AN EXPERIMENTAL INVESTIGATION OF GAZE BEHAVIOR MODELING IN VIRTUAL REALITY

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

ELİF ESMER

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

NOVEMBER 2022

AN EXPERIMENTAL INVESTIGATION OF GAZE BEHAVIOR MODELING IN VIRTUAL REALITY

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

Prof. Dr. Banu Günel Kılıç
Dean, **Graduate School of Informatics** ____________________

Dr. Ceyhan Temürcü
Head of Department, **Cognitive Science** ____________________

Asst. Prof. Dr**.** Murat Perit Çakır
Supervisor, **Cognitive Science Dept., METU** ____________________

Assoc. Prof. Dr. Cengiz Acartürk
Co-Supervisor, **Cognitive Science Dept., METU** ____________________

Examining Committee Members:

Assoc. Prof. Dr. Barbaros Yet
Cognitive Science Dept., METU ____________________

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

Asst. Prof. Dr. Murat Ulubay
Management Dept., AYBU ____________________

Date: 29.11.2022

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 : Elif Esmer

Signature : __________

ABSTRACT

AN EXPERIMENTAL INVESTIGATION OF GAZE BEHAVIOR MODELING IN VIRTUAL REALITY

Esmer, Elif
MSc., Department of Cognitive Science
Supervisor: Asst. Prof. Dr. Murat Perit Çakır
Co-Supervisor: Assoc. Prof. Dr. Cengiz Acartürk

November 2022, 55 pages

The purpose of this study is to investigate if gaze behavior modeling on a robot avatar makes any difference in the engagement of human interlocutors. To that end, visual interaction patterns and impressions of participants during a one-on-one active conversation setting in the form of a mock-up job interview with an avatar agent have been analyzed. Experiments were conducted through eye-tracking methodology, questionnaires, and open-ended post-experimental evaluations. A robot avatar as an artificial agent programmed with pre-recorded speech and pre-planned gaze behavior, based on human gaze patterns, took the role of the human resource manager, in other words, the interviewer. The eye-tracking data of the participants, namely interviewees, were collected and analyzed to investigate whether the human-like gaze behavior of the artificial agent had any effect on the gaze allocation of the human interlocutors. Results of the questionnaires and the TF-IDF analysis of the post-experimental evaluations were inspected to detect reflections of participants' gaze behavior and further patterns for the perception of the avatar.

Keywords: Human-Robot Interaction, multimodal interaction, eye-tracking, virtual reality, TF-IDF

ÖZ

GÖZ HAREKETİ MODELLEMESİNİN SANAL GERÇEKLİKTE DENEYSEL BİR İNCELEMESİ

Esmer, Elif
Yüksek Lisans, Bilişsel Bilimler Bölümü
Tez Yöneticisi: Dr. Öğretim Üyesi Murat Perit Çakır
Eş-Danışman: Doç. Dr. Cengiz Acartürk

Kasım 2022, 55 sayfa

Bu çalışmanın amacı, robot bir avatara programlanmış göz hareketi modellemesinin insan muhatapların etkileşimi üzerinde herhangi bir etkisinin olup olmadığını araştırmaktır. Bu amaçla, robot bir avatarla iş görüşmesi senaryosu dahilinde gerçekleştirilen birebir konuşma ortamında katılımcıların görsel etkileşim örüntüleri ve izlenimleri araştırılmıştır. Deneyler, göz takip metodolojisi, anketler ve deney sonrası sözlü görüşmeler yoluyla gerçekleştirilmiştir. Önceden kaydedilmiş konuşmayla programlanmış ve insan göz hareketleri örüntülerine dayılı biçimde bakışları modellenmiş bir avatar robot, insan kaynakları yöneticisi -dolayısıyla görüşmeci- olarak deneyde yer almıştır. Katılımcıların -görüşülen kişilerin- göz izleme verileri, robot avatarın insan benzeri bakış davranışının, insan muhataplarının göz teması ve göz kaçırma davranışları üzerinde herhangi bir etkisinin olup olmadığını araştırmak üzere analiz edilmiştir. Anket sonuçları ve deney sonrası değerlendirmelerin TF-IDF analizi, katılımcıların göz takibi verilerinin yansıtılışı ve avatarın algılanışına ilişkin örüntüler tespit etmek için incelenmiştir.

Anahtar Sözcükler: İnsan-Robot Etkileşimi, çok-kipli etkileşim, göz takibi, sanal gerçeklik, TF-IDF

To mind flowers

ACKNOWLEDGMENTS

First and foremost, I would like to express my gratitude to my co-supervisor Assoc. Prof. Dr. Cengiz ACARTÜRK for his mentorship, for granting me a place in his laboratory, broadening my horizon and giving me a lesson every step of the way over the years. I would also like to give my sincerest thanks to my supervisor Asst. Prof. Dr. Murat Perit ÇAKIR and the members of the Examining Committee, Assoc. Prof. Dr. Barbaros YET and Asst. Prof Dr. Murat ULUBAY, for their positive attitude and valuable feedback. Another thanks should be dedicated to TUBITAK for the project “117E021/A Gaze4HRI”

I would like to extend my gratitude to Assoc. Prof. Dr. Murat ÖZGEN for his sincere help, for igniting my curiosity and also paving the way for me, which led me to go to places I did not think I could.

My deepest thanks go to my mother Feriha ESMER for her serene but inspirational nature and endless love; I feel so blessed. I am very grateful to my parents for always encouraging me.

Next comes my colleagues and friends from the Department of Cognitive Science at METU. I thank Gizem ÖZEN for being the Louise to my Thelma; Efecan YILMAZ for his go-getter nature and patient helps; Gamze EŞDUR for helping me overcome formalities; and Ekinsu ÖZKAZANÇ for being a comrade to me on this journey from day one.

It would not go without mentioning my friends, in other words, my second family; so big thanks goes to these names starting from the members of “the gang” Berk ÖZERGÜN for his bright and lovely nature; Rıfat Can SALMAN for providing me with the time, space and much-needed pep talks for years; Tayfun KARAKAVUZ and Zeynep TABAKOĞLU for their kindness and joy. I also thank Sevdanur ÇETİN for providing me with her witty mind, her cats and a sunny place. I owe a lot to these names for always being there for me with their generous hearts. I truly appreciate them for being patient with me, which, I imagine, can be hard at times.

Lastly, I would like to acknowledge the support of dear M. Hakan GÜLER for his patience and supreme jokes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

HRI	Human-Robot Interaction
HCI	Human-Computer Interaction
VR	Virtual Reality
ASR	Automated Speech Recognition
STT	Speech-to-Text
VAD	Voice Activity Detection
NLU	Natural Language Understanding
TTS	Text-to-Speech
HHI	Human-Human Interaction
HMD	Head Mounted Display
TIPI	Ten Item Personality Inventory
NARS	Negative Attitudes and Anxiety toward Robots
RAS	Robot Anxiety Scale
TF-IDF	Term Frequency-Inverse Document Frequency
TF	Term Frequency
IDF	Inversed Document Frequency
HMM	Hidden Markov Model
WoZ	Wizard of Oz

CHAPTER 1

INTRODUCTION

Non-verbal signals play a crucial role in our daily lives. Body language, eye gaze, intonation, and the like can reveal much about our internal state of mind. The way we automatically pick up and process these signals has been a topic of interest for the science community for a long time (Ekman, 1957). This second channel of communication comprises such an essential aspect of human communication that, for example, from data collected in two minutes, some studies used "social sensors" such as eye gaze and posture to successfully predict which people trade contact information at a conference without any previous information about the attendees (Pentland & Heibeck, 2010). Another striking example that highlights the essential part of this second channel is that during a marital conflict discussion, couples that would break up in six years gave themselves away in three minutes with their bodily functions and paralinguistic behavior (Carrère & Gottman, 1999).

With the developments in technology and science, deciphering the representations that underlie the state of affairs of the human mind is as attractive as ever. Considering that partners exchange mental states and representations of the world throughout social interactions (Marchetti et al., 2018), one might wonder what happens when one of the partners is an artificial agent in the form of a humanoid robot. In the present study, we aimed to contribute to the literature by investigating the subtle meanings of new interpretations introduced to the human mind due to social interactions with artificial agents. Our motives were bifold; we aimed to arrive at conclusions regarding the multimodal aspects of human communication and test alternative tools for measuring Human-Robot Interaction (hereafter, HRI).

The present thesis employed eye tracking with a head-mounted display (hereafter, HMD) device to investigate aspects of HRI in a virtual reality (hereafter, VR) environment. The environment had a human-robot active conversation setting to explore non-verbal communication cues in HRI. With this aim, we designed a VR setting in which a robot as an artificial agent interviews participants in the context of a mock-up job interview. The multimodal design of the setting employs pre-planned speech, gestures, and modeled gaze behavior, allowing us to investigate the following research questions:

1. Does gaze behavior modeling on an artificial agent as a robot influence subtle gaze signals of the human interlocutor in an active conversation?

2. Do traditional HRI questionnaires reflect the observed gaze patterns of human interlocutors?

3. Does linguistic analysis of open-ended post-experimental evaluations reflect the gaze behavior of the human interlocutors?

While investigating the above questions, traditional HRI questionnaires and post-experimental evaluations were applied alongside with the experiments in VR.

The thesis is made up of five chapters in total. In the first chapter, the scope of the study, why it is worth investigating, motivations, a brief explanation of the research questions and methods, and a basic outline are presented to the reader. The following chapter explains the field of HRI, the multimodality of human communication, and current research in this context. In the third chapter, the study's methodology, the data acquired, and the strategies adopted for the analysis are described in detail. In the fourth chapter, the collected results from the experiments are stated. In the fifth and last chapter, acquired results are discussed, and key findings are concluded. Limitations of the current study and recommendations for future studies are stated.

CHAPTER 2

LITERATURE REVIEW

As stated by Duguleana et al. (2011), in the literature, the term Human-Robot Interaction (HRI) tends to be used to refer to the process of understanding and shaping the interaction between humans and robots. As a multidisciplinary field, HRI is heavily influenced by studies in many fields such as human-computer interaction, robotics, artificial intelligence, cognitive science, linguistics, psychology, philosophy, and design.

As robots' place in society grows and changes, several problems relating to the interaction between humans and robots arise (Bartnecket al., 2020). With the need to advance how robots interact with people, the relatively new field of HRI has emerged. Human-Computer Interaction (HCI), which can be broadly defined as a multidisciplinary field focused on developing computer technologies concerning human aspects (Magnenat-Thalmann et al., 2016), does not cover the field of HRI, despite it concerning robots. This is because interactions with robots usually include physically embodied robots; therefore, they are fundamentally distinct from other computing technologies because of this embodiment. Hence, the problem "to understand and shape the interactions between one or more humans and one or more robots" (p.215) defined the challenge that served as the foundation for HRI as a research domain (Goodrich & Schultz, 2008).

HRI seeks to create robots that can interact with humans in various real-world settings. Robots are constructed from several software modules joined by sensors and actuators. HRI expertise is needed both for software and hardware design (Bartneck et al., 2020). The various elements of a robot must be selected and integrated concerning the requirements of the particular HRI application for it to succeed.

In this chapter, the milestone robots regarding HRI, the main requirements for socially interactive robots, the humanoid robot Pepper, the multimodality aspect of human communication, and the state-of-art technologies, eye-tracking, and virtual reality (VR) technologies in HRI are presented to the reader.

2.1. Evolution of Human-Robot Interaction

Even when robots were only in fiction, their interaction with humans has been a topic of pondering. Every new technological or conceptual advancement in robotics and new findings in social sciences made scientists reevaluate their perception of and connection with robots.

The first industrial robot, called Unimate, was first manufactured in 1961. Following this benchmark, entrepreneurs saw the potential, which inspired the development of robots with quick reactions and a sense of social presence (Gasparetto & Scalera, 2019).

One of the first examples of social robots was Kismet, the robot was comprised of only a head and a neck mounted on a tabletop box. Kismet used posture and facial expression to interact with people and engage with them physically, emotionally, and socially to eventually learn from them (Breazeal, 2003).
One of the most influential robots in the study of social robotics was Nao. Humanoid robot Nao was a manifestation of worldwide robot manufacturers working together to provide a platform for social and scientific purposes (Shamsuddin, 2011).

Kinetically built differently, the Baxter robot was an industrial robot with common sense, manufactured from 2011 to 2018. Baxter was a unique robot with abilities such as working with humans safely without needing protective cages, collaborating through an unrivaled, user-friendly interface, being trained manually with no programming required, and responding adaptively to the changes in its environment (Fitzgerald, 2013).

As a result of technological advancements and easily accessible tools such as 3D printing and laser cutting, researchers can now construct and alter robots quickly and affordably, resulting in a wide variety of studies in HRI. Our study benefitted from these advancements by conducting an experiment in VR with a robot avatar as an artificial agent whose eye gaze behavior had been modeled.

2.2. Socially Interactive Robots

Amid the fourth industrial revolution, the new machine age, social robots are gradually transitioning from fiction to reality. Social robots, also known as socially interactive robots, are those that interact with people and other robots in a way that is considered acceptable in society, transmit their intentions in a manner that people can understand, and are given the authority to work with other agents -human or robot- to achieve their objectives (Daily et al., 2017). Service robots, educational robots, companion robots for older people, and personal assistant robots are a few examples of robot application areas where socially interactive robots, therefore HRI, play a significant role. With the growing interest in robots and advancements in technology and science, robotics markets are expanding more and more (Gonzalez-Aguirre et al., 2021). Yet, many of the products on the market today still have limited social interaction skills (Kaminka, 2013). HRI studies have been trying to clarify the way in the complex context of human-robot communication and Human-Robot Collaboration (HRC) by laying a strong emphasis on examining the structure of social interactions between people and between people and robots not only in dyads but also in groups.

A robot's overall appearance, mobility, and posture play an essential role in how it functions and how it collaborates as well as how it is perceived by humans. In the instance of socially interactive robots, HRI research has also demonstrated how crucial it is to contemplate human-like shape and function in the designing process for the sake of improving the perceived interaction quality, HRI acceptability, and engagement (Fink, 2012). Studies have shown that a robot is seen as more compelling and scored higher on perception than an animated figure due to its physical embodiment and tactile communication even when the robot is not present in the same physical environment as humans (Kidd & Breazeal, 2004; Lee et al., 2006), suggesting that people are prone to seeing robots as social agents. A study has revealed that even perceived social categorization and subsequent differential social evaluations of robots matter (Eyssel & Kuchenbrandt, 2012). A study showed that when a humanoid robot aids the teacher in a classroom setting, children are interested in exploring the robot's identity by paying attention to details such as its voice, actions, and reactions, which indicates that a humanoid robot that embodies human-like social signals happens to be capable of being quite engaging (Chang et al., 2010)

Since social interaction requires verbal and non-verbal cues; in addition to its appearance, a robot's spatial interaction behavior, voice, linguistic choices, gestures, and facial expressions also comprise the critical considerations in the designing process not only for achieving socially acceptable and pleasant HRI but also for providing a safe space for engagement where people can comprehend the robot's intentions and interact with them comfortably in their physical space. In this present study, we programmed our robot avatar as an artificial agent with gaze behavior modeling and gestures to account for the requirements listed above.

2.3. The Humanoid Robot: Pepper

Pepper is a 1.2 meters tall, wheeled, industrially produced humanoid robot, developed by SoftBank Robotics and launched in 2014, intending to play critical roles in everyday life and coexist with humans. Pepper's operating system is called NAOqi; software development kits are also provided to control and develop it like Phyton, C++, and Java. It also has a tablet attached to its upper body, serving as an additional communication channel.

Pepper was initially intended for a specific application of business-to-business uses in SoftBank stores, yet the robot quickly gained popularity around the globe for a variety of other applications -including business-to-costumer, business-to-academics, and business-to-developers areas- and in a diverse range of use cases such as healthcare, education, entertainment, and business (Pandey & Gelin, 2018).

Pepper is capable of moving around, detecting and reacting to its environment, and displaying body language. Pepper can also evaluate facial expressions and speech tones by using voice and emotion recognition algorithms to facilitate conversations. The robot has features and interfaces enabling multimodal interaction with the people around it.

Its design was influenced by several factors, including a pleasing appearance, safety, affordability, interaction, gender-neutralness, and good autonomy (Pandey & Gelin, 2018). A study showed that secondary school students' trust was positively affected as they got to know Pepper more after three different types of exposure; furthermore, when they had a live interaction with Pepper, their willingness to have Pepper in their house increased due to its embodiment (Rossi et al., 2018). When taken to a shopping mall for one day, humans were keen to interact with and trust Pepper (Aaltonen et al., 2017). In this study, we chose Pepper as our artificial agent for the role of Human Resource manager in the mock-up job interview in the VR study.

2.4. Multimodal Interaction

Multimodality of human interaction refers to the multi-channel nature of human communication; multimodality represents both a field of study and a domain to be theorized (Kress, 2009). It is an inter-disciplinary approach to exploring interaction. Multimodal interaction is a combination of verbal and non-verbal interaction, in which the latter emphasizes the meaning and intent of the former. These elements are introduced in 2.4.1. and 2.4.2.

Human interaction with the world is inherently multimodal (Bunt et al., 1998; Quek et al., 2002). Humans employ their senses sequentially and simultaneously to actively and passively explore their surroundings, validate their assumptions about the outside world, and comprehend new information (Turk, 2014). Given how humans experience external stimuli, interpret and take action with their surroundings via their senses, it is not surprising that their communication also involves multiple senses.

In order to provide a more natural and effective interaction, HCI studies have sought to grant computers with comparable capabilities. In addition to its intuitiveness, multimodal interaction systems have many advantages compared to unimodal systems. They can offer better flexibility and reliability, more engagement, improved efficiency, greater precision, and can reach a wider variety of users (Oviatt et al., 2000).

Studies show that humans may engage with robots that have faces in a manner similar to how they do with other living beings with faces. A person may speak to it, make gestures at it, smile at it, and do other things. If a person uses a computer or another machine that can interpret spoken commands, they could talk to it and may assume it has proficiency in spoken language (Perzanowski et al., 2001). These findings suggest that multimodal interaction not only offers many advantages but also is expected. Therefore, multimodality of human interaction is regarded as one of the essential keys in the designing process of socially interactive robots since the goal is to provide humans with comfortable, comprehensible, and safe engagement with the robots.

The following phases make up a typical pipeline for a multimodal interaction system: (1) Individual low-level sensing modules, (2) multimodal tracking and fusion to combine data from various trackers to draw high-level conclusions about the environment and the user state, (3) decision making and dialogue management to choose what to say and do given the limited sensory information, actions taken in the past, and internal state of the artificial character, (4) planning and synchronization of the output behavior to render the output decisions, and (5) actual execution of the planned behaviors (Yumak & Magnenat-Thalmann, 2016).

In our study, we employed several strategies for the artificial agent to account for the multimodality aspect of human interaction.

2.4.1. Verbal Interaction

Verbal interaction is the interaction that happens through the usage of words or messages in a linguistic form. Verbal communication is enhanced by paralinguistic signals like prosody and intonation and non-verbal behavior such as gaze, gestures, posture, and facial expressions.

As it is maybe the most prominent communication type for humans, a crucial objective for HRI is to develop robots that can interact effectively with verbal interaction. In order to provide natural language interaction with robots, several technical requirements must be met. These include the robot's capacity to transcribe spoken language into written language, formulate suitable answers, and produce spoken language, which is more complicated than using only written text (Bartneck et al., 2020).

Speech recognition, also known as automated speech recognition (ASR) or speech to text (STT), is often used to enable natural HCI and HRI. It is the process of converting an acoustic speech data sequence into a word or other linguistic symbol sequence and has been an intensive research area for decades (Yu & Deng, 2016). This process requires a speech recognition software trained to transcript a specific language. Some of these softwares are trained for only detecting certain commands, while others are unrestrained. As speech recognition is mainly used for controlling digital devices, in the context of HRI, some difficulties arise. Since robots are generally physically present when engaging with human partners, distance can be challenging, or various other environmental sounds can cause beclouding. Signal processing and specific microphone arrays can eliminate distance problems; as for environmental noise -for example, other people talking, music, or machines- there are technologies such as voice activity detection (VAD). In recent years VAD technology has been dramatically improved by deep learning studies (Wang et al., 2019). Although these improvements may cause VAD to give the impression of mutual understanding to the human partner, speech recognition does not comprehend the meaning of the utterance; it only converts the speech data into a text form for further processing. For excerpting the semantic content from speech utterances, sentiment analysis or natural language understanding (NLU) methods and tools can be utilized; these have been developed to identify affective feelings using keywords, prosody, and amplitude (Salvendy & Karwowski, 2021). Despite these developments, an open-ended understanding of natural language still remains to be one of the biggest challenges in HCI and HRI.

After analyzing the speech input, there needs to be a speech production to make the interaction. In this production process, dialogue management systems may be utilized; these systems can be roughly categorized as task-oriented systems and non-task-oriented systems. The former aims to assist the user in completing certain tasks, while the latter's -the more commonly used one- objective is to provide reasonable responses and entertainment (Chen et al., 2017). The most prevalent dialogue managers are event-based since they offer enough flexibility for the majority of language-based business interactions. However, using dialogue managers to conduct a natural and open conversation is not viable. Free linguistic conversation necessitates a wide variety of language rules, and the dialogue script quickly grows cumbersome (Bartneck et al., 2020).

The last step in natural language interaction is to turn the response of the system into speech. For this part of the speech production process, text-to-speech (TTS) or also known as speech synthesis, is applied. This field of study is the process of converting written text into speech (Taylor, 2009). By training generative deep neural networks (DNNs), most of the limitations of the field have been overcome. A DNN model called WaveNet -also adopted by Google for the voice of its digital assitance- produces speech that is rated more natural sounding for both English and Mandarin than the best TTS systems of the time (Oord et al., 2016).

Overall, verbal HRI may be achieved to some extent using technical tools for speech analysis, synthesis, and production. There have been some studies aiming to make the verbal interaction between a human and a robot as fluent as possible, which utilizes verbal commands and state conveying utterances in collaboration on simple tasks (Nikolaidis et al., 2018). Even though they had some success, open-ended, fluid communication problem still remains to be unsolved as this type of interaction is exceptionally complex and vastly varied (Mavridis, 2015).

For our study, we utilized pre-recorded speech for the artificial agent. Since a mock-up job interview scenario comprised our study, we were able to choose pre-recorded speech as roles in these kinds of interactions are pre-fixed, and utterances (for the interviewer part) are generally the same.

2.4.2. Non-verbal Interaction

Humans coordinate their actions effectively through both verbal and non-verbal interaction. While what is being conveyed linguistically comprises the verbal channel of the interaction; posture, gestures, facial expressions, and gaze comprise the other part, namely the non-verbal channel. In the absence of this channel, it is tough to comprehend the true meaning of the utterance and the intention and emotional state of the sender. That is why it is an essential topic for HRI research to study interpreting, using, and appropriately responding to non-verbal cues. Otherwise, establishing common ground and mutual understanding between humans and robots may not be possible.
Non-verbal cues people produce can serve as an indicator of attitude and engagement for robots. On the other hand, cues presented by robots through this channel can decide whether the interaction is smooth, awkward, or even non-existent. A study found that non-verbal interaction positively affects HRC regarding the understandability of the robot, the efficiency of task performance, and robustness to miscommunication errors (Breazeal et al., 2005).

Non-verbal interaction cues are realized through various modalities, such as sound, gaze, and movement. Combining these different types of cues enables partners to experience a context-appropriate HRI.

Eye gaze, which is an example of a visual sensor input channel, is one of the most fundamental ways; in addition to providing visual information about a certain location in the environment, staring in a particular direction also communicates to other people that we are interested in that particular area, which can be picked up by users in the observer's immediate physical area (Moniri et al., 2016).

Eye gaze happens automatically and signals shared attention, indicating that both parties are speaking about the same topic, and it also recognizes the conversation partner (Bartneck et al., 2020). Compared with pointing, body posture, and other non-verbal cues, the gaze is a particularly significant non-verbal cue as research in psychology suggests that eyes are cognitively exceptional stimuli with certain "hard-wired" neural pathways in the brain devoted to their interpretation (Emery, 2000). Another study adding to the growing evidence of the particular importance of eye gaze included young children (24-35 months old). When presented with videos of a robot attempting to manipulate an object and failing, children reacted differently regarding whether or not the robot made eye contact with the human adult in the video. Children imitated the robot’s intended actions if it made eye contact (Itakura et al., 2008). This study also suggests that when it comes to improving engagement between robots and humans, it is human-like behavior, not human-like appearance, that matters.

The integration of gaze and eye movements into human-robot interaction may take many different forms. Studies on the impact of eye gaze on HRI cover the disciplines of robotics, virtual agents, artificial intelligence, and psychology. Robots are used as stimuli by certain researchers to understand the limitations of human perception better. Others experiment with aspects of robot behavior and appearance in order to understand better how the robot gaze affects human responses. Others concentrate on the supporting technology needed to create believable social eye contact (Admoni & Scassellati, 2017).

Research has shown that people expect human-like manners from anthropomorphic robots (Fong et al., 2003). Hence, in order for a humanoid robot to provide comfortable HRI with a human partner, it must participate in interactions with human-like behavior. Therefore, conceiving communication strategies in human-human interaction (HHI) as a modeling framework, such as studying human conversation patterns like back-and-forth regulations, plays a crucial role in achieving the desired outcome.

A study on a humanoid robot that participates in conversational, collaborative interactions with engagement gestures has reported that people become more engaged with robots when engagement gestures are present; they direct their attention to the robot interlocutor more often. In addition, they find the interactions more appropriate than when gestures are absent (Sidner et al., 2005).

A classification of speech-related hand gestures, which categorizes gestures into four groups as deictic, symbolic, iconic, and pantomimic, is provided (Rimé & Schiaratura, 1991). Deictic gestures are pointing gestures that are used to deliver spatial information or refer to an object. Symbolic gestures have a certain cultural meaning and thus have limited use in HRI. Iconic gestures are conventional symbols that have specific meanings, commonly used in HRI. Lastly, pantomimic gestures mimic a desired action or behavior (Burke & Lasenby, 2015). Gesture recognition, which can be defined as the mathematical interpretation of human motions by a computing device, and gesture production algorithms are utilized to provide more effective and natural HRI and HRC (Liu & Wang, 2021).

The research and technologies introduced in this chapter indicate that humans view an interaction with an artificial partner as social when the artificial agent is capable of expressing intentions and states through behavioral signals typical in HHI. To control for this proposition and observe its effects, in the present study, we implemented our artificial agent -a robot- with gaze behavior modeling and pre-planned gestures to give non-verbal cues to the participants.

2.5. Eye-Tracking in Human-Robot Interaction

Eye tracking technology allows a computer to track and record eye movements in real time with high precision. From eye movements, where a subject is looking, how their gaze moves and pauses can be determined. Findings of this technology are used in various fields, from medical and psychological research to marketing and product design. Utilizing eye-tracking methodology to assess gaze patterns can provide insights not only for elucidating information processing and cognition but also for creating modeling frameworks for communication strategies.

In robotics, eye-tracking is used for the teleoperation of robots with gaze (Minamoto, Suzuki, et al., 2017). Also, it can also be used by people with degenerative neuromuscular diseases or neurological or developmental disorders to improve access in daily life (Pasqualotto et al., 2015). For example, a study used an eye-tracking interface for artistic purposes for people with movement disabilities (Scalera et al., 2021). The results of another study showed that adolescents and young adults with autistic spectrum disorder exhibit less diminished eye contact with robots than humans, suggesting that they might prefer social robots in experimental settings (Damm et al., 2013).

In the context of HRI, eye-tracking technology can be utilized for building models for natural gaze patterns to better the interaction. To enable robots to interact with human partners more naturally, reading eye gaze becomes a crucial element. A study showed that in a collaborative building task with a humanoid robot with eye gaze tracking abilities and a human partner, compared to head-tracking methods, eye tracking provides a more efficient HRC (Palinko et al., 2016). When a robot's interactional gaze behavior is close to that of humans, human gaze patterns typical for HHI can be observed. A study showed that inappropriate robot gaze behavior leads to slower response times and disruptions in the usual distribution of fixations (Staudte & Crocker, 2008). According to another study, participants performed significantly better in remembering the story told by ASIMO when the robot looked at them more (Mutlu et al., 2006).

A key aspect of human communication is smooth management of conversational floor or also known as turn-taking. The findings of a study with the robot Nao implemented with purposeful gaze aversions specific to conversational functions found that gaze aversions can be utilized to appear more thoughtful and more effectively manage turn-taking (Andrist et al., 2014). In another study, the robot used human-like gaze patterns to signal conversational roles to its human partners, and the results showed that participants conformed to these signaled roles 97% of the time; also, their conversational roles affected their impressions of the robot, feelings of groupness and attention to the task (Mutlu et al., 2012). The findings of these studies also add to the growing evidence that there are many commonalities between human-human communication and human-robot communication; thus, HRI necessitates embodying manners similar to HHI. For our study, we utilized eye-tracking technology to explore if people make a similar kind of gaze contact with an artificial agent with a gaze behavior model.

2.6. Human-Robot Interaction in Virtual Reality

VR technology can be utilized as a particularly convenient tool to simulate HRI as it is less resource-consuming, such as time and cost, less risky, and provide more spatial freedom compared to a real environment. Furthermore, it allows for a convenient iteration of software and hardware designs at lower cost and effort in a shorter amount of time.

In comparison with on-screen conditions, through two user studies, VR improves performance under certain task conditions (Liu et al., 2017). However, since VR has technical limitations, such as the comfort of the hardware, limited field of view, and display resolution, the degree of immersiveness might affect the perception of and reaction to the robot's presence when compared to real life. A study investigating whether running user studies in VR yields realistic results through a social robot and a secret-keeping scenario produced similar findings (Wijnen et al., 2020). Another study with the robot Pepper showed that proxemic preferences are not consistent between live and VR conditions such that people will allow a robot to get closer to them in a real environment than in a virtual environment (Li et al., 2019). On the contrary, a complex interaction task study revealed very similar results between VR and real environment conditions (Villani et al., 2018). A manipulation experiment performed in Cave Automatic Virtual Environment (CAVE) results suggest that fully immersive VR might be a good alternative to classical HRI testing (Duguleana et al., 2011). Likewise, our experiment was conducted in a VR environment with a robot avatar as the artificial agent.

In this chapter, the literature and the concerning technological advancements for the present study were briefly introduced to the reader. In the next chapter, the methodology employed was described, and the underlying motives were stated.

CHAPTER 3

METHODOLOGY

This study is conducted to investigate the effect of gaze behavior modeling on an avatar robot in a virtual reality environment in the context of Human-Robot Interaction. There are three main questions sought to be answered:

1. Does gaze behavior modeling on an artificial agent as a robot influence subtle gaze signals of the human interlocutor in an active conversation?

2. Do traditional HRI questionnaires reflect the observed gaze patterns of human interlocutors?

3. Does linguistic analysis of open-ended post-experimental evaluations reflect the gaze behavior of the human interlocutors?

In order to elucidate the study's research questions, an investigation of participants' gaze behavior was employed by establishing a one-on-one interview setting replicated from the second experiment in Acartürk et al. (2021). For this purpose, a mock-up job interview scenario was adopted. This mock-up job interview scenario was applied for two reasons. Firstly, we wanted to fix a common goal for all the participants. Secondly, we wanted to reduce the effect of the conversation context on gaze behavior as much as possible. Since job interviews are generally quite similar to each other and the roles of the interlocutors are fixed beforehand, they are fit for an active conversation. Robot avatar, Pepper, took the role of a human resource manager, hence the interviewer. Pepper was implemented with a pre-recorded speech and dependent, pre-planned gaze behavior. The interviewee, the human interlocutor, wore Head Mounted Display (HMD) device, which doubles as VR glasses and an eye-tracker.

To investigate the gaze behavior of the participants, a VR environment, presented in Figure 1, was created by Gizem Özen for the 117E021 coded TUBITAK project via Unity Engine (5.6.7f1). The VR environment is comprised of minimalist designed office space and the avatar agent, humanoid robot Pepper. The room is designed to keep the visually salient objects in the environment to a minimum with no shadows, white walls, a grey desk, wood floors, and a grey curtain located behind the participants. We wanted to keep the visually salient objects in the environment to a minimum of no shadows, white walls, a grey desk, and wooden-textured floors. This minimalist space design was created with the aim of eliminating visual distractions.

A between-subjects design method was applied for the study. Three different aversion strategies for Pepper comprised the conditions, which were used as a between-subjects factor. Further details of the three distinct aversion strategies are given in 3.1.3. Experiments were conducted in a VR environment. During the Virtual Reality (VR) part of the sessions, audio recordings and eye-tracking data of the participants were obtained. Participants also filled out questionnaires before and after the experiment and joined experiment evaluation interviews, audio of which were also recorded. The language used throughout the experiment was Turkish.

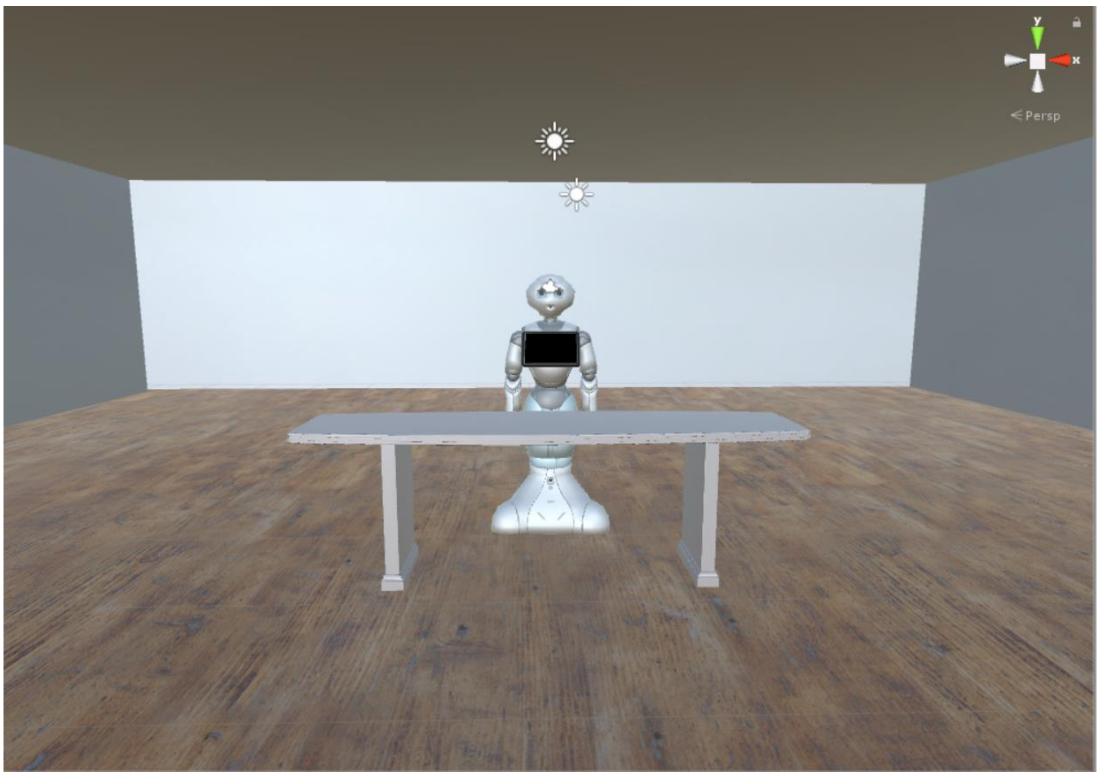

Figure 1 Virtual Reality environment in Unity Editor

3.1. Material and Experimental Procedure

Preceding the experimental procedure, participants signed the consent form (see Appendix A). Then they were informed about the study and presented with screenshots from the VR environment. The experimenter gave instructions to the participants, which are further detailed in 3.1.2. Following the instructions, participants were asked to fill out a demographic survey (see Appendix B) and three pre-experiment questionnaires: a TIPI (Ten Item Personality Inventory) test (Gosling, Rentfrow & Swannl, 2003), a NARS (Negative Attitudes and Anxiety toward Robots) test (Nomura, Kanda, Suzuki & Kato, 2004), and a RAS (Robot Anxiety Scale) test (Nomura, Suzuki, Kanda & Kato, 2006). The questionnairers can be found in Appendix C, D and E respectively.

The questionnaires were used to account for any potential personality impacts of the participants, to control for any potential pre-existing negative attitudes toward robots, and to determine whether the participants were anxious in the presence of a robot. The NARS test was also carried out after the interview sessions in VR to account for potential shifts in participants' attitudes toward robots. Details of the questionnaires can be found in 3.1.1.

After filling out the pre-experiment questionnaires, the environment in Figure 2 was introduced to participants through an HMD device. For each experiment condition, participants took three VR sessions. The first scene of all VR sessions was always the four-point eye-tracker calibration. Following the calibration, participants were asked to adjust the resolution by scrolling the wheel on top of the HMD device. Prior to the conversational interaction with the avatar agent, the experimenter made sure to allow some time for the participants to get used to the environment and feel comfortable by taking a look.

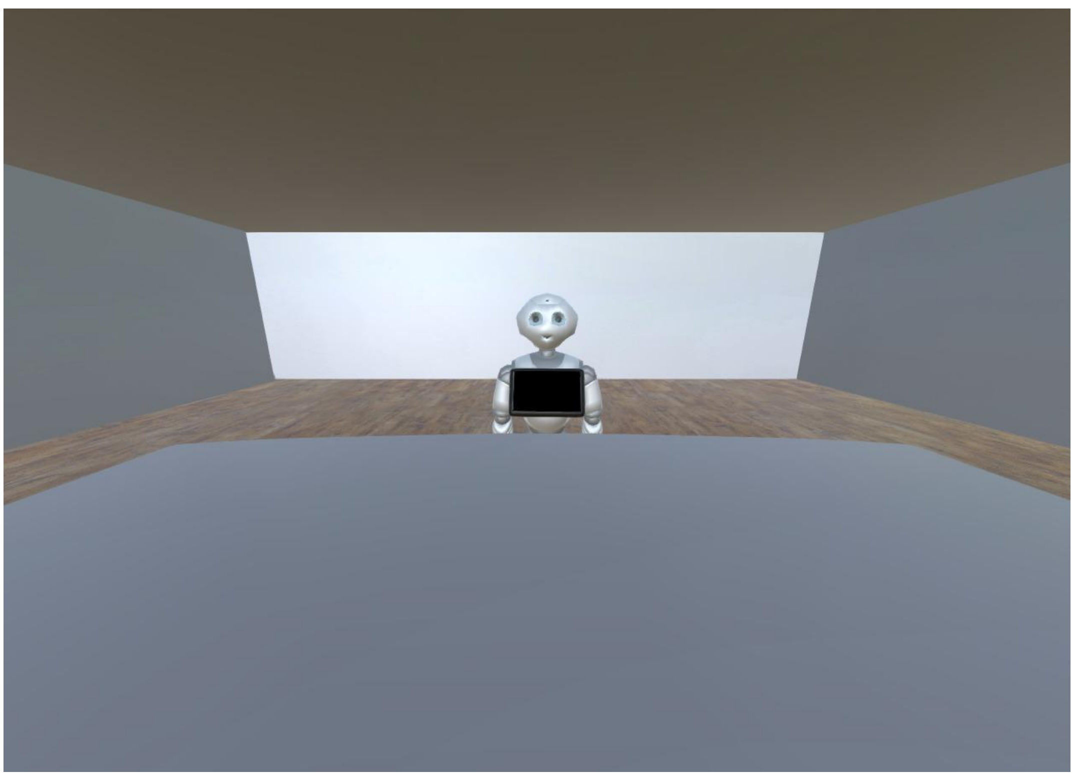

Figure 2 First screen of the experiment and robot avatar Pepper

For all conditions, the experiment commenced with the demo part, in which the artificial agent, the humanoid robot Pepper, started with a greeting and advanced by giving instructions and emphasizing the key points. The demo part continued with Pepper asking questions. The Wizard of Oz (WoZ) technique was utilized in the experiment flow. The term WoZ refers to someone (often the experimenter or a confederate) directing a robot remotely (Riek, 2012). It can include controlling the robot's movement, navigation, voice, gestures, etc. In our case, it was turn-taking; the experimenter controlled the interview flow via a video game controller without the knowledge of the human interlocutors. The motive behind adopting this technique was threefold. Firstly, it prevented participants from pushing a wrong button and disturbing the course of the experiment. Secondly, it provided a more immersive experience by eliminating a possible distraction and contributed to the smoothness and naturalness of the interaction. Thirdly, it allowed participants to think about the artificial agent as more autonomous than he actually was. The avatar agent was not capable of speech recognition; however, this strategy gave the impression that he might have. In managing the conversational floor, the experimenter followed the rules below to decide the timing of proceeding with the following question of the interview:

If 30 seconds have passed after Pepper finishes the question, then pass to the next question even if the participant continues their utterance.
If verbal and paralinguistic signals indicate an end for the reply, wait for 2 seconds and proceed with the next question.

The job interview consisted of 10 questions in total, the first two of them were demographic-type inquiries. The rest of the questions were divided into two sets, and the presentation order of them was adjusted interchangeably to apply counterbalancing. Eight common job interview questions are adapted from (Villani, Repetto, Cipresso & Riva (2012), and the translations are collected from Acartürk et al., (2021). Interview questions are listed in Table 1, and Turkish version and all utterances of Pepper can be found in (Appendix F).

While giving his utterances, Pepper also performed some pre-planned gestures. These included looking straight up and down, left and right, right up, hand waving, nodding for agreement, hands held in the air, and hands in the belly. A screenshot from the animation controller in Unity Engine can be seen in Figure 3.

Table 1 Mock-Up Job Interview questions asked by Pepper

The order for the odd-numbered participants	The order for the even-numbered participants	Question
1	1	Can you tell us a little about yourself?
2	2	How did you find out about this position we propose to you?
3	7	Let us start talking about you. What are your expectations and goals? What do you like doing?
4	8	Why are you interested in this position that we propose?
5	9	Why do you think you could be suitable for this position?
6	10	Where do you see yourself in 5 years time?
7	3	Let us talk about your personality. What are three adjectives that better describe you?
8	4	What do you think of your leadership skills? What type of leader would you be?
9	5	Can you describe an experience you have had that highlights your leadership qualities?
10	6	How do you feel about the possibility of being transferred

Following the job interview session, participants were taken to a different room and the NARS questionnaire was repeated. Then, the experimenter had a post-experiment evaluation session with participants. During the interview, audio recordings were taken with the participant's permission, which were then transcribed into text. Post-experiment participant evaluation interviews are further detailed in 3.1.4.

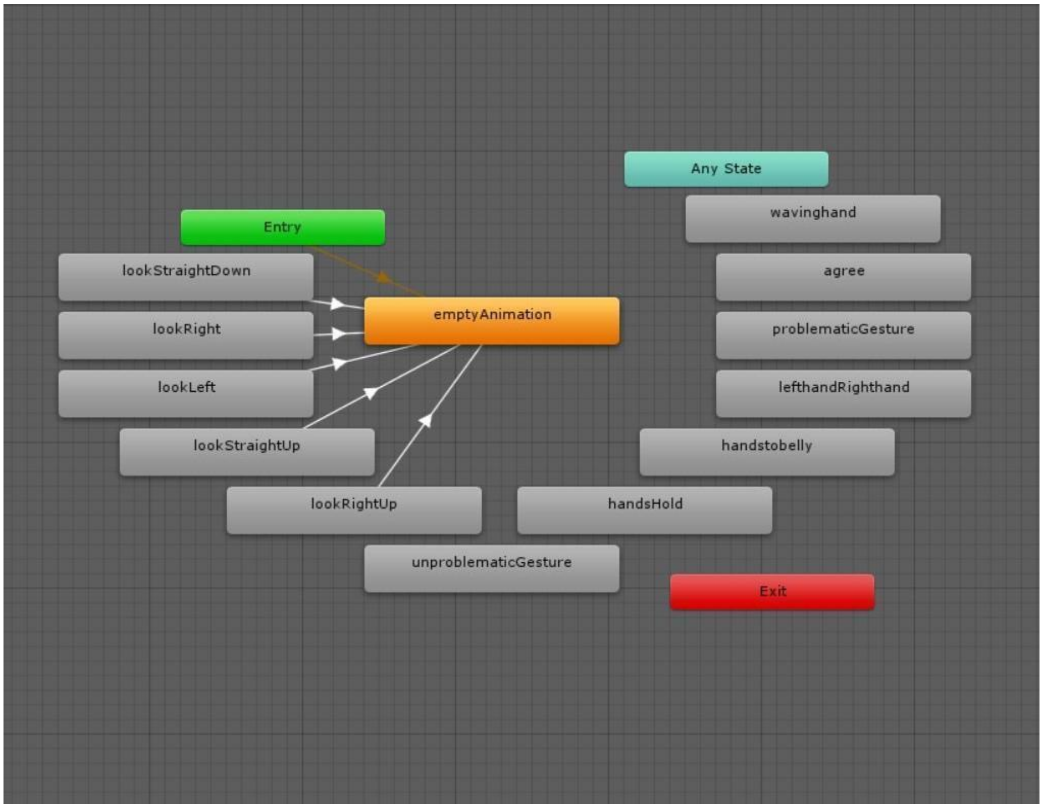

Figure 3 A screenshot from the animation controller in Unity Editor

3.1.1. Questionnaires

Prior to the VR sessions of the experiment, participants were given three pre-experiment questionnaires: A TIPI test (Gosling et al., 2003), a NARS test (Nomura et al., 2004), and a RAS test (Nomura et al., 2006).

TIPI is a brief assessment of the Big Five personality dimensions: (1) Extraversion, (2) Agreeableness, (3) Conscientiousness, (4) Emotional Stability, and (5) Openness to Experience. The test items are graded by the test-taker on a 7-point Likert scale; 1, disagree strongly, to 7, agree strongly. The Turkish translation by Atak (2013), which the creator of the original questionnaire confirms, was used in the study, see Appendix X. This test was utilized in our experiment to monitor any potential effects of the participants' personalities on the experiment.

The NARS test aims to scale the negative attitude towards robots in a group of subordinate scales: (1) Negative Attitude toward Situations of Interaction with Robots consisting of 6 items, (2) Negative Attitude toward Social Influence of Robots, 5 items, (3) Negative Attitude toward Emotions in Interaction with Robots, 3 items. The questionnaire consists of 14 items in total, with one of them being inversed. NARS utilizes a 5-point Likert scale: From 1, strongly disagree to 5, strongly agree. This test was employed to explore whether or not participants had any foregoing negative demeanors against robots. The test, with its translation, is presented in Appendix X. NARS was conducted twice, preceding and following the job interview session of the experiment, to account for the potential change of attitude participants had toward robots.

The goal of the RAS test is to scale the anxiety that prevents individuals from interacting with communicative robots, particularly in a dyadic form between a human and a robot. Like the NARS test, RAS is also sub-categorized: (1) Anxiety toward Communication Capability of Robots, with 4 items, (2) Anxiety toward Behavioral Characteristics of Robots, 4 items, and (3) Anxiety toward Discourse with Robots, 3 items. The test consists of 11 items in total and uses a 6-point Likert scale: From 1, "I do not feel anxiety at all" to 6, "I feel anxiety very strongly". We applied this questionnaire to detect if participants were anxious in the prsence of a robot.

3.1.2. Instructions and Training

The participants were informed that they would attend a job interview, in which they had applied for a position in the next manned Mars mission for a role specific to their own interests. The participants were introduced to Pepper as the Human Resources expert and shown images of the robot. Screenshots from the calibration screen and the experiment were also presented to familiarize the participants. A reminder to leave at any stage of the experiment if needed or wanted was also made.

For all conditions, the experiment commenced with the demo part so as to train the participants. In the demo, the humanoid robot Pepper starts with a greeting and advances by giving instructions and emphasizing the key points. The demo part continues with Pepper asking demographic questions about the interviewees. The avatar agent ends the demo by thanking the participants; see Appendix X for all demo utterances Pepper gave.

3.1.3. Conditions of the Experiment

The present study employed a between-subjects design, in which the aversion strategies of the robot avatar differ. Avatar agent Pepper was programmed to utilize one of the three different gaze conditions only for each participant.

The first condition, the model aversion condition, required the robot to control his gaze as a function of the eye movements of the human partner. Online data (i.e., the gaze data of the participant) sent from the eye-tracker was fed to the Android device mounted in HMD, which then executes Unity and eye-tracker libraries and controls the avatar with model input. Model executes the function with respect to the computational framework developed by Acartürk et al. (2021), in which data collection and analysis of human-human experiments for the same set of questions were conducted to create a probabilistic model for determining parameters such as time and location of gaze behavior using Hidden Markov Models (HMM). The first condition was the only one; in that Pepper did not perform any gaze misbehavior.

The second condition, the constant aversion condition, was an equal distribution for Pepper to avert its gaze at random intervals to one of the eight locations; in that, leftwards, rightwards, upwards, downwards, or a diagonal aversion were each given a 20% weight in randomization. This constant aversion condition functioned as a control case for the first condition (i.e., model aversion based on the framework derived from human interviewers' gaze patterns).

The third condition, the continuous gaze contact condition, used the strategy of the Pepper robot continuously fixating on the participant.

3.1.4. Post-Experimental Participant Evaluations

Posterior the VR sessions and second NARS questionnaires, the experimenter had an open-ended oral interview with the participants. Audio recordings of these interviews were taken with the participants' permission. These recordings were stored as a separate document for each participant in the folders their job interview recordings were stored. All participants were asked the questions in Table 2 in the respective order. When participants provided answers to all the questions, the experimenter thanked them and stopped the recording. Since it is an open-ended conversation, participants interfered, and the wording of the questions sometimes changed, yet the semantic contents and the order in which they were presented remained the same. We conducted these oral evaluations because the participants who joined the pilot sessions of the study had shown great eagerness to give input about the job interview they had with the avatar agent Pepper. As a consequence, we felt the need to structure and record these engagements to establish a better understanding of the participants' HRI experience.

Table 2 Post-experimental evaluation questions asked by the experimenter

The Order	Question
1	How did you feel, in general, during the course of the conversation?
2	Have you felt any discomfort?
3	Have you felt like your replies were being comprehended?

3.2. Participants

32 participants, most of them from the Middle East Technical University, participated in the experiment for monetary compensation of 50 Turkish liras. One of the participant's data was removed due to technical reasons. The ages of the participants varied between 18 and 33 (M=23.74, SD=3.04). There were 18 men, and 13 women participants. 21 of the participants were college students, 7 were graduates from college, and the remaining 3 had a master's degree.

The participants were divided into three groups, and each group was presented with one of the conditions. In the end, 12 participants participated in the model aversion condition, 10 participated in the constant aversion, and 9 in the continuous gaze contact condition. other two conditions (i.e. constan aversion and continuous gaze contact). All participants were native Turkish speakers, and the language used throughout the experiment was also Turkish. 17 of the participants stated that they had not had an experience with VR previously, while the rest of them said otherwise. All of the participants had normal, or corrected-to-normal vision.

3.3. Technical Specifications of the Experiment Environment

For VR sessions and the recording of eye gaze, the Senso-Motoric Instruments (SMI) mobile eye tracking HMD device is utilized (Figure 4a and 4b). The HMD device is built on the Samsung Gear VR platform.The HMD is used with The HMD Samsung Galaxy S7 Android smartphone with 2560 by 1440 resolution on an AMOLED diamond pen-tile display. The software of this Android device was modified by SMI to accommodate the requirement for APIs to access to the backend and operating system level. In order to ensure that the experiment would operate for extended periods of time without the HMD device encountering overheating and performance concerns, the Android device was further modified for greater cooling capabilities. Thus, further problems with data integrity were also prevented.

SMI HMD device provides a 96° field of view vision in the VR environment and has a capacity of 60 Hz refresh rate, binocular eye tracking for gaze direction, and an inter-pupil distance measurement with 0.5 accuracy. VR environment for the experiment was built with Unity Engine, version 5.6.7f1 (licensed for non-profit, academic use), and Microsoft Visual Studio Community 2019.

Figure 4a Device front: Reflectors and cameras for eye-tracking and phone socket[11a]

Figure 4b Device back: Infrared lights around both lenses and the proximity sensor[21b]

Each group of questions (namely demo, 1, and 2) was developed as an individual application, so each experiment session was comprised of 3 applications total. The recorded experiment data was stored automatically on the local storage of the Android device and kept in separate locations in the file system.

[1] 1a Excerpted from Yılmaz, 2018

[2] 1b Excerpted from Yılmaz, 2018

In order for the execution of the model condition, the Python server and Android device communicate through Eclipse Mosquittto (MQQT broker) using system call and system queue strategies. A router is used for this communication. The server receives the outputs of the human interlocutors' eye gaze as inputs and sends outputs to the smartphone. The Android device then receives the inputs and executes Unity and eye-tracker libraries, and controls the avatar with model input. This is reflected to the participants in the form of gaze behavior adjusted to theirs. For the experimenter to control the turn-taking floor management, a gamepad is utilized; see Figure 5 for devices used in the set-up. The experimental environment was set by Efecan Yılmaz for the TUBITAK project coded 117E021.

Figure 5 The devices used in experiment set-up: HMD device, gamepad, Android smartphone and the casing

The oral answers of the participants both in the VR interview sessions and post-experiment evaluations were recorded via an Android smartphone (Samsung A02S) and stored in the local drive of the device.

The collected data are further described in 3.4.

3.4. The Description of the Collected Data

The left column in Table 3 lists the variables of the data acquired from the participants. The right column comprises the description of the variable on left.

Table 3 List of the data variables retrieved

Variable name	Variable details
Gaze sample #	The number indicating the concerning sample
Gazed object	For each gaze sample, the object in the virtual reality environment on which the participant's gaze fixates, it's one of 1,2,3, Ceiling, Front, Head, Table, Left, Right
Binocular Position X Y	Vectors of the eyes
PositionsValid	It's either True or False
InterOcular Distance	The distance between the eyes
InterPupil Distance	The distance between the pupils
DistancesValid	It's either True or False
Avatar Aversion:	0 for constant aversion and continuous aversions from Pepper, has no value otherwise

3.5. Analysis Procedure

The files that contained data obtained on the local storage of the Android device which worked with the HMD device were collected at the end of each experiment day and stored together with each participant's job interview and post-experiment evaluation recordings. The audio files of the evaluations were listened by the experimenter and transcribed into text. The audio recordings of the job interviews were stored only as a backup in case there are any problems in the synchronization of the eye and audio data for the interview questions, and speech segmentation is required.

One participant's data had to be removed from the analysis due to technical problems with data storage.

The controlled nature of the virtual reality environment provides a perfect environment for eye-tracking experiments. Contrary to the experiments conducted with traditional wearable eye-tracking devices, matching the dynamic movements resulting from the participant's environmental interaction with visual interests are realized directly inside the objects called collision boxes while coding the virtual environment, thus eliminating the need for manual annotation. These areas of visual interest can be exemplified with environmental objects, such as curtains, table, and ceiling, or the parts of the avatar such as the head and torso. These collider-type objects work with a target market that will collide with them, which is not visible to the participant. Through this process, the point the participant gazed at in the virtual environment can be recorded both as a coordinate and directly as an object. Data analysis can be almost completely automated by syncing these data with markers such as experiment blocks or stimuli.

The SMI Gear VR Eye Tracker was developed by Sensomotoric Instruments by modifying a Samsung Gear VR first generation, Google Android OS-based device and retrofitting a reflector-type eye tracker. The device allows eye tracking by illuminating the eye using a reflector and infrared light placed in front of the participant's view of the virtual reality device, and by logging into the virtual reality environment after a 4-point calibration process of the black pupil detection algorithm working with it. The data recorded with the SMI device with time, experiment, and object markers are then analyzed by defining the shortest eye fixation time parameter in time, and answering the question of how long to look at an object at least defines an eye fixation. Since the movement of the virtual reality world viewed on the VR device and the camera movement are moved on the same axis, only the shortest eye fixation parameter is sufficient for fixation detection.

For the eye-tracking data, we used factory settings for the SMI HMD device. The .csv files retrieved from the Android device were transferred to a spreadsheet in terms of the gazed object (1, 2, 3, Ceiling, Front, Head, Table, Top, Left, and Right) and vectors. Figure 6 shows the concerning objects in the VR environment. Then, for each participant, a spreadsheet for further analysis was created, which shows how many times a participant gazed upon a certain object, and the percentage of it compared to the total eye gazes for that participant.

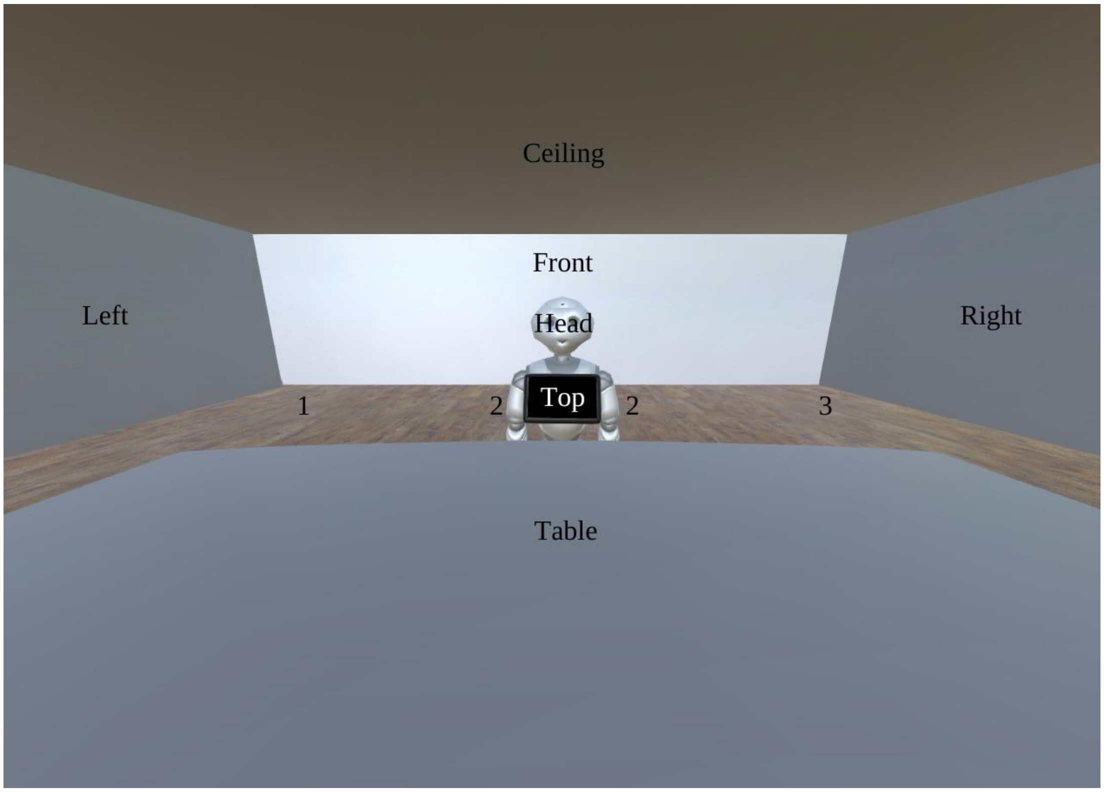

Figure 6 Marked locations of the gazed object categories

For the eye gaze data, several analyses were done via IBM SPSS Statistics Data Editor (v.28). Test of Normality (Shapiro-Wilk) was performed to check for the normality of the data distribution. Compare Means for comparing the percentages between different aversion groups.

Independent t-tests were performed to investigate the possible personality effects or attitudes toward robots; also to take into the potential changes of attitude by participants through the course of the experiment. ANOVA tests were conducted for detecting the between group distributions of questionnaires TIPI, RAS, and NARS. In addition, descriptive details of all participants were retrieved for each questionnaire. Since NARS was employed prior to and posterior to the job interviews for each participant, these were analyzed as paired samples results for exploring if the participants' attitudes towards robots had any significant changes. The results of the analyses were discussed in the next chapter.

Lastly, the collected audio of post-experimental evaluations was transcribed into text manually. Afterward, the utterances produced by the experimenter were deducted, and the remaining text data were subcategorized into three distinct experimental conditions for further processing. These data were then analyzed using the Term Frequency-Inverse Document Frequency (TF-IDF) measure. TF-IDF is a statistical measure which is consisted of two parts; Term Frequency (TF) and Inverse Document Frequency (IDF). TF refers to the frequency of a particular term relative to the document. IDF stands for how frequent a word is across a collection of documents. Therefore, this statistical measure explores how relevant a word is for a document in a series of documents. This analysis is performed by multiplication of the following metrics: How many times a word is seen in a document and the inverse document frequency of the concerning word in relation to a set of documents.

The function for TF-IDF is as follows (the word *w* in the documents *i* and *j)*:

$$w_{i,j} = tf_{i,j} \times log\left(\frac{N}{df_i}\right)$$

where

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k \quad n_{i,j}}$$

$$idf(w) = log\left(\frac{N}{df_t}\right)$$

TF-IDF is mainly preferred to represent text-based contents of documents; it eliminates the most common terms and extracts only the most relevant terms from the concerning series of documents (Bafna et al., 2016).

Before TF-IDF analysis, we preprocessed the text data so that it had no punctuation. We also removed stop words to reduce the dimensionality of the input space. Stop words are words that have no contextual meaning and only have a function. These are commonly used words in any language; determiners, auxiliaries, conjunctions, degree adverbs, pronouns, and prepositions. According to Khosrow (2009), stop words have no significant semantic relation to the context in which they exist. We also lemmatized the words manually by removing the inflectional suffixes. In this process, subordinating suffixes were also removed, even though some s them derivations in the literature since they change the category of a word in terms of its parts of speech properties (Aronoff & Fudeman, 2022). In our study, we adopted the approach by Göksel & Kerslake (2004) and treated subordinate suffixes as inflectional. While performing the TF-IDF analysis, a Python code was utilized (Maklin, 2019).

We decided to benefit from the TF-IDF measure in terms of its power in extracting keywords and giving an idea about how important a word or a phrase is in our dataset of participants' post-experiment oral evaluations. The results of the TF-IDF analysis can be found in Chapter 4.

We also made a simple sentiment analysis on the post-experiment oral evaluation data by utilizing a Turkish NLP pipeline pre-built model based on Turkish Bert (Yıldırım, 2020). The model's dataset was taken from movie and product reviews (Demirtas & Pechenizkiy, 2013) and tweets (Hayran & Sert, 2017). Stop words were not removed in the preprocessing step to provide enough context information. The results of the sentiment analysis are presented in Chapter 4.

CHAPTER 4

RESULTS

In this chapter, the analysis results for experiment conditions are presented. Firstly, statistical results of the TIPI, NARS (pre and post), and RAS test scores are reported in 4.1. Secondly, statistical analysis of the gaze aversions of the participants is reported in section 4.2. Finally, the TF-IDF and sentiment analysis results of the post-experiment evaluations are presented in 4.3.

4.1. Questionnaire Results

The ANOVA test results, which were performed to explore if the three experimental conditions showed a homogenous distribution of participants by personality traits, showed no significant difference between the groups, as shown in Table 4. In addition, the descriptive results of the TIPI test are presented in Table 5.

Table 4 Results of ANOVA on TIPI between groups of participants

	df	F	Sig.	η^2
Extraversion	2	1.422	0.258	0.092
Agreeableness	2	1.381	0.268	0.090
Conscientiousness	2	0.063	0.939	0.004
Emotional Stability	2	0.021	0.979	0.001
Openness	2	0.955	0.862	0.433

Table 5 Descriptive results of TIPI tests for all participants

	N	Min.	Max.	Mean	Std. Deviation
Extraversion	31	1.50	7.00	5.306	1.370
Agreeableness	31	2.50	6.50	4.887	1.174
Conscientiousness	31	2.00	7.00	5.177	1.228
Emotional Stability	31	1.00	6.50	3.774	1.425
Openness	31	3.50	7.00	5.726	1.047

The statistical analysis of RAS was employed for between group distribution, similarly with an ANOVA test and results showed that there was not a significant difference between participant groups in regards of anxious feelings toward robots, as shown in Table 6. Descriptive details are also given in Table 7.

Table 6 Results of ANOVA on RAS between groups of participants

	df	F	Sig.	η^2
S1	2	1.060	0.360	0.070
S2	2	3.111	0.060	0.182
S3	2	1.489	0.243	0.096

Table 7 Descriptive results of RAS tests for all participants

	N	Min.	Max.	Mean	Std. Deviation
S1	31	4.00	15.00	8.000	2.380
S2	31	4.00	22.00	12.290	3.504
S3	31	4.00	20.00	9.967	4.029

Lastly, an ANOVA test for participants' distribution in the three experimental conditions for the NARS test resulted, again, in no significant difference between the groups for both of the NARS tests. The concerning numbers are shown in Table 8, and Table 9 for pre and post-NARS tests, respectively. The descriptive test results for both of the NARS tests are shown in Table 10.

Furthermore, paired samples test was applied to pre-experiment and post-experiment NARS tests in order to see whether or not the participants' attitudes toward robots were significantly distinct before and after the VR job interview session with the robot agent. Results did not indicate any significant difference for any of the subsections: S1, $t(30) = 1.209$, $p = 0.236$, $r = 0.550$; S2, $t(30) = 1.123$, $p = 0.270$, $r = 0.737$; S3, $t(30) = 0.256$, $p = 0.800$, $r = 0.425$.

Table 8 Results of ANOVA on pre-experimental NARS test

	df	F	Sig.	η^2
S1	2	0.541	0.588	0.037
S2	2	7.860	0.591	0.037
S3	2	3.129	0.422	0.060

Table 9 Results of ANOVA on post-experimental NARS test

	df	F	Sig.	η^2
S1	2	0.536	0.591	0.037
S2	2	1.527	0.235	0.098
S3	2	1.333	0.280	0.087

Table 10 Descriptives of both NARS tests for all participants

	N	Min.	Max.	Mean	Std. Deviation
Pre-NARS S1	31	7.00	19.00	12.451	3.404
Pre-NARS S2	31	6.00	19.00	13.097	3.771
Pre-NARS S3	31	7.00	15.00	10.323	1.869
Post-NARS S1	31	7.00	22.00	11.742	3.483
Post-NARS S2	31	5.00	19.00	12.548	3.722
Post-NARS S3	31	6.00	14.00	10.226	2.045

4.2. Gaze Interaction Results

To test whether the distribution of our data is normal, the Shapiro-Wilk test is performed on the data, and the results showed a significant non-normality for the constant aversion condition ($W = 0.82$, $p < 0.03$). However, since our sample size is larger than 20 ($N = 31$), a normal distribution for our data can be presupposed.

With the aim of reaching observations about the participants' gaze distribution on the environment, data retrieved from the head-mounted eye-tracker was subcategorized in terms of the gazed objects. Then these subcategories were analyzed statistically. When means are compared for the gazed objects, the gaze allocation represented with numbers can be seen in Table 11. Graphically represented gaze distribution means can also be seen in Figure 7.

The following inferences can be made by the results obtained: For every condition, participants averted their gaze to the head of the avatar agent more than the rest. Top and front came in second and third respectively in all conditions. On the assumption that fixations made on the head indicate gaze contact, the following inferences can be made: In the model condition, where the gaze behavior model controlled Pepper's gaze, participants made more eye contact. They also made more eye contact when Pepper stared at them compared to when he made constant aversions. For more clarity, objects in the environment can be seen in Figure 6.

We also tested if having experienced VR made any difference in the gaze contact. Out of 31 participants, 14 of them expressed that they had used VR before. We performed a two-tailed independent t-test and found that on average, participants who had a previous VR experience made more eye contact (M = .440, SE = .037) compared to those who had not (M = .409, SE = .040). This difference was not significant $t(29) = .569$, $p > .05$; however, it did represent a small-sized effect $r = .105$.

The results are discussed more in the next chapter.

Table 11 Comparison results of the gazed objects between groups

Conditions	**Front**	**Head**	**Top**	**Ceiling**	**Table**	**Tile 1**	**Tile 2**	**Tile 3**
Modeled aversion	.079	.423	.355	.028	.037	.004	.066	.005
Std. Deviation	.052	.157	.156	.040	.061	.007	.041	.008
Constant aversion	.080	.459	.315	.020	.044	.027	.068	.009
Std. Deviation	.073	.124	.104	.027	.075	.048	.049	.009
Continuous gaze contact	.178	.383	.274	.044	.034	.004	.073	.015
Std. Deviation	.136	.184	.190	.112	.056	.007	.060	.023
Total	.108	.423	.319	.032	.038	.011	.069	.009
Std. Deviation	.098	.153	.151	.070	.063	.027	.048	.015

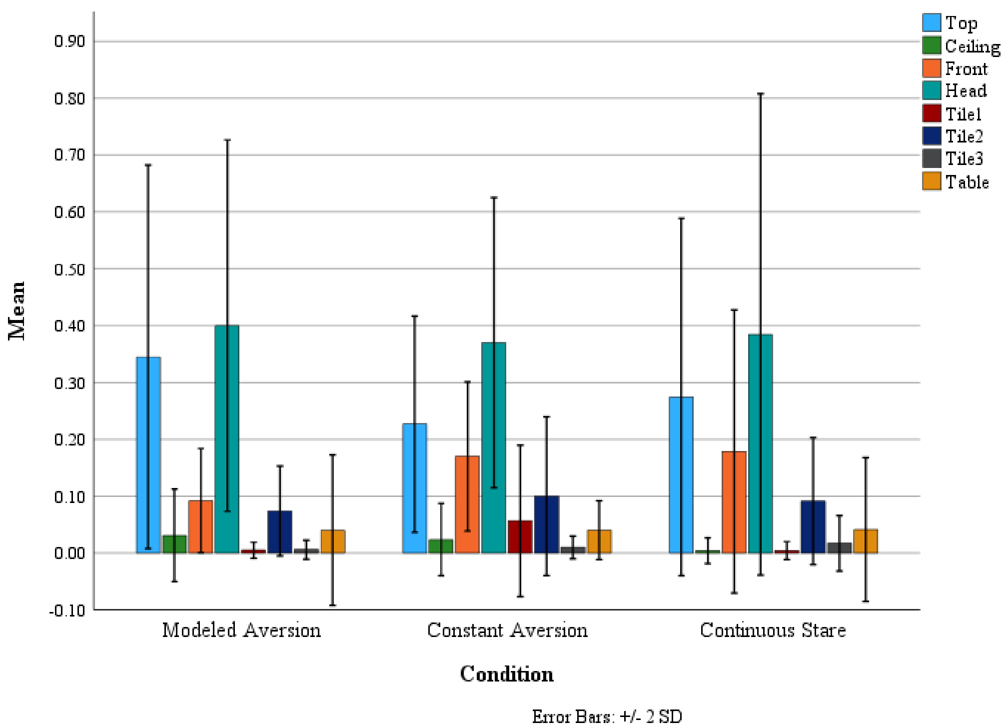

Figure 7 Participants' gaze distribution on the environment regarding different aversion conditions

4.3. TF-IDF Results of the Post-Experimental Evaluations

In order to explore keywords and have an idea about how important a word or a phrase is in our dataset of participants' post-experiment oral evaluations, the statistical method of Term Frequency-Inverse Document Frequency (TF-IDF) is performed.

TF-IDF is a statistical measure which is consisted of two parts; Term Frequency (TF) and Inverse Document Frequency (IDF). TF refers to the frequency of a particular term relative to the document. It is calculated by the number of times a word appears in a document divided by the total number of words in the document. IDF stands for how frequent a word is across a collection of documents. It is calculated by the log of the number of documents divided by the number of documents that contain the word w. Therefore, this statistical measure explores how relevant a word is for a document in a series of documents.

From the results of the TF-IDF analysis, the 25 most used contextual words and their scores for each group are visualized in Figure 8. The words followed by a hyphen indicates that these are verbs.

Apart from TF-IDF analysis, basic word count on the raw transcripts resulted with the numbers 676, 406, and 624 for modeled, constant aversion and continuous stare conditions respectively.

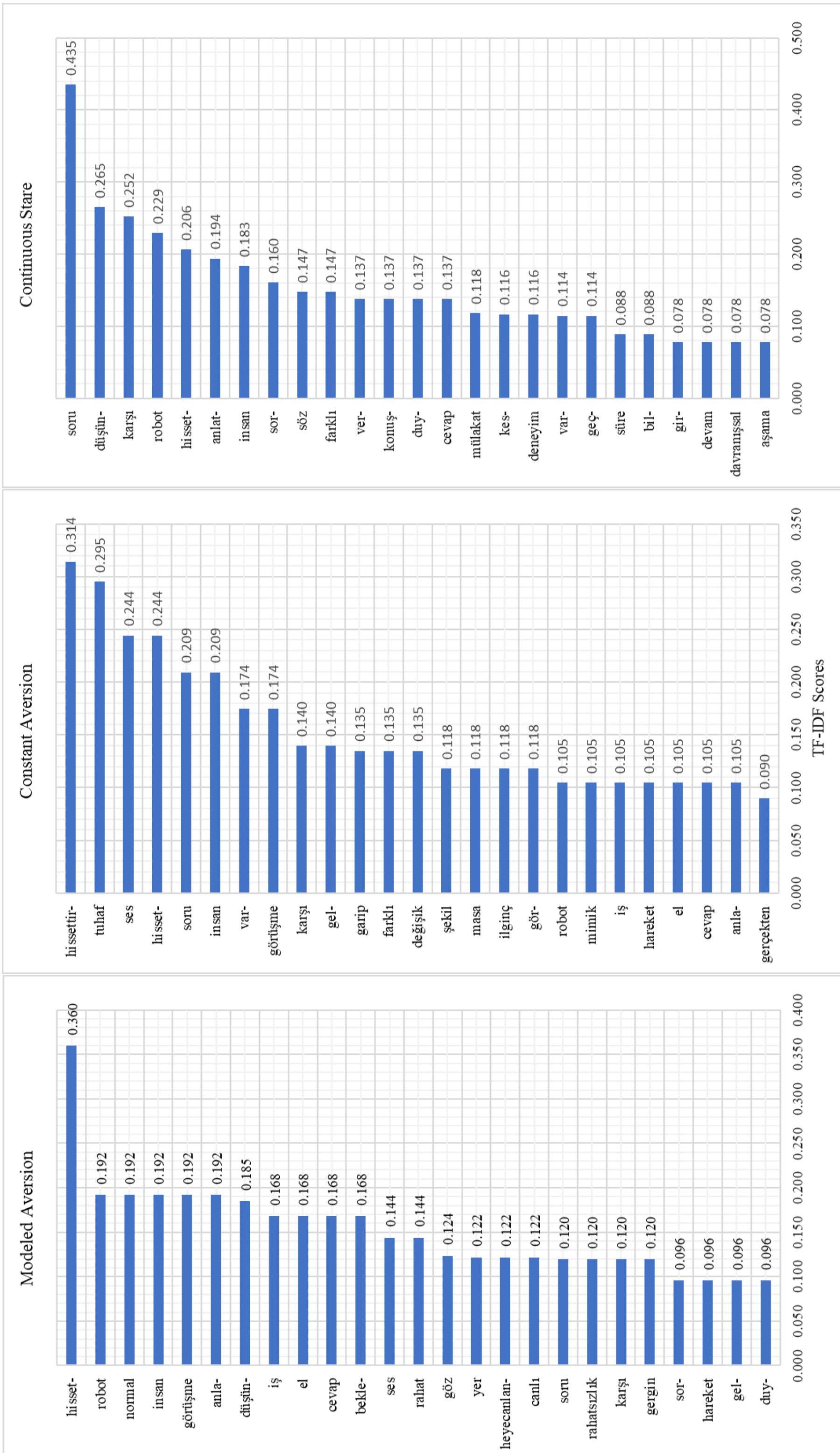

Figure 8 25 Highest Ranked Words of The Evaluations and Their TF-IDF Scores

Results show that the inflections of ‘hisset-‘ (*ing.* feel) were the most prominent for the condition in which robot avatar Pepper averted his gaze as a function of the eye movements of the human partner. ‘hisset-‘ was also fourth and fifth most relevant word for the constant aversion and continuous stare conditions respectively. Similarly, ‘hissettir-‘ (*ing.* make feel, evoke) was the highest-ranked word for the constant aversion condition; however it was not on the list for the other two conditions. ‘soru’ was the highest-ranked for the third condition in which Pepper stared at the participants, the same word was also on the list for the other conditions. ‘insan’(*ing.* human), ‘robot’ (*ing.* robot), ‘cevap’ (*ing.* answer) and ‘karşı’ (*ing.* opposing) were on the list for all conditions. The lemma ‘robot’ was higher on the list, second and fourth for modeled and stare conditions respectively, compared to the constant aversion (eighteenth) condition. An eye-catching finding is that the word ‘normal’ (*ing.* normal) and ‘canlı’(*ing.* alive) were only prominent for the condition in which the avatar did not perform any gaze misbehavior. Concordantly, the words ‘tuhaf’ (*ing.* odd, strange), ‘farklı’ (*ing.* different), ‘değişik’ (*ing.* different), ‘garip’ (*ing.* odd, strange), and ‘ilginç’ (*ing.* interesting) were on the list for these two conditions where there is gaze misbehavior.

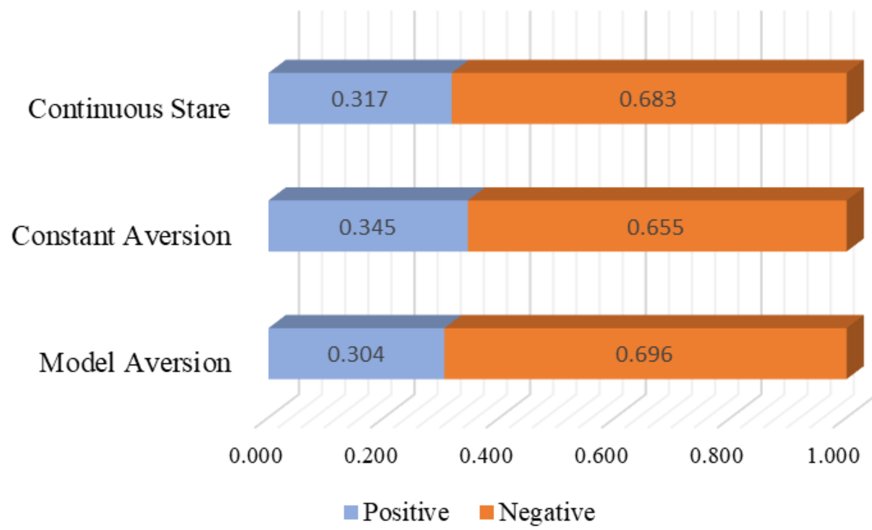

Figure 9 Sentiment Analysis Results by Different Conditions

The results of the sentiment analysis of post-experiment oral evaluations are visualized in the Figure 9 and 10 for comparison. We found that the utterances produced by the participants were mostly negative. Model condition has the most negative while constant aversion has the least.

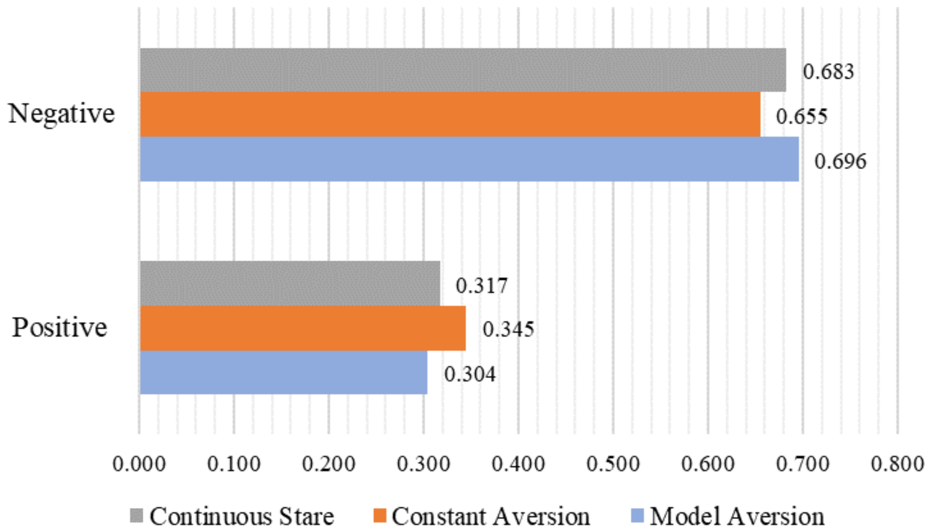

Figure 10 Sentiment Analysis Results by Polarity

The results of the sentiment analysis are discussed more in Chapter 5.

CHAPTER 5

CONCLUSION AND DISCUSSION

Even the most simple human interaction involves various cues from different channels. Neurotypical people execute these so automatically and effortlessly that most of the time no one is aware. However, if these cues are interrupted by any means, the interaction loses its smooth nature and becomes more challenging. In fact, following the Covid-19 pandemic, the term “zoom fatigue” has entered our vocabulary, partly due to the non-verbal overload videoconferencing causes (Bailenson, 2021). Since non-verbal channel comprises such an essential part of human interaction, it constitutes a widely-researched scientific area from various points. In the study at hand, we approached this area through artificial agents.

In the present study, our main research question was whether or not gaze behavior modeling on an artificial agent makes any difference in the eye behavior of human interlocutors. If so, do traditional HRI questionnaires, and open-ended post-experiment evaluations reflect any difference comprised the sub-questions. To explore these questions, we conducted a one-on-one conversation setting in VR with a mock-up job interview scenario. The humanoid robot Pepper as the avatar agent took the role of the Human Resource Manager, hence the interviewer. To control the effect of the gaze behavior model, we constructed two more conditions. In one of them, the artificial agent averted his gaze randomly; in the other, he made continuous gaze contact. The turn-taking approach of the artificial agent was managed by WoZ technique while interviewing the humans. Participants also filled out some of the most common HRI questionnaires and a demographic survey. One of the questionnaires was employed prior to and posterior to the VR job-interview session. Finally, they had an open-ended conversation with the experimenter to evaluate their job interview with the avatar agent. The experimenter recorded the audio of the evaluations for further processing. This chapter states the key findings of the present study and its limitations.

5.1. Key Findings

Statistical results of the traditional HRI questionnaires and demographic surveys showed no difference in terms of distribution across different gaze conditions. The participants had no preexisting anxiety or negative attitude toward robots. NARS test also resulted in no difference in when it was conducted before and after the interview with the avatar agent. Hence, it can be said that participants’ anxiety levels did not differ after their interaction with the avatar agent Pepper.

The statistical analyses of the gaze data showed that the gaze allocation of the participants differed slightly. Participants made more eye contact with the artificial agent when he performed human-like gaze patterns. These findings can indicate more engagement for humans when the artificial agent shows humanlike behavior regarding its gaze. Other than the artificial agent's head, the gaze distribution of the human interlocutors on the environment did not result in any significant difference.

Compared to the results of the same mock-up job scenario with human-human settings , the study in question resulted in a lower percentage of eye contact occurrences. 73% of participants' gaze behavior in the human-human setting comprised eye contact, while it was 42% in the human-avatar setting.

Results obtained from the gaze data were not reflected in the questionnaires. However, linguistical analysis of the open-ended post-experimental participant evaluations resulted in differences between groups. A TF-IDF weighting analysis was performed to investigate the open-ended evaluation conversations held with participants. The model condition differed from the control conditions regarding agency effects reflected in participants' evaluations. Participants who had a job interview with Pepper when it was gaze modeled used words like 'normal' and 'alive'; these words were not prominent for the two control conditions in which the artificial agent Pepper performed gaze misbehavior. Agency attribution can be linked to human-like appearance and human-like behavior (Kamewari et al., 2005). As in our study, only the behavior of the artificial agent was controlled; it can be inferred that human-like gaze behavior may be a leading factor for humans to attribute agency to artificial agents. Concordantly, words referring to oddness and differentness were only absent when Pepper showed human-like gaze behavior.

The negative attitudes of the participants observed in the sentiment analysis can be an indicator of the uncanny valley, which refers to the relation between the human likeness and a viewer's affinity toward it (Mori, 1970). The fact that TF-IDF analysis has also resulted in words referring to oddness and eeriness reaching higher scores can support this view. If so, from the data at hand it is unclear if mostly the behavior or the appearance of the avatar agent caused it.

To conclude, perceptions of humans regarding artificial agents may differ when the agent shows human-like gaze behavior; however, further research is needed. Traditional HRI questionnaires may not always reflect the interaction experience of humans in terms of engagement and the perception of the artificial agent. While questionnaires and surveys provide researchers with valuable insight and constitute one of the most used tools for the field of psychology, they have their limitations. For example, some of them require the respondents to reflect on their beliefs and feelings, which may not be possible to report accurately since humans do not have conscious access to the majority of such information (Kosslyn & Rosenberg, 2014). Sometimes humans are not even aware of the beliefs that motivate their behavior. On the other hand, intentional or unintentional linguistic choices humans make in their utterances might grant access to more reliable inferences along with the overt ones.Therefore, it can be stated that researchers may also utilize linguistic analysis as a measurement tool for more clarity and accuracy in HRI studies.

5.2. Limitations of the Study

Some of the limitations of the study were related to the hardware used. Using a VR eye-tracker instead of a desktop had several advantages and disadvantages. Some advantages are the immersiveness it provides and the easier and more reliable analysis procedure. It also led us to reach more participants since it can be used by people with various vision problems, contrary to desktop eye-trackers, except for those with prosthetic eyes. On the other hand, the physical discomfort of the HMD device was a disadvantage compared to a desktop eye-tracker.

Another hardware-related limitation was the immersiveness of the VR environment we provided. In our study, the HMD VR device we utilized was relatively old in terms of the technology it offers, concerning the fact that it was released in 2015 and the experiments were conducted at the end of 2021. The noncompetent immersiveness, in turn, may have affected how engaging the VR experience was.

The following limitation concerning the hardware was related to HMD device design. Due to its form, it was easy to accidentally push the buttons on the HMD device for someone unfamiliar. In the mock-up job interview part of the experiment, initially, it was planned for participants to control proceeding to the next question by pushing a button on the HMD device when they finished providing an answer for the concerning question in the interview. However, following the mishaps in pilot sessions, we chose to control the turn-taking of the interview with a gamepad. As we had made the experiment sessions beforehand, this change in the experimental flow resulted in the artificial agent giving misleading instructions at the beginning of the VR sessions, which the experimenter corrected. This change may have caused less agency attribution on behalf of the artificial agent. However, this change also led participants to think that the artificial agent was capable of speech recognition even though he was not.

The presence of a third person as a spectator, in our case the experimenter, in the experiment room during the mock-up job interview with the avatar agent may have been a confounding factor in terms of environmental validity. Some studies on social facilitation implied restraint by the presence of spectators rather than facilitation for intellectual or implicit responses of thought (Zajonc, 1965). Further studies of social presence effects can be investigated in the domain of HRI.

The surveys we used, NARS (Negative Attitudes Towards Robots) and RAS (Robot Anxiety Scales), were not developed or validated for VR robots. However, they were employed by some previous studies with VR robots (Holmes et al., 2018; see also Parenti et al., 2021; Babel et al., 2022; Björling et al., 2019).

There was a considerable amount of compound utterances in the data. However, the sentiment analysis model used in our study did not differentiate the neutral sentences. As a result, these utterances may have been wrongly classified as negative or positive by the model. A more detailed sentiment analysis that includes neutral classification could be employed to understand the impressions with the avatar agent better.

5.3. Future Work

In the future, different questionnaires measuring a particular robot after an interaction could be utilized alongside general perception surveys. The Godspeed questionnaire series (Bartneck et al., 2009), which measures human impression of robots using five concepts, such as anthropomorphism, animacy, likeability, intelligence, and safety, could be an example of such questionnaires.

Another future work could introduce a different voice for the avatar agent to make sure whether the effect is caused prominently by the gaze behavior.

Further future work could include different conversational settings. As job interviews tend to cause anxiety for the applicant (McCarthy & Goffin, 2004), other active conversation scenarios could be designated for eliminating the context effect as much as possible.

Combining some paralinguistic analysis, such as intonation, on the linguistic data might provide more insights into the attitude and perception of the participants.

Analyzing TF-IDF results of the post-experimental evaluations with a rank similarity matrix could comprise an opportunity for future work.

Lastly, sentiment analysis for post-experimental evaluations is needed to be investigated again with an improved model.

REFERENCES

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

Gonzalez-Aguirre, J. A., Osorio-Oliveros, R., Rodríguez-Hernández, K. L., Lizárraga-Iturralde, J., Morales Menendez, R., Ramírez-Mendoza, R. A., ... & Lozoya-Santos, J. D. J. (2021). Service robots: Trends and technology. *Applied Sciences*, *11*(22), 10702.

Magnenat-Thalmann, N., Yuan, J., Thalmann, D., & You, B. J. (Eds.). (2016). *Context aware human-robot and human-agent interaction* (pp. 1083-4435). Springer International Publishing.

Bartneck, C., Belpaeme, T., Eyssel, F., Kanda, T., Keijsers, M., & Šabanović, S. (2020). *Human-robot interaction: An introduction*. Cambridge University Press.

Kaminka, G. A. (2013, May). Curing robot autism: A challenge. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems* (pp. 801-804).

Eyssel, F., & Kuchenbrandt, D. (2012). Social categorization of social robots: Anthropomorphism as a function of robot group membership. *British Journal of Social Psychology*, *51*(4), 724-731.

Fink, J. (2012, October). Anthropomorphism and human likeness in the design of robots and human-robot interaction. In *International conference on social robotics* (pp. 199-208). Springer, Berlin, Heidelberg.

Daily, S. B., James, M. T., Cherry, D., Porter III, J. J., Darnell, S. S., Isaac, J., & Roy, T. (2017). Affective computing: historical foundations, current applications, and future trends. *Emotions and affect in human factors and human-computer interaction*, 213-231.

Gasparetto, A., & Scalera, L. (2019). From the Unimate to the Delta robot: the early decades of Industrial Robotics. In *Explorations in the History and Heritage of Machines and Mechanisms* (pp. 284-295). Springer, Cham.

Breazeal, C. (2003). Toward sociable robots. *Robotics and autonomous systems*, *42*(3-4), 167-175.

Shamsuddin, S., Ismail, L. I., Yussof, H., Zahari, N. I., Bahari, S., Hashim, H., & Jaffar, A. (2011, November). Humanoid robot NAO: Review of control and motion exploration. In *2011 IEEE international conference on Control System, Computing and Engineering* (pp. 511-516). IEEE.

Fitzgerald, C. (2013, April). Developing baxter. In *2013 IEEE Conference on Technologies for Practical Robot Applications (TePRA)* (pp. 1-6). IEEE.

Nikolaidis, S., Kwon, M., Forlizzi, J., & Srinivasa, S. (2018). Planning with verbal communication for human-robot collaboration. *ACM Transactions on Human-Robot Interaction (THRI)*, *7*(3), 1-21.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & biobehavioral reviews*, *24*(6), 581-604.

Admoni, H., & Scassellati, B. (2017). Social eye gaze in human-robot interaction: a review. *Journal of Human-Robot Interaction, 6*(1), 25-63.

Pandey, A. K., & Gelin, R. (2018). A mass-produced sociable humanoid robot: Pepper: The first machine of its kind. *IEEE Robotics & Automation Magazine*, *25*(3), 40-48.

Perzanowski, D., Schultz, A. C., Adams, W., Marsh, E., & Bugajska, M. (2001). Building a multimodal human-robot interface. *IEEE intelligent systems*, *16*(1), 16-21.

Goodrich, M. A., & Schultz, A. C. (2008). Human–robot interaction: a survey. *Foundations and Trends® in Human–Computer Interaction*, *1*(3), 203-275.

Kidd, C. D., & Breazeal, C. (2004, September). Effect of a robot on user perceptions. In 2004 IEEE/RSJ international conference on intelligent robots and systems (IROS)(IEEE Cat. No. 04CH37566) (Vol. 4, pp. 3559-3564). IEEE.

Lee, K. M., Jung, Y., Kim, J., & Kim, S. R. (2006). Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction. *International journal of human-computer studies*, *64*(10), 962-973.

Chang, C. W., Lee, J. H., Chao, P. Y., Wang, C. Y., & Chen, G. D. (2010). Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. *Journal of Educational Technology & Society*, *13*(2), 13-24.

Rossi, A., Holthaus, P., Dautenhahn, K., Koay, K. L., & Walters, M. L. (2018, December). Getting to know Pepper: Effects of people's awareness of a robot's capabilities on their trust in the robot. In *Proceedings of the 6th international conference on human-agent interaction* (pp. 246-252).

Aaltonen, I., Arvola, A., Heikkilä, P., & Lammi, H. (2017, March). Hello Pepper, may I tickle you? Children's and adults' responses to an entertainment robot at a shopping mall. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 53-54).

Mavridis, N. (2015). A review of verbal and non-verbal human–robot interactive communication. *Robotics and Autonomous Systems*, *63*, 22-35.

Bunt, H., Beun, R. J., & Borghuis, T. (Eds.). (1998). *Multimodal human-computer communication: systems, techniques, and experiments* (Vol. 1374). Springer Science & Business Media.

Quek, F., McNeill, D., Bryll, R., Duncan, S., Ma, X. F., Kirbas, C., ... & Ansari, R. (2002). Multimodal human discourse: gesture and speech. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *9*(3), 171-193.

Turk, M. (2014). Multimodal interaction: A review. Pattern recognition letters, 36, 189-195.

Oviatt, S., Cohen, P., Wu, L., Duncan, L., Suhm, B., Bers, J., ... & Ferro, D. (2000). Designing the user interface for multimodal speech and pen-based gesture applications: State-of-the-art systems and future research directions. Human-computer interaction, 15(4), 263-322.

Yu, D., & Deng, L. (2016). *Automatic speech recognition* (Vol. 1). Berlin: Springer.

Wang, M., Huang, Q., Zhang, J., Li, Z., Pu, H., Lei, J., & Wang, L. (2019, February). Deep learning approaches for voice activity detection. In *The International Conference on Cyber Security Intelligence and Analytics* (pp. 816-826). Springer, Cham.

Salvendy, G., & Karwowski, W. (Eds.). (2021). *Handbook of human factors and ergonomics*. John Wiley & Sons.

Chen, H., Liu, X., Yin, D., & Tang, J. (2017). A survey on dialogue systems: Recent advances and new frontiers. Acm Sigkdd Explorations Newsletter, 19(2), 25-35.

Taylor, P. (2009). Text-to-speech synthesis. Cambridge university press.

Oord, A. V. D., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., ... & Kavukcuoglu, K. (2016). Wavenet: A generative model for raw audio. arXiv preprint arXiv:1609.03499.

Yumak, Z., & Magnenat-Thalmann, N. (2016). Multimodal and multi-party social interactions. In Context aware human-robot and human-agent interaction (pp. 275-298). Springer, Cham.

Moniri, M. M., Valcarcel, F. A. E., Merkel, D., & Sonntag, D. (2016, September). Human gaze and focus-of-attention in dual reality human-robot collaboration. In 2016 12th International Conference on Intelligent Environments (IE) (pp. 238-241). IEEE.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems*, *42*(3-4), 143-166.

Sidner, C. L., Lee, C., Kidd, C. D., Lesh, N., & Rich, C. (2005). Explorations in engagement for humans and robots. *Artificial Intelligence*, *166*(1-2), 140-164.

Rimé, B., & Schiaratura, L. (1991). Gesture and speech. In R. S. Feldman & B. Rimé (Eds.), *Fundamentals of nonverbal behavior* (pp. 239–281). Cambridge University Press; Editions de la Maison des Sciences de l'Homme.

Burke, M., & Lasenby, J. (2015). Pantomimic gestures for human–robot interaction. *IEEE Transactions on Robotics*, *31*(5), 1225-1237.

Liu, H., & Wang, L. (2021). Latest developments of gesture recognition for human–robot collaboration. In *Advanced Human-Robot Collaboration in Manufacturing* (pp. 43-68). Springer, Cham.

Pasqualotto, E., Matuz, T., Federici, S., Ruf, C. A., Bartl, M., Olivetti Belardinelli, M., ... & Halder, S. (2015). Usability and workload of access technology for people with severe motor impairment: a comparison of brain-computer interfacing and eye tracking. *Neurorehabilitation and neural repair*, *29*(10), 950-957.

Minamoto, M., Suzuki, Y., Kanno, T., & Kawashima, K. (2017, August). Effect of robot operation by a camera with the eye tracking control. In *2017 IEEE International Conference on Mechatronics and Automation (ICMA)* (pp. 1983-1988). IEEE.

Scalera, L., Seriani, S., Gallina, P., Lentini, M., & Gasparetto, A. (2021). Human–robot interaction through eye tracking for artistic drawing. Robotics, 10(2), 54.

Damm, O., Malchus, K., Jaecks, P., Krach, S., Paulus, F., Naber, M., ... & Wrede, B. (2013, August). Different gaze behavior in human-robot interaction in Asperger's syndrome: An eye-tracking study. In 2013 IEEE RO-MAN (pp. 368-369). IEEE.

Palinko, O., Rea, F., Sandini, G., & Sciutti, A. (2016, October). Robot reading human gaze: Why eye tracking is better than head tracking for human-robot collaboration. In

2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 5048-5054). IEEE.

Staudte, M., & Crocker, M. (2008, February). The utility of gaze in spoken human-robot interaction. In Proceedings of Workshop on Metrics for Human-Robot Interaction 2008, March 12th (pp. 53-59).

Andrist, S., Tan, X. Z., Gleicher, M., & Mutlu, B. (2014, March). Conversational gaze aversion for humanlike robots. In 2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 25-32). IEEE.

Mutlu, B., Kanda, T., Forlizzi, J., Hodgins, J., & Ishiguro, H. (2012). Conversational gaze mechanisms for humanlike robots. ACM Transactions on Interactive Intelligent Systems (TiiS), 1(2), 1-33.

Kosslyn, S. M., & Rosenberg, R. S. (2014). *Introducing Psychology:Brain, Person, Group*. Pearson Higher Ed. (pp. 39).

Mutlu, B., Forlizzi, J., & Hodgins, J. (2006, December). A storytelling robot: Modeling and evaluation of human-like gaze behavior. In 2006 6th IEEE-RAS International Conference on Humanoid Robots (pp. 518-523). IEEE.

Breazeal, C., Kidd, C. D., Thomaz, A. L., Hoffman, G., & Berlin, M. (2005, August). Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In 2005 IEEE/RSJ international conference on intelligent robots and systems (pp. 708-713). IEEE.

Wijnen, L., Bremner, P., Lemaignan, S., & Giuliani, M. (2020). Performing human-robot interaction user studies in virtual reality. In 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) (pp. 794-794). IEEE.

Li, R., van Almkerk, M., van Waveren, S., Carter, E., & Leite, I. (2019, March). Comparing human-robot proxemics between virtual reality and the real world. In 2019 14th ACM/IEEE international conference on human-robot interaction (HRI) (pp. 431-439). IEEE.

Liu, O., Rakita, D., Mutlu, B., & Gleicher, M. (2017, August). Understanding human-robot interaction in virtual reality. In 2017 26th IEEE international symposium on robot and human interactive communication (RO-MAN) (pp. 751-757). IEEE.

Villani, V., Capelli, B., & Sabattini, L. (2018, August). Use of virtual reality for the evaluation of human-robot interaction systems in complex scenarios. In 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) (pp. 422-427). IEEE.

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. Journal of Research in personality, 37(6), 504-528.

Nomura, T., Kanda, T., Suzuki, T., & Kato, K. (2004, September). Psychology in human-robot communication: An attempt through investigation of negative attitudes and anxiety toward robots. In RO-MAN 2004. 13th IEEE international workshop on robot and human interactive communication (IEEE catalog No. 04TH8759) (pp. 35-40). IEEE.

Nomura, T., Suzuki, T., Kanda, T., & Kato, K. (2006, September). Measurement of anxiety toward robots. In ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication (pp. 372-377). IEEE.

Acarturk, C., Indurkya, B., Nawrocki, P., Sniezynski, B., Jarosz, M., & Usal, K. A. (2021). Gaze aversion in conversational settings: An investigation based on mock job interview. Journal of Eye Movement Research, 14(1).

Villani, D., Repetto, C., Cipresso, P., and Riva, G. (2012).May i experience more presence in doing the same thing invirtual reality than in reality? an answer from a simulatedjob interview.Interacting with Computers, 24(4):265 – 272.Special Issue on Presence and Interaction.

Hasan, Atak. (2013). The Turkish adaptation of the ten-item personality inventory. *Nöro Psikiyatri Arşivi*, *50*(4), 312.

Alex Pentland and Tracy Heibeck.Honest signals: How they shape our world.MIT Press, Cambridge, MA, 2010. ISBN 978-0262515122.

Ekman, P. (1957). A methodological discussion of nonverbal behavior. *The Journal of psychology*, *43*(1), 141-149.

Carrère, S., & Gottman, J. M. (1999). Predicting divorce among newlyweds from the first three minutes of a marital conflict discussion. *Family process*, *38*(3), 293-301.

Kress, G. (2009). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

Marchetti, A., Manzi, F., Itakura, S., & Massaro, D. (2018). Theory of mind and humanoid robots from a lifespan perspective. *Zeitschrift für Psychologie*, *226*(2), 98.

Itakura, S., Ishida, H., Kanda, T., Shimada, Y., Ishiguro, H., & Lee, K. (2008). How to build an intentional android: Infants' imitation of a robot's goal-directed actions. *Infancy*, *13*(5), 519-532.

Riek, L. D. (2012). Wizard of oz studies in hri: a systematic review and new reporting guidelines. *Journal of Human-Robot Interaction*, *1*(1), 119-136.

Maklin, C. (2019).*tfidf* [source code]. Retrieved from https://github.com/corymaklin/tfidf
Bailenson, J. N. (2021). Nonverbal overload: A theoretical argument for the causes of Zoom fatigue.

Kamewari, K., Kato, M., Kanda, T., Ishiguro, H., & Hiraki, K. (2005). Six-and-a-half-month-old children positively attribute goals to human action and to humanoid-robot motion. *Cognitive Development*, *20*(2), 303-320.

Bafna, P., Pramod, D., & Vaidya, A. (2016, March). Document clustering: TF-IDF approach. In *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)* (pp. 61-66). IEEE.

Khosrow-Pour, D. B. A. (Ed.). (2005). *Encyclopedia of information science and technology*. IGI Global.

Khosrow, M. (2009). *Encyclopedia of Information Science and Technology*. Information Sci. 2 Ed.

Aronoff, M., & Fudeman, K. (2022). *What is morphology?*. John Wiley & Sons.

Demirtas, E., & Pechenizkiy, M. (2013, August). Cross-lingual polarity detection with machine translation. In *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining* (pp. 1-8).

Hayran, A., & Sert, M. (2017, May). Sentiment analysis on microblog data based on word embedding and fusion techniques. In *2017 25th Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE.

Yıldırım, Savaş (2020). *Turkish-Bert-NLP-Pipeline.* Retrieved from https://github.com/savasy/Turkish-Bert-NLP-Pipeline

Holmes, A., Sherwani, A., Kok, E., Ghattas, F., Pisanelli, G., Wen, X., ... & Cameron, D. (2018, July). Intuitive interfaces in human-robot interaction. In *Proceedings of the 19th Towards Autonomous Robotic Systems (TAROS) Conference* (pp. 462-464). Springer, Cham.

Parenti, L., Marchesi, S., Belkaid, M., & Wykowska, A. (2021, November). Exposure to robotic virtual agent affects adoption of intentional stance. In *Proceedings of the 9th International Conference on Human-Agent Interaction* (pp. 348-353).

Zajonc, R. B. (1965). Social Facilitation: A solution is suggested for an old unresolved social psychological problem. *Science*, *149*(3681), 269-274.

Babel, F., Vogt, A., Hock, P., Kraus, J., Angerer, F., Seufert, T., & Baumann, M. (2022). Step Aside! VR-Based Evaluation of Adaptive Robot Conflict Resolution Strategies for Domestic Service Robots. *International Journal of Social Robotics*, 1-22.

Björling, E. A., Xu, W. M., Cabrera, M. E., & Cakmak, M. (2019, October). The effect of interaction and design participation on teenagers' attitudes towards social robots. In *2019 28th IEEE international conference on robot and human interactive communication (RO-MAN)* (pp. 1-7). IEEE.

Mori, M. (1970). The uncanny valley: the original essay by Masahiro Mori. *IEEE Spectrum*.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, *1*(1), 71-81.

McCarthy, J., & Goffin, R. (2004). Measuring job interview anxiety: Beyond weak knees and sweaty palms. Personnel Psychology, 57(3), 607-637.

APPENDICES

APPENDIX A

CONSENT FORM

Gönüllü Katılım Formu

Katılacağınız çalışmanın amacı insanların bir iş görüşmesi senaryosu sırasında yaptıkları göz hareketleri temel alınarak geliştirilen bir insan gözbakış modelinin insan-avatar etkileşimi kapsamında değerlendirilmesidir. Bu çerçevede size bir sanal gerçeklik (VR) gözlüğü takılacak ve deney sırasında göz hareketleriniz kaydedilecektir.

Çalışma kapsamında temas edeceğiniz tüm yüzey ve cihazlar, el ve yüzey dezenfektanlarıyla ve ultraviyole ışık ile temizlenmektedir. Çalışma boyunca gösterilecek materyallerin herhangi biri kişisel rahatsızlık verecek ya da stres yaratacak içeriğe sahip değildir. Sizden kimlik belirleyici herhangi bir bilgi istenmemektedir. Katılımınız süresince gösterilen materyal nedeniyle ya da herhangi başka bir nedenden ötürü kendinizi rahatsız hissederseniz çalışmayı yarıda bırakıp çıkmakta serbestsiniz. Böyle bir durumda uygulayıcıya çalışmayı tamamlamadığınızı söylemek yeterli olacaktır. Çalışma sonunda, dilerseniz bu çalışmayla ilgili sorularınız cevaplanacaktır.

Bu çalışmaya katıldığınız için şimdiden teşekkür ederiz. Çalışma hakkında daha fazla bilgi almak için Enformatik Enstitüsü Bilişsel Bilimler Ana Bilim Dalı öğretim görevlilerinden Doç. Dr. Cengiz Acartürk (Oda: B-203; Tel: 210 7704; e-posta: acarturk@metu.edu.tr) ile iletişim kurabilirsiniz.

Bu çalışmaya tamamen gönüllü olarak katılıyorum ve istediğim zaman yarıda kesip çıkabileceğimi biliyorum. Verdiğim bilgilerin bilimsel amaçlı yayınlarda kullanılmasını kabul ediyorum. (Formu doldurup imzaladıktan sonra uygulayıcıya geri veriniz).

İsim Soyisim	Cinsiyet	Doğum Yılı	Tarih	İmza
			----/----/----	

APPENDIX B

DEMOGRAPHIC SURVEY

Demografik Bilgi Formu

Yaş :
Cinsiyet :
Eğitim durumu :

(Eğer hala öğrenciyseniz lütfen devam ettiğiniz seviyeyi işaretleyip, yanına "devam ediyor" şeklinde not düşünüz)

☐ İlköğretim	☐ Lise	☐ Lisans
☐ Yüksek lisans	☐ Doktora	☐ Doktora-sonrası

Anadiliniz nedir?

……………………………………………………………………………………………………
……..

Herhangi bir göz rahatsızlığınız var mı? Yanıtınız evetse, ne olduğunu belirtiniz.

……………………………………………………………………………………………………
………

Daha önceden sanal gerçeklik gözlüğü kullanmışlığınız var mı?

☐ Evet ☐ Hayır

APPENDIX C

TEN ITEM PERSONALITY INVENTORY (TIPI)

On-Maddeli Kişilik Ölçeği-(OMKÖ)

Aşağıda sizi tanımlayan ya da tanımlamayan birçok kişilik özelliği bulunmaktadır. Lütfen her bir ifadenin yanına, o ifadenin sizi tanımlama düzeyini dikkate alarak, o ifadeye katılıp katılmadığınızı belirtmek için 1 ile 7 arasında bir rakam yazın. İfadelerde size en çok tanımlayan özelliği dikkate alarak, uygun gördüğünüz rakamı yazın.

Dr. Hasan Atak

1 = Tamamen katılmıyorum
2 = Kısmen katılmıyorum
3 = Biraz katılmıyorum
4 = Kararsızım
5 = Biraz Katılıyorum
6 = Kısmen katılıyorum
7 = Tamamen katılıyorum

Kendimi ………………………………… olarak görürüm:

1. _____ Dışa dönük, istekli
2. _____ Eleştirel, kavgacı
3. _____ Güvenilir, öz-disiplinli
4. _____ Kaygılı, kolaylıkla hayal kırıklığına uğrayan
5. _____ Yeni yaşantılara açık, karmaşık
6. _____ Çekingen, sessiz
7. _____ Sempatik, sıcak
8. _____ Altüst olmuş, dikkatsiz
9. _____ Sakin, duygusal olarak dengeli
10. _____ Geleneksel, yaratıcı olmayan

APPENDIX D

NEGATIVE ATTITUDES AND ANXIETY TOWARD ROBOTS (NARS)

Katılımcı ID:
Tarih:

Aşağıda robotları içeren birtakım ifadeler bulunmaktadır. Lütfen ifadeleri okuyup, ne derece katıldığınızı 1-5 arasından uygun gördüğünüz rakama göre kutucukları işaretleyerek bildiriniz.

	Kesinlikle katılmıyorum (1)	Katılmıyorum (2)	Kararsızım (3)	Katılıyorum (4)	Kesinlikle katılıyorum (5)
Eğer robot kullanmam gereken bir iş verilse huzursuz hissederdim.	☐	☐	☐	☐	☐
“Robot” kelimesi benim için hiçbir şey ifade etmiyor.	☐	☐	☐	☐	☐
Diğer insanların önünde robot kullanırken kendimi gergin hissederdim.	☐	☐	☐	☐	☐
Robotların veya yapay zeka işleten varlıkların bir şeyler hakkında karar vermesi fikrinden nefret ederdim.	☐	☐	☐	☐	☐
Bir robotun karşısında sadece ayakta durmakta bile gergin hissederdim.	☐	☐	☐	☐	☐
Bir robotla konuşurken paranoyakça hisler beslerdim.	☐	☐	☐	☐	☐
Eğer robotlar gerçek duygulara sahip olsalardı huzursuz hissederdim.	☐	☐	☐	☐	☐
Eğer robotlar canlı varlıklara	☐	☐	☐	☐	☐

dönüşselerdi kötü şeyler olabilirdi.					
Eğer robotlara çok bağlı olursam kötü hissederdim, çünkü kötü şeyler olabilirdi.	☐	☐	☐	☐	☐
Robotların çocuklara kötü etkileri olacak diye endişeleniyorum.	☐	☐	☐	☐	☐
İlerde toplumun robotlar tarafından domine edileceğini hissediyorum.	☐	☐	☐	☐	☐
Bir robotla konuşurken rahat olurdum.	☐	☐	☐	☐	☐
Eğer robotların duyguları olsaydı onlarla arkadaş olabilirdim.	☐	☐	☐	☐	☐
Duyguları olan bir robotla olmak bana kendimi rahat hissettirirdi.	☐	☐	☐	☐	☐

APPENDIX E

ROBOT ANXIETY SCALE (RAS)

Katılımcı ID:
Tarih:

Aşağıda robotları içeren birtakım sorular bulunmaktadır. Lütfen soruları verilen ölçeği temel alarak, 1-6 arasından uygun gördüğünüz rakama göre işaretleyerek yanıtlayınız.
1 = hiç endişelenmezdim 2 = neredeyse hiç endişelenmezdim 3 = az da olsa endişelenebilirdim
4 = endişelenirdim 5 = çok endişelenirdim 6 = şiddetle endişelenirdim

	1	2	3	4	5	6
Robotun bir konuşmanın ortasında alakasız şeyler hakkında konuşması size nasıl hissettirirdi?	☐	☐	☐	☐	☐	☐
Robotun konuşmanızın yönündeki değişimleri takip edememesi size nasıl hissettirirdi?	☐	☐	☐	☐	☐	☐
Robotun anlaşılması zor konuları anlayamaması size nasıl hissettirirdi?	☐	☐	☐	☐	☐	☐
Robotun yapacağı hareketleri düşünmek size nasıl hissettirirdi?	☐	☐	☐	☐	☐	☐
Robotun etkileşim sırasında ne yapacağını düşünmek size nasıl hissettirirdi?	☐	☐	☐	☐	☐	☐
Robotun fiziksel olarak ne kadar güçlü olacağını düşünmek size nasıl hissettirirdi?	☐	☐	☐	☐	☐	☐
Robotun ne kadar hızlı hareket edeceğini düşünmek size nasıl hissettirirdi?	☐	☐	☐	☐	☐	☐
Robotla nasıl konuşmanız gerektiğini düşünmek size nasıl hissettirirdi?	☐	☐	☐	☐	☐	☐
Robot sizinle konuştuğunda nasıl cevap vereceğinizi düşünmek size nasıl hissettirirdi?	☐	☐	☐	☐	☐	☐
Robotun dediğinizi anlayıp anlayamayacağını düşünmek size nasıl hissettirirdi?	☐	☐	☐	☐	☐	☐
Robotun ne dediğini anlayıp anlayamayacağınızı düşünmek size nasıl hissettirirdi?	☐	☐	☐	☐	☐	☐

APPENDIX F

Utterances by Pepper during the Mock-Up Job Interview

Demo

Merhaba ben Pepper, bugün sizinle bir iş görüşmesi yapacağım. Ama ilk önce birkaç soruyla deneme yapalım. Tüm görüşme boyunca sorduğum sorulara sesli olarak cevap vermeniz, ardından size yeni bir soru sormam için de sanal gerçeklik gözlüğünün sağ tarafına dokunmanız gerekmektedir.

1. Biraz kendinizden bahsedebilir misiniz?
2. Size önerdiğimiz bu pozisyondan nasıl haberdar oldunuz?

Teşekkürler, demomuz sona ermiştir. Şimdiyse iş görüşmemize başlayacağız.

Group 1:

3. Biraz sizden bahsedelim. Neler yapmaktan hoşlanırsınız? Beklentileriniz, hedefleriniz nelerdir?
4. Önerdiğimiz bu pozisyonla neden ilgileniyorsunuz?
5. Bu pozisyonun hangi açılardan size uygun olduğunu düşünüyorsunuz?
6. Önümüzdeki beş yıl içinde kendinizi nerede görüyorsunuz?

Group 2:

7. Biraz da kişisel özelliklerinizden konuşalım. Sizi tanımlayan en belirgin üç özelliğiniz nelerdir?
8. Liderlik yetenekleriniz hakkında ne düşünüyorsunuz? Nasıl bir lider olurdunuz?
9. Liderlik özelliklerinizi ön plana çıkaran, yaşadığınız bir deneyimi paylaşır mısınız?
10. Başka bir işyerinden transfer teklifi alsanız, bu size nasıl hissettirir?