DEICTIC GAZE IN VIRTUAL ENVIRONMENTS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

EFECAN YILMAZ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF MASTER OF SCIENCE
IN
THE DEPARTMENT OF COGNITIVE SCIENCE

AUGUST 2018

DEICTIC GAZE IN VIRTUAL ENVIRONMENTS

Submitted by Efecan Yılmaz in partial fulfillment of the requirements for the degree of **Master of Science in Cognitive Science Department, Middle East Technical University** by,

Prof. Dr. Deniz Zeyrek Bozşahin
Dean, **Graduate School of Informatics** _____

Prof. Dr. Cem Bozşahin
Head of Department, **Cognitive Science** _____

Assoc. Prof. Dr. Cengiz Acartürk
Supervisor, **Cognitive Science Dept., METU** _____

Examining Committee Members:

Prof. Dr. Cem Bozşahin
Computer Engineering Dept., METU _____

Assoc. Prof. Dr. Cengiz Acartürk
Medical Informatics Dept., METU _____

Prof. Dr. Deniz Zeyrek Bozşahin
Cognitive Science Dept., METU _____

Asst. Prof. Dr. Murat Perit Çakır
Cognitive Science Dept., METU _____

Asst. Prof. Dr. Özkan Kılıç
Computer Engineering Dept., Yıldırım Beyazıt University _____

Date: *28.08.2018*

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name : Efecan Yılmaz

Signature :

ABSTRACT

DEICTIC GAZE IN VIRTUAL ENVIRONMENTS

Yılmaz, Efecan

MSc., Department of Cognitive Sciences

Supervisor: Assoc. Prof. Dr. Cengiz Acartürk

August 2018, 53 pages

The research in human-robot interaction (HRI) involve topics, such as interlocutor collaboration in joint action, deixis in HRI, or the properties of shared environments. Moreover, referring expressions are particularly studied in joint action from both expression generation and resolution perspectives. Selective visual attention in gaze interaction and saliency patterns are also active topics in HRI. The present thesis investigated in a virtual reality (VR) environment an HRI and joint action situation with the assistance of eye tracking in a head-mounted display device in order to explore the augmentation of non-verbal communication in HRI. For this purpose, we employed a multimodal approach in communication with both non-verbal deictic expressions (gaze) and explicit verbal references in a multi-robot agent, single human experiment setting. The number of robot agents varied during experiments in order to investigate the social robotics influence of this measure on our metrics. We also utilized two distinct robot agent designs to explore an interaction effect with the number of robot agents as we evaluated participants' deixis resolution time and accuracy, as well as their gaze interaction patterns. The results of the research showed that the participants' accuracy, gaze interaction, and response time in deixis resolution were significantly influenced by the varying number of robot agents. However, this effect was not present when the participants were presented with explicit verbal references. Participants' gaze interaction results also showed that the number of robot agents significantly influence the saliency of the robot agents. Moreover, the participants interacted with the robot agents even when the joint task did not require gaze interaction.

Keywords: Human-robot interaction, virtual reality, eye tracking, deixis, social robotics

ÖZ

SANAL GERÇEKLİK ORTAMINDA YÖNLENDİRİCİ BAKIŞLAR

Yılmaz, Efecan

Yüksek Lisans, Bilişsel Bilimler Bölümü

Tez Yöneticisi: Doç. Dr. Cengiz Acartürk

Ağustos 2018, 53 sayfa

İnsan-robot etkileşimi araştırmaları muhattapların ortak eylem durumlarındaki işbirliği, insan-robot etkileşiminde gösterim özellikleri veya ortak etkileşim ortamlarının özellikleri gibi konular içerir. Özellikle yönlendirme ifadeleri ortak eylem durumlarında hem ifade üretimi ve ifade çözümlemesi bakış açılarından çalışılmaktadır. Bunların yanında, insan-robot etkileşiminde bakışsal etkileşimlerdeki görsel dikkatte seçicilik ve göze çarpma örnekleri de bazı diğer aktif araştırma alanları arasındadır. Bu tezde insan-robot etkileşiminde ve ortak eylem durumlarında, sanal-gerçeklik ortamında göz takip teknolojisinden de destek alarak, sözel olmayan iletişimin arttırılmasını araştırılmıştır. Bu amaçla, deneylerimizde çok kipli bir yaklaşımla sözel olmayan ifadeler (gözbakışı) ve sözel referanslar ile çoklu robot avatar, tek insan deney ortamı kullanılmıştır. Robot avatarların sayıları robotların ölçümlerimiz üzerindeki sosyal etkilerini araştırmak amacıyla deney boyunca değişken tutulmuştur. Ayrıca, katılımcıların gösterim çözümleme zamanlarını ve doğruluklarını, ek olarak da gözbakış ile etkileşimlerini değerlendirirken, robot avatarlarının tasarımında bir etkileşim etkisini araştırmak için iki ayrı robot avatar tasarımı kullanılmıştır. Araştırmanın sonuçları katılımcıların gösterim çözümleme zamanları, doğrulukları ve robotlar ile gözbakışı ile etkileşimlerinin değişken robot sayılarından etkilendiği görülmüştür. Ancak, bu etkiler katılımcılardan sözel referanslara cevap vermeleri istenildiğinde devam etmemiştir. Ek olarak, katılımcıların gözbakışı verileri ortak etkileşimin robotlarla görsel etkileşim gerektirmediği durumlarda bile robot avatarların sayılarındaki değişimin etkisi olduğunu göstermiştir.

Anahtar Sözcükler: İnsan-robot etkileşimi, sanal gerçeklik, göz takip, gösterim (İng. deixis), sosyal robotlar

To my mother,

my brother,

my partner,

and my best friend;

Filiz, Eşref, Begüm, and Taylan

ACKNOWLEDGEMENTS

I would like to express my most heartfelt gratitude to my thesis advisor Assoc. Prof. Dr. Cengiz Acartürk, for his endless guidance, encouragement, constructive criticism, as well as for him going the extra mile every chance he had in helping me work towards my dream of becoming an academic and a researcher. Thank you, sincerely.

I would also like to thank all instructors in my graduate institute for helping me both in and out class studies and in my endless inquiries.

I also wish to thank my friends from graduate school Yasemin Göl, Emre Erçin, İpek and Utku Havuç for their support both inside and outside the laboratories. You never made a big deal of your contributions but be it in motivation, help with data gathering, or the heated brain storming in the analysis period, you were always there.

I would like to thank all my friends, teachers, and colleagues from the department of Computer Technologies and Information Systems of Bilkent University; particularly chair Dr. Erkan Uçar, co-chair Serpil Tın, and, of course, Hatice Zehra Yılmaz for their teachings, patience, support, and endless efforts in helping me grow up, become a professional and an academic.

Finally, I would like to thank my mother, my brother, my partner in life, and my best friend for their patience and understanding in the countless opportunities of socializing and events I missed due to my studies the past two years.

This thesis is dedicated to you all…

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AMOLED	Active-Matrix Organic Light-Emitting Diode
API	Application Programming Interface
HCI	Human-Computer Interaction
HLM	Hierarchical Linear Model
HMD	Head Mounted Display
HRI	Human-Robot Interaction
SDK	Software Development Kit
SMI	Senso-Motoric Instruments
VR	Virtual Reality

CHAPTER 1

INTRODUCTION

The phenomena of joint action can take place through a number of methods, such as joint attention, action observation, task sharing, and action coordination (Galantucci, 2005; Sebanz, Bekkering, & Knoblich, 2006). These methods are often areas of focus in human-robot interaction (hereafter, HRI) studies. HRI is a recent research and development field (Rus, 2017). The research in this area involves a variety of topics, ranging from investigations of the properties of shared HRI environments (Haddadin & Croft, 2017) to how humans and robots collaborate in joint actions by means of shared tasks and/or deictic expressions (Fang, Doering, & Chai, 2015; Lemaignan, Warnier, Sisbot, Clodic, & Alami, 2017; Piwek, 2009). These deictic expressions allow for multimodal approaches involving a variety of communication methods to be employed in joint attention and task sharing in HRI. Some of these methods are where deictic expressions are accompanied by gaze (Admoni & Scassellati, 2017; Ruhland et al., 2015; Yücel et al., 2013), by verbal references (Fang et al., 2015; Lemaignan et al., 2017), or by gestures and other embodied referring expressions (Imai, Ono, & Ishiguro, 2001). Topics on how humans and robots share communication environments in collaborative tasks or by means of referring expressions have been previously investigated by researchers (Dağlarlı, Dağlarlı, Günel, & Köse, 2017; Imai et al., 2001).

Referring expressions specifically have been investigated from both the perspectives of expression generation and expression resolution (Brooks & Breazeal, 2006; Clark & Wilkes-gibbs, 1986; Devault, Kariaeva, Kothari, Oved, & Stone, 2005; Eldon, 2015; Whitney, Eldon, Oberlin, & Tellex, 2016), and also within the framework of discourse analysis and linguistics (Grosz, Joshi, & Weinstein, 1995; Levelt, Richardson, & La Heij, 1985).

Selective visual attention and saliency of objects both have pivotal roles in human – human interaction through gaze (Duchowski, 2017, p. 3). For instance, Morales, et al., (2000; 1998) investigated the effect of 6 to 24 month old infants' skill responding to joint attention in a multimodal setting, which included gaze also. The researchers claimed infants' joint attention skills influence their language acquisition rate significantly. Saliency, in particular, can be used as a method of detection through the inclusion of items and objects of varying saliency in the HRI environment (Breazeal, 2004). Common methods in research are object detection and/or recognition, while eye tracking is also frequently employed. Eye tracking can be utilized in HRI to investigate where and how gaze vectors are allocated in the communication environment, thus providing robust data for the role of saliency during the course of interaction (Duchowski, 2017, pp. 49–53; Yu, Scheutz, & Schermerhorn, 2010).

Along with the emergence of virtual reality (hereafter, VR) in the past two decades, new areas of research in human-computer interaction (hereafter, HCI) and

HRI emerged. Furthermore, deictic expressions and gaze interaction are relatively novel domains of research in both HCI and HRI research that take place in VR environments. The improvements in quality of technical aspects, such as higher resolution and higher refresh rate displays available in head-mounted VR equipment, also allowed these novel domains to gain importance. The increase in quality also expands to other aspects of VR; such as better motion tracking in body tracking, and three dimensional motion sensors, or more natural input methods in better haptic engines. These VR enabling equipment shifted from being limitations against immersion to factors that allow for better human immersion in VR environments. As a result of these limitations being overcome, while not yet fully mature, the research on VR has been expanding (Baizid, Li, Mollet, & Chellali, 2009; Wang, Giannopoulos, Slater, & Peer, 2011; Witmer & Singer, 1998). While virtual reality has also been studied from the perspective of human cognition (Duguleana, Barbuceanu, & Mogan, 2011; Rizzo & Buckwalter, 1997; Wickens & Baker, 1995). As Admoni and Scassellati (2017) claim, the communication between the interlocutors in HRI, in addition to verbal communication, can take place in non-verbal modalities, such as gaze and gestures.

The present thesis employed eye tracking in a head mounted display (hereafter, HMD) to investigate aspects of HRI in a VR environment. The environment had a human-robot joint attention setting in order to explore the augmentation of non-verbal communication in HRI. For this purpose, we designed a VR setting such that multiple robot agents and a single human participant shared a communication environment. The multimodal design of our setting employs both verbal, explicit referring expressions and deictic expressions assisted by robot agents' gaze vectors, thus allowing us to investigate VR HRI from a social robotics joint action perspective within the scope of the following research questions:

1. Does the number of robot agents have an influence on the resolution of deictic referring expressions by human participants (measured in terms of accuracy and response time of the participants)?
2. How is the human participant's gaze distributed on the robot agents?
3. Does the design of a robot agent (i.e. body and face morphology) have an influence on gaze measures (accuracy, response time, and gaze distribution)?

The first research question is directed at the deixis in the discourse within the virtual reality environment. For this first condition, a question was replayed to the human participant and the robots of varying numbers were configured to assist the participant in answering the questions. The questions involved resolution of an implicit deictic expression, such as "what is the object *here*?", as the robot(s) fixated with their gaze on an object on the table within the VR environment. The participants were provided no clues other than the gaze expression used as a deictic reference by the group of robot agents. Our hypothesis in the scope of the first research question, firstly, is that there will be a significant effect on the resolution of deictic expressions in terms of accuracy and response time of the participants. Secondly, it is that this effect will be partially explained by the amount of available data, as specified by the number of robots (thus the number of gaze vectors that refer to the object on the table) will have an influence on accuracy and response time. Specifically, as the number of robot agents increases the amount of gaze vectors available to the human participant will increase linearly, which may cause the participant to have a significantly different accuracy and/or speed in resolving deictic references.

The second research question is directed at the role of visual attention and saliency, as well as social aspects of a communication environment in VR. We investigated how salient robot agents in a VR environment are, without a pre-existing requirement given for the participants to commit into gaze interaction with the robot agents. For this, robot agents of varying numbers were configured in the experiment setting as the human participant answered questions. The questions were in forms of referring expressions that utilized geometric locations of objects on the table in the VR environment. These questions were "what is the object is at left top" (and similarly "what is the object at the left bottom / right top / right bottom"). The number of robots and the robots' gaze vectors were randomized for each new scene. As a result of the explicit verbal description, the participants were not required to commit to gaze interaction with the robot agents. This setting was that the participants were able to answer the questions directly, based on the provided geometric description in the question rather than an implicit reference resolution. Our hypothesis is that even without a requirement for gaze interaction in HRI, the participants will interact with the robots and that they will do so in significantly differing numbers depending on the number of robots in the VR environment due to the affected saliency of the group of robot agents.

Finally, we investigated how the *design* of the robot agents would interact with the first two research questions in the VR environment. In order to bring out this likely effect, two robot designs were used in the experiment. Our hypothesis is that the design of the robots will have a significant effect on accuracy and response time of the participants in their answers about object locations.

In summary, this thesis employed a multimodal setting (i.e., verbal references and non-verbal deictic expressions) in a VR HRI environment where a multi-agent, single human participant joint attention and joint action situation were investigated. With this aim, the joint action situation was evaluated from the perspectives of expression resolution accuracy and response time of the participants, as well as the human participants' gaze interaction patterns.

The following chapter two presents a literature review of multi-party interactions, deictic references, and joint action in HRI, as well as the contribution and/or the effect of VR in HRI. In chapter three and four our experiment methodology and results respectively are presented. In the former of these two chapters we present the experiment conditions, experiment sessions, our VR enabling device's technical specification, and our analysis procedure. In the latter chapter we present our experiment results in our experiment conditions' respective sections. Finally, in chapter five, we present a discussion and conclusion of this thesis that evaluate various aspects and details of our findings from our experiment results, and conclude with the perspectives gained via this thesis and the research involved for each of our research questions.

CHAPTER 2

LITERATURE REVIEW

2.1. Social Robotics, Multi-Agent Settings and Collaboration in HRI

Social robotics is the study of robots that are capable of interacting with humans in joint environments, often as partners with human interlocutors. Communication modalities, methods, as well as interaction patterns between the interlocutors are major fields of research in social robotics (Breazeal, Dautenhahn, & Kanda, 2016). Admoni and Scassellati (2017) reported this interaction between humans and robots within the framework of the concept of intuitive interactions in their review of social gaze in HRI. Admoni and Scassellati, firstly, propose three categories of research in HRI: human-focused, design-focused, and technology-focused approaches. Secondly, the researchers categorize how social eye gaze is investigated in various studies in mutual gaze interaction, deictic gaze referencing, gaze aversion and joint visual attention.

Our study is compatible with the first two categories reported by Admoni and Scassellati (2017). *Human-focused* studies are those that investigate human actions in HRI situations. Our study, firstly, investigated human cognition in a virtual reality HRI environment from the perspectives of deictic expression resolution time and accuracy in resolution. Secondly, the participants' gaze interaction patterns were also investigated. The researchers report that *design-focused* studies investigate the design choices for the robot agent or agents within HRI environments. Our study explored robot agents' design by means of comparing two design choices in the third research question by focusing on whether the design choices have a significant effect on communication in the HRI setting or not. In terms of the second list of categories that Admoni and Scassellati (2017) reported, our experiment design included *social gaze* in our joint attention setting that consisted of interlocutor agents in the VR HRI environment and multi-agent focus of attention on the table.

Admoni and Scassellati (2017) report that a study in the *technological-approach* category should provide a systematic approach into how robot agents in virtual environments operate their gaze during the course of an interaction. A similar survey study offered a set of guidelines about how the eyes, eyelids, gaze vectors, and facial features of robot agents should be animated (Ruhland et al., 2015), where the researchers approached the topic from the perspectives of computer graphics and motor animation technologies. They state that as computer technologies have developed, the gap between the appearances and the motor movements of virtual characters have widened.

Clark and Wilkes-gibbs (1986) stated that in multi-party communication environments, joint effort must be minimized in virtual environments for the robot agents and the human participant to engage in a joint task in collaboration. We followed this suggestion in our study, by means of keeping the variance in gaze

interaction between robots and human participants minimal such that the robot agents utilize their gaze vectors on the objects or the participants in mutual exclusivity.

Ruhland, et al., (2015) approach their survey from the perspectives of eye anatomy and physiology, in order to investigate how eye gaze is utilized as a deictic reference. They report that utilizing gaze as a non-verbal tool of communication serves various functions, such as relaying information or indicating and/or directing visual attention, as well as facilitate references in a shared visual space. In our study, the robot agents' gaze vectors were utilized by means of these two functions reported research literature. Our second experiment session, where the deictic references were given through implicit, verbal expressions, the participants were assisted by gaze expressions of the varying number of robot agents in the joint action situation. Furthermore, Ruhland, et al., (2015) report that utilization of human participants' gaze and visual attention play roles of engagement and distraction in multi-party communications. They also claim that the design of the robots play a significant role in our exploration. In our study, we follow a categorical approach in the design process of our robots; in that humanoid robots and non-humanoid robots are the two distinct agent types, as described in the next chapter.

Another novel aspect of our study is that we designed a single-human multi-robot environment where the number of robot agents also varies between the two experiment sessions. Our approach in the multi-robot single-human setting allowed us to compare and contrast the aforementioned extra layer of information reported by Ruhland, et al., (2015). In the process of creating a common goal for a joint action, our experiment conditions followed an approach that either openly allowed for ambiguity in deictic gaze with implicit references and gaze references utilized in combination, or eliminated it entirely with explicit references.

Deictic expressions may be non-verbal, such as gaze or gesture expressions as well as in verbal expression form. Devault, et al., (2005) evaluated collaborative reference from a linguistics point of view. In order to propose a dialog system in human-human communication, the researchers analyzed collaborative reference as a supervised model alternating between descriptions and affirmations between the interlocutors. The ultimate aim of the study was to replicate a collaborative communication behavior; as the researchers investigated collaborative reference from the perspectives of information states and linguistic references in both utterance planning and understanding. Devault, et al., give a referring task example. For the success of the example task, a target object must be identified correctly by means of participant supervision for the robot agent. The joint task is similar to the task in our study; the difference is that in our references, there is neither a verbal nor a non-verbal supervision. Our experiment conditions were always presented to the human participant in the communication environment and the robot agents did not provide any instruction other than a non-party-alternating gaze reference to the participant. Therefore, the joint task's success in our experiments solely relied on the validity of the common goal in the communication environment. The methodological approach being created by means of the joint task was, as a result, for the participants to supervise their own answers. Therefore, existence of more gaze data are expected to assist resolution of implicit deictic references.

In their second perspective of linguistic references Devault, et al., (2005) investigated utterances from the perspective of semantic representations and

utterances' grammatically specific constraints. As mentioned in chapter 1 joint tasks require both parties of the interaction to contribute to the success of the common goal. According to Devault, et al., the contributions must also be shared between the interlocutors, and be contextually appropriate for a joint task. The researchers claim that these properties create a knowledge-interface and is a must for a robust collaborative reference situation to take place. Our approach is that in our experiment the robot agents continuously refer to the objects in two types of verbal expressions. The first condition was an explicit referring expression that utilized the geometric locations of objects within the VR environment. In the context of the joint task, the robot agents and the human participant always followed a set of questions in the explicit questions. Allowing the constraints to be realized as geometric locations, and be shared between the HRI interlocutors. On the other hand, the second condition was an implicit deictic expression referring to the objects by means of gaze vectors of robot agents in the VR HRI environment. This second condition always had a single question replayed to the participants tasked with resolving implicit references in the context that the robot agents were providing them with gaze data in collaboration. The geometric locations in both of these conditions were static and albeit that the objects in the environment shifted between the locations of one another, as further explained in chapter 3, the locations never shifted from four corners of the table. Allowing further context to be realized that the four corners would be solely referred.

Finally, another relevant study on gaze interaction in multi-agent HRI environments was conducted by Mutlu, et al., (2009). They investigated how gaze cues can be utilized in HRI, more specifically in information exchange tasks. The researchers aimed to investigate what forms of cues a robot agent can provide to human participant(s) in order to regulate conversational roles. They investigated the gaze interaction and joint attention in participants in three distinct conversational roles. Firstly, a robot agent addresses participants in a one-to-one HRI situation. Secondly, a robot agent acknowledges the existence of by-standers (non-participating actors) by means of cues during an ongoing conversation. Finally, a robot agent also acknowledges over-hearers (non-acknowledged by-standers) in the HRI environment. Mutlu, et al. employs a situation where the robot agent needs to provide gaze cues in managing turn-taking, and showing active listening behaviors, as well as to use gaze in social communication.

Mutlu, et al., (2009) designed gaze cues for their robot agent to signal the three conversational roles to the participants. The results showed that gaze cues significantly manipulated the information in regards to whom it is being addressed in a multi-party conversation; this is observed when the robot agent, firstly, initiates a greeting period where both the addressees and the bystanders are greeted with gaze cues in acknowledging all parties a part of the information exchange. The robot agent, then, directs its gaze towards the addressees mainly and away from the bystanders, causing this significant manipulation. Mutlu, et al. also investigated how these gaze cues can be and/or are utilized in turn exchanging in shared tasks. They claimed that by means of only using gaze cues, a robot agent is able to manipulate information flow in an HRI environment as a function of who attends the conversation, which parties are grouped, as well as their liking of the robot agent. The researchers also state that the addressees are more likely to commit to the joint attention. Our study utilized gaze cues of the robot agents similarly with the research by Mutlu, et al., (2009). In our investigation, gaze cues (as deictic expressions) were partially utilized to direct attention, as a role exchange between the robot agents' and the human participants' does not take place.

In the first experiment session of our experiment joint task was executed by means of robot agents utilizing geometric locations of objects with explicit referring expressions. In the first session, to engage the participants further with the joint attention task, the robot agents fixated their gaze on the participants' location in the VR environment instead of the objects in approximately 25% of the time.

2.2. Deictic References and Joint Action in HRI

The previous research on HRI in social robotics and joint attention tasks utilize multi-modal approaches, where the conditions may involve the interaction of a user with a robot through deictic expressions. More generally, in a multimodal setting, the focus of research is usually verbal communication or social gaze.

Lemaignan, et al., (2017) claim that there exist certain requirements for a robot agent not only to share an HRI environment with a human, but also to commit to a joint action and multi-modal communication. They claim that a robot needs to be able to both reason with, as well as assess and represent the HRI environment to the human participant in communication environments. Meanwhile, in terms of joint action, there are three aspects on which the basis for joint action is built. The first is a joint goal for the human participant and robot agents to set a target within, such as answering true or false questions or exchanging information about the HRI environment. Secondly, an HRI environment setting where both the robot agent and the human participant can identify the objects and/or a situation. Finally, a belief state where the human and the robot share common knowledge and sense. Clodic, Alami and Chatila (2014, p. 172) name these processes and requirements as "intention, goal, plan, knowledge, skills, [and] a model of the current reality".

With the aim of contributing into social robotics from the perspective of joint action in HRI, Lemaignan, et al. (2017) developed a cognitive processing architecture to deploy interactive robots in a goal oriented joint setting. They took into account verbal communication, deictic gestures, and social gaze in their multi-modal communication model. The model involved a robot participant to obtain a representation of the physical world by means of the methodology offered by Sisbot, Ros and Alami (2011). As a result Lemaignan, et al., based their model on description logistics, logical disambiguation of objects, and mental modelling of the environment. By employing this model, Lemaignan, et al. aimed for their physical robot agent to interact with human participants through semantics and cognitive skills. Another novel property of the model is that it allows the robot agent to take limitations into account for the human participant, such as whether the objects in the environment being referred to are reachable by the addressed participant or not. Our approach to designing the virtual reality HRI environment was similar to the research by Lemaignan, et al., (2017) in terms of how the joint action was formed between the robot agents and the human participant. The communication model in our study was also similar to that of Lemaignan, et al. in that verbal communication was assisted by a deictic reference in social gaze. However, in our study the HRI environment was set in virtual reality, and the joint attention situation was accomplished by a question and answer based process, in a minimalistic environment where only salient objects and no distractions exist. This allowed us to investigate our research questions with a sole, pre-set task, which did not involve the objects being interacted by anything other than referring expressions, collaborative gaze and deictic references.

There are also similarities between the architecture developed for cognitive processing of Lemaignan, et al., (2017) and our experiment methodology. The first of these was a module in Lemaignan, et al. for robot agents to structuralize goals and manage joint plans in HRI environments. Our robots were not interactive with the participant, however they were programmed to provide information through gaze vectors to the participant in the VR HRI environment. As a result our robot agents had a fixed goal of providing gaze data to the participants and to direct the participants' attention to the relevant object on the table in questions. Furthermore, the participant also shared the fixed common goal of communicating with the robot agents through gaze interaction. Similarly to the experiment setting of Lemaignan, et al., geometric locations were utilized as the referred points in the VR environment space for the explicit referring expressions. Finally, the contextual setting between the participants and the robot agents was fixed in that the participant was responsible for answering correctly to the questions asked for validity of the communication.

Another similar research was conducted by Fang, et al., (2015), where the researchers developed a model for generating referring expressions in collaborative tasks in HRI. The researchers designed a joint action environment where the robot's gesture behavior was supervised by either or both the human participants' gaze vectors and verbal feedback. Similarly, the robot agent, in its referring expression generation task, utilized either or both deictic expressions and gesture based expressions. Fang, et al. explored collaborative referencing in HRI from the perspectives of embodied deictic expressions, such as the ones accompanied by gestures, and gaze interaction patterns of the interlocutors. The researchers argue that in most approaches the generation of referring expressions are solely based on linguistic approaches. They claim that only with recent contributions to computer vision studies the approaches to studying HRI have been widened with generation of multi-modal expressions. Fang, et al. follow a similar perspective to that of Lemaignan, et al., (2017) in that humans and robot agents do not have the same representation of the shared world and this is an issue in need of investigation. In order to address this problem, Fang, et al. investigated their multi-modal approach in HRI by making the robot agent come up with referring expressions in forms of questions. The researchers claim that this allowed an investigation of collaborative referencing in HRI from a broader perspective.

In forming their questions, Fang, et al., (2015) utilized geometric locations of the objects on the table with respect to the participant with the context of information exchange as their joint action. Our methodology was designed in a similar approach to that of by Fang, et al. In our virtual reality HRI environment, where the varying numbers of robots and a single human participant engaged in joint action, the multi-modal setting is either established through verbal communication only, or through deictic expression assisted gaze references.

Another research conducted in this field is Imai, Ono & Ishiguro (2001), where the researchers investigated situated utterance generation in HRI environment, while also focusing on the factor of achievement in joint attention. They also conducted a psychological experiment on effective methods of establishing joint attention in HRI. The researchers claim that there are three issues their system needed to overcome in developing this communication model: directing the participant's attention to the robot's focus of attention, letting the participant receive robot's communicative intention, and dealing with the participant's formed attention towards information in

the environment. These were challenges we also faced. The first difficulty of directing the participant's attention was handled through a categorical approach of two types of referring expressions; a verbal, explicit reference in linguistic discourse and an implicit, gaze based referring expression. The regularity in our approach formed a common goal the participant expected the robot agents to initiate communication with each new scene. Moreover, the fixed geometric locations of the objects in our VR environment builds up a situated environment where the objects have clear saliency over the rest of the environment. This also addresses Imai, et al.'s final issue. Our methodology was that we shifted the objects between each other's locations randomly in each scene to prevent the participant forming a predisposition.

A study in visual saliency by Piwek (2009) investigated what is necessarily a property of an environment or an object for being a referring expression "pointing is primarily a means for changing the saliency of objects" (Piwek, 2009, p. 4). The researcher mainly focused on the changes in computational efficiency when saliency was introduced into an object. The main outcome of this research for our study is that it is a compatible approach to utilize referring expressions within the VR environment in order to increase the objects' and the referents' saliency.

Similar to Piwek (2009), another gaze based HRI research was conducted by Yücel, et al., (2013) to investigate saliency in joint attention environments. The researchers studied joint attention in HRI in a physical environment, to develop an image based method for a robot agent. The robot agent's aim was to establish joint attention with a human participant. They followed a novel approach by using saliency as a metric in estimation of gaze direction of the human participant, allowing their robot agent to commit into visual joint attention without the need of eye tracking. Yücel, et al. utilizes not only images of the human and the salient object, but also uses the head pose of the human to predict the salient object. We designed an experimental condition similar to this for the human participant to establish joint attention with the robots. In our study, the utilization of eye tracking technology in the VR environment allow us to accurately predict the human participant's gaze vector in real time, which allows us to analyze how the human participants' gaze vectors are distributed. However, the robot agents in our environment do not commit to any communication modality other than utilizing their gaze vectors. The interaction our robot agents make are fixed coded and randomized in each new scene in the experiments, as described in chapter 3.

2.3. Interaction and Robot Agent Design in Virtual Reality

Virtual reality is a computational set of tools put together in order to create an environment in which humans can be better immersed in three dimensional virtual environments (Rizzo & Buckwalter, 1997; Wickens & Baker, 1995). State-of-the-art virtual reality equipment, such as head mounted displays with high resolution and high refresh rates, along with intuitive input and movement enabling devices allow human participants to be immersed in VR environments from all their senses. This interaction is not exclusively limited to any of human-computer, human-robot or human-human interaction. However, virtual reality (VR) must not be confused with virtual environments designed in three dimensional graphics alone without the immersion enabling devices. VR involves not only a three dimensional design of an environment, but utilizes devices with stereoscopic displays, spatial sound and haptic input

mechanisms as reported by Adams (1999), and Wickens and Baker (1995). These equipment allow for the participants to be immersed in the three dimensional environment through most of our senses, in addition to vision.

Studies in HRI that utilize VR environments often focus on psychological or psychiatric therapy report (Anderson, Zimand, Hodges, & Rothbaum, 2005; Grealy, Johnson, & Rushton, 1999; Rizzo & Buckwalter, 1997; Schuemie, van der Straaten, Krijn, & van der Mast, 2001).

Moreover, cognitive issues in virtual reality have been discussed in a survey by Wickens and Baker (1995) where the researchers investigated approaches and perspectives in VR that enable users to be immersed in VR environments. They also report on the importance of multi-modal interaction with the environment. They claim that in the real world, humans interact with their environment through all their senses, and that virtual reality enabling devices work in towards simulating the realistic interaction with kinesthetic feedback, spatial audio, stereoscopic vision, and even bodily movement through auxiliary cameras. The conclusion of this survey is that an appropriate implementation of VR environment depend on both the environmental and equipment side requirements. In our study, we followed a similar, compatible approach with Wickens and Baker's reportings in our experimental design process.

Bartneck, et al., (2018) investigated how the visual design of robot agents might interact with racial perception[1]. The researchers claim that our first impressions of both other humans and non-humans in interaction environments are often based on visual cues, such as age, gender, and race. They argue that because of the visual cues, for instance, of body shapes or stereotypical hair styles the design of robot agents tend to give rise to ethical and social issues. The researchers claim that these issues of racial perceptions are likely to have implications in HRI. To investigate this perception and its effects, they manipulated the appearance of a Nao robot in different racial characteristics and utilized shooter bias paradigm[2] to explore the effects on reaction times and racial perception of the participants.

Bartneck, et al., (2018) report that there is a significantly more positive bias towards white characters and that the majority of their participants attributed racial features to the robot agents. In our experiment, we used two distinct robot agent designs that inherently have variance in color also: the humanoid robot agent design is a robot agent with a body, arms, head, and eyes with simple details, such as buttons on the body. On the other hand, the non-humanoid robot agent design is a simple sphere with eyes. We aimed to have neutral colors for both of our robot agents and as such the design decisions were done accordingly. The humanoid robot agent is a mixture of colors and the non-humanoid robot agent is a neutral gray color. Both designs are lit from an angle from the skybox in the VR environment graphics. The robot designs are presented in Figure 2 in chapter 3.

In designing non-human agents another important topic is the phenomenon of *uncanny valley*. Originally described by Mori (1970), the term uncanny valley is used to refer to the feeling of discomfort people experience when humans look at virtual

[1] Identification of robot agents as being racialized (Bartneck et al., 2018).
[2] A method employed to investigate the implicit tendencies for racial categorization and biased responses (Bartneck et al., 2018).

humanoid agents (Brenton, Gillies, Ballin, & Chatting, 2005). This phenomenon is considered in the design process of our robots particularly in the VR environment because of the immersion VR creates in a three dimensional perception of the graphical virtual environment. The phenomenon has been investigated by Seyama and Nagayama (2007), who claim that uncanny valley can apply to any human-like object, such as avatars in virtual reality. The researchers claim that facial features, while being the most unpleasant features of an uncanny design, are not by themselves enough for uncanny valley to be confirmed. The finding in the study is that for uncanny valley to develop there is also a requirement for abnormal features, such as bizarre eyes.

In our study, we followed a neutral color palette for the robot agents that are either balanced in their coloring between dark and light colors (humanoid design) and a neutral gray color (non-humanoid design). We do not investigate whether there is the effect of uncanny valley or not in our VR environment.

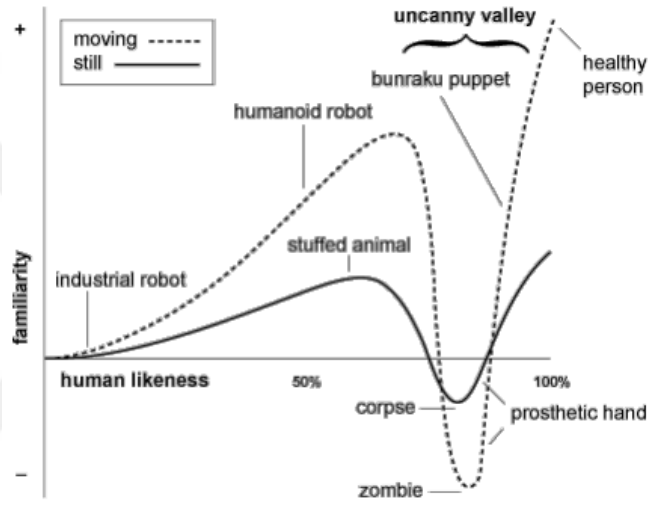


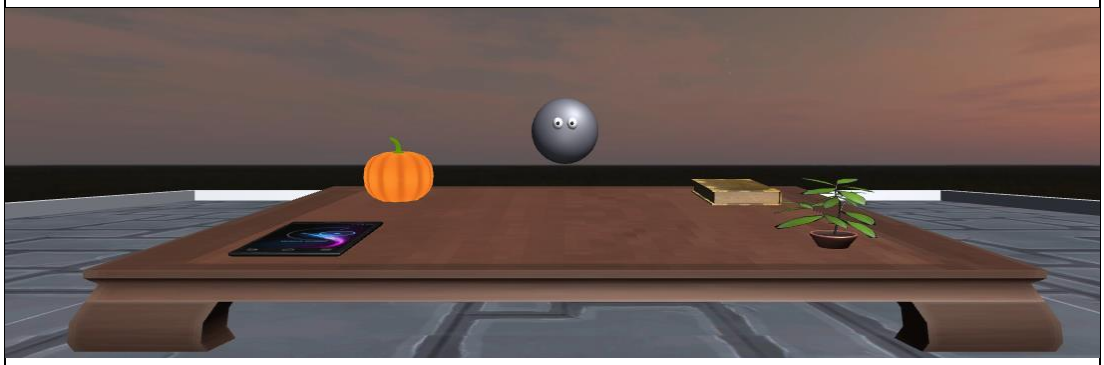
Figure 1 - Representation of Uncanny Valley from *(Mori, 1970)*

CHAPTER 3

METHODOLOGY

The experiments in the present study were conducted in an immersive virtual reality environment (VR). We employed the VR environment for the purpose of a multimodal investigation of joint action and gaze assisted deictic expressions within the context of Human Robot Interaction (HRI). The VR environment was implemented with a Head Mounted Display (HMD) device. The experiment setting is a 2x2 design: Two robot-design conditions were used as a between-subjects factor. The within-subject factor consisted of two distinct joint attention tasks (viz. reference resolution under the two experimental conditions). Further details of the two distinct joint tasks are given in 3.2.2 and 3.2.3. This experiment was designed in order to investigate our third research question (whether the design of a robot agent have an influence on gaze measures). The two designs of humanoid robot agents and non-humanoid robot agents are presented in Figure 2.

(a) Humanoid Robot Agent



(b) Non-humanoid Robot Agent

Figure 2 - Robot agent design as a between-subject factor.

The two within subject experiment conditions (explicit referring expressions and deictic expressions with gaze assistance) are identical between the two distinct robot agent designs. For the purpose of experimental control, robots' locations and the gaze vector angles were fixed. Figure 2 exemplifies our experiments' one-robot agent

condition, where the robot agents fixate their gaze on the participant (see 3.2.2 and 3.2.3 for details).

3.1. Participants

Forty-two participants from the Middle East Technical University and Bilkent University, Turkey, participated in the experiment for monetary compensation (approximately 5 EUR). The participants were either students or academic personnel, such as assistants, or instructors. The ages of the participants varied between 18 and 50 (M = 24.38, SD = 5.801).

The participants were randomly divided into two groups of 21 participants each. The two groups were presented either the humanoid robot agent condition or the non-humanoid robot agent as a between-subject factor. All participants were native Turkish speakers and the experiment was conducted in Turkish.

In the beginning of each experiment the first scene shown to the participants is a test screen used primarily for lens (focal) adjustments of the HMD device, which are presented in APPENDIX A. The test scene was also employed so that the participants could report if they had issues with myopia (near-sightedness) or hypermetropia (long-sightedness) with the HMD device worn. The test screen had every object in the VR environment visible at once for this purpose. No participants reported any issues in focusing the lenses or with their eye-sight.

3.2. Experiment Procedure

Each participant was greeted with a form of consent[3] and instruction sheets. The instruction sheets are presented in APPENDIX A. The sheets had screenshots of the experiment environment in different stages, and the related per-scene explanations of the experiment procedure in written text. This experiment flow is shown in Figure 3 in the next page.

The experiment sessions always started with the experimenter providing instructions to the participant. The instructions are further detailed in section 3.2.1. For each robot design and experiment condition the participants were asked forty questions in the VR environment. The first four questions were training questions and therefore are not recorded in the data files. The remaining thirty-six questions comprised the experiment session.

[3] The ethics committee approval was obtained from METU Ethics Committee.

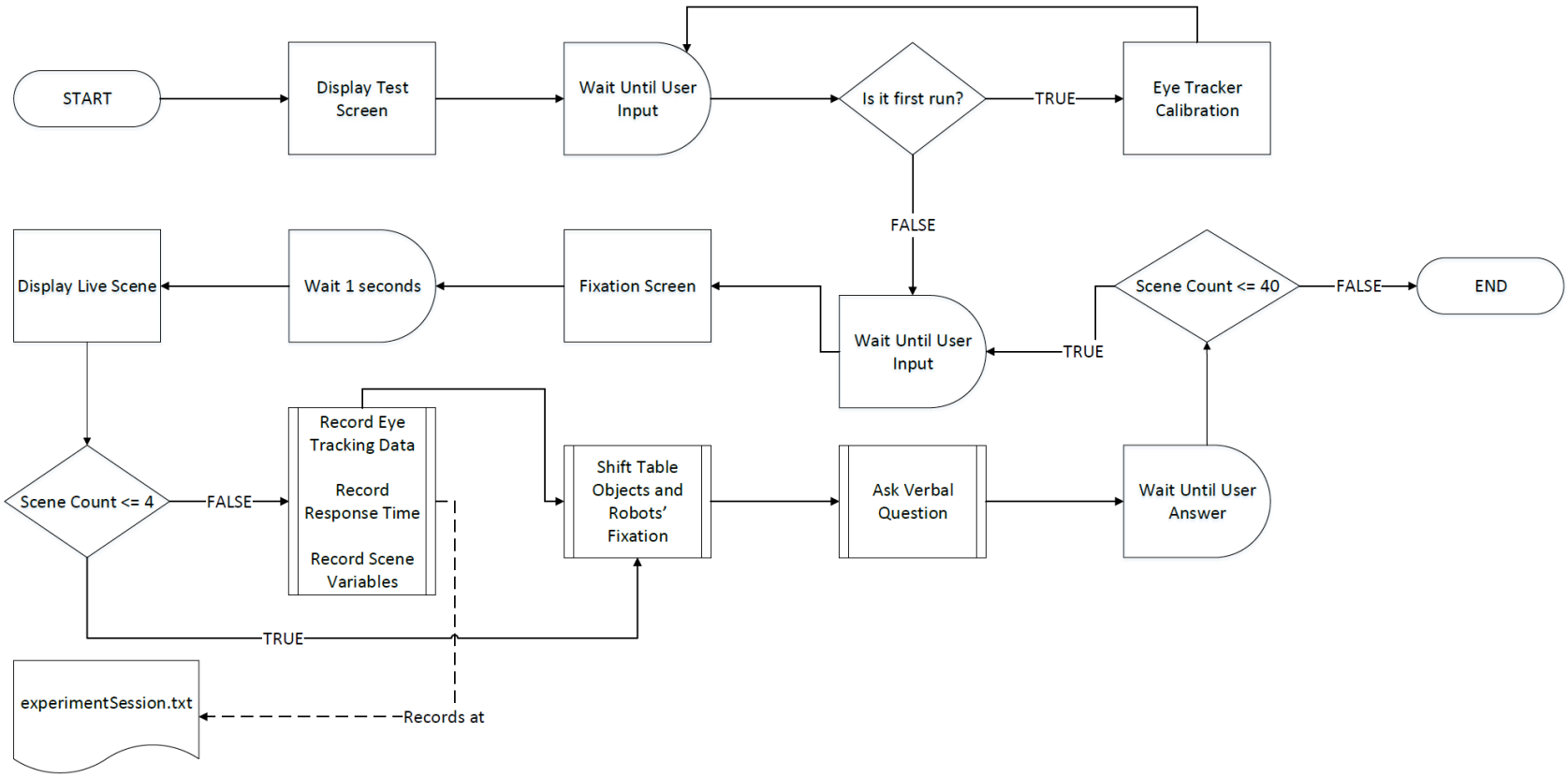


Figure 3 - Experiment execution flow

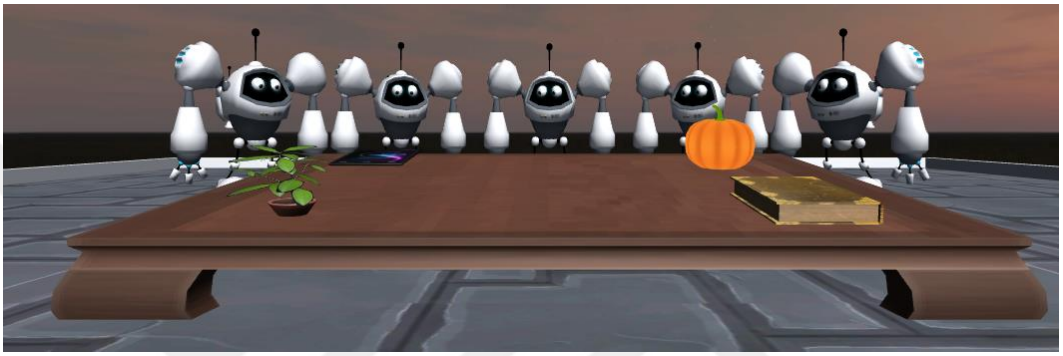
The experimental variables were all randomized in terms of the order of presentation, including the number of robot agents, the placement of the objects on the table. The questions asked in the communication environment were also randomized but only for the first experiment session, because the second session always had a single question replayed. Moreover, in experiment conditions where there were more than one robot agent in the environment, the gaze vectors of all robots were fixated at the same point in the VR environment; in other words, multiple robots always fixated on the same object altogether in collaboration.

The VR environment scenes are demonstrated in Figure 4 with screen images. In these examples, the differences in robot agents' gaze vectors, numbers of appearance, and the shifting of the objects' placement are shown. Also a video demonstration of the VR environment and the experiment flow is presented on YouTube (Yilmaz, 2018).

(a) No robots condition in the first session of the experiments.

(b) One robot condition, occurs in both sessions of the experiments. The robot agent is fixating its gaze on the right-bottom corner object (pumpkin).

(c) Three robots condition in both the first and the second session of the experiments. The robot agents are fixating their gaze on the top-right corner object (book).



(d) Five robots condition in both the first and the second session of the experiments. The robot agents are fixating their gaze on the top-left corner object (tablet).

Figure 4 - Virtual reality environment

3.2.1. Instructions and Training

The first part of the instructions given to the participant involved introducing the virtual reality HRI environment, with further explanations of the experiment flow and the names of the objects on the table were (pumpkin, tablet, plant, and book). This introduction was done partially with the HMD taken off, and partially in the "Display Test Screen", where a static scene was displayed to the participant in the VR environment. The test screen consisted of all objects in the VR environment that would appear in various stages in the flow of the experiment.

The objects used for the joint task were located in each corner of a table in the VR environment. These objects are visible in all screen images of the VR environment, such as Figure 2, Figure 3, as well as in Figure 5. The aforementioned information was given to the participant because geometric locations of objects were used in explicit, verbal referring expressions in the first experiment sessions and the participant needed to be on a contextual common ground with the robot agents in order to provide valid answers.

Figure 5 – Objects' placements on the table in the VR environment

The test scene also aimed to allow a static display for making physical adjustments of the HMD. As each participant's eye and head physiology is unique, therefore the HMD had to be adjusted for the screen to be in visual focus for each participant, and for the participants to feel comfortable with the HMD device on their heads. At the right hand side of the HMD device there is an area physically marked with a round cornered rectangle and a touch enabled button at its center. This trackpad area was the sole input accepted in our experiments, with touches in any location registering the same input command. Once the participant confirmed comfort and a focused, sharp image in the test scene in the HMD device, the experimenter instructed the participant to proceed with an input given by means of tapping to the right hand side of the HMD device.

(a) Device front: Eye tracking components (reflectors, cameras) and phone mount.

(b) Device back: Infrared lights (around each lens), proximity sensor (in the middle, and the lenses.

Figure 6 - SMI mobile eye tracking HMD (Hayden, 2016)

Following the first input from the user, the test scene disappeared and an intermediary resting screen was displayed. In the instruction sheets, the resting screen was described as an intermediary resting point for the participants, as there was no data recording in this scene. This resting scene appeared in between each new scene and allowed the participant to rest if needed.

After the participant progressed by means of giving another input on the trackpad of the HMD device, the first scene of all experiments was always the four-point eye tracker calibration. This calibration screen required participants to fixate their gaze on a red circle inside a white circle. The eye tracker calibration starts as soon as gaze fixation on the red circle is detected and is completed automatically. In the end of the eye tracker calibration the participants are directed back to the resting screen, and the experimenter notified them that the experiment was about to start with trials. The experimenter monitored the flow of this pre-experiment preparation of the participant through the participant's inputs to the HMD device visually and with verbal feedback from the participant.

3.2.2. The First Experiment Session

This first part of the experiment was where the participants replied questions in forms of *explicit referring expressions*. The referring expressions referred to the geometric locations of the objects on the table. The objects were rendered as the most salient objects in this virtual reality HRI environment as the communication between the human participant (verbal) and the robot agents (gaze vectors) always took place about the objects. The geometric locations of the objects on the table did not change throughout the experiment, however in each scene the objects had their places shifted in order to prevent the human participant from forming a predisposition towards the locations of the objects.

The utilization of geometric locations allowed the questions to be formed via verbal referring expressions as follows:

Table 1 - Verbal referring expressions utilizing geometric locations

Turkish question in the experiment	English Translation
Sol üstte ne var?	What is the object at left top?
Sol altta ne var?	What is the object at left bottom?
Sağ üstte ne var?	What is the object at right top?
Sağ altta ne var?	What is the object at right bottom?

The first experiment sessions had three experiment conditions investigated as the number of robot agents varied between 0, 1, 3, and 5. The first experiment condition aimed at investigating the differences between 0 robots (Figure 4a) versus having 1 (Figure 4b), 3 (Figure 4c), and 5 (Figure 4d) robots in the VR environment in two groups. The first experiment session was where the verbal referring expressions utilizing geometric locations did not a requirement of interacting with the robot agents for the success of the joint task. This condition was set up to investigate whether the robot agents were salient enough for the participant to commit to gaze interaction without a requirement. The first condition was followed by within group variances of 1, 3, and 5 in a single group. The setting in the current analysis was to investigate whether the further increasing number of robot agents increased the saliency of the robot agents or not.

Finally, only for the first experiment session; and only for the cases other than when 0 robots were present in the environment, the robot agents gazed at the participant at a random order (in approximately 25% of the scenes), instead of the four corners of the table. This randomly occurring condition was set in order to further differentiate the two experiment conditions (explicit questions and implicit deictic references).

3.2.3. The Second Experiment Session

The second part of the experiment was where participants replied questions that involved *implicit referring expressions* while the robot agents assisted the participants

with deictic gaze expressions. The participants always replied to the same question "burada ne var (what is the object here)?". The question was a deictic reference utilizing the word "burada (here)" as its pointer.

Geometric locations in the second experiment session were also always absolute and the robot agents always fixate their gaze on the objects on the table throughout the second session of the experiment. The robot's gaze was utilized as a deictic expression, assisting the implicit deictic pointer and directing the participant's attention towards one of the objects. The second session of the experiments was designed and conducted to investigate whether deictic gaze in virtual reality environments had a significant effect on the metrics of response times and accuracy of the participants in the VR environment. This investigation was conducted by means of analyzing the variation of the number of robot agents in the VR environment between 1 (Figure 4b), 3 (Figure 4c), and 5 (Figure 4d) robot agents in every new scene. The object locations on the table were being shifted similarly to that in the first session of the experiments also. The no robot or 0 robot condition was absent in the second session of the experiments as the robot agents were inherently forced to be interacted with the participant and, also, due to their role in assisting the deictic expressions with their gaze. The absent condition was, in short, due to the fact that the joint task could not succeed in the present experiment condition without gaze data from the robot agents.

3.3. Experiment Environment Technical Specification

We used Senso-Motoric Instruments (hereafter, SMI) mobile eye tracking HMD device. The HMD device is based on Samsung Gear VR. The HMD works with a modified Samsung Galaxy S7 (hereafter, S7) smartphone device of 2560 (in width) by 1440 (in height) resolution on a diamond pen-tile matrix AMOLED display. This S7 was modified in software by SMI, which is on Google's Android operating system version 6.0.1, for the device to provide the necessary backend and operating system level access for the SMI eye tracking APIs. We modified this device further for better cooling capabilities in order to ensure that our experiment would run for longer periods without the HMD device experiencing overheating and performance throttling issues. Both of these issues can also cause further issues in data integrity.

The HMD device with the S7 attached has a 96° field of view on the VR environment and is capable of 60Hz, binocular eye tracking for gaze direction, and inter pupil distance measurement with eye tracking accuracy of 0.5. Our experiment environment was designed using SMI SDK for the device using the Unity game engine version 5.x (licensed for non-business, academic use), and Microsoft Visual Studio 2015 Community. Each experiment session is built as a separate application to a total of four applications as; humanoid robot – explicit questions, humanoid robot – implicit questions, non-humanoid robot– explicit questions, and non-humanoid robot– implicit questions. The four applications stored the recorded experiment data in local flash storage of the S7. The experiment data was kept in distinct locations in the file system for the four applications.

Due to the between-subjects design, each participant attended only one of the distinct robot designs, and had two files, in that there are one file created per experiment session. The experiment applications stored an array of the data structure that store the variables recorded. The variables of the data structure are further

described in the next pages, with Table 2. The vocal answers the participant gave were recorded by the experimenter via an audio recording device and the files were similarly stored in a two files per participant in one file per experiment session manner.

Table 2 - Raw data structure as stored by the experiment applications

Variable name	Variable definition
System time	This value was used as a validation for the participant response time variable per each response sample the participant answered.
Response sample	A counter for the current participant response; 36 for each experiment session and/or file.
Gaze samples	For each response sample, a gaze sample is recorded approximately every 16.6 milliseconds or at a rate of 60 samples per second.
Gazed object (Participant)	For each gaze sample, the object in the virtual reality environment on which the participant gaze fixates is recorded.
Gazed object (Robot)	Only active for the first session of the experiments; where for each response sample, the gaze group of the robot agents randomly shifted between the objects on the table and the participant.
Active Robots	For each response sample, the number of robot agents active in the environment is recorded.
Question Played	For each response sample, the focused object of the robot agent determines which geometric location was referred to.
Response Time	For each gaze sample recorded, the response time variable stored at what time within the current response sample does the participant gaze at a certain object. This variable is used to also store what the total time spent in each response sample is.

3.4. Analysis Procedure

The data files stored on the local storage of the S7 device were retrieved at the end of each experiment day and stored along with each participant's respective audio files. In the analysis procedure, the audio files were listened to by the experimenter and transcribed into a spreadsheet file with the structure below, in Table 3.

Table 3 - Participant responses sheet structure

		Experiment Session One	Experiment Session Two
Participant ID Number	For each response sample	Explicit question by the robots	x
		Participant Response	Participant Response

After the transcribing process the spreadsheet was expanded with the information obtained through the raw data files recorded in the structure given in Table 2. As shown in Table 3, the second session only had the answers recorded as the same question is asked repeatedly in the communication environment due to the experiment setting. As a result of the present setting, the correct answers for the second session were checked by means of the composition between the numbers of robot agents, with "Active Robots" variable in the data structure, and the geometric location the robot agents fixated their gaze, with "Question Played" variable in the data structure being reused to store these.

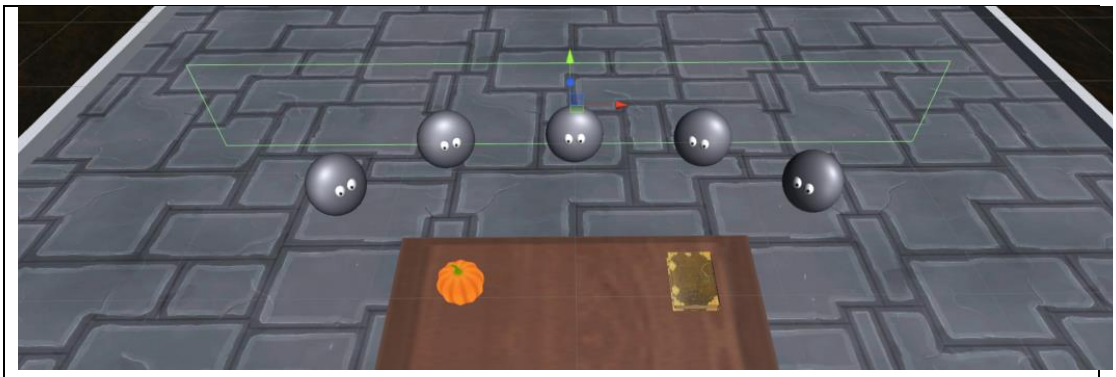
At this stage in the analysis, the answers were checked for whether they were correct or not for both sessions of the experiment. In the first experiment session, correct / incorrect check on both the experiment sessions was done through a parser using the answer key in Table 4. The explanations of the parsing application are given in APPENDIX B.

Table 4 - The answer key for the information exchange joint task

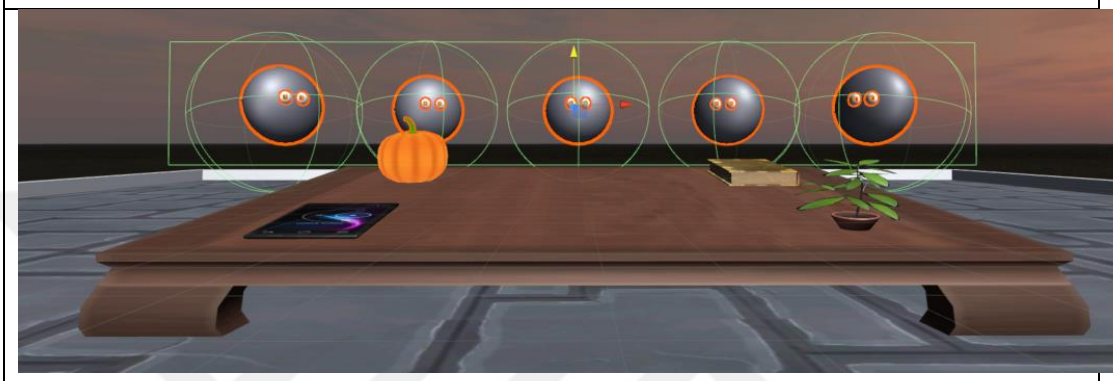
Number of Robot Agents	Left Top	Left Bottom	Right Top	Right Bottom
Robot 1	Plant	Book	Tablet	Pumpkin
Robot 3	Book	Pumpkin	Plant	Tablet
Robot 5	Pumpkin	Tablet	Book	Plant
Robot 0	Tablet	Plant	Pumpkin	Book

Finally, the raw data from the experiment recordings were processed in order to obtain the following list of parameters:

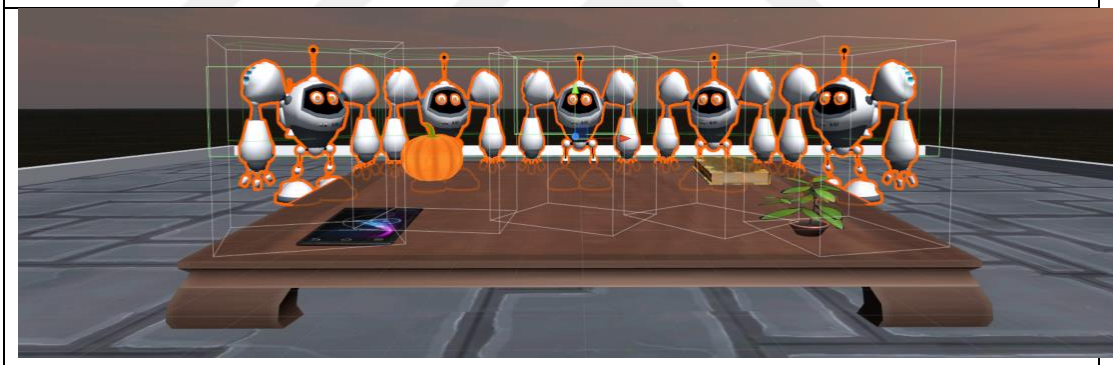
- The number of robot agents present for each response sample
- Participant's gaze distribution for each response sample
    - The gaze distributions were recorded by means of collision detection between a gaze cursor, which tracked participants' gaze fixation and the objects in the VR environment.
    - Total gaze distribution on robot agents
        - Gaze distribution on each robot agent, as well as the total of all gaze fixation in percentages for center, center left, center right, left-most, and right-most robot agents.
    - Total gaze distribution on the robot agents' surroundings
        - As shown in Figure 7a, a rectangle shaped object is placed behind the robot agents in the VR environment. This object is not rendered visible in the runtime in any of the experiment applications. The occurrences of the participants' gaze fixation colliding with the rectangle shaped object were marked as errors in eye tracking and were not evaluated as gaze interaction with the robot agents in statistical analysis. The first purpose of the object was to catch any gaze fixation of the participant that may have been on a robot agent, however deviated too much outside the collision detection area of the robot agents. The collision detection areas for robot agents as well as the rectangle shaped object behind the robot agents are shown from the participant's perspective in the VR environment in Figure 7b-c. The second purpose of this area was to detect whether participants gaze at the location in the VR environment where the robot agents used to be in the first session of the experiments when the 0 robots condition is in effect.
    - Participants' total gaze distribution on the objects in the VR environment
        - Table, book, pumpkin, tablet, plant, and others
            - The "others" in the set includes everything other than what is explicitly named, such as the robot agents, the screen behind the robot agents, the table, and the objects on top of the table.
- Participants' response time for each gaze sample.
    - Because there are many gaze samples for each response sample, the parsing application is used to select the last gaze sample for all response samples to determine the final response time.



(a) Green rectangle object, exemplified behind the non-humanoid robot agents.



(b) Spherical collision detection areas; non-humanoid design.



(c) Rectangular prism (inner-green boxes around the robot agents) detection areas; humanoid design.

Figure 7 - Eye tracking collision areas in the virtual reality environment.

The data refinement procedure was repeated for both the humanoid and non-humanoid robot agents design and two spreadsheets were obtained. The spreadsheets are considered to be processed, gold data and are used throughout the statistical analysis process.

In the statistical analysis process, the data was filtered from the spreadsheets in order to obtain gaze distribution and response time data for both robot designs and both experiment sessions, as well as accuracy in resolution of deictic expressions in the second session of the experiments. Firstly, the filtered data were analyzed for outliers using z-scores of ±2 as the acceptable margins. Secondly, the data were analyzed for the effect sizes of each dependent variable of accuracy, gaze distribution on robot agents, and response time in a repeated measures ANOVA. Finally, each variable was analyzed in a paired sample t-test in order to investigate the mean differences for each dependent variable of the number of robot agents, explicit – verbal

referring expressions in the first experiment sessions, and implicit – deictic gaze assisted referring expressions in the second experiment sessions.

The three analyses were done using IBM SPSS Statistics Data Editor version 24, with further information in regards to analysis described below, and discussed in the next chapter.

- General linear model with repeated measures analysis used for effect sizes analysis.
    - Within subject factor: NumOfAgents
        - The NumOfAgents factor is a 3 level factor in the second session of the experiments; while for the first session, it is tested for both 4 levels in investigating the within subject effect of the number of robots variable; with 0 robot agents versus the rest) and 3 levels (for investigating changes in the effect size between 1, 3, and 5 robot agents).
    - Between subjects factor: Group
        - The two robot designs were used to divide the groups of 42 experiment participants into their respective groups.
- Paired samples t-test
    - Pairs of samples were set using the robot agent numbers as variables:
        - 1 robot – 3 robots
        - 3 robots – 5 robots
        - 1 robot – 5 robots
        - 0 robots – 1 robot (first experiment session only)
        - 0 robots – 3 robots (first experiment session only)
        - 0 robots – 5 robots (first experiment session only)
- Descriptive statistics (Z-scores)
    - For each variable of accuracy, gaze distribution on robot agents, and response time the standardized scores are used in order to eliminate outliers.

CHAPTER 4

RESULTS

In this chapter the analysis results for our experiment conditions are presented. Firstly, the accuracy and response time of the participants in their resolution of the deixis are reported in sections 4.1 and 4.3, as the participants response time in answering explicit questions is reported in 4.2. Secondly, the patterns of gaze interaction in the first and the second experiment sessions are reported in sections 4.4 and 4.5. Each of the sessions also have the between subjects effects of robot agent designs reported.

4.1. Resolution of Deictic Expressions – Accuracy

In the first experiment session, the communication situation was formed by utilizing referring expressions with geometric locations of objects on the table. This joint task was set up through questions that explicitly refer to object locations. As a result it does not requisite a resolution process in deixis for the participants, because there are no deictic expressions. The accuracy rates in the first session, as shown in Table 5, revealed that almost all the participants were successful in answering these explicit questions.

Table 5 – Mean accuracy ratios for varying numbers of robot agents and in between robot designs in the first experiment sessions

Number of Robot Agents	Humanoid Robot Agents	Non-humanoid Robot Agents
1	100.0% (SD = 0.00%)	97.89% (SD = 5.88%)
3	99.47% (SD = 2.42%)	98.90% (SD = 3.85%)
5	99.46% (SD = 2.72%)	98.95% (SD = 3.71%)
No robots (0)	100.0% (SD = 0.00%)	98.84% (SD = 2.62%)

The results in Table 5 are those without the outliers eliminated. This is our approach in reporting the results because all the occurrences of as many as one incorrect answer was marked as outliers by the descriptive values of z-scores ±2. As a result, once the outliers were eliminated, all scores were 100% (SD = 0.00%) accurate for all participants ($N_{humanoid}$ = 19, $N_{non\text{-}humanoid}$ = 15). The situation persisted in both of the robot-agent design groups (humanoid, non-humanoid).

The resolution accuracy of the participants in the second session for each robot design were investigated in the scope of the first research question. The results in Table 6 present the second session data after the outliers were eliminated. 41 participants

remained in the dataset when one outlier from the humanoid robot design was removed ($N_{humanoid}$ = 20, $N_{non\text{-}humanoid}$ = 21).

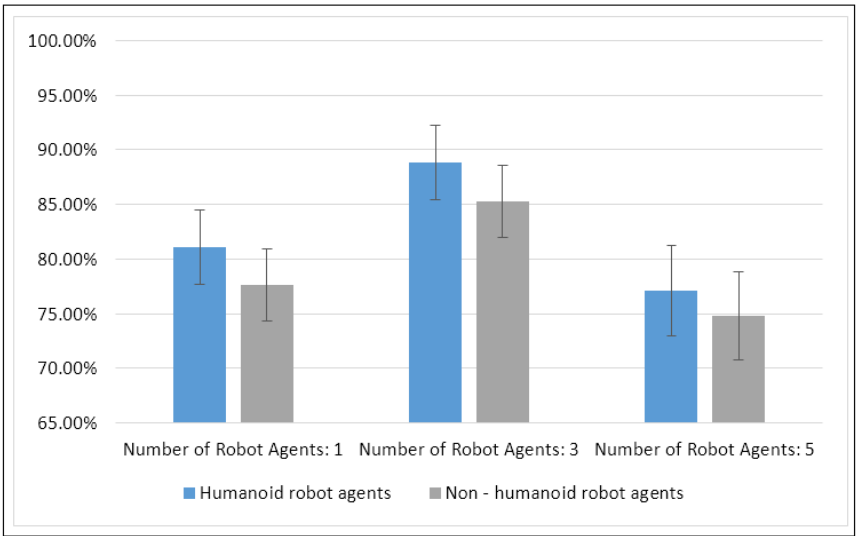
Table 6 – Mean accuracy ratios for varying number of robot agents and in between robot designs in the second experiment sessions

Number of Robot Agents	Humanoid Robot Agents	Non-humanoid Robot Agents
1	81.19% (SD = 14.98%)	77.64% (SD = 15.15%)
3	88.85% (SD = 11.72%)	85.39% (SD = 17.68%)
5	77.11% (SD = 17.75%)	74.80% (SD = 18.54%)

The analysis of the dataset was conducted on a 3x2 model with three robot agent within-subject conditions and the two robot agent designs as between-subjects. A Mauchley's test indicated that the sphericity was not violated for the main effect of the varying number of robot agents χ^2 (2) = 0.195, $p > .05$. Therefore, the results of the sphericity assumed values reported. The results indicated a significant effect [F (2, 78) = 7.880, $p < .05$] between the number of agents in the VR environment for the participants' deictic resolution accuracy. However, there was no significant effect [F (2, 78) = 0.029, $p > .05$] for the interaction effect introduced by the design of the robot agents.

Follow-up pairwise comparisons showed a significant effect in mean differences between 1 agent and 3 agents ($M_1 - M_3 = \pm 0.077$, SE = 0.029, $p < .05$), as well as between 3 and 5 agents ($M_3 - M_5 = \pm 0.112$, SE = 0.028, $p < .05$). However, between 1 agent and 5 agents ($M_1 - M_5 = \pm 0.035$, SE = 0.029, $p > .05$) did not show a significant effect. The results showed that as the numbers of robot agents in the VR HRI environment varied between 1, 3, and 5, the participants' accuracy in resolving gaze cues as deictic references were affected significantly, while this effect did not persist between the variation in numbers between 3 and 5 robot agents.

Following these results a paired sample t-test was conducted in order to further investigate the accuracy rates between the respective pairs of 1, 3, and 5 agents. The results (presented in Figure 8 in the next page) showed that there was a significant difference in the mean accuracy rates of the participants when the number of robot agents increased to 3 agents (M = 0.870, SD = 0.149) from 1 agent (M = 0.793, SD = 0.149) in the second experiment session (non-humanoid robot agents); t (40)= -2.684, $p < .05$; as well as when the increase is to 5 agents (M = 0.759, SD = 0.179) from 3 agents (M = 0.870, SD = 0.149); t (40)= 4.065, $p < .05$. These significant differences did not persist between 1 agent (M = 0.793, SD = 0.149) to 5 agents (M = 0.759, SD = 0.179); t (40) = 1.181, $p > .05$.



1, 3 Robot Agents:	t (40) = -2.684, p < .05 *
1, 5 Robot Agents:	t (40) = 1.181, p > .05
3, 5 Robot Agents:	t (40) = 4.065, p < .05 *

* significant differences

Figure 8 – Mean differences in participant accuracy of deictic expression resolution between the varying numbers of robot agents and the two robot agent designs

4.2. Response Times in the First Experiment Session (Explicit Questions)

In the first session of the experiment where the participants replied to questions utilizing explicit geometric locations in the VR environment. This experiment condition did not employ the cognitive process of resolving referring expressions in discourse. However, by investigating our virtual reality HRI environment in the scope of our first research question, the analyses provided an insight to how the varying number of robot agents in the joint task affect the response times of the participants. The analysis was based on a 4x2 model, with the variation of the numbers of robot agents at four instances as within-subject conditions and the two robot agent designs as between-subjects. The mean response time results are presented in Table 7.

Table 7 - Mean participant response times for the two robot agent designs and their varying numbers (all values reported in seconds and milliseconds) in the first experiment sessions

Number of Robot Agents	Humanoid Robot Agents	Non-humanoid Robot Agents
1	2.19 (SD = .432)	2.11 (SD = .454)
3	2.32 (SD = .536)	2.11 (SD = .442)
5	2.22 (SD = .429)	2.10 (SD = .451)
No robots (0)	2.20 (SD = .432)	2.18 (SD = .447)

Using the z-scores of ±2 being used as acceptable margins, three participants were eliminated; two from the humanoid robot agents group, while the third from non-humanoid robot agents ($N_{humanoid}$ = 19, $N_{non\text{-}humanoid}$ = 20).

A Mauchley's test indicated that the sphericity had been violated for the main effect of the number of robot agents $\chi 2$ (2) = 30.493, $p < .05$. Therefore, the results of the Greenhouse-Geisser correction are reported. There was no significant effect [F (1.923, 71.145) = 1.061, $p > .05$] between the varying number of robot agents and the participants' response times to the questions they were requested to answer. The interaction effect analysis of the distinct robot agent design and the number of robot agents resulted in [F (1.923, 71.145) = 1.880, $p > .05$] that there was no significant interaction effect.

4.3. Response Times in the Second Experiment Session (Implicit Deictic References)

The present condition in the second session was analyzed in a 3x2 model, with a variation of three discrete numbers of robot agents and the two robot agent designs. The mean response time results are presented in Table 8.

Table 8 - Mean participant response times in regards to the two robot agent designs and their numbers (all values reported in seconds.milliseconds) in implicit expressions

Number of Robot Agents	Humanoid Robot Agents	Non-humanoid Robot Agents
1	2.19 (SD = .407)	1.96 (SD = .536)
3	2.40 (SD = .560)	2.25 (SD = .576)
5	2.61 (SD = .790)	2.26 (SD = .518)

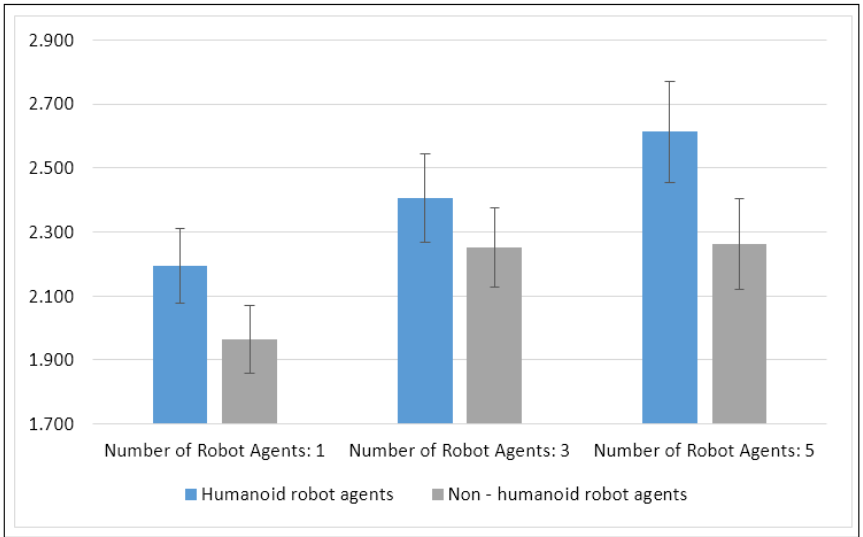
Using the descriptive statistics obtained in z-scores of the same acceptable margins as the other analyses of ±2; four participants from the humanoid robot agents' design group were eliminated in total, resulting in 38 participants remaining in the dataset ($N_{humanoid} = 17$, $N_{non\text{-}humanoid} = 21$).

A Mauchley's test indicated that the sphericity had been violated for the main effect of the number of robot agents $\chi 2$ (2) = 4.676, $p > .05$. Therefore, the results of the sphericity assumed values are reported. The results showed that there was a significant effect [F (2, 72) = 14.380, $p < .05$] between the number of agents in the VR environment in the participants' response times. However, the interaction effects of the number of robot agents together with the two groups of agent types [F (2, 72) = 1.060, $p > .05$] did not have a significant effect on the response times of the participants.

In the follow-up pairwise comparisons the analyses showed a significant effect in mean differences between 1 agent and 3 ($M_1 - M_3 = \pm 0.250$, SE = 0.055, $p < .05$), as well as between 1 and 5 robot agents ($M_1 - M_5 = \pm 0.359$, SE = 0.073, $p < .05$). However, the mean differences for the number of agents between 3 and 5 agents ($M_3 - M_5 = \pm 0.109$, SE = 0.076, $p > .05$) were not significantly different.

The results of the paired samples t-test are presented in Figure 9. These results showed that there was a significant difference in the response times of the participants with 1 agent (M = 2.06, SD = 0.491) and 3 agents (M = 2.321, SD = 0.567); t (37) = -4.647, $p < .05$; and likewise with 1 agent (M = 2.06, SD = 0.491) and 5 agents (M =

2.419, SD = 0.668); t (37) = -4.896, p < .05. However, there was no significant difference between the gaze interaction made with 3 agents (M = 2.321, SD = 0.567) and 5 agents (M = 2.419, SD = 0.668); t (37) = -1.289, p > .05.



1, 3 Robot Agents:	t (37) = -4.647, p < .05 *
1, 5 Robot Agents:	t (37) = -4.896, p < .05 *
3, 5 Robot Agents:	t (37) = -1.289, p > .05

* significant differences

Figure 9 - Response times for the second experiment sessions with the three as the variation of robot numbers of agents and the two designs.

4.4. Gaze Interaction in the First Experiment Session (Explicit Questions)

The model of analysis for the present condition was a 3x2 model, with the variance of robot agents (1, 3, and 5 robot agents) and the two robot agent designs (humanoid, non-humanoid). The amounts of gaze interaction participants committed in to in the entire session are presented in Table 9Table 11. The amounts of gaze interaction participants committed in to with individual robots are presented in Table 10.

The human participant was replayed questions in form of referring expressions utilizing the explicit geometric locations of objects. By means of investigating the inherent saliency of the robot agents' without a joint task requirement for interaction, the analysis in this section was conducted to explore our scope with the second research question (how is the human participant's gaze distributed on the robot agents).

Table 9 - Ratios in which the participants interacted with the robot agents in the two designs in the first experiment session.

Number of Robot Agents	Humanoid Robot Agents	Non-Humanoid Robot Agents
1	14.00% (SD = 13.48%)	25.31% (SD = 12.55%)
3	17.22% (SD = 15.52%)	30.57% (SD = 14.54%)

5	15.97% (SD = 15.32%)	28.59% (SD = 15.91%)

Table 10 - Gaze distribution between the varying number of robot agents, their locations, and the two robot agent designs in the first experiment session

	Robot Location	Humanoid Robot Agent	Non-humanoid Robot Agent
1 Robot Agent	Center	14.00% (SD = 13.48%)	26.60% (SD = 15.52%)
3 Robot Agents	Left	3.67% (SD = 4.62%)	5.41% (SD = 7.08%)
	Center	11.72% (SD = 12.29%)	23.60% (SD = 15.39%)
	Right	1.85% (SD = 2.50%)	2.64% (SD = 2.65%)
5 Robot Agents	Left-Most	0.64% (SD = 1.42%)	0.14% (SD = 0.35%)
	Left	1.79% (SD = 1.86%)	2.49% (SD = 2.04%)
	Center	10.17% (SD = 11.17%)	22.22% (SD = 14.36%)
	Right	1.60% (SD = 2.27%)	2.58% (SD = 5.49%)
	Right-Most	1.64% (SD = 3.22%)	0.53% (SD = 0.97%)

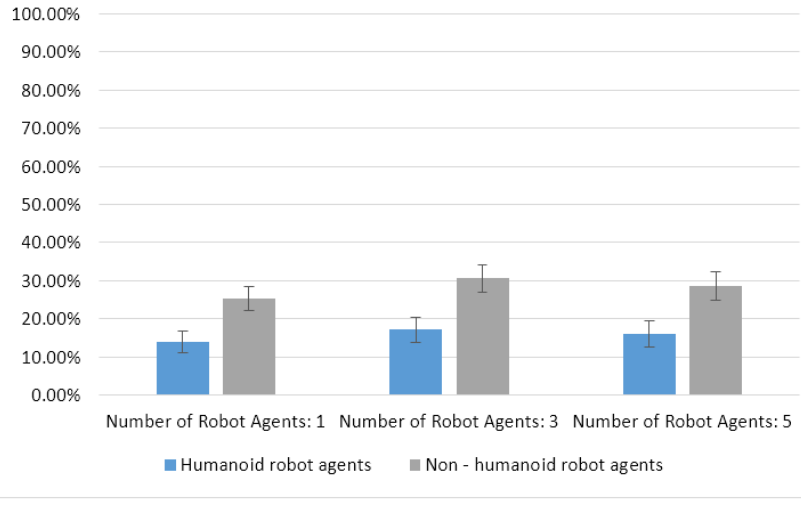
After the outliers were eliminated using z-scores ±2 as the acceptable margins, 38 participants remained from the 42. The four participant data removed were all from the second participant group of non-humanoid robots ($N_{humanoid}$ = 21, $N_{non\text{-}humanoid}$ = 17).

A Mauchley's test indicated that the sphericity had not been violated for the main effect in the number of robot agents $\chi 2$ (2) = 0.364, p > .05. Therefore, the results of the sphericity assumed values reported. A significant effect [F (2, 72) = 3.708, p < .05] was obtained between the number of robot agents and the participants' gaze interaction with the robot agents. In the interaction effects between the number of agents with the two groups of agent types, the results showed that [F (2, 72) = .218, p > .05] there was not a significant effect.

The pairwise analysis showed a significant effect of the gaze interaction participants committed to between 1 and 3 robot agents ($M_1 - M_3 = \pm 0.042$, SE = 0.016, p < .05). In contrast, the mean differences between 1 and 5 agents ($M_1 - M_5 = \pm 0.026$, SE = 0.015, p > .05), as well as between 3 and 5 agents ($M_3 - M_5 = \pm 0.16$, SE = 0.016, p > .05) were not significant effects. Following these results a paired samples t-test was conducted in order to see mean differences between the respective pairs of 0, 1, 3 and 5 agents.

The paired samples t-test are presented in Figure 10. The results showed that the no agents (0 agent) condition (M = 0.002, SD = 0.001) and the three robot agent conditions: 1 agent (M = 0.190, SD = 0.141), 3 agents (M = 0.231, SD = 0.163), and 5 agents (M = 0.216, SD = 0.166) revealed significant differences; 0 agent – 1 agent: t (37) = -8.302, p < .05; 0 agent – 3 agents: t (37) = -8.731, p < .05; and 0 agent – 5 agents: t (37) = -7.984, p < .05. The results showed that having robot agents in the environment significantly affects the gaze interaction the participants make with the robot agents.

Moreover, the t-tests also showed that there was a significant difference in the amount of gaze interaction the participants made with 1 agent (M = 0.190, SD = 0.141) and 3 agents (M = 0.231, SD = 0.163) in the VR environment; t (37) = -2.550, p < .05. However, there were no significant differences between the gaze interaction made with 1 agent (M = 0.190, SD = 0.141) and 5 agents (M = 0.216, SD = 0.166), t (37) = -1.726, p > .05; as well as between 3 agent (M = 0.231, SD = 0.163) and 5 agents (M = 0.216, SD = 0.166); t (37) = 1.028, p > .05.



0, 1 Robot Agent:	t (37) = -8.302, p < .05 *
0, 3 Robot Agents:	t (37) = -8.731, p < .05 *
0, 5 Robot Agents:	t (37) = -7.984, p < .05 *
1, 3 Robot Agents:	t (37) = -2.550, p < .05 *
1, 5 Robot Agents:	t (37) = -1.726, p > .05
3, 5 Robot Agents:	t (37) = 1.028, p > .05
	* significant differences

Figure 10 - Participants' gaze interaction patterns with the robot agents in explicit referring expressions in varying robot agent counts and two robot designs

4.5. Gaze Interaction in the Second Experiment Session (Implicit Deictic References)

The gaze interaction between the participant and the robot agents was an inherent requirement in the joint task of the second experiment session, because the participant requires assistance from gaze vectors in resolving deictic references. The robot agents in the second session appear in numbers varying between 1, 3, and 5; with the two robot agent designs as a 3x2 model. The amounts of gaze interaction participants

committed in to in the entire session are presented in Table 11Table 11. The amounts of gaze interaction participants committed in to with individual robots are presented in Table 12.

The outliers were eliminated using z-scores ±2 as the acceptable margins. 39 participants remained from the total of 42; one participant from the first group of humanoid robot design and two from the latter group of non-humanoid robot design were removed ($N_{humanoid} = 20$, $N_{non\text{-}humanoid} = 19$).

Table 11 - Ratios in which the participants interacted with the robot agents in the two designs in the second experiment session.

Number of Robot Agents	Humanoid Robot Agents	Non- Humanoid Robot Agents
1	67.17% (SD = 9.16%)	61.92% (SD = 11.88%)
3	76.78% (SD = 9.06%)	79.57% (SD = 8.79%)
5	77.33% (SD = 9.70%)	78.48% (SD = 7.84%)

Table 12 - Gaze distribution between the varying number of robot agents, their locations, and the two robot agent designs in the second experiment session

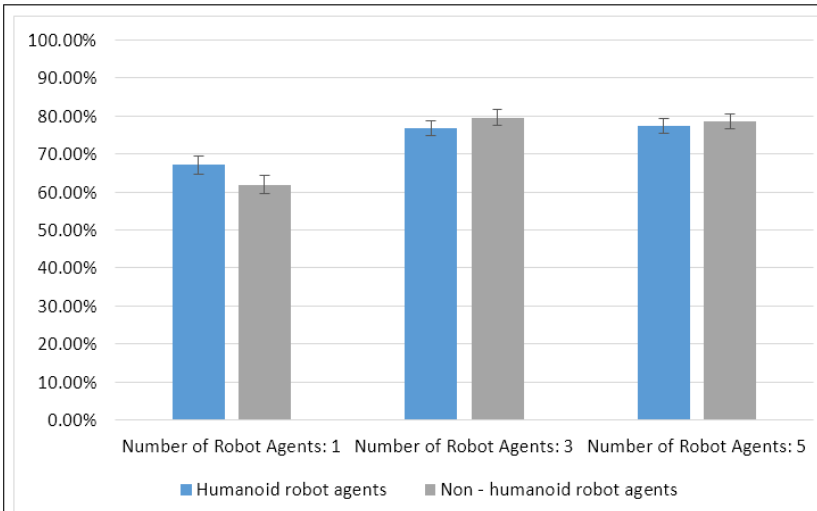
	Robot Location	Humanoid Robot Agent	Non-humanoid Robot Agent
1 Robot Agent	Center	65.23% (SD = 14.47%)	61.47% (SD = 13.95%)
3 Robot Agents	Left	20.08% (SD = 7.79%)	21.06% (SD = 10.34%)
	Center	32.46% (SD = 7.50%)	39.44% (SD = 12.40%)
	Right	21.31% (SD = 7.77%)	16.94% (SD = 9.12%)
5 Robot Agents	Left-Most	8.45% (SD = 6.59%)	5.97% (SD = 7.38%)
	Left	12.93% (SD = 6.23%)	11.34% (SD = 6.27%)
	Center	30.39% (SD = 9.89%)	39.12% (SD = 15.37%)
	Right	12.92% (SD = 6.04%)	17.14% (SD = 6.71%)
	Right-Most	10.18% (SD = 6.44%)	7.20% (SD = 5.77%)

A Mauchley's test indicated that the sphericity had been violated for the main effect of the varying number of robot agents χ^2 (2) = 12.889, $p < .05$. Therefore, the results of the Greenhouse-Geisser correction are reported. A significant effect [F (2, 74) = 59.187, $p < .05$] was formed between the varying number of robot agents in the

participants' gaze interaction with the agents. The results of our interaction effect between the number of robot agents and the two agent designs [F (2, 74) = 4.410, p < .05] indicated that the robot agent design in the VR environment had a significant effect on the gaze interaction patterns of the participants.

Follow-up pairwise comparisons showed a significant effect in mean differences between 1 agent and 3 agents ($M_1 - M_3$ = ±0.136, SE = 0.014, p < .05), as well as ($M_1 - M_5$ = ±0.134, SE = 0.017, p < .05) between 1 and 5 agents. However, the mean differences for between 3 and 5 agents ($M_3 - M_5$ = ±0.003, SE = 0.010, p > .05) were not significant effects.

A paired samples t-test was conducted in order to see whether the mean values for gaze interaction were significantly different or not between the respective pairs of 1, 3 and 5 agents. The results showed that there were significant differences in the amounts of gaze interaction the participants committed with 1 agent (M = 0.646, SD = 0.107) and 3 agent (M = 0.781, SD = 0.089) in the VR environment; t (38) = -8.769, p < .05; and likewise with the 1 agent (M = 0.646, SD = 0.107) and 5 agent in (M = 0.778, SD = 0.087); t (38) = -7.381, p < .05. However, there was no significant difference between the amounts of gaze interaction made with 3 agents (M = 0.781, SD = 0.089) and 5 agents (M = 0.778, SD = 0.087); t (38) = .241, p > .05. The differences in the amounts of gaze interaction and the t-test results are presented in Figure 11.



1, 3 Robot Agents:	t (38) = -8.769, p < .05 *
1, 5 Robot Agents:	t (38) = -7.381, p < .05 *
3, 5 Robot Agents:	t (38) = 0.241, p > .05

* significant differences

Figure 11 - Participants' gaze interaction patterns with the robot agents in resolving implicit referring expressions with the assistance of gaze cues in varying robot agent counts and two robot designs

4.6. Summary of Results

- Our investigation of deixis resolution time and accuracy showed that the participants were more accurate in their answers until a threshold level.

- The participants' response time as the number of robot agents increased were affected significantly until a threshold level observed in their accuracy.
- The participants committed to more gaze interaction with the robot agents as the number of robot agents increased until a threshold of five robot agents. The threshold value was observed similarly in all deixis resolution conditions.
- The robot agents were significantly salient in the VR environment that they were interacted with even in experiment conditions that did not require gaze interaction with the robot agents.
- The participants' accuracy and response time in answering explicit referring expressions were not affected by the varying number of robot agents in any condition.

CHAPTER 5

DISCUSSION AND CONCLUSION

5.1. Referring Expression Resolution in VR HRI

Our investigation of deictic expression resolution accuracy and time showed that the number of robot agents in the environment and partially the design of the agents both have influence on the accuracy and time of the participants responses. The deictic references were studied in the second experiment sessions where the robot agents provided gaze cues for the resolution of the deixis. The accuracy in which the participants responded explicit questions in the first experiment sessions shows that the VR HRI environment is compatible with the communication task. The first session's results also provided a base-line for response time metrics. On the other hand, the implicit questions in the second experiment sessions allowed us to investigate the likely effects of varying numbers of robot agents in the environment.

In the first experiment session, the number of robot agents in the environment did not have a significant effect on the response times of the participants. This was an expected result as the robot agents were not directly involved in the joint task in the first experiment session.

In the second experiment session, the number of the robot agents in the environment did have a significant effect on the accuracy that the participants were able to solve deictic references with robot agents' gaze cues. However, the effect was not a linear increase or decrease and the significant difference was particularly observed between 1 robot agent in the environment and 3 robot agents. This showed that the participants were better in resolving deictic references with the increased amount of gaze cues up to a certain point. A reverse effect was observed when the number of robot agents increased from 3 agents to 5 agents; the accuracy in deictic expression resolution decreased. This situation likely points at a threshold in available amount of data in the communication environment from the perspective of cognitive workload (Cain, 2007; Casali & Wierwille, 1983; Gopher & Donchin, 1986). Despite that Cain claims that there is not a sole consensus on the definition among researchers, the present results are likely related to the phenomenon.

Another metric of the second experiment session was the response time of the participants in answering the questions about object locations. The results showed that the number of robot agents did have a significant effect on the response times of the participants. The pairwise comparisons also showed that the robot agent number increasing from 1 agent to 3 agents resulted in significantly higher response times. Again, however, the effect was reversed when the number of robot agents went from 3 robot agents to 5 robot agents.

To summarize, the results from both accuracy and response time in the second experiment session show that increasing the amount of gaze cues assists up until a threshold. After the threshold point in the amount of gaze data in the environment this effect is reversed. This conveys that even though the gaze data increases linearly from 1 to 3 and from 3 to 5 robot agents, the increase of data does not translate linearly into accuracy and response time in deictic expression resolution. These findings from both perspectives can also be credited to the limitations discussed in detail in section 5.4, as well as to the suggestions towards a minimalist design in communication environments by Devault, et al., (2005). Moreover, the threshold appearing in the present findings might point at a likely issue with cognitive workload in regards to how many gaze cues can be utilized positively and how more gaze data than the amount that can be utilized cognitively cause an overload in cognitive processing.

5.2. Human Gaze Interaction in VR HRI

Our scope in investigating gaze interaction included a sole metric of the number of robot agents in the VR HRI environment. The experiments exploring this scope were conducted in both sessions. In the first session with the explicit questions, the numbers of robot agents varying between 0, 1, 3, and 5 were included in the experiment setting. These conditions were grouped in to two for the statistical analysis; in that no robots versus robots, and the varying number of robot agents between each other are the two within subject analyses.

The analyses showed that the existence of the robot agents in the VR HRI environment has a significant effect on gaze interaction patterns of our participant. The human participants were not obligated to commit to gaze interaction with the robot agents in the first session of the experiments (due to the questions requiring no gaze cues for answering). The findings show that the existence of robot agents convey high saliency in the environment enough to draw visual attention. Morales, et al., (1998, 2000) claimed that gaze as a tool for facilitating visual attention is also a valid influence in infants' language acquisition process. Therefore, the significant influence of the amount of gaze data in implicit references (in terms of accuracy, gaze interaction, and deixis resolution time) may be credited to the likely higher success of correctly indicating visual attention (Ruhland, et al., 2015) between different objects with similar saliency in the VR environment.

A similar effect was observed in that the participants' gaze interaction with the robot agents increased only when the number of agents went from 1 to 3 robot agents. However, there was not a significant increase nor decrease when the number of agents went from 3 to 5 robot agents. Therefore, a similar threshold point in the number of robot agents is observed too in the gaze interaction patterns of human participants with the HRI environment.

In the second experiment session, the number of the robot agents directly correlated with the amount of participants' gaze interaction in the communication environment. Our experiment setting always had the robots collaborate in providing the gaze cues, as the robot agents altogether fixated only one of the objects at a time. Therefore, there is not a confusion nor a competition condition in the experiment setting. The results showed that a similar threshold point persisted in the gaze interaction patterns as in explicit and deictic expression communication conditions.

These results showed that the gaze interaction between human participant and the robot agents increased significantly when the number of agents went from 1 robot agent to both 3 and 5 robot agents. However, when the number of agents went from 3 robot agents to 5 robot agents, this did not cause a significant change in the amount of gaze interaction human participants committed.

5.3. Robot Agent Design in VR HRI

As studied by Bartneck, et al., (2018) the robot agents' design were kept as neutral as possible with neutral gray as the color being used for the non-humanoid, and a mixture of both white and black being used for the humanoid design. Devault, et al., (2005) stated that the interlocutors of a communication environment must share a minimalistic, static knowledge-base. In our experiment environment the number of salient objects were limited to only the gaze vectors of the robot agents and the objects on the table. This experiment setting was similar to that of Fang, et al., (2015) and Lemaignan, et al., (2017) also, in that it was an HRI communication compatible environment.

The between subjects condition of a humanoid robot agent design versus a non-humanoid robot agent in the VR HRI environment showed that there is not a significant effect introduced in to the communication environment by this factor. Specifically, deictic expression resolution, accuracy and response time metrics are analyzed and resulted in this outcome. However, in regards to participants' gaze interaction with the robot agents, the design of the robot agents did have a significant effect only in the second experiment sessions.

The present results might point to a likely situation that the combination of the static knowledge state and the fixed variables have caused the participants filtering out the robot agent design and focusing solely on the gaze direction and the objects on the table. This situation resulted in an effect for robot agent design in interaction conditions in the second experiment session, as well as a significant effect in all of the first experiment session. Another likely reason may be that the robot agents were not sufficiently distinct from one another. The robot agents might have been perceived as more human-like or less human-like by the participants. In other words, our designs of humanoid and non-humanoid robot agents might have been similar in terms of their uncanny-valley effects (Mori, 1970; Seyama & Nagayama, 2007).

5.4. Limitations

The limitations we experienced during both the design and the execution periods of our experiments begin with immersion issues related to our HMD device. Our HMD device utilized a smartphone device as its screen. The first limitation to immersion was caused by this smartphone device and its effective low resolution display. The smartphone, S7, has a tiled matrix display, where the pixels of the display are arranged in a pen-tile diamond shape. As a result of this pixel matrix arrangement being pen-tile instead of a true red-green-blue (RGB) display, the true resolution of the display is 66% of the 2560x1440 resolution. This screen is, then, also shared between both eyes, resulting in a theoretical 1280x1440 resolution per eye with only 66% of its specified pixels being utilized. The resolution is only a theoretical maximum due to the unknown

utilization of the screen behind the lenses, because the device when observed does not utilize the entirety of the vertical 1440 pixels. In normal terms this resolution is by no means low, however when gazed at through magnifying optical lenses as HMD devices all have, the low resolution resulted in a pixelated image of the VR environment.

The second limitation of the HMD display was its isolation, since it does not isolate light sources from outside the HMD sufficiently. This may have caused users of the HMD device to see the interior of the HMD device, instead of only seeing the VR environment through the lenses, which in turn may have led to distraction.

Another next limitation is related to the experiment procedure. The two experiment sessions for each robot design were always presented to the participants in the same order. Counter balancing the experiment sessions may improve the overall validity of our findings.

5.5. Conclusion

Our first research question and our first hypothesis (that the number of robot agents in the VR HRI environment will have a significant effect on the resolution of deictic references in regards to the metrics of response time and response accuracy) is validated by our experiments. Our second research question and our second hypothesis (that the number of robot agents in the environment will have a significant effect on the saliency and thus the amount of gaze interaction the human participants commit to) is also validated by our experiments. However, in our investigation of both research questions we observed likely threshold points, in terms of the number of robot agents.

The threshold points shows that there is a likely limit to not only how much of the available gaze cues are utilized within the cognitive workload, but also how much gaze interaction the human participants make with the robot agents in the environment. In other words, having between too many robot agents in the environment may provide too much information to the human participants than they can cognitively process, possibly due to cognitive overload, and/or, secondly, can cause a likely social discomfort.

Our final research question and our third hypothesis were investigated by means of a between-subjects design (two distinct robot agents). We observed that the human participants were influenced by the robot design as suggested by the differences in gaze distribution between the conditions. There are two perspectives this result can be interpreted. The first is that because of the thresholds observed in the analyses of the first two hypotheses the interaction between designs of the robot agents are not realized significantly by the participants. The results showed that if and only if the threshold of too many robot agents is passed, then the gaze interaction patterns are significantly affected by the robot agent designs. Secondly, the results may also be interpreted that the robot designs are not sufficiently distinct from one another for a significant effect to be realized.

5.6. Future Work

The final research question requires further investigation. Revisiting the present study with a third robot agent design that is more human-like may improve our understanding of the influence of agent designs in communication. Due to the likely insufficient amount of distinction between the designs of the robot agents, the current experiments have not provided the insight in to how the design of a robot agent in VR HRI environment might interact with the first two research questions. The utilization of a human avatar design is, therefore, a likely follow up study. Also, the addition of a competition setting in how the robot agents provided gaze data may likely allow us to explore the gaze interaction with the robot agents further. The gaze interaction results showed that the robot agents were not equally gazed in the environment for the given joint task. This varying gaze distribution and a competition setting where the robot agents of varying numbers compete for the participants' visual attention would allow us to explore social gaze interaction in HRI further.

The first and the second experiment sessions require further investigation also. A repeated measures ANOVA did not reveal valid results when the two within-subject conditions (one from each session) were analyzed as a single group. A hierarchical linear model (HLM) is required for this analysis for the entirety of the three (1, 3, and 5 as the numbers of robot agents) by two (explicit references versus implicit deictic references) by two (humanoid, non-humanoid robot agent design) model. This analysis may reveal the effects of deixis resolution in response time and accuracy.

The use of an HMD device with higher resolution displays is likely to increase accuracy of the deictic expression resolution, and in return lower the response times also. Even though no participant complained of the sharpness of images, a better HMD device may increase immersion and likely allow us to observe VR related effects, such as reinforcing the distinction between the two robot agent designs, and a likely human avatar design.

Utilizing multimodal referring expressions, such as gestures and gaze vectors at the same time and using bodily movements, similar to Imai, et al. (2001), is an area this experiment has not investigated.

REFERENCES

Adams, R. J. (1999). Stable Haptic Interaction with Virtual Environments, 1–132.

Admoni, H., & Scassellati, B. (2017). Social Eye Gaze in Human-Robot Interaction: A Review. *Journal of Human-Robot Interaction*, *6*(1), 25. https://doi.org/10.5898/JHRI.6.1.Admoni

Anderson, P. L., Zimand, E., Hodges, L. F., & Rothbaum, B. O. (2005). Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure. *Depression and Anxiety*, *22*(3), 156–158. https://doi.org/10.1002/da.20090

Baizid, K., Li, Z., Mollet, N., & Chellali, R. (2009). Human multi-robots interaction with high virtual reality abstraction level. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5928 LNAI*, 23–32. https://doi.org/10.1007/978-3-642-10817-4_3

Bartneck, C., Yogeeswaran, K., Ser, Q. M., Woodward, G., Sparrow, R., Wang, S., & Eyssel, F. (2018). Robots and Racism. *ACM/IEEE International Conference on Human-Robot Interaction*, *Part F1350*, 196–204. https://doi.org/10.1145/3171221.3171260

Breazeal, C. (2004). Social interactions in HRI: The robot view. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, *34*(2), 181–186. https://doi.org/10.1109/TSMCC.2004.826268

Breazeal, C., Dautenhahn, K., & Kanda, T. (2016). Social Robots. In B. Siciliano & O. Khatib (Eds.), *Springer Handbook of Robotics* (pp. 1935–1961). Springer.

Brenton, H., Gillies, M., Ballin, D., & Chatting, D. (2005). The uncanny valley: Does it exist and is it related to presence. *Proc. of British HCI Group Annual Conference: Human-Animated Characters Interaction Workshop*, (2004), 1–8.

Brooks, A. G., & Breazeal, C. (2006). Working with robots and objects. *Proceeding of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction - HRI '06*, 297. https://doi.org/10.1145/1121241.1121292

Cain, B. (2007). A Review of the Mental Workload Literature. *Defence Research and Development Toronto (Canada)*, (1998), 4-1-4–34. Retrieved from http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA474193

Casali, J. G., & Wierwille, W. W. (1983). A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communication load. *Human Factors*. https://doi.org/10.1177/001872088302500602

Clark, H. H., & Wilkes-gibbs, D. (1986). Referring as a collaborative process, Cognition, 22 (1986) l-39 1, *22*, 1–39.

Clodic, A., Alami, R., & Chatila, R. (2014). Key Elements for Joint Human-Robot Action. *Sociable Robots and the Future of Social Relations*, 23–33. https://doi.org/10.3233/978-1-61499-480-0-23

Dağlarlı, E., Dağlarlı, S. F., Günel, G. Ö., & Köse, H. (2017). Improving human-robot interaction based on joint attention. *Applied Intelligence*, *47*(1), 62–82. https://doi.org/10.1007/s10489-016-0876-x

Devault, D., Kariaeva, N., Kothari, A., Oved, I., & Stone, M. (2005). An Information-State Approach to Collaborative Reference. *Proceedings of the ACL 2005 on Interactive Poster and Demonstration Sessions.*, (June), 1–4. https://doi.org/10.3115/1225753.1225754

Duchowski, A. T. (2017). *Eye tracking methodology: Theory and practice: Third edition*. *Eye Tracking Methodology: Theory and Practice: Third Edition*. https://doi.org/10.1007/978-3-319-57883-5

Duguleana, M., Barbuceanu, F. G., & Mogan, G. (2011). Evaluating human-robot interaction during a manipulation experiment conducted in immersive virtual reality. In *International Conference on Virtual and Mixed Reality* (pp. 164–173). Berlin.

Eldon, M. (2015). Incrementally Interpreting Multimodal Referring Expressions in Real Time, 1–22.

Fang, R., Doering, M., & Chai, J. Y. (2015). Embodied Collaborative Referring Expression Generation in Situated Human-Robot Interaction. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 271–278. https://doi.org/10.1145/2696454.2696467

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science*, *29*(5), 737–767. https://doi.org/10.1207/s15516709cog0000_34

Gopher, D., & Donchin, E. (1986). Workload - An examination of the concept. *Handbook of Perception and Human Performance, Vol. 2, Cognitive Processes and Performance*.

Grealy, M. A., Johnson, D. A., & Rushton, S. K. (1999). Improving cognitive function after brain injury: The use of exercise and virtual reality. *Archives of Physical Medicine and Rehabilitation*, *80*(6), 661–667. https://doi.org/10.1016/S0003-9993(99)90169-7

Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A Framework for

Modelling the Local Coherence of Discourse, (1986). https://doi.org/10.21236/ADA324949

Haddadin, S., & Croft, E. (2017). Springer Handbook of Robotics. In *Springer Handbook of Robotics* (2nd ed., pp. 1835–1869). Springer.

Imai, M., Ono, T., & Ishiguro, H. (2001). Physical relation and expression: Joint attention for human-robot interaction. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, *50*(4), 512–517. https://doi.org/10.1109/ROMAN.2001.981955

Lemaignan, S., Warnier, M., Sisbot, E. A., Clodic, A., & Alami, R. (2017). Artificial cognition for social human–robot interaction: An implementation. *Artificial Intelligence*, *247*, 45–69. https://doi.org/10.1016/j.artint.2016.07.002

Levelt, W. J. M., Richardson, G., & La Heij, W. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language*, *24*(2), 133–164. https://doi.org/10.1016/0749-596X(85)90021-X

Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Messinger, D., Neal, R., & Schwartz, H. K. (2000). Responding to Joint Attention Across the 6- Through 24-Month Age Period and Early Language Acquisition. *Journal of Applied Developmental Psychology*, *21*(3), 283–298. https://doi.org/10.1016/S0193-3973(99)00040-4

Morales, M., Mundy, P., & Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior and Development*, *21*(2), 373–377. https://doi.org/10.1016/S0163-6383(98)90014-5

Mori, M. (1970). The uncanny valley. *Energy*. https://doi.org/10.1109/MRA.2012.2192811

Mutlu, B., Shiwa, T., Kanda, T., Ishiguro, H., & Hagita, N. (2009). Footing in human-robot conversations: how robots might shape participant roles using gaze cues. *Human Factors*, *2*(1), 61–68. https://doi.org/10.1145/1514095.1514109

Piwek, P. (2009). Salience in the generation of multimodal referring acts. *Proceedings of the 2009 International Conference on Multimodal Interfaces - ICMI-MLMI '09*, 207. https://doi.org/10.1145/1647314.1647351

Rizzo, a a, & Buckwalter, J. G. (1997). Virtual reality and cognitive assessment and rehabilitation: the state of the art. *Studies in Health Technology and Informatics*, *44*, 123–145. https://doi.org/10.3233/978-1-60750-888-5-123

Ruhland, K., Peters, C. E., Andrist, S., Badler, J. B., Badler, N. I., Gleicher, M., … McDonnell, R. (2015). A Review of Eye Gaze in Virtual Agents, Social Robotics and HCI: Behaviour Generation, User Interaction and Perception. *Computer Graphics Forum*, *34*(6), 299–326. https://doi.org/10.1111/cgf.12603

Rus, D. (2017). Springer Handbook of Robotics. In *Springer Handbook of Robotics* (2nd ed., pp. 1785–1788). Springer.

Schuemie, M. J., van der Straaten, P., Krijn, M., & van der Mast, C. A. P. G. (2001). Research on Presence in Virtual Reality: A Survey. *CyberPsychology & Behavior*, *4*(2), 183–201. https://doi.org/10.1089/109493101300117884

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009

Seyama, J., & Nagayama, R. S. (2007). The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces. *Presence: Teleoperators and Virtual Environments*, *16*(4), 337–351. https://doi.org/10.1162/pres.16.4.337

Sisbot, E. A., Ros, R., & Alami, R. (2011). Situation assessment for human-robot interactive object manipulation. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, *11*, 15–20. https://doi.org/10.1109/ROMAN.2011.6005258

Wang, Z., Giannopoulos, E., Slater, M., & Peer, A. (2011). Handshake: Realistic Human-Robot Interaction in Haptic Enhanced Virtual Reality. *Presence: Teleoperators and Virtual Environments*, *20*(4), 371–392. https://doi.org/10.1162/PRES_a_00061

Whitney, D., Eldon, M., Oberlin, J., & Tellex, S. (2016). Interpreting multimodal referring expressions in real time. In *Proceedings - IEEE International Conference on Robotics and Automation* (Vol. 2016–June, pp. 3331–3338). https://doi.org/10.1109/ICRA.2016.7487507

Wickens, C. D., & Baker, P. (1995). Cognitive Issues in Virtual Reality. *Virtual Environments and Advanced Interface Design*, (2), 514–541.

Witmer, B. G., & Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240. https://doi.org/10.1162/105474698565686

Yilmaz, E. (2018). Deictic Gaze in Virtual Environments Experiment Flow Demo [Video File]. Retrieved August 12, 2018, from https://youtu.be/1roYlDwtDqY

Yu, C., Scheutz, M., & Schermerhorn, P. (2010). Investigating multimodal real-time patterns of joint attention in an HRI word learning task. *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 309–316. https://doi.org/10.1109/HRI.2010.5453181

Yücel, Z., Salah, A. A., Meriçli, Ç., Meriçli, T., Valenti, R., & Gevers, T. (2013). Joint attention by gaze interpolation and saliency. *IEEE Transactions on Cybernetics*, *43*(3), 829–842. https://doi.org/10.1109/TSMCB.2012.2216979

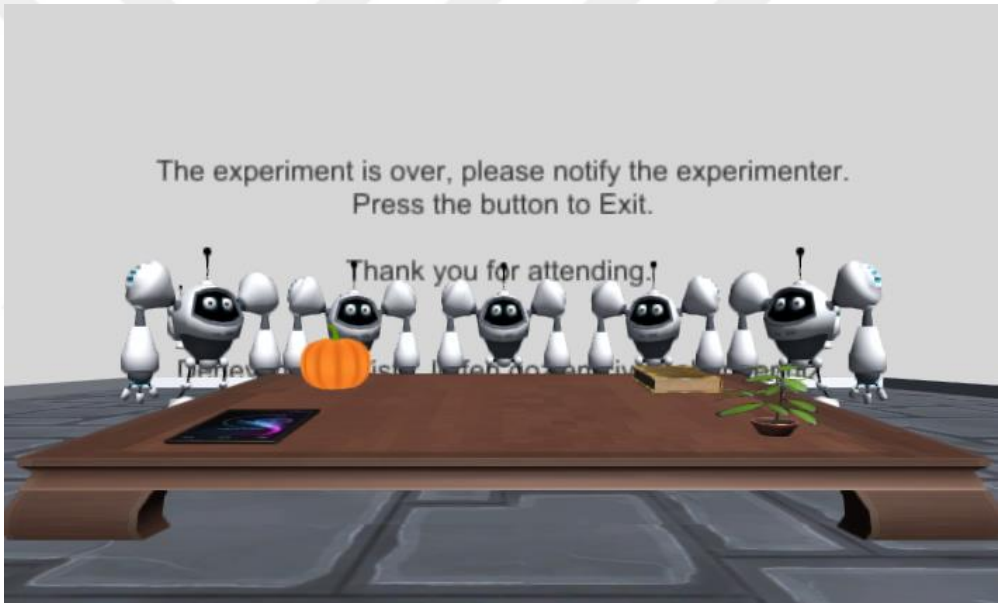
APPENDICES

APPENDIX A

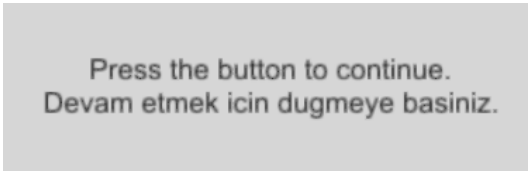
EXPERIMENT INSTRUCTION SHEET 1 – INSTRUCTIONS FOR HUMANOID ROBOT AGENT

Merhabalar,

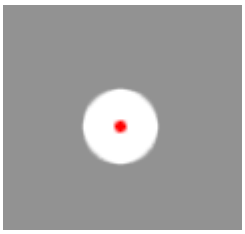
Oncelikle gozlugu takarken sag tarafina dokunmamaya lutfen dikkat ediniz. Sanal gerceklik ortaminda goz takip cihazi ile yapilan bu calismada oncelikle asagidaki test ekraniyla karsilasacaksiniz:



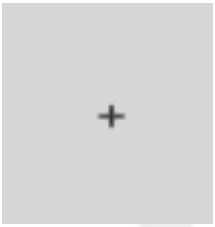
Bu sahnede cihazin ust kisminda bulunan tekerlek araciligi ile netlik ayari yapmaniz bekleniyor; robotlari ve masanin uzerindeki 4 objeyi olabildigince net gorecek sekilde bu ayari yapiniz. Bir miktar netlik bozuklugu kalabilir, burada onemli olan ortamin **olabildigince** net olmasidir. Bunu bitirip hazir oldugunuzda cihazin **sag tarafina** hafifce **1 kere dokunmaniz** sizi asagidaki ekrana tasiyacaktir:



Bu gecis sahnesi deney boyunca hazir oldugunuzda tekrar cihazin sag tarafina hafifce 1 kere dokunmanizla ilerleyecektir. Ilk sahne olarak asagidaki ekrani goreceksiniz.



Bu sahnede ortasinda kirmizi nokta olan cemberi gozlerinizle takip etmeniz gerekiyor, goz takip cihazinin kalibrasyonu icin lutfen bu kirmizi noktayi olabildigince keskin bir sekilde izlemeye calisiniz. Bu islem bittiginde tekrar gecis ekranina, ilerlediginizde ise sonraki sayfadaki '+' isareti ekranina tasinacaksiniz.



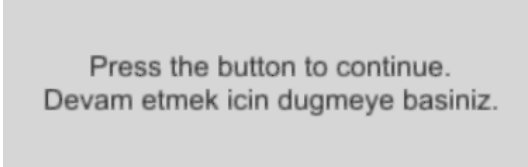
Soldaki bu '+' isaretin amaci sizin her yeni sahneye ayni noktaya bakarak girmenizdir. Lutfen "Devam etmek icin dugmeye basiniz" yazisi sonrasinda bu noktaya bakmaya ozen gosteriniz.

'+' Isareti sonrasinda 1 saniyelik bir bekleme suresi ardindan asagidaki sahne veya benzeri ekranda gosterilecek:

Bu ornek sahnede masa uzerinde birer adet tablet, saksi, kitap ve kabak gormektesiniz; ayrica masanin uzak tarafinda da 1 adet robot aktor gozukmekte. Deneyimiz sirasinda robot ya da robotlar size sunlardan birisini soracak:

- Sol ustte ne var?
- Sol altta ne var?
- Sag ustte ne var?
- Sag altta ne var?

Yapmaniz gereken sorulara olabildigince hizli ve dogru cevap vermek, ancak kendi hizinizda gitmeniz en onemlisi; kisaca, hizli cevap vermek ugruna yanlis cevap vermeyin, ancak cevabinizi da kesinlestirir kesinlestirmez yuksek sesle soyleyin ve soyledikten sonra dugmeye tekrar dokunarak ilerleyin. Lutfen cevap vermeden dugmeye tekrar basmayin. Bu asamada cevabinizi verir vermez tekrardan dugmeye tiklayarak (1. basma) asagidaki bekleme ekranina ile devam edin ve tekrar dugmeye basarak (2. basma) yeni sahneye gecin.

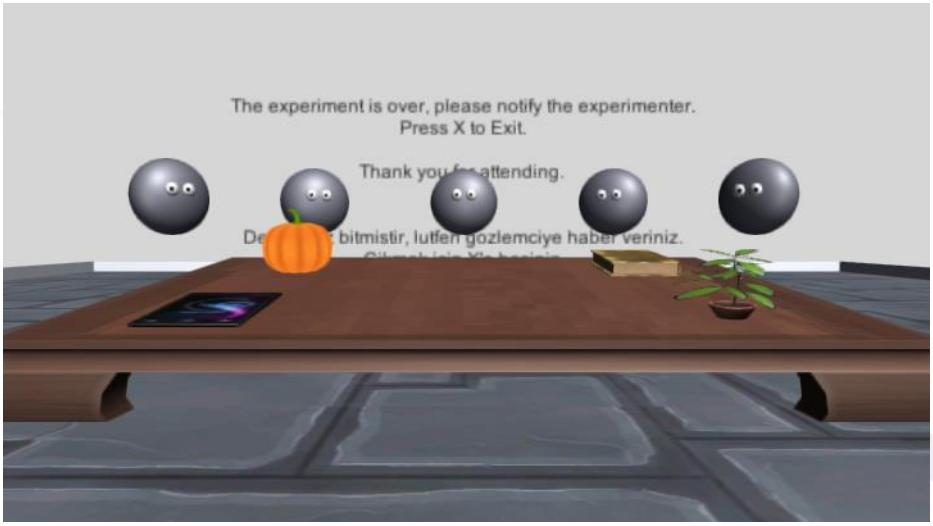


Calismanin ikinci asamasinda robotlarin sorulari “Burada ne var?” seklinde olacak. Bu asamada robotlarin hangi objeyi kastettiklerini goz isaretleriyle takip edebiliyor olacaksiniz.

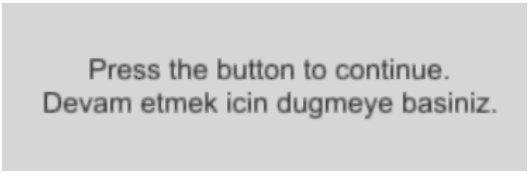
EXPERIMENT INSTRUCTION SHEET 2 – INSTRUCTIONS FOR NON-HUMANOID ROBOT AGENT

Merhabalar,

Oncelikle gozlugu takarken sag tarafina dokunmamaya lutfen dikkat ediniz. Sanal gerceklik ortaminda goz takip cihazi ile yapilan bu calismada oncelikle asagidaki test ekraniyla karsilasacaksiniz:

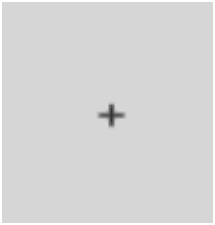


Bu sahnede cihazin ust kisminda bulunan tekerlek araciligi ile netlik ayari yapmaniz bekleniyor; robotlari ve masanin uzerindeki 4 objeyi olabildigince net gorecek sekilde bu ayari yapiniz. Bir miktar netlik bozuklugu kalabilir, burada onemli olan ortamin **olabildigince** net olmasidir. Bunu bitirip hazir oldugunuzda cihazin **sag tarafina** hafifce **1 kere dokunmaniz** sizi asagidaki ekrana tasiyacaktir:



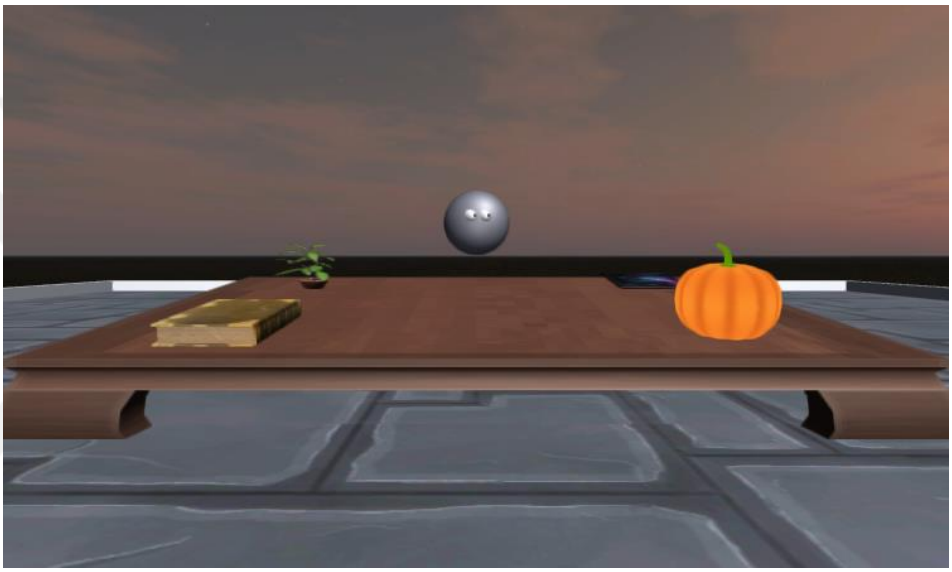
Bu gecis sahnesi deney boyunca hazir oldugunuzda tekrar cihazin sag tarafina hafifce 1 kere dokunmanizla ilerleyecektir. Ilk sahne olarak asagidaki ekrani goreceksiniz.

Bu sahnede ortasinda kirmizi nokta olan cemberi gozlerinizle takip etmeniz gerekiyor, goz takip cihazinin kalibrasyonu icin lutfen bu kirmizi noktayi olabildigince keskin bir sekilde izlemeye calisiniz. Bu islem bittiginde tekrar gecis ekranina, ilerlediginizde ise sonraki sayfadaki '+' isareti ekranina tasinacaksiniz.



Soldaki bu '+' isaretin amaci sizin her yeni sahneye ayni noktaya bakarak girmenizdir. Lutfen "Devam etmek icin dugmeye basiniz" yazisi sonrasinda bu noktaya bakmaya ozen gosteriniz.

'+' Isareti sonrasinda 1 saniyelik bir bekleme suresi ardindan asagidaki sahne veya benzeri ekranda gosterilecek:

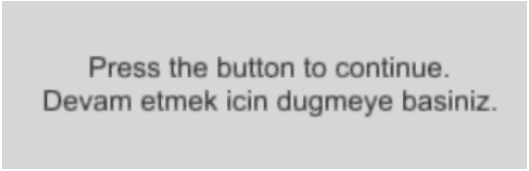


Bu ornek sahnede masa uzerinde birer adet tablet, saksi, kitap ve kabak gormektesiniz; ayrica masanin uzak tarafinda da 1 adet robot aktor gozukmekte. Deneyimiz sirasinda robot ya da robotlar size sunlardan birisini soracak:

- Sol ustte ne var?
- Sol altta ne var?
- Sag ustte ne var?
- Sag altta ne var?

Yapmaniz gereken sorulara olabildigince hizli ve dogru cevap vermek, ancak kendi hizinizda gitmeniz en onemlisi; kisaca, hizli cevap vermek ugruna yanlis cevap vermeyin, ancak cevabinizi da kesinlestirir kesinlestirmez yuksek sesle soyleyin ve soyledikten sonra dugmeye tekrar dokunarak ilerleyin. Lutfen cevap vermeden dugmeye tekrar basmayin.

Bu asamada cevabinizi verir vermez tekrardan dugmeye tiklayarak (1. basma) asagidaki bekleme ekranina ile devam edin ve tekrar dugmeye basarak (2. basma) yeni sahneye gecin.



Calismanin ikinci asamasinda robotlarin sorulari “Burada ne var?” seklinde olacak. Bu asamada robotlarin hangi objeyi kastettiklerini goz isaretleriyle takip edebiliyor olacaksiniz.

APPENDIX B – PARSING APPLICATION

The raw data generated by our experiment applications were cleared by means of a parsing application written in C programming language. The parsing application took each participant's data for each experiment session and performed the necessary data refinement by selecting specific data pieces, such as the final response time for each response sample, or by calculating the ratios in which the participant's gaze was distributed in the environment for each response sample. In programming this parsing application, the following computational approaches were followed:

- A stack type data structure is used in storing each gaze sample for a given response sample.
- An array of counters is used for calculating both the amount of total gaze samples in each response sample and what specific object was gazed for how many gaze samples. Percentages of which are calculated by means of the ratio between each object and the total gaze samples in each response sample.
- Each participant data was stored as a text file and was parsed using native C file scan function and scan-set expressions, which are similar to regular expressions.
- Each participant file processed are stored in output files in the naming convention of "initialFileName.txt" (input) → "initialFileNameResult.txt" (output).
- The application automatically opens and processes all files in the given input naming convention and records various versions of refined data to allow for easier processing afterwards, as selected by the experimenter, in their respective output files.