. All questions were 7-point Likert scales in order to accurately determine intensity of opinion.

Further questions asked were:

- Would you prefer to interact with Loomo or a human in reception situations?
- Was the last interaction fun?
- Did you feel entertained by your counterpart?
- Was your counterpart empathetic?
- Were you satisfied with the help you received?
- Was the interaction confusing?
- Did the interaction feel natural?

The time needed for the tasks was recorded. This was kept constant by starting the timer once the participant had entered the front doors of iteratec GmbH. The questions were structured in a way as to not lead the participant unduly, not giving the exact keywords necessary to interact with Loomo.

4.3 Result Analysis

In total there were 20 participants who took part in the study. The average age was 28.1 with a standard deviation of 8.8. 10 of the participants were male and 10 female. Any Likert scales used shall be interpreted with 1 as a yes and a 7 as a no.

4.3.1 Hypotheses Analysis

Hypothesis 1: Affective State before Interaction Strongly Influences Perception of the Reception after Interaction

In order to effectively judge the perception of participants, data was correlated from SAM and questions from the questionnaire. The questionnaire questions asked if the last interaction was fun and if the person felt entertained by their counterpart. A 1 was labeled yes and a 7 was labeled with a no. From SAM the pleasure and arousal figures were taken.

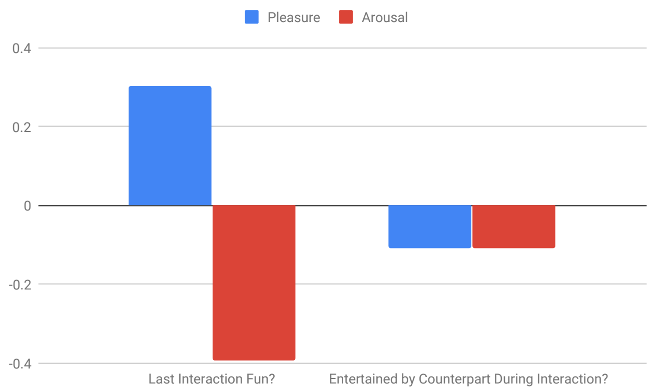

Figure 4.2: Correlation figures of SAM arousal and pleasure and questionnaire data

A 9 on the pleasure scale is a sad figure and a 7 on the questionnaire's scale was a no. Therefore, with a positive correlation of p=0.3, we can claim the happier the participant was before the interaction, the more likely they were to have had fun during their interaction. On SAM's arousal axis a 9 was a sleeping figure, a 1 an excited figure. There was a negative correlation of p=- .4, meaning the less a person was aroused, the more the interaction was perceived as fun.

There was a slight negative correlation, p=-.1, for being entertained by their respective dialogue partner. In the case of pleasure, this would indicate that people were more entertained if they were feeling more sad. In comparison with arousal, where the less stimulated a person was feeling, the more likely they would be to find the interaction entertaining. A definitive correlation could be seen between fun and pleasure and arousal. A lesser effect was noticed between entertainment value and pleasure and arousal.

Interestingly, if the participants were less happy, they were more entertained; in contradiction with the fun had. Therefore this hypothesis can be seen as accepted: people interpret interactions more fun if they are feeling happy and are not feeling stimulated. In addition, people are more entertained when they are feeling sad and unstimulated although to a lesser extent.

Hypothesis 2: The Younger a User, the More Likely They Are to Be Able to Intuitively Interact with the Robot

The amount of time needed to successfully complete a task and perceived naturalness of interaction is correlated with the age of the participant here as a measure of the intuitiveness of the device.

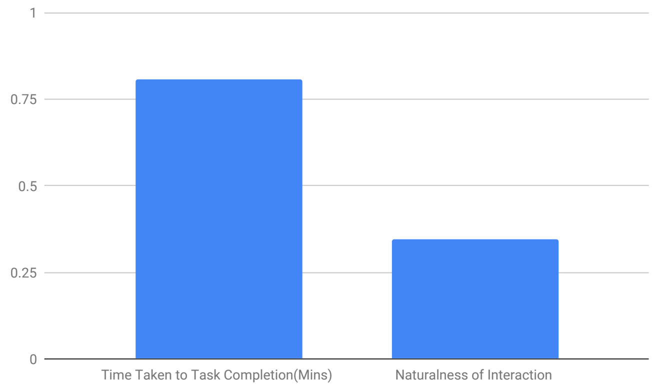

Figure 4.3: Correlation figures of age and time taken and age and perceived naturalness of interaction

A very strong correlation between age and time taken can be seen(p=0.8) and a significant correlation between age and perceived naturalness of interaction can also be seen(p=0.35). Taking these results into account, a clear correlation between age and ability to intuitively interact with the robot can be seen. Therefore, this hypothesis can be accepted.

Hypothesis 3: A Robotic Concierge is on Average Faster than a Human
The tasks as a singular list:

Task 1: Posing as a client, look for meeting room using Loomo

Task 2: Posing as a spouse, look for spouse (the examiner) using Loomo

Task 3: Posing as a client, look for meeting room by talking to the receptionist

Task 4: Posing as a spouse, look for spouse (the examiner) by talking to the receptionist

In order to assess the difference between human and machine, two sample t-tests were prepared; a two tailed test with p = 0.05 was created with Tasks 1 and 2 and Tasks 3 and 4, task 1 and task 3 and task 2 and task 4. Comparing human and robot[3], one notices a significant difference in results. A human receptionist was able perform the task much faster, requiring no time for a person to learn how to operate them. Tasks 1 and 3 also showed a statistically significant difference between human and robots. Tasks 2 and 4, however, did not conform to this trend.

[3]Group 1: task 1, task 2, Group 2: task 3, task 4

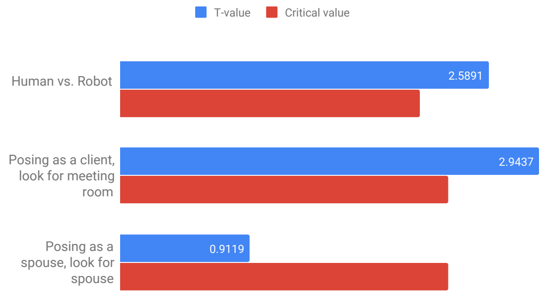

Figure 4.4: Comparison values across groups

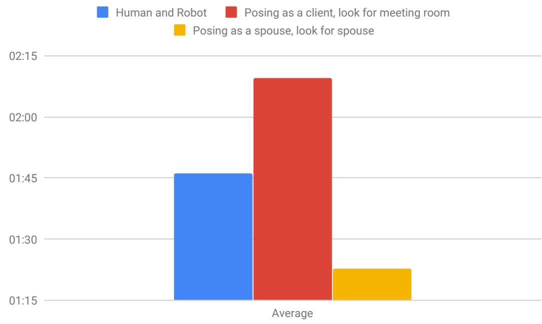

Figure 4.5: Average time taken

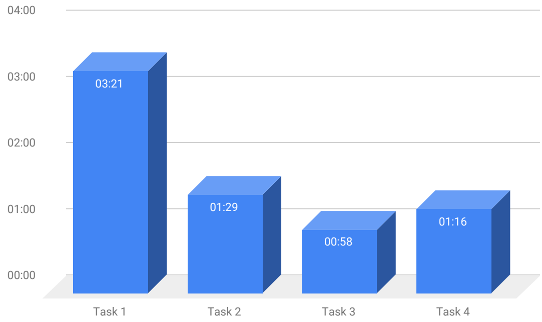

Figure 4.6: Average time taken per task

The first task took much longer due to an outlier taking double the time of most of the other participants. However the second task is closer to the time taken for a real receptionist with only 13 seconds difference. Despite these reasonably promising results, a human receptionist was much faster and slightly faster respectively. This hypothesis counts as partially disproven, as certain tasks proved not to have statistically significant differences.

Hypothesis 4: Human - Human Interaction Still Preferential to Human - Robot Interaction(HRI)

Each person was expected to interact with either the receptionist or Loomo as above, afterwards they were asked what they would prefer to interact with in further similar situations. From this data we created three one sample t-tests with p = 0.05 and a hypothetical mean of 4. Once again, a two tailed test was used.

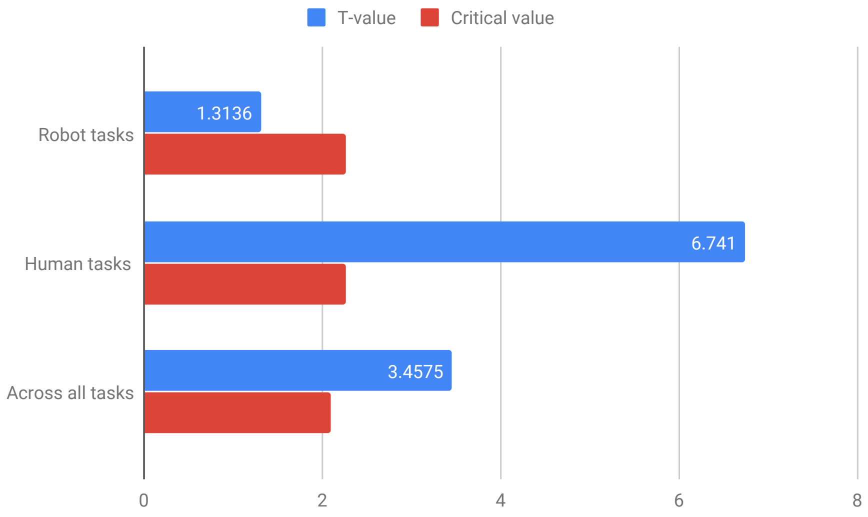

Figure 4.7: T-value and critical value

The group using the robot showed no statistically significant difference. After talking to a receptionist however, the results were strongly skewed in favour of a human receptionist. Across all tasks a statistically significant difference could also be noticed. These results show a tendency for people to like human interaction more, although those who actually interacted with Loomo didn't significantly prefer a human. With an available alternative, this hypothesis counts as validated. In the absence of an alternative, however, people seem to be satisfied with a robotic concierge.

4.3.2 Explorative Analysis

Many interesting results that were not hypothesis worthy came as a result of the user studies, which will be discussed here.

AttrakDiff
The results of the usability evaluation were as follows:

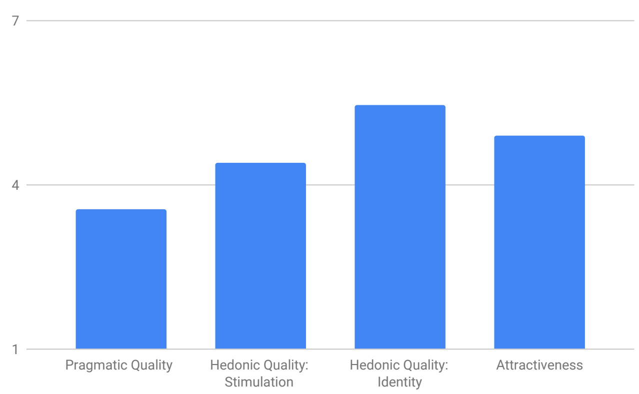

Figure 4.8: AttrakDiff results

The highest possible rating here is a 7, the lowest result a 1, hereby the results are very average:

people didn't think using Loomo was pragmatic or practical. They felt slightly stimulated by the interaction with the robot. Although users were able to identify with Loomo and found the GUI[4] attractive.

Results of Questionnaire

All extra questions not directly pertaining to hypotheses or not plainly illustrated have been collected and concisely explained:

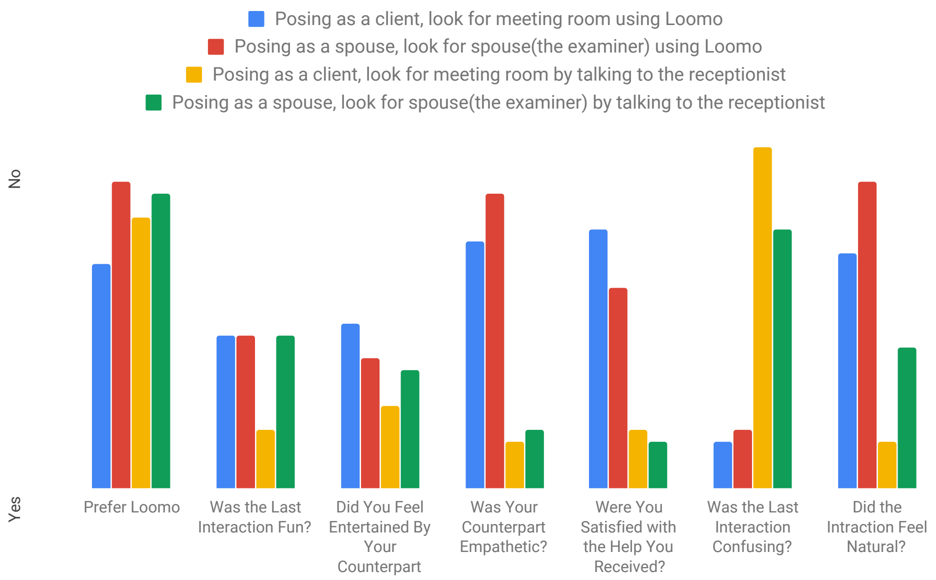

Figure 4.9: Average results of questionnaire

The results show that people strongly and consistently prefer a human receptionist in similar situations.

People who interacted with Loomo found he was moderately fun and was moderately entertaining, although the results more closely approximated ambivalence. When contrasted with a human receptionist, people found their experience on average more fun and felt more entertained.

Participants did not find Loomo empathetic and were moderately dissatisfied with the help they received. Whereas the receptionist was rated as being empathetic and guests were satisfied with the help they received. People were confused by Loomo and didn't feel that the interaction was particularly natural in contrast with the human counterpart, who people logically found was not at all confusing and felt the situation was moderately natural.

Time Taken according to Age Comparison

Interestingly there was an almost linear slope when comparing each participants age and time taken to task completion.

[4]Graphical User Interface

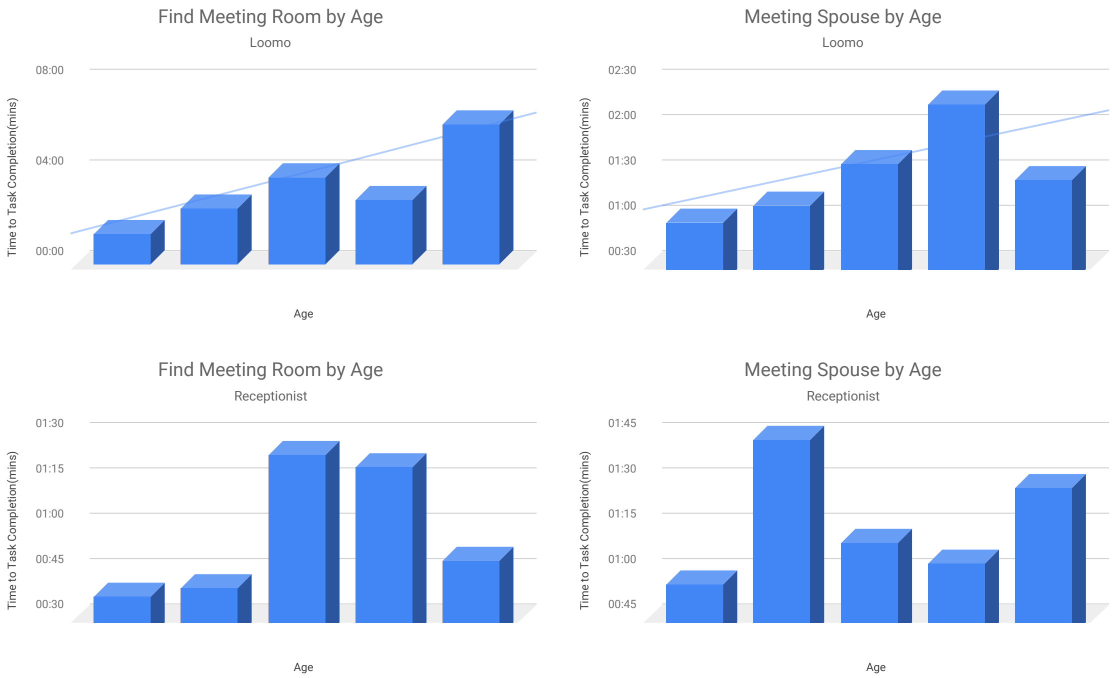

Figure 4.10: Time taken per task per individual arranged according to age

This is strongly contrasted with human - human interactions, wherein participants took arbitrarily long.

4.4 Discussion

One major problem found was the cleft between how the developer imagined users would interact with the software and how users expected the interaction to behave. The passive state was changed to an active one after seeing a face but what if the person began talking before their face had been seen? Some became frustrated that it was not doing what it was supposed to.

Another facet was the speech detection; as no open source continuous listening libraries were available, the standard android speech to text was used and abused in a way to give the illusion of continuous listening. This was achieved through an intelligent control system as well as restarting the listening procedure through regularly timed threads. Unfortunately, this also meant that if the listening process had not yet been started and people talked to Loomo only understood half of what was said. This was further exacerbated by some people not hearing the trigger[5] or merely for having an accent that did not conform to the norm[6]. However these types of problems are common with disruptive technologies[9]: a technology that is ground breaking or challenges norms, such as the personal computer at its debut. Potentially with time and ironing out of a few minor conceptual and practical problems, this technology could become as ubiquotous as the personal computer.

Another factor which could have played a role is the learning effect: after two or three runs the human receptionist knew what the task was and was able to more quickly respond to those needs. This as opposed to Loomo, with an inability to become practiced at a task.

[5] Standard android listening noise
[6] A notable example being an Austrian accent

5 Conclusion

The history of seemingly autonomous robots is storied and the future is rapidly becoming a reality. However this kind of technology still needs to be trialed more and integrated with a more powerfu

Many applications already exist to approximate some of the dimensions needed to create the impression of autonomy. What is lacking is the integration into one global program to give all the facets expected of intelligent life. Although perhaps not yet able in its own right, a robot like this definitely has the capability to become a low cost supporter of reception staff. The responses received from participants also shows that younger generations show less reluctance and difficulty in intuitively using an alien software. Given no alternative, people also seemed pleased with a robotic concierge.

This technology is unfortunately still in its infancy, more time and man power would be necessary in order to make this kind of technology robust and stable. Although as a proof of concept, this paper proves itself to be able. Some challenges still facing this technology are the natural variation of speech that is common in natural speech; as well as the natural disinclination to interact with an alien software or format. The question of whether or not the human experience can truly be replicated in order to give interactions a human touch is still debatable.

Enclosed CD Content

Original questionaire(German): As this work was written in english and everything has been translated in the body of this work, a copy of the original questionaire has been added.

Copy of Automated Concierge Android Program: Source code

Copy of Automated Concierge Android Program: Android .apk file

Spreadsheet with Collected Data

Copy of PDF

LaTeX File

References

[1] Known issues — Aldebaran 2.8.5.10 documentation. http://doc.aldebaran.com/2-8/news/known_issues.html, Dec 2018. [Online; accessed 28. Jan. 2019].

[2] H A. Yanco, Jill Drury, and Jean Scholtz. Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition. *Human-Computer Interaction*, 19:117–149, 06 2004.

[3] Evan Ackerman. Honda Halts Asimo Development in Favor of More Useful Humanoid Robots. *IEEE Spectrum: Technology, Engineering, and Science News*, Jun 2018.

[4] ASIMO - A.I. https://sites.google.com/a/siena.edu/a-i/e-asimo/inside-asimo, Nov 2018. [Online; accessed 10. Dec. 2018].

[5] ASIMO Specifications | ASIMO Innovations by Honda. http://asimo.honda.com/asimo-specs, Dec 2018. [Online; accessed 10. Dec. 2018].

[6] Two legs good | The Engineer. https://www.theengineer.co.uk/issues/november-2000-online/two-legs-good/#ixzz1TF8K9jyc, Nov 2000. [Online; accessed 10. Dec. 2018].

[7] Soledad Bedu, Clément Bidot, and Sandrine Tourcher. *How To Create a Great Experience with Pepper*. SoftBankRobotics, September 2017.

[8] Retail Service Robots to Deliver Data | About Bossa Nova. http://bossanova.com/about/about-company, Jan 2019. [Online; accessed 14. Jan. 2019].

[9] Joseph L Bower and Clayton M Christensen. Disruptive technologies: catching the wave. *Harvard Business Review*, 1995.

[10] JIBO, The World's First Social Robot for the Home. https://www.indiegogo.com/projects/jibo-the-world-s-first-social-robot-for-the-home#, Dec 2018. [Online; accessed 10. Dec. 2018].

[11] Akkharaphong Eksiri and Tetsuya Kimura. Restaurant service robots development in thailand and their real environment evaluation. *Journal of Robotics and Mechatronics*, 27(1):1, 2015.

[12] Dennis Green. Walmart reveals why it has robots roaming the aisles in 50 of its stores. https://www.businessinsider.de/walmart-robots-in-50-stores-2018-3?r=US&IR=T, Mar 2018.

[13] Meet Jibo, the world's first family robot - built by MIT's social robotics master - ExtremeTech. https://www.extremetech.com/extreme/186401-meet-jibo-the-worlds-first-family-robot-made-by-mits-social-robotics-master, Jul 2014. [Online; accessed 10. Dec. 2018].

[14] Khari Johnson. Bossa Nova Robotics raises $29 million for its store inventory robot. *VentureBeat*, Jun 2018.

[15] Lora Kolodny. Bossa Nova just raised another $29 million for its grocery store robots used by Walmart. https://www.cnbc.com/2018/06/20/bossa-nova-raises-29-million-for-grocery-store-robots-walmart-uses.html, Jun 2018.

[16] Gustavo López, Luis Quesada, and Luis A. Guerrero. Alexa vs. siri vs. cortana vs. google assistant: A comparison of speech-based natural user interfaces. In Isabel L. Nunes, editor, *Advances in Human Factors and Systems Interaction*, pages 241–250, Cham, 2018. Springer International Publishing.

[17] Patricia Marx. Learning to Love Robots. `https://www.newyorker.com/magazine/2018/11/26/learning-to-love-robots`, Nov 2018. [Online; accessed 14. Jan. 2019].

[18] Parmy Olson. Softbank's Robotics Business Prepares To Scale Up. `https://www.forbes.com/sites/parmyolson/2018/05/30/softbank-robotics-business-pepper-boston-dynamics/#6f2d41054b7f`, May 2018.

[19] Replacing humans with robots in the service sector has its limits - The Mainichi. `https://mainichi.jp/english/articles/20170311/p2a/00m/0na/005000c`, Dec 2018. [Online; accessed 10. Dec. 2018].

[20] *Pepper - User Guide — Aldebaran 2.4.3.28-r2 documentation*, Apr 2018. [Online; accessed 10. Dec. 2018].

[21] Christian Rohrer. When to use which user-experience research methods. *Nielsen Norman Group*, 2014.

[22] *Social Robotics: 8th International Conference, ICSR 2016, Kansas City, MO, USA, November 1-3, 2016 Proceedings*, volume 9979. Springer, 2016.

[23] Jeffrey Van Camp. Review: Jibo social robot. `https://www.wired.com/2017/11/review-jibo-social-robot`, July 2017. [Online; accessed 10. Dec. 2018].

[24] Viswanath Venkatesh, Michael G Morris, Gordon B Davis, and Fred D Davis. User acceptance of information technology: Toward a unified view. *MIS quarterly*, pages 425–478, 2003.

[25] Laura Wood. Global Consumer Robot Market 2017-2023. `https://www.prnewswire.com/news-releases/global-consumer-robot-market-2017-2023-market-to-grow-at-a-cagr-of-2235-to-reach-14` `html`, Dec 2017.

[26] Holly A Yanco and Tucker Balch. The american association for artificial intelligence mobile robot competition and exhibition. *AI Magazine*, 24(1):45, 2003.