THE EFFECT OF KINEMORPHS ON F-FORMATION SHAPES: AN INVESTIGATION ON HUMAN ROBOT INTERACTION IN VIRTUAL REALITY

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

GİZEM ÖZEN

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

SEPTEMBER 2019

THE EFFECT OF KINEMORPHS ON F-FORMATION SHAPES: AN INVESTIGATION ON HUMAN ROBOT INTERACTION IN VIRTUAL REALITY

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

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

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

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

Examining Committee Members:

Prof. Dr. Cem Bozşahin
Cognitive Science Dept., METU

Assoc. Prof. Dr. Cengiz Acartürk
Cognitive Science Dept., METU

Assist. Prof. Dr. Burcu Ayşen Ürgen
Psychology Dept., Bilkent University

Date: 12 September 2019

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: Gizem ÖZEN

Signature :

ABSTRACT

THE EFFECT OF KINEMORPHS ON F-FORMATION SHAPES: AN INVESTIGATION ON HUMAN ROBOT INTERACTION IN VIRTUAL REALITY

Özen, Gizem
MSc., Department of Cognitive Sciences
Supervisor: Assoc. Prof. Dr. Cengiz Acartürk

September 2019, 66 pages

The main goal of this thesis is to study how kinemorphs influence F-formation shapes by focusing on the occurrences of one-to-one F-formation shapes in a virtual environment social interaction setting. The problem statement relies three aspects of this investigation. Firstly, it has been known that the proximity is affected by kinemorphs; however, it is not clear whether the shape of a social group's spatial arrangement is also influenced by kinemorphs. In the present study, F-formation shape of two facial kinemorphs were comparatively analyzed. Secondly, the interactions between human participants and robots in virtual reality were analyzed by focusing on likely similarities and differences to the real life human-human interactions in similar F-formation settings. Thirdly, the role of F-formation shapes was investigated on joining a dyadic interaction scenario. The results reveal systematic findings that address likely postures in a triadic interaction setting, for a robot agent in a virtual environment.

Keywords: F-formation, kinesics, kinemorphs, virtual environment, human-robot interaction

ÖZ

SÖZDIŞI İLETİŞİM BİÇİMBİRİMLERİNİN F-FORMASYON ŞEKİLLERİNE ETKİSİ: SANAL ORTAMDA İNSAN ROBOT ETKİLEŞİMİ ÜZERİNE BİR ÇALIŞMA

Özen, Gizem

Yüksek Lisans, Bilişsel Bilimler Bölümü

Tez Yöneticisi: Doç. Dr. Cengiz Acartürk

Eylül 2019, 66 sayfa

Bu tezin temel amacı sözdışı iletişim biçimbirimlerinin F-formasyon şekillerine etkisini, iki karakterli F-formasyon şekillerinin sanal ortamda kurgulanan sosyal etkileşim senaryosunda oluşumuna odaklanarak araştırmaktır. Bu araştırmanın üç yönü önemlidir. Öncelikle, F-formation içinde etkileşimde bulunan iki kişinin birbirine uzaklığı literatürde yeri olan bir araştırma konusudur. Bu uzaklığın sözdışı iletişim biçimbirimlerinden etkilendiği bilinmektedir; fakat, uzaklığın yanı sıra, sosyal bir grubun uzamsal düzeninin şeklinin de sözdışı iletişim biçimbirimlerinden etkilenip etkilenmediği açık değildir. Bu çalışmada, yüze ait iki sözdışı iletişim biçimbiriminin F-formasyon şekilleri üzerindeki etkisi karşılaştırmalı olarak analiz edilmiştir. Bunun dışında, insan katılımcılar ve sanal ortamdaki robotların arasındaki etkileşim, gerçek hayatta F-formasyon kurgusundaki insan-insan etkileşimiyle olan farkı ve benzerliklerine odaklanarak analiz edilmiştir. Son olarak, F-formasyon şekillerinin, iki karakterli bir etkileşime üçüncü olarak katılma üzerindeki rolü araştırılmıştır. Sonuçlar, sanal ortamdaki robotlar için, üç karakterli bir sosyal etkileşimde en olası duruşu niteleyecek sistematik bulgular ortaya koymaktadır.

Anahtar Sözcükler: F-formasyon, sözdışı iletişim, sözdışı iletişim biçimbirimleri , sanal ortam, insan robot etkileşimi

To Science

ACKNOWLEDGMENTS

First of all, I would like to express my gratitude to my supervisor Cengiz Acartürk for giving me a place in his laboratory, allowing me to use all the equipment I wanted to use and leading me to an intellectual survey including both philosophical and technical aspects. I also would like to thank TÜBİTAK for the project 117E021/A gaze-mediated framework for multimodal human-robot interaction.

Besides my supervisor, I would like to thank Faruk Büyüktekin to discuss a wide range of topics every day without getting bored even for a single moment and feeding these discussions with the literature from cognitive science, anthropology and philosophy. I also want to thank my partner Tunç Tuğcu for motivating me to learn coding, helping me every time I stuck somewhere with his magical computation skills and being there through the whole painful and joyful processes in writing a thesis.

I would also like to thank all of the colleagues from the department of Cognitive Science at METU, starting from Efecan Yılmaz for his technical advices, Bengisu Çağıltay for being a great friend to "get loose", Emre Erçin for his visual performances and never-ending psychological supports, Samet Albayrak for taking all the classes with me, sharing the funniest moments and having ridiculous conversations, Kerem Usal for creating a beautiful environment in lab with his good chats, Çağrı Şakiroğulları for his laughter.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ECA	Embodied Conversational Agents
US	United States

CHAPTER 1

INTRODUCTION

In either verbal or non-verbal social interaction, the signals that we pick up from the other interactants have different implications in different contexts. Think about Neo's first encounter with Agent Smith in the movie "Matrix Reloaded" (Silver & Wachowski, 2003). If the direction of Agent Smith's body were not towards Neo when he was standing in front of him, his threats might not seem as serious as it is. Exaggeratively, if Agent Smith were standing in a way that his back is turned to Neo, so, looking at the very opposite direction from Neo, Neo would think that something is wrong with the Agent Smith or there is 'a glitch in Matrix'. If they were waiting in a bank queue, this gathering form of Agent Smith and Neo would be meaningful; however, in a discussion scene, it does not seem so. Therefore, where to stand and how to stand is an important que in social interactions. This study is nothing but a step about the decision on Agent Smith's position and rotation in his communication with Neo. More specifically, it is about the position and rotation of a virtual agent relative to other agents in a virtual environment setting.

Our use of space can be explained by environmental, subjective and social parameters. As in the comprehensive study on proxemics conducted by Ciolek and Kendon (1980), how people stand in a crowd, in between buildings and in front of a building are different from each other. Thus, the arrangement of physical entities around us affects how to stand and where to stand. On the other hand, what individuals try to express and whom they want to interact with change their decision on their rotation and position. Finally, cultural differences also affect the use of space. In a pagan culture, for Cauldron rite, members need to construct circular form wide enough to include all of the members that joined the ritual (Slater, 1978, p.11); in Christian culture, prayers are performed in lines because of the seating plan of the churches. As culture affects how we stand in an interaction, all other means of relationship between human-beings affects our use of space. This study is about much more smaller parameters than the ones described in this paragraph. On the other hand, it is not about human-human proxemics in real life; it is about human-robot proxemics in a virtual environment. The reason why human's proxemics behavior with robots in a virtual environment is important is that social interaction with robots defines new representations served to human mind and these new representations are needed to be investigated. If environment and social parameters are affecting human-human proxemics, as new social and environmental stimuli, robots and virtual environment should be investigated in terms of their effects on human proxemic behavior. By investigating this, whether there is a difference between artificial agents, namely, human-like virtual agents such as Agent Smith and virtual robot agents, in terms of their allowed distance from humans (i.e. natural agents) will be revealed.

In order to strengthen the relationship between robot/virtual agents and humans, robot/virtual agents need to process complex behavior of humans with their limited sensory input. Therefore, if robots are expected to demonstrate a proxemics act convenient to social requirements of human relations, a pattern must be introduced to them in order to make sense of spatial organization of humans. In the present study, by using taxonomy of Ciolek and Kendon (1980) (which is demonstrated in Chapter 1), one-to-one F-formation types are investigated on triads in a virtual environment by looking at their relations with two facial kinemorphs (smallest unit of kinesic behavior). There are two main questions about the relationship between F-formation shapes and facial kinemorphs in the present study. First question is whether these kinemorphs determine the focus of the participants in a virtual social interaction or not. The second question is whether kinemorphs have an effect on the rotation of the participants. These two questions are linked to each other in the sense that if the answer of the first question is negative, then the probability of the second question' answer to come out negative is very high. Therefore, the second question can be reformed in the following way, 'if kinemorphs determine the focus of the people, do they also have an effect on the rotation of the participants?'. More precisely, what are the F-formation shapes people form, when they focus on a robot in a virtual interaction?

1.1 Outline

This thesis is made up of five chapters. In the first chapter, what is the thesis about is stated briefly, what is importance in terms of science and life is mentioned, the motivations for the research is explained and a basic outline is introduced to the reader.

In the second chapter, what is proxemics and what are the important studies in the literature is presented. Then, the place of the proxemics' in human-robot interaction is explained. Finally, distributed cognition is demonstrated with its relation with the human-robot interaction.

In the third chapter, methodology of the experiment is explained. As a summary, participants join three different experimental stages and they are expected to approach the humanoid robots in a virtual environment as if they are approaching in order to have a conversation with them. The first one of the stages is for making the environment more familiar to the participants, the second one is to observe the gaze's (first kinemorph) effect on F-Formation, the third one is to observe lip movement's (second kinemorph) effect on F-Formation.

In the fourth chapter, experimental results are stated and discussed. As a brief summary, in the second stage of the experiment, participants tend to rotate their virtual bodies (first person screen perspectives) to the robots who look at them. In the third stage of the experiment, participants tend to rotate their virtual bodies to the robots who speaks (moves their lips). In both stages, when people are encountered two robots and join the joint interactional space as a third person, they create one-to-one F-Formation shapes with the one they look at. When both of the robots are having the same action, distribution of the F-Formations is wider. When their action is different from each other, distribution of F-Formation is much more specific.

In the last chapter, key findings are summarized. Then, they are interpreted and discussed briefly. Finally, recommendations for further studies are introduced.

CHAPTER 2

LITERATURE REVIEW

Virtual agents and robots are aimed to be able to communicate with humans in social settings; however, social interaction is a very complex phenomenon which has many aspects like verbal and non-verbal interactive behaviors, high level mental processes, low level sensory perception, setting of a goal and etc. The main challenge in strengthening the interaction between humans and virtual agents/robots is to make virtual agents/robots perceive and understand complex human behavior via their limited sensory input capacity and give meaningful responses using their multimodal features. Yumak et al. describes the standard route that leads researchers to a multimodal interaction system in five steps:

> "(1) Individual low-level sensing modules (e.g., face recognition, skeleton tracking, sound localization, speech recognition), (2) multimodal tracking and fusion to combine information from individual trackers for making high-level inferences about the situation and the user state, (3) decision making and dialogue management to decide what to say and what to do given the partial sensory information, history of actions and the artificial character's internal state, (4) planning and synchronization of the output behavior to render the output decisions, and (5) actual realization of the planned behaviors at the level of motor controls for the robots and using computer animation techniques for virtual humans." (2016)

In this research, I proposed contributions to the third steps by focusing on non-verbal communication as a part of face-to-face interaction. When people decide to have an interaction with each other, they create a spatial-orientational arrangement in which they can exchange sentences or more visual or paralinguistic features of communication like gestures, posture, tone of voice and etc. Kendon (2010) call this spatial-orientational arrangement "F-Formation". According to him, F-Formation is such a joint interaction space that people who are not interested in the interaction seek to avoid intersecting each other's personal field, conversely, people who are interested in the interaction resist staying apart and tend to approach each other. Moreover, if they tend to approach each other, they group themselves in lines, circles or other kinds of visual patterns (Ciolek & Kendon, 1980). In these patterns, they can be only two people as well as thirty-six people or more; however, as it gets crowded, it is much more difficult to analyze visual structure of the gatherings. Previous works on one-to-one F-Formation deal with both structure of the gatherings and social or physical factors affecting it; however, in researches about multi-party F-Formation focuses only on social or physical factors affecting the multi-party interactions because of this difficulty.

For humans, it is not even a matter how to decide where to join the F-Formation. Decision on where to join the formation when one wants to have a role in a conversation has very implicit nature in the human mind. However, for virtual agents or robots, choosing one space over the other, being closer to one person than the other requires a lot of explicit antecedent decision-making steps. Therefore, in order to design an agent who is able to choose a space in a conversation appropriate for the communication norms, more explicit knowledge about the usage of the space in humans is required. Thus, people's behavior in F-Formations needed to be understood in systematic way that virtual agents and robots will perceive it using their partial sensory information.

According to Ciolek (1977), a formation may be analyzed in eight ways:

> "(a) its basic shape; (b) the spatial-orientational arrangements used; (c) the type of spaces generated; (d) the degree of orientational 'polarization' of the participants; (e) the positioning of the focal points of a formation; (f) sharpness of boundaries; · (g) type and degree of movement possible in a formation; and (h) the ways in which a formation may gain and lose its participants."(p.18)

Main purpose of this thesis is to analyze the basic shape of the F-Formations in the context of human-robot interaction. Ciolek (1977) gives a detailed and systematic account to the effect of physical settings, the effect of individuals' characteristics, the effect of group's characteristics, the effect of spatial arrangements, the effect of basic line of activities and the effect of social relationship on proxemics in human-human interaction. Except for one of these effects, all of them are external factors influencing the size of the interactional space or the distance between interactants. Only 'the effect of unit's characteristics' examined by Ciolek is about the individuals in an F-formation, therefore, it is an effect about the internal structure of an F-formation. Also, the topic of unit's characteristics can be classified under 'postural-sex identifier' variable (see 1.1) which is described by Hall (1968). According to Hall, in addition to 'postural-sex identifier' there are seven more internal parameters of proxemic behavior and they give a way to understand the behavior of the people in the interaction:

Table 1: The table taken from Agnus (2012) putting Hall's the eight variables of proxemics into picture (p.5-6).

Postural-Sex Identifier	It refers to the postural status and sex identities of the participants in the interaction
Sociofugal and sociopetal Axis	It refers to the positioning of one's shoulders and face (sociofugal space) which will encourage or discourage the

	process of communication (sociopetal space)
Kinesthetic Factor	It refers to the distance between the individuals which provides capability to touch each other
Touch Code	It deals with the individual's manner of touching one another
Visual Code	It refers to the eye contact with reference to the space between the individuals
Voice Loudness	It refers to the loudness of the voice with reference to the space between the individuals
Thermal Code	It refers to the heat transmitted from the individuals' bodies with reference to the space
Olfaction Code	It refers to the individual's degree of breath and odors that can be felt by the other individual standing next to him

The unit's (individual's) characteristics in Ciolek's study is a part of the posture-sex identifier researches. The reason is that it specifically based on the gender differences. According to him, "in all cases the presence of a male tends to increase the group's buffer zone, or conversely, the presence of a female tends to reduce the width of the space-bubble tree from passers-by." (1977, p.140) It means that when a group of people is having a communication, they are sensitive to the pedestrians who are passing by and the size of their joint interactional space varies with respect to the gender ratio in the group. However, the issue is that the knowledge about the parameters which determines the rotation of the individuals is as important as the knowledge about the parameters which determine the size of the joint interactional space. In that case, we do not know whether the shape of an F-formation is affected by the gender differences, although we know that size of the F-Formation is affected by it. If it is affected, it would be nice to know it. If it is not affected by it, might other variables or parameters which are described by Hall be affecting the shape of the formations? At that moment, human-computer interaction studies are important because of two reasons:

First of all, in order to investigate the topic above, it is hard to come up with an experiment design in real-time human-human interactions because codes that are described by Hall are deep-seated in humans. It is so deep-seated that loudness of the voice or eye contact comes naturally in a social interaction. They are working almost like reflexes. Therefore, it is hard to control them and so manipulating a participant's actions by controlling them is harder. The best way to make an analysis is observing the F-formation from an air view without interfering participants' interaction. Edmonson and Han (1983) used a camera standing high above the participants in their study and it was a very important strategy for measuring the distance in an accurate way. It is also an important strategy for analyzing the shapes of the F-formations. On the other hand, interactional space is something very dynamic since it is affected by the environmental arrangements (position of the artifacts), social interactions and individual's personal states. This dynamic structure is making it harder to set up an experiment. Conversely, in a virtual environment, it is easier to manipulate the actions of the virtual agents. Every action that is needed to be reacted can be controlled easily.

Secondly, Hollan et al. (2000) proposes a framework for new researches in the human-computer interaction combining ethnographic observations and controlled experiments. They suggest a distribution of cognitive processes over members of a group and relationship between internal and external structures and how products of previous events can affect the nature of the later events. The present study follows this framework except for ethnographic observations. As an internal structure, visual codes on robots are controlled. As an external structure, F-formation shapes are observed. Therefore, the effect of an internal structure to an external structure is aimed to be resolved.

2.1 Structure or Shape-Based Analysis in F-formation

Ciolek and Kendon (1980) gives a taxonomy of spatial arrangements for a one-to-one F-Formation:

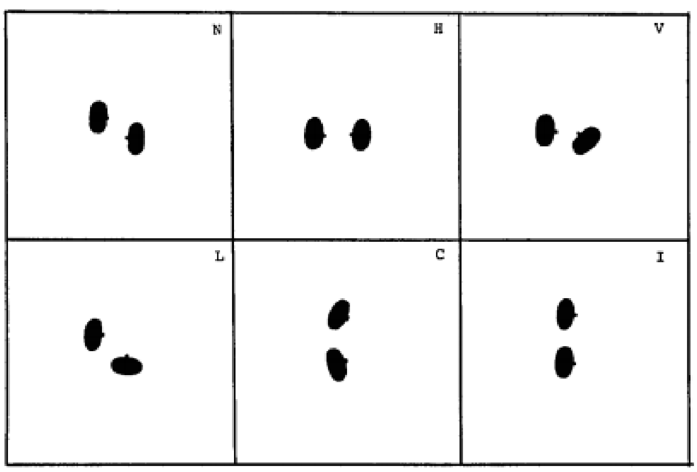

Figure 1: Six Basic Types of F-Formation (Ciolek & Kendon, 1980)

They made observations on a large set of data coming from a camera in a mall and found that close arrangements like N, H, V shaped formations are used more often than the others in open, empty spaces. L shaped formation is found to be used in semi open areas with a heavy human traffic. Therefore, physical setting is not only affecting the size of the joint interactional space but also the shape in the F-Formations.

Same shapes are used in investigating seating arrangements and their social implications. In Harris and Sherblom (2011), there is a figure (6.1) which reflects the seating patterns that are identified by Sommer (1969). In that figure, H-shaped formations are said to define a competitive social interaction, while N-shaped formations define a co-active, L-shaped formations define a cooperative and I-shaped formations define a modified cooperative social interaction. (p.119) Therefore, also social relationships have an effect not only on joint interactional space but also on the shape of the F-Formations.

While the shape of the human clusters is affected by both social context and physical settings, clarifying how individuals' interactive behaviors described by Hall affect the shape of the F-formations is very important.

2.2. Territoriality of proxemics

Territoriality is a concept in the study of animal behavior and it refers to all the behaviors of animals concerning claiming of a geographical area and securing it for food and reproduction. By using multi-modal aspects of their communication abilities like auditory, visual or olfactory skills, they mark their territory and interact with each other by giving signals. Agnus (2012) proposes that proxemics is a concept of territoriality. (p.6) Although people are not aware of the features of proxemics behavior in their culture, they become conscious of it when it is violated by different features of foreign cultures.

Vine (1975) proposes it that through the animal history, different animals and cultures have different understanding of territories. There are four kinds of definitions in terms of territories, namely, home range, territory, individual space and social space. Firstly, home range is a fixed area that animals uses regularly through an extended time for doing some routines of activities; however, does not feel to have an obligation to defend it. Secondly, territory means a fixed area which is controlled by an individual or group. In territories, residents detect unfamiliar animals rapidly and give violent responses to defend it. Thirdly, individual space means a mobile and body centered area in which individuals tend to externalize other individuals. Lastly, social space means, again a mobile and body centered area in which animals can tolerate the others. All animals use these four kinds of territories in different ways. For example, gems have their own territory in which other gems can't enter into while dogs share their territory with their relatives.

In that sense, people are getting into each other's social space in order to communicate. By this way, they become able to produce clusters that is called F-Formations.

However, once they form a cluster, their spatial organization is no longer a part of the social space studies. As in the Agnus' claim mentioned above, a person tends to become uncomfortable when her/his social space which is shared with another person is disrupted by an intruder. This situation makes F-Formation look like a term of territoriality, although territoriality defines a fixed area, while F-formation defines a mobile area. Therefore, it seems that when a person accepts other people into his/her social space, social space might turn into a mobile territory.

In the present study, conditions of how social spaces of individuals come together and produce a territory for them is investigated with a multi-modal approach. For a full investigation, the effects of the items in the list of Hall (1968) on the shape of F-Formations should be investigated one by one. The reason is that all of the items in the list defines different behavioral signals in territoriality. By manipulating them one by one, much more systematic results may be achieved. However, this study, as a starter, is about on visual code which stands as the fifth item in the list of Hall. Other items, namely, olfaction code, thermal code, voice loudness, touch code, kinesthetic factor, socio-fugal and socio-petal axis, sex identifier is were left for further research.

2.3. A conceptual discussion on Proxemics and Kinesics

Knapp (2013) has a classification about non-verbal social interaction. In this classification, kinesics is defined as all the body movements including gestures, body movements, facial and eye behaviors, while proxemics is defined as the use and perception of space such as in seating and spatial arrangements. In that sense, it can be claimed that by investigating the effect of kinesics (visual codes in Hall's classification) on the shapes of the formations, one is also investigating the relationship between kinesics and proxemics. However, Harrigan (2005) proposes that proxemics is mostly determining 'approach distance' (Hayduk, 1983) and the orientation of bodies with respect to other interactants (which creates the shape of the F-formation) is the topic of kinesics. In that case, it is hard to decide which study area (proxemics or kinesics?) that F-formation shapes are included in.

Ciolek (1977), mentions Wilkinson's study (1975) in which an Eskimo's hunting techniques are described. In this description, there is a part in which the shape and the distance is tightly linked and both is the terms of territoriality.

> "When threatened by predators, musk oxen usually adopt a loose, semi-circular or linear formation facing the predators. The mature bulls occupy the most vulnerable positions at the front and sides. Calves are protected between the flanks…" (1975, p.20-21)

In these sentences, it is implied that the shape of these gatherings helps animals to protect themselves from the danger. Therefore, it seems that the orientation of animal bodies is doing something different than a kinesic act. It is not simply the bodies' way of communication, but it is a way of protection. It does not aim to give a signal to the opponent, it is just an instinctive action to exhibit protection or aggression. With the discussion of Agnus (2012) on territoriality of proxemics described in 2.2, this condition makes F-formation closer to proxemics. On the other hand, the variables of proxemics (described in Table 1 and stated by Hall) might have an effect on not only the distance between the interactants but also the F-formation shapes between them. If they have such effects, it would not be plausible to separate F-formation shapes from proximity.

In Harrigan's sense, such acts like aggression and protection might be included in the communication. Therefore, the magnitude of the F-formations is related with proxemics but the shapes of the F-formations might be related with the kinesics. Therefore, F-formation might be a hybrid study including both of these study areas. It is also quite logical to see F-formation as a hybrid study encapsulating both kinesics and proxemics. Therefore, there are two options:

1. Seeing F-formation shapes as the topic of proxemics and claiming that the study is on facial kinemorphs' effects on proxemics.

2. Seeing F-formation shapes as the topic of kinesics and claiming that the study is on the relationship between facial kinemorphs and the trunk orientation kinemorphs.

2.4. Interactive Agent Design

A non-verbal social interaction should be examined in an interactive environment. The following sections describes what kind of interactive environment is used and why it is used in the present study.

2.4.1 Social Interaction

What is implied by the social interaction is a focused interaction in Goffman's terms. According to him, there are two kinds of social interactions and one of them is focused, while the other is unfocused. In unfocused interactions, people are gathering information about other people either by peripheral awareness or by quick glancing them. Conversely in focused interactions, people are cooperating in order to keep a single focus of attention in the conversation, simply by taking turns and etc. (1963, p.24) Therefore, a focused social interaction through visual code is the main theme of the research.

2.4.2 Design with multi-modal aspects

There are two facial kinemorphs in the present study. Firstly, gaze (a result of the movements of the eyes and the head) is the major modality in non-verbal communication. By looking at another person's eyes and perceiving the direction of his/her gaze, we direct our attention to gather needed information from the environment. (Frischen et al.2007) It's effect on personal distance in proxemics also stated by numerous researchers (e..g, McCall, 2015; Argyle & Dean, 1965; Argyle & Cook, 1976 among others) What about its effect on the shape of the F-Formation shapes? Might gaze be one of the parameters by which people determines their place and rotation in the joint interactional space?

Secondly, lip movement, as another visual, communication modality and an element of facial kinesics, is an important implicature which tells people that a person is in the center of the focus of attention. In other words, lip movements are the indication of the person who is doing an exchange during the course of communication. If someone took the turn in a conversation, his/her lips start to move in a certain way that a corresponding sound goes along with it. Then, words, sentences that are intended to be exchanged with other individuals occur. However, as I mentioned in the above section (section 1.2), in order to boost the effect of the visual code, vocal code is not used in the present study, therefore, lip movements have no vocal companion. Since it has no vocal companion, it might be claimed that lip movement is used as a “regulator” in Ekman and Friesen’s notions (1969a). The reason is that if there is no vocal companion, there will be no verbal interaction and the only information conveyed will be who’s turn in the interaction. According to Ekman and Friesen, regulators ‘convey information necessary for the pacing of the conversation’ (p.82). Harrigan claims that regulator actions are perceived from the periphery of awareness because they are so deep-seated, or ‘overly-learned’ (p.82) habits in Ekman and Friesen’s terms and they can be defined as ‘conversational exchange behaviors’ (2005, p.165).

While people are having a face-to-face interaction, it is impossible eliminate one of the modalities which are used by interlocutors to communicate with each other. Either in a focused or an unfocused way, people are getting information about all of the modalities in the list of Hall (1968). As long as one has not a physical impairment in his/her brain or ears, it is impossible in real life to see the lip movements but somehow, close sounds of the words. In other words, human beings are living with kinemorphic constructions (unity of kinemorphs) and it is impossible to separate kinemorphs. In order to separate kinemorphic constuctions into its kinemorphs, in order words, in order to get rid of the physical limitations and exclude the effects of other codes than visual code, two different kinemorphs are used separately to investigate their effects on the shape of the spatial arrangements. The key point in investigating these effects is to see that gaze and lip movement might have different influences on spatial arrangements. The reason is that gaze implies a context in which someone is in an interaction with you; however, lip movement implies a context in which someone is in an interaction with someone else. The body orientation habits might have been developed in different ways for these two contexts.

2.4.3 Embodied conversational agents

In order to understand the term "embodied conversational agents", the notion of agency and the theory of embodied cognition must be understood. Firstly, agency is a term which implies a capacity to make choices. It takes different meanings in different contexts. In ethics, it means a person who can act morally; in feminism, if someone is an agent, s/he is treated equally with other individuals. In the present study, agents are software entities with autonomous interactive features. Secondly, embodied cognition is a theory depending on the idea that cognition is dependent upon the physical features of an agent. According to these theory, not only brain structure but also physical structures of the body plays an important role in cognitive processes. For example, Lakoff and Johnson (2001) proposed that our notions like "back" and "front" implies the embodiment of cognition in a great way. According to them, creatures with longer and flatter body which moves backwards would have a different conception of "in front of". In that sense, embodied conversational agents refer to the agents with physical characteristics that are coherent with the capability to have a conversation.

Ruttkay et al. (2004) defines embodied conversational agents (ECAs) as "autonomous software entities with human-like appearance and communication skills". According to Isbister et al. (2004), at first, the reason why ECA research is produced is that intelligent agents wanted to be capable of executing certain social actions and leave an impression that it is a social entity. However, using them in order to reveal their impact on humans or in order to create a social interaction setting which is impossible to create in real life is also a good strategy in investigating accepted social actions among humans. Recently, the research literature on the interaction between ECAs and humans has been growing rapidly.

Louwerse et al. (2009) studies on how humans distribute eye gaze towards embodied conversational agents in virtual tutoring systems by analyzing fixation times of the participants. They found that the tutoring agent received the attention in during the whole interaction in a single-agent tutor setting. On the other hand, in a multiple-agent setting, participants fixate their gaze on the related agents when all the agents are speaking. Lusk and Atkinson (2007) proposes that ECAs are better than vocal tutoring in increasing learning performance in humans.

Embodied conversational agents are mostly human avatars because they are generally designed to be assistants or companions to help humans in their domestic or professional works. Therefore, using humanoid robots instead of human avatars opens a whole different study area. In this area, what effects the human decision can be observed, what robots need to do in a conversation can be discovered and what are the differences between human-human and human-robot interactions can be found at the same time. Yılmaz (2018) looks at how language and vision interact in virtual reality environment. In his study, he uses humanoid robot avatars imitating human gaze and language to have a joint attention with the human participants. Analogously, this study is on Human-Robot proxemics as a part of

Human-Robot interaction and aims to reveal how kinesics and proxemics interact in virtual reality environment using humanoid robots.

2.4.4 Human-Robot Proxemics

Nomura et al. (2007) examined the effect of the robot size on proxemic behavior between humans and robots. They found that smaller size humanoid robots (29 cm) make people anxious and people tend to keep their distance. Koay et al. (2007) investigated the difference between physical, verbal and no interaction situations in terms of proxemics. They found that physical interaction requires closer distances than verbal interaction and no interaction.

As it is stated before, F-Formation is a combination of the location and rotation of interactants relative to each other. Therefore, a shape-based analysis which includes rotation information is very important for HRI studies.

2.5.Multi-Agent Setting and Distributed Cognition

According to Yumak et al. (2016), previous works on interactive virtual agents and robots focus on only one-to-one interaction, namely, F-Formations with two people; however, multi-agent settings are not investigated enough. Therefore, this research is partially to make sense of one-to-one F-Formation shapes in human-robot interactions; however, its main focus is to make sense of F-Formation shapes with multi-agents.

As I mentioned at the start of this chapter, it is harder to analyze the shape (or visual structure) of a more crowded population. In Ciolek's study (1977), there are 24 shapes of spatial-orientational patterns which are recorded in different places of the world and put into a scheme (Diagram 1). Each of them is so familiar that it is easy to visualize the contexts:

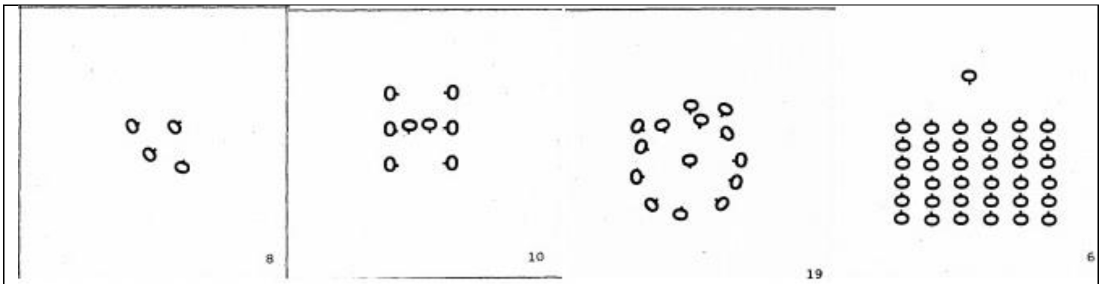

Figure 2: These figures are four patterns among twenty-four spatial organizational patterns in Ciolek's study (1977)
The first figure(8.th figure in Ciolek's) is a group of people in front of a painting in an exhibition, second figure (10.th figure in Ciolek's) is bride and groom leaving the chapel with military guards, third figure (19.th figure in Ciolek's) is a ring of people listening and watching the person in the center, last figure (6.th figure in Ciolek's) is a group of people in a lecture room. (p.19-22)

Although the shapes in Fig 2. are all familiar, they haven't been titled in the way that one-to-one interaction shapes are titled. They are recorded from real-life and mentioned only by their contexts. On the other hand, as it can be seen from Fig 2, it is difficult to say why a person chooses to stand where s/he stands. The reason why it is difficult is that it is affected by many parameters including people's internal processes and external conditions. In that case, distributed cognition perspective could be useful to mention.

The idea that cognition is not determined or limited by only an individual is called distributed cognition. In this view, cognition is thought to be distributed over the environment including the individual, artefacts, natural entities, time and other individuals. Hutchins describes distributed cognition as in the following sentences:

> "…cognitive processes may be distributed across the members of a social group, cognitive processes may be distributed in the sense that the operation of the cognitive system involves coordination between internal and external structure, and processes may be distributed through time in such a way that the products of earlier events can transform the nature of later events." (2000, p.1-2)

According to Hutchins, cognitive processes are distributed across a social group. For this study, it means that every member of the social group has an effect on decision on where to stand and how to stand and this is the reason why more crowded formations are harder to analyze. Moreover, 'products of earlier events can transform the nature of later events' (p.2). Therefore, the position and rotation of every member might transform the position and rotation of the new member. In that case, instead of analyzing more crowded formations at once, starting the research from smallest groups and continuing by adding one interactant at a time seems like a plausible strategy.

Hutchins went to stay with US navy ship called "the Palou" in order to investigate their organization during the navigation of the ship. In his book "Cognition in the Wild", he describes how the ship is organized in order to navigate. According to him, places, behaviors, tools are working and evolving together and producing the computation. Hutchins (1995) also proposes that every computation or every task is a part of a larger computational system. In that sense, he thinks that a task (including all the tools, methods and places) in navigation is a part of an organization of a team performance. Therefore, not only abstract properties of navigation matters in the implementation of the navigation but also tools, methods, places and people matter. In his book, he claims that the perspective of "cognition as computation" of Marr is applied to navigation and a different metaphor, namely, "navigation as computation" can be suggested. The reason why he proposes this metaphor is that he explains everything about marine life in terms of Marr's levels of analysis of an information processing task (1982):

1. Computational theory

2. Representation and algorithm

3. Hardware implementation (figure1-4) (p.25):

According to Marr, the first level which is computational theory requires the answer of the question of what is the goal of the computation? In this level, there should be a mapping of the goal and the abstract properties of the map should be defined. Then, its relevance, adequacy and logic behind it should be demonstrated. Second level is about the strategies of implementation including the representations of input and output or computational architecture of the goal. The last level is about physical realization of the representation and algorithm (p.24-25). What Hutchins does is following this path. According to him, the goal of the computation is to predict one's position and to determine how to end up at a specific position. The computational architecture is the combination of one-dimensional constraints (linear positional, circular positional, position-displacement, distance-rate-time). He also illustrates the difference between western navigation and Micronesian navigation by claiming that the difference can be explained by the creation of the tools.

The computational theory behind this thesis is the effect of kinemorphs on the F-formations in a virtual environment. For representation and algorithm level, Ciolek's and Kendon's taxonomy (1980) is used. Therefore, the representation of an input and output can be exported from the result of this study. However, hardware implementation level is not in the scope of the study.

This study is to make a contribution to the issue of virtual F-Formations in order to achieve a notion like "F-formation as computation" in the future. A lot of researches contributed to the production of this notion consciously or unconsciously through the history. For example, firstly, Kendon (2010) has a great example supporting Hutchins' historical perspective to events and representations. He uses a photograph of a group of people posing to the camera in Fig. 16 (p.12) in his article "Spacing and Orientation in Co-present Interaction". In this photograph, people are in a special kind of spatial arrangement that does not exist before 1839. He claims that the existence of the arrangement of people in the photograph is meaningful only after the invention of the photography which happens in 1839.

Figure 3: Kendon's Fig.16 (p.12)

Without photograph machines, this alignment of people might not exist. Secondly, the notion of F-formation itself is a support to another claim of Hutchins because it implies that cognitive process is distributed across the other member in the joint interactional space. Thirdly, Hall's variables of proxemics describe the internal structure of a gathering, while all the studies on how the environmental arrangement effects the social interaction (Sommer 1959: Sommer & Becker 1969: Ciolek 1977: Hall 1966, 1974) describe the external structure of a gathering. It means that all kinds of distribution of cognitive process are studied on in different places and times.

CHAPTER 3

METHODOLOGY

The experiment is conducted in order to investigate kinesics' effect on F-formation shapes in a virtual environment within the context of Human-Robot Interaction. There are two main questions (and sub-questions) in terms of the relationship between kinesics and proxemics:

1. Do the kinemorphs (i.e., head rotation, lip movements) determine the focus of the participants in a virtual social interaction?

2. If kinemorphs determine the focus of the participants, do they also have an effect on the rotation of the participants? More precisely, what are the F-Formations they form, when they focus on a robot in a virtual interaction?

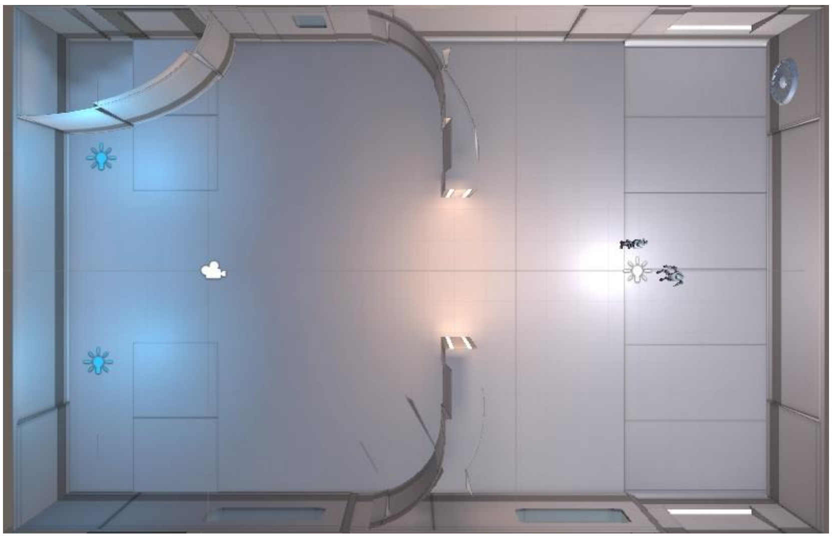

Figure 4: The environment produced in Unity Engine. White camera (on the left) represents the starting point of the participants. Two characters (on the right) are robots.

To search for an answer of these questions, the environment in Fig. 4 made by Unity Engine (5.6.6f2) was introduced to participants through a regular computer screen. The white camera icon in the Fig. 4 indicates the location where the participants are situated at the start of each session. There are at least one, at most two robots at the other side of the room in each session. Participants are expected to approach the robots by using keyboard and mouse as in all FPS games and situate themselves as if they want to have a dialog with these robots. Once they find the right spot to have a communication, they are required to use the left click of the mouse which send them back again to where they start. 'The right spot' is not determined by the researcher, depending solely on the participants' choice and it provides the main data for this study. The other important data which was taken from participants to find their focus in the interaction during the experiment is their gaze location and it was collected via Eye Tribe eye tracker.

There are two conditions in terms of the number of the robots, namely, one robot conditions and two robot conditions. In the one robot conditions, participants approach to a single robot and aim at initiating a dyadic interaction. Accordingly, the goal of this design of this condition is to investigate whether one-to-one F-Formation shapes (Fig. 1) occur in the virtual environment or not.

Figure 5: One-robot condition

Figure 6: Two-robots condition

In the two robots conditions, participants approach to two robots and form triads. This condition aims at examining the occurrence of one-to-one F-Formation shapes in triads. The main feature of these robots is that they are completely identical in order to suppress a likely effect which is caused by the saliency factor. Salient features of robots are beyond the scope of this study. Only kinemorphs are investigated in terms of their relation with F-Formation shapes.

3.1. The Procedure

Before starting the experiment, a consent form (see appendix) is introduced to the participants. Each participant participated all of the sessions in the experiment. There were three sessions of the experiment in which eye tracking calibration is done at the start of each.

3.1.1 The first session: Formations Only

First phase of the experiment includes both one-robot and two-robots condition. In one-robot condition, there are no formations; there is only one robot without any kinemorph, waiting for the participant to approach. In two-robots condition, the robots stand in different F-Formation shapes (Ciolek's taxonomy in Fig 1 above and Fig 7 below) in each session. F-formation shapes are worn by the robots as postures. Since it is not a motion study, there is only one F-formation shape in each session and there are no body movements in robots. Every F-Formation is introduced to participants once in a random order. Therefore, the total amount of sessions in the first phase of the experiment is seven, including one one-robot condition and six two-robot conditions. After the third and seventh sessions, a one-point calibration screen is interrupting the experiment for one second in order to check whether there is a slip in the calibration of the eyes or not.

This phase is designed to make participants see the robots, get used to the environment, realize the fact that robots are not standing in the same way but changing their rotations in every session. The data gathered from this session of the experiment was not included in the analysis. The reason is that all of the seven sessions in this part of the experiment is equal to the neutral sessions (the sessions without kinemorphs) of the other two part of the experiment.

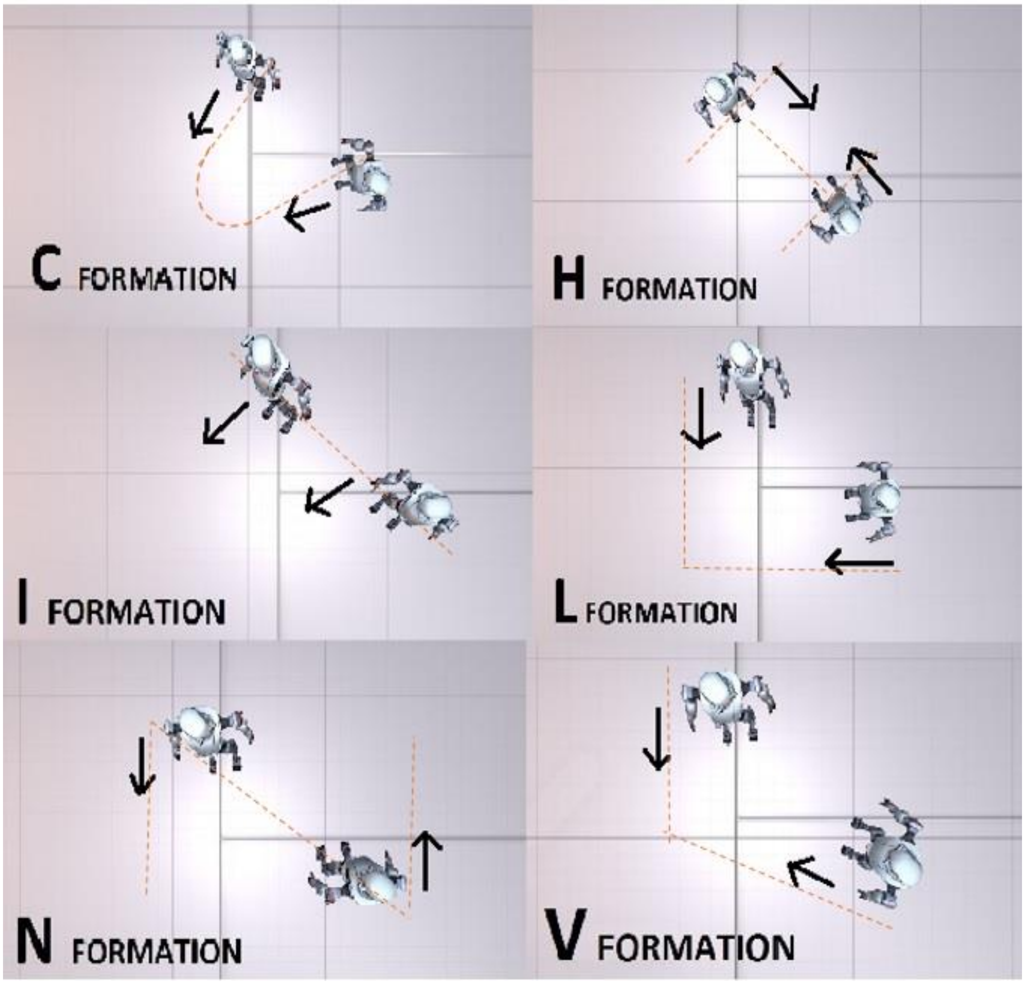

Figure 7: Robots on F-Formation shapes. Arrows demonstrate the trunk orientation of the robots. Red dashed lines demonstrate the F-formations

3.1.2 The second session : Formations and Eye Gaze

In the second phase of the experiment, only one of the kinemorphs, namely, eye gaze is used. However, since participants start each session at the other side of the room and it is impossible to see robot gazes from such distance, head rotation is used instead of eye movement. There are two states of head rotation in one-robot condition:

1. Robot is looking at the participant (Fig. 9)

2. Robot is not looking at the participant (Fig. 8)

Figure 8: One-robot condition/ robot is not looking at the participant (the robot which is closer to the camera is the participant)

Figure 9: One-robot condition/robot is looking at the participant (the robot which is closer to the camera is the participant)

In the state in which the robot is looking at the participant, the robot follows the participant wherever s/he goes. It is able to rotate its head in 360 degrees; therefore, if a participant places himself/herself behind the robots, robots are able to look at even that direction. (Fig. 10) Again, all instances were presented the participants only once, therefore, there are two sessions for the one-robot condition. These sessions are there in order to check the occurrence of one-to-one F-Formation shapes under the effect of head rotation.

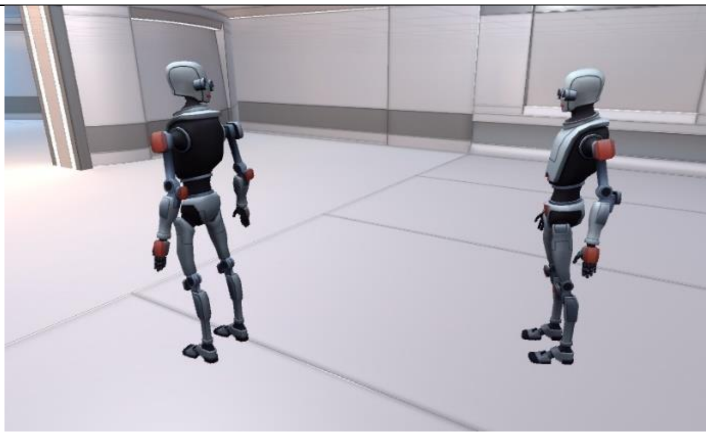

Figure 10: The robot in the left is ECA and the robot in the right is the participant. ECA is looking at the participant by means of head rotation.

In the two-robot conditions, the robots are standing in different F-Formations shapes (Fig.7) in each session. Head rotation is used as a kinemorph and there are four states of head rotation for two-robots setting:

1. Both robots are looking at the participant (Fig.11)

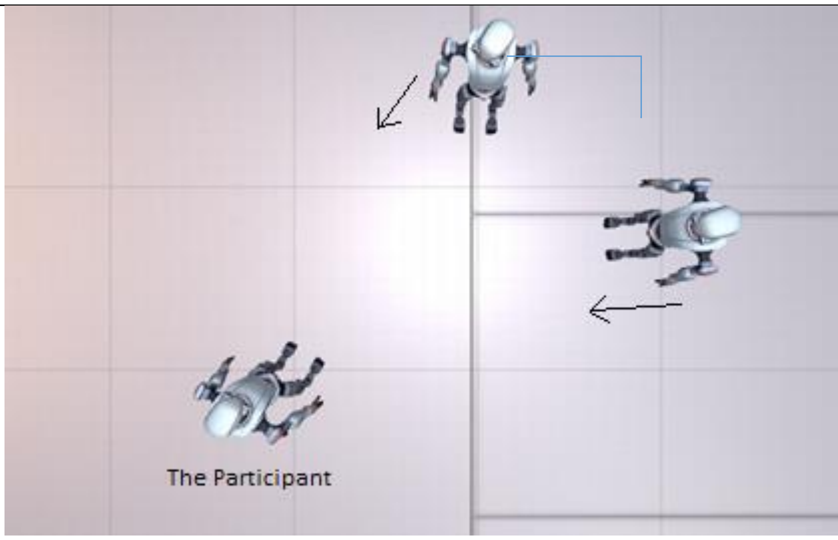

Figure 11: Both robots (right) are looking at the participants (leftmost) (robots are forming L shaped F-Formation and it is demonstrated with the blue line.)

2. None of the robots are looking at the participant (Fig.12)

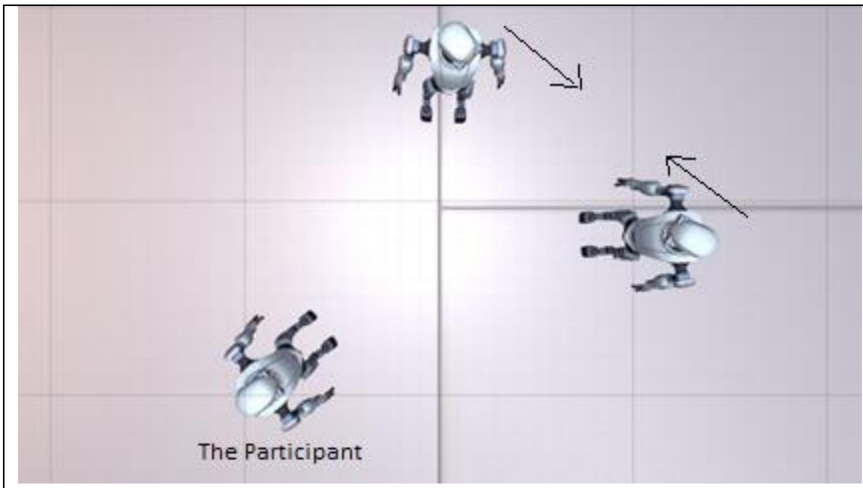

Figure 12: Both robots (right) are looking at the participants (leftmost) (robots are forming L shaped F-Formation)

3. The robot on the left is looking at the participant but the robot on the right is not looking at the participant (Fig. 13)

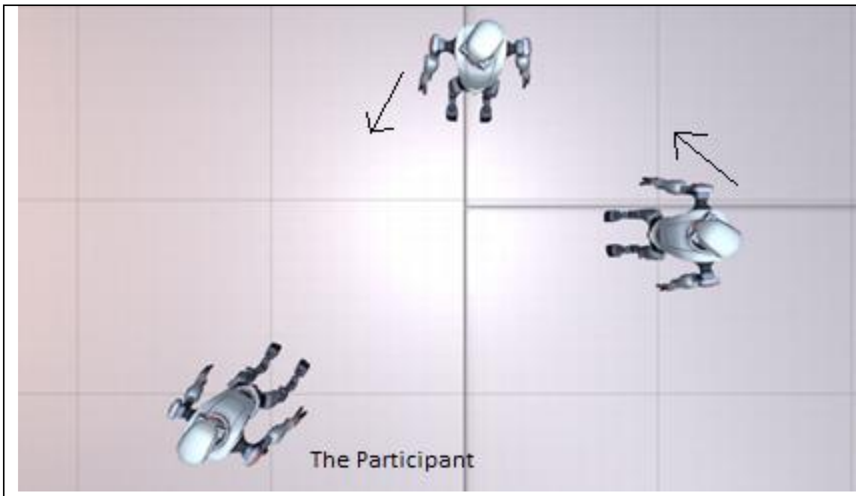

Figure 13: Robot in the up location is looking at the participant, but the robot in the down location is not looking at the participant (leftmost) (robots are forming L shaped F-Formation)

4. The robot on the right is looking at the participant but the robot on the left is not looking at the participant (Fig.14)

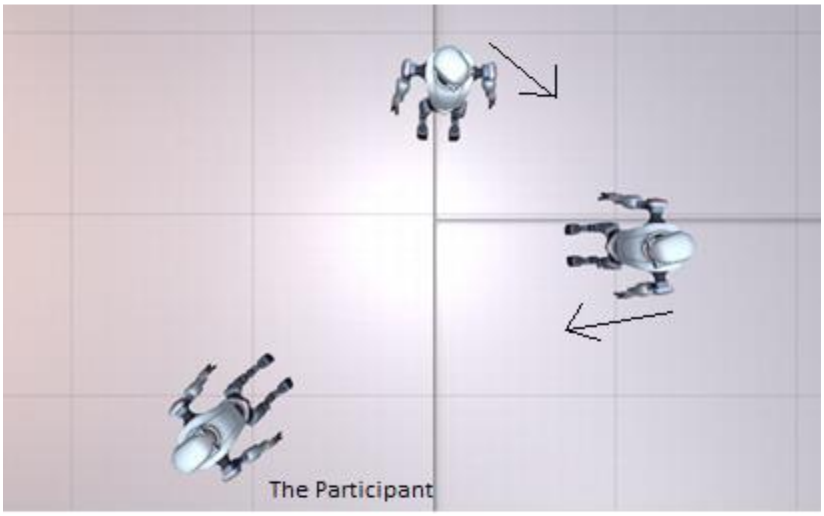

Figure 14: Robot in the up location is not looking at the participant but the robot in the down location is looking at the participant (leftmost) (robots are forming L shaped F-Formation)

With six formation shapes and four head rotation conditions, there are 6 x 4 = 24 two-robot conditions. Each condition was presented to participants only once in a random order; therefore, the total number of the sessions is 26, including two one-robot conditions and 24 two-robot conditions. In the seventh, eighteenth and twenty sixth sessions, experiment was interrupted by a one-point calibration screen.

This phase was designed to investigate the relationship between head rotation and F-Formation. More precisely, it is designed to answer the question of 'where to join a joint interactional space?' by introducing head rotation as a parameter. The conditions in which both robots are looking and not looking to the participants are there as neutral conditions. The actual effect is expected to achieve by the other conditions in which one robot is looking and the other one is not looking at the participants.

3.1.3 The Third Session : Formations and Lip Movement

The same details with eye gaze parameter (head rotation parameter) are the case for lip movement parameter. In one-robot condition, there two states of the lips:

1. Talking lips (Fig. 15)

2. Motionless, baseline lips (Fig. 15)

Figure 15: The rightmost robot is not talking and has the baseline scale of the lips. The leftmost robot is talking.

Each state was introduced to participants only once. Therefore, there are two sessions devoted to one-robot conditions. These sessions are there in order to check the occurrence of one-to-one F-formation shapes under the effect of lip movements. There is no rotation in the head; the only parameter is the lip movement.

In the two-robot conditions, again, six F-Formation shapes are used together with the lip movement states. There are four lip movement states in the two-robot conditions:

1. Both robots are talking
2. No robots are talking
3. The left robot is talking but the right robot is not talking
4. The left robot is not talking but the right robot is talking

Including the combination of four lip movement states and six F-formation shapes, the total number of the sessions is 24 in two-robots condition. If one-robot conditions are also added, it becomes 26 sessions. In the seventh, eighteenth and twenty sixth sessions, the experiment was interrupted by a one-point calibration screen.

3.1.4. Eye Tracking During the Experiment

At first, the environment was designed to be introduced to participants through a VR headset. However, since VR glasses might cause people to feel sick or dizzy while they are in the virtual reality, the experiment was changed and transformed into an experiment with a computer screen and an Eye Tribe eye tracker. Although Eye Tribe company was acquired by Facebook and the product is not available anymore, Eye Tribe Development Kit for unity and Eye Tribe Unity SDK for the integration of the unity with the Eye Tribe eye trackers is still available[1]. The scene which had been designed by the company and included in the Eye Tribe Development Kit conducted the calibrations at the start of each phase (Formation only, Formations and Eye Gaze, Formations and Lip Movements). Inspired by this calibration scene, a one-point calibration scene was produced and used between some sessions of these phases. As it is mentioned above, in the second and the third phases, after seventh, eighteenth and twenty fourth sessions, one-point calibration scene interrupted the experiment and gaze location was checked according to the calibration point. Since the number of the sessions in the first phase is different than the other two, one-point calibration canvas interrupted the experiment after the third and seventh phase at the first phase.

Eye gaze is collected only for detecting which robot was looked by the participants. Gaze distribution analyses was not conducted. The reason why eye gaze is important is that even though location and rotation of the participant is enough to see the shapes of the F-Formations, without eye gaze data it would be impossible to understand which robot was focused by the participant in a two-robots condition. Which robot was chosen to be interacted with is a crucial information because one-to-one F-Formation in a triad is thought to be between the participant and the robot that s/he looked.

3.2. Participants

Fourteen participants including students and academic personnel from Middle East Technical University and Bilkent University, Turkey join the study. There were seven male, seven female participants; however, gender differences were ignored and both genders were accepted to join the experiment. Each participant joined all the three phases of the study.

[1] Eye Tribe Development Kit and Eye Tribe Unity SDK is available in the following website: https://github.com/EyeTribe/tet-unity-devkit

3.3. The Description of the Collected Data

The left column in the table 2 includes the list of the data taken from the experiment. The right column includes the description of the data and some further information on how the data was collected.

Table 2: All the data that are collected during the two phases of the experiment

1. Robot that is looked	It is either “First” or “Second”. It reflects the robot that participants looked.
2. Time	Directly taken by ‘stopwatch.Elapsed’ command from Unity. It is collected to see how long participants stay in an F-formation
3. Robot’s Formation	It is valid only when participants encounter two robots in the environment. It reflects robot’s F-formation that participants are seeing on the screen. It changes in every session.
4. Participants’ location	Directly taken by ‘transform.position’ command from Unity. It reflects the position of the participant in vectoral form. It is updated every second.
5. Participants’ rotation	Directly taken by ‘transform.rotation’ command from Unity. It reflects the rotation of the participant in vectoral form. It is updated every second.
6. Robot’s speaking status	It is either “mouth-” (not speaking) or “mouth+” (speaking). It reflects the status of the robots. It is collected from the third phase of the experiment.
7. Robot’s eye status	It is either “eyes-” (not looking) or “eyes+” (looking). It reflects the status of the robots. It is collected from the second phase of the experiment.
8. Fixation point location	Fixation point belongs to the one-point calibration screen that interrupts the experiment after some of the sessions.

	This data reflects the position of the fixation point on the screen in a vectoral form.
9. Eye gaze location	It reflects the position of the 'indicator' that indicates the position of the eye gaze. The location of the eye gaze is provided by Eye Tribe.
10. The distance between fixation point and eye gaze location	This data is only collected when the one-point calibration screen, in other words, fixation point interrupts the experiment. It is there to check whether there is a slip in the calibration or not. It is expected to be below the value 2.0.
11. Keyboard tracking (up, down, left, right arrow buttons)	This data reflects the button which is pressed by the participants. It is collected in order to achieve the route that they used. It is meaningful when it is used with the location and rotation information.

3.4. Analysis

The data was collected from the last two phases of the experiment since the first phase was to make participants get used to the environment and the robots. The data which was collected from the last two phases and described in the above section is written on txt. files. Each phase had its own txt. data file, therefore, there were two files for each participant. Since there were fourteen participants, there were 28 txt. files; 14 for F-Formations and Eye Gaze phase, 14 for F-Formations and Lip Movement phase.

3.4.1. F-Formation and Eye Gaze analysis

After dividing each data file into 26 pieces, last location and rotation information in each session was taken from them. According to these information, participants last posture was reanimated in a robot form in the environment. Then, from bird's eye perspective, a photograph was taken. Reanimation and photography steps were repeated 26 times for each participant. Then, all the photographs were gathered up and categorized according to their F-Formation shapes. Finally, a visual data was achieved in the form of Fig. 16 which shows some of the photographs in the category of H-Formation. In this figure, there are three horizontal planes in the figure and each of them indicates different participants. The red dots indicate the robot which was looked by the participants. The smiling emojis indicate the participants. “--” means that both robots are not looking at the participant. “++” means that both robots are looking at the participant. “-+” means that the second robot looks at but the first robot does not look at the participant. “+-” means that the first robot is looking at the participant but the second robot is not looking at the participant.

After the final visual is provided, the F-formation which occurs between the participant and the robot s/he looks at is manually determined (self-annotated) and written down. Then, written F-Formations are counted to be put into a scheme in Microsoft Excel (Microsoft Office Professional Plus 2016, Version 16.0.11929.20144) file.

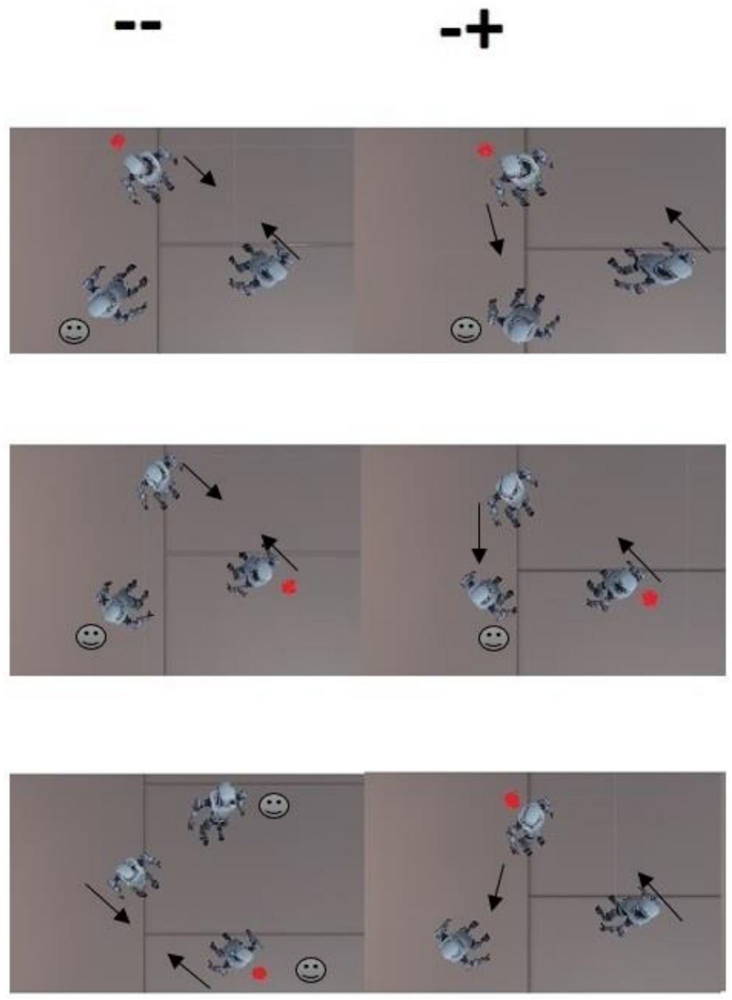

Figure 16: This figure describes the final version of the data for H-formation including the conditions in which none of the robots are looking at the participant and only the second robot is looking at the participant.

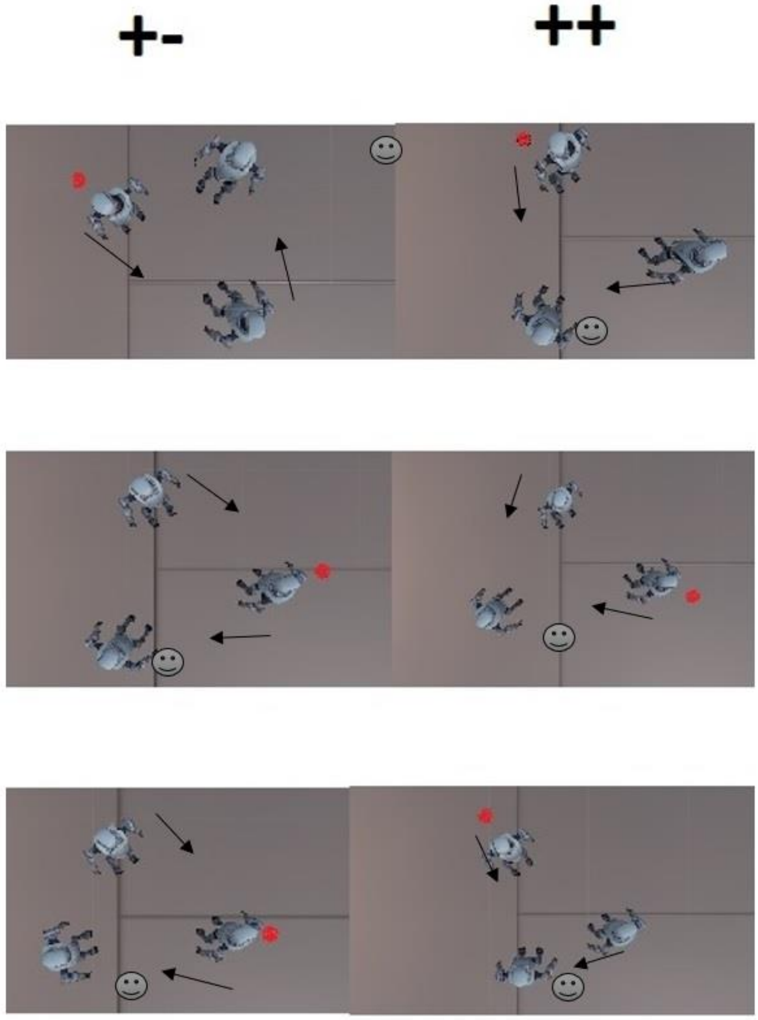

Figure 17: This figure describes the final version of the data for H-formation. It includes the condition in which the first robot is looking at the participant and the condition in which both robots are looking at the participant

3.4.2. F-Formation and Lip Movement analysis

The same procedure in the 3.4.1 was implemented for the F-Formations and Lip Movement phase of the experiment.

3.4.3. Statistical analysis

Schemes from lip movement and eye gaze analysis were reorganized in an Excel sheet (Fig. 17) with the other data describing independent variables and gaze location data. Although there were independent variables like robot states and robot formations in the experiment; in analysis, independent variables and dependent variables were treated as the same, since log-linear regression analysis is the most suitable analysis to the data.

	A	B	C	D	E	F	G
1	experiment phase	participantNR	robot State	symmetry	robot (participants looked at)	robot formations	participant formations
2	looking	0	0--0	symmetrical	first	C	V
3	looking	0	1--0	asymmetrical	second	C	V
4	looking	0	0--1	asymmetrical	second	C	V
5	looking	0	1--1	symmetrical	first	C	C
6	looking	0	0--0	symmetrical	second	H	L
7	looking	0	1--0	asymmetrical	first	H	V
8	looking	0	0--1	asymmetrical	second	H	L
9	looking	0	1--1	symmetrical	second	H	C
10	looking	0	0--0	symmetrical	second	I	N
11	looking	0	1--0	asymmetrical	first	I	V
12	looking	0	0--1	asymmetrical	second	I	H
13	looking	0	1--1	symmetrical	second	I	N
14	looking	0	0--0	symmetrical	first	L	N

Figure 18: Excel sheet including whole data

There are two goals of the analysis. The first goal is to see the distribution of the participants' gaze on robots according to the experiment phases and robot states. The second goal is to see the likelihood of the occurrence of a one-to-one F-Formation in a triad in which other two members of the triad are also forming a F-Formation shape. Basically, it is about the distribution of participants on F-Formation shapes in three different conditions, namely, robot states, experiment phase (lip movement and eye gaze phases), robots' F-formations. Therefore, the sheet was processed in JASP[2] (JASP 2019, Version 0.10.2) by two separate log-linear regression analysis.

The first log-linear analysis (saturated) was conducted with four factors, namely, robot states, experiment phase, robot's F-Formations and symmetricity. Symmetricity part includes the same data with the robot state part; however, it is a different expression of it. In this part, conditions in which robot states are similar are ascribed as symmetrical, while conditions in which one robot states are different are ascribed as asymmetrical. The second saturated log-linear analysis was conducted with three factors. These factors are, experiment phase (lip movement phase, eye gaze phase), robot states (0—0, 0—1, 1—0, 1—1) and the robot that participants looked at (first robot or second robot). After these log-linear analyses, significant associations were searched further by means of contingency tables. Both contingency tables and odds values are reported in Chapter 4.

[2] JASP 0.10.2 is available in the following website: https://jasp-stats.org/

CHAPTER 4

RESULTS AND DISCUSSION

The research questions of the present study and the methodology were presented in the third chapter. This chapter reports the results of the data analyses. The first research question was:

1. Do the kinemorphs influence the focus of the participants in a virtual social interaction?

The second question is:

2. If kinemorphs influence the focus of the participants, do they also have an effect on the rotation of the participants? More precisely, what are the F-Formations participants form with the agent they focus on in a virtual environment? (is there a difference between the types of F-formations occurred in the lip movement phase and the eye gaze phase of the experiment? Is there a difference between the types of F-formations occurred in real-time human-human interaction and in virtual human-robot interaction?)

4.1. The first Question: Do the kinemorphs influence the focus of the participants?

A three-way loglinear analysis produced significant partial associations. Firstly, the experiment phase and the robot that the participants focused on has an association (x^2 (1) = 5.062, p< .05). Secondly, the robot states and the robot that participants looked at has an association (x^2 (3) =17.705, p =< .001). In order to understand these associations, separate chi-square tests were conducted.

Table 3: Contingency table that describes the relation between robot state and the robot that participants focus on. In this table, all the data coming from eye gaze phase and the lip movement phase is included.

	Robot (looked at)		
Robot State	**First**	**Second**	**Total**
0--0	96	70	166
0--1	40	124	164
1--0	133	33	166
1--1	88	73	161
Total	357	300	657

Chi-Squared Tests

	Value	**df**	**p**
X^2	104.577	3	< .001
N	657		

Table 3 is a contingency table which describes the frequency between the robot states and the robot that participants focused on. "1—1" in the "Robot State" column means that both robots are looking at the participant; "0—1" means that the first robot is not looking at the participant; "1—0" means that the second robot is not looking at the participant; "0—0" means that none of the robots are looking at the participant. This table is broken down into 2x2 contingency tables in order to see the odds ratio between the robot states and the participants' decisions:

- Odds of focusing on the first robot in the state 0—0 (both robots are not looking or both robots are not speaking, together) is %77 higher (4.28 higher) than the state 0—1 (first robot is not looking or speaking)

- Odds of focusing on the first robot in the state 0—0 is %66 lower (0.34 lower) than in the state 1—0 (second robot is not looking or speaking).

- Odds of focusing on the first robot in the state 0—0 is %14 higher (1.14 higher) than in the state 1—1 (both robots are looking or speaking)

- Odds of focusing on the first robot in the state 0—1 is %92 lower (0.08 lower) than in the state 1—0.

- Odds of focusing on the first robot in the state 0—1 is %75 lower (0.25 lower) than in the state 1—1.

- Odds of focusing on the first robot in the state 1—0 is %70 higher (3.35 higher) than in the state 1--1.

Accordingly, the results reveal that there was not much difference between the odds of focusing on the first robot in the states 1—1 and 0—0; however, the difference was large between the state 0—1 and 1—0. The first robot was focused on more than the second robot in the states 1—0, 0—0 and 1—1. For the state 1—0, this result could be expected; however, for the states 0—0 and 1—1, it was not expected. A discussion of the results presented in chapter 5. As a general scheme, odds of looking at the first robot in symmetrical states (0—0 and 1—1) is much lower than 1—0 state in which the first robot is active (i.e., either gazing at the participant or the lips moving). On the other hand, the difference between the number of participants who looked at the first robot and the number of the participants who looked at the first robot was so small. In addition to this, there was a large difference between the states 1—0 and 0—1. These two results reveal that participants tend to look at the active participants.

Table 4 (below) is the contingency table which describes the frequency between the experiment phase and robot that participants focused on. This is a 2x2 contingency table, therefore, it does not require to be broken down into pieces.

Table 4: Contingency table that describes the relationship between experiment phase and the robot that participants focused on.

	Robot (looked at)		
Experiment phase	**First**	**Second**	**Total**
looking	166	163	329
speaking	191	137	328
Total	357	300	657

Chi-Squared Tests

	Value	**df**	**p**
X^2	4.003	1	0.045
N	657		

- Odds of looking at the first robot in the eye gaze phase of the experiment is %22 lower (0.78 lower) than in the lip movement phase of the experiment.

In the eye gaze phase of the experiment the number of the participants who looked at the first robot was almost equal to the number of the participants who looked at the second robot. Conversely, in the lip movement phase, the odds of the participants who looked at the first robot was %39 higher than the number of the participants who looked at the second robot. Therefore, the effect between the robot state and the robot that participants focused on might be influenced by this condition. More frankly, it is not clear that if there is a difference between the effect of lip movement and the effect of eye gaze on participants focused. Therefore, robot state-robot(focused) association needs to be broken down into two pieces in order to see whether the effect is different in the eye gaze phase than in the lip movement phase.

As it can be seen from the table 5, eye gaze phase results are different from the result which includes eye gaze phase and lip movement phase together (Table 3.). When two phases of the experiment are analyzed together, there is a tendency to look at the first robot. Only in the state of 0—1, where the second robot is active, this tendency disappears. In the state 1—0, where the first robot is active, the frequency of looking at the first robot increases when it is compared with the states 0—0 and 1—1. However, when only the eye gaze phase is analyzed, there is no tendency to look at the first robot.

Table 5: Contingency table that describes the relationship between the robot state and the robot that participants focused on only for the eye gaze phase of the experiment.

	Robot (looked at)		
Robot State	**First**	**Second**	**Total**
0--0	44	39	83
0--1	17	66	83
1--0	67	16	83
1--1	38	42	80
Total	166	163	329

Chi-Squared Tests

	Value	**df**	**p**
X^2	60.744	3	< .001
N	329		

- In state 0—0, the frequency of the first robot is %13 higher than the second robot.

- In state 0—1, the frequency of the first robot is %74 lower than the second robot.

- In the state 1—0, the frequency of the first robot is %77 higher than the second robot.

- In the state 1—1, the frequency of the first robot is %10 lower than the second robot.

As it can be seen from the ratios, in the states 0—0 and 1—1, participants frequency to look at the first robot was different than each other; however, still, the frequency of looking at the first robot and the frequency of looking at the second robot was so close to each other as expected. The difference increases in states 1—0 and 0—1. When the first robot is looking at the participants, participants look at the first robot in %74 of the time; when the second robot is looking at the participants, participants look at the second robot in %77 of the time.

When two phases of the experiment were analyzed together the effect described in the above paragraph decreases and the tendency to look at the first robot occurs. Therefore, in the lip movement phase of the experiment, the effect was lower than the eye gaze phase. Table 6. demonstrates this affect.

Table 6: Contingency table that describes the relationship between the robot states and the robot that participants focused on only for the lip movement phase of the experiment.

	Robot (looked at)		
Robot State	**First**	**Second**	**Total**
0—0	52	31	83
0—1	23	58	81
1—0	66	17	83
1—1	50	31	81
Total	191	137	328

Chi-Squared Tests

	Value	df	p
X^2	46.183	3	< .001
N	328		

Table 6 demonstrates the relationship between robot states and the robot that participants looked at for the lip movement phase of the experiment:

- Odds of looking at the first robot in the state 0—0 is %68 higher (1.68 higher) than looking at the second robot.

- Odds of looking at the first robot in the state 0—1 is %60 lower (0.4 lower) than looking at the second robot.

- Odds of looking at the first robot in state 1—0 is %74 higher (3.88 higher) than looking at the second robot.

- Odds of looking at the first robot in state 1—1 is %60 higher (1.6 higher) than looking at the second robot.

These results reveal that the tendency to looking at the first robot is coming from the lip movement phase. Both asymmetrical states (0—1 and 1—0) have lower differences and symmetrical states (0—0 and 1—1) have higher differences. The effect of focusing on the active robot decreases in this phase, although it is still significant. As a result, both kinemorphs seem to determine the focus of the participants in a social interaction; however, the effect of gaze is higher than lip movement.

4.2. The Second Question: what are the F-Formations participants form with the agent they focus on in a virtual environment?

4.2.1. Two-robots condition

A four-way loglinear analysis revealed significant partial associations. Firstly, the experiment phase and the participant formations had a significant interaction (X^2 (5) = 11.751, p< .05). Secondly, symmetricity and robot formations had a significant interaction (X^2 (5) = 26.061, p< .001). Thirdly, robot formations and participant formations had an association (X^2 (22) =139.230, p =< .001). In order to understand these associations, separate chi-square tests were conducted.

According to table 7, V and L formations occurred more frequently than other formations in the eye gaze phase of the experiment, while H, V and C formations occurred more frequently than other formations in the lip movement phase of the experiment. This result shows that F-Formations shapes are affected by the kinemorphs. In order to clear the effect of the kinesics and see other effects by themselves, other interactions (robot formations*participant formations, experiment phase*participant formations) need to be examined by separating the data file into two pieces (one data file for eye gaze, one data file for lip movement).

Table 7: Contingency table that describes the relationship between the distribution of formations and the phases of the experiment

	Participant Formations						
Experiment Phase	**C**	**H**	**L**	**N**	**V**	**none**	**Total**
Eye gaze	55	57	70	35	101	11	329
Lip movement	71	88	39	39	87	4	328
Total	126	145	109	74	188	15	657

Chi-Squared Tests

	Value	df	p
X^2	22.000	5	< .001
N	657		

Table 8 shows the relationship between robot formations and the participants' formations with the robot they look at in the eye gaze phase of the experiment. According to this table:

- While robots are standing in C formation, participants tend to use V formation.
- While robots are standing in H formation, participants tend to use L formation.

- While robots are standing in I formation, participants tend to use V and N formations.

- While robots are standing in L formation, none of the F- formations reach %30; however, participants tend to use V, H and C formations above %20 of the times.

- While robots are standing in N formation, participants tend to use L formation.

- While robots are standing in V formation, participants tend to use V formation.

Total amount of L and V formations are higher than the other formations and I formation does not even have a single instance.

Table 8: Contingency table that describes the relationship between the participant formations and the robot formations in the eye gaze phase of the experiment

	Participant formations						
Robot formations	**C**	**H**	**L**	**N**	**V**	**none**	**Total**
C	9	16	2	8	19	0	54
H	10	0	26	2	13	4	55
I	1	16	0	18	20	0	55
L	11	14	8	5	15	1	54
N	14	3	22	1	11	5	56
V	10	8	12	1	23	1	55
Total	55	57	70	35	101	11	329

Chi-Squared Tests

	Value	df	p
X^2	138.067	25	< .001
N	329		

Table 9 demonstrates that different robot formations lead participants to form different F-formations with the robot they choose to interact.

- While robots are standing in C formation, participants tend to form H formation.

- While robots are standing in H formation, participants tend to form V formation.

- While robots are standing in I formation, participants tend to form V and N formations.

- While robots are standing in L formation, participants tend to form H formation.

- While robots are standing in N formation, participants tend to form C formation.

- While robots are standing in V formation, participants tend to form H formation.

Table 9: Contingency table that describes the relationship between the participant formations and the robot formations in the lip movement phase of the experiment

	Participant formations						
Robot formations	**C**	**H**	**L**	**N**	**V**	**none**	**Total**
C	8	24	1	11	11	0	55
H	13	3	15	1	18	0	50
I	2	13	0	18	23	0	56
L	15	17	3	6	14	1	56
N	18	12	15	1	8	2	56
V	15	19	5	2	13	1	55
Total	71	88	39	39	87	4	328

Chi-Squared Tests

	Value	**df**	**p**
X^2	118.450	25	< .001
N	328		

The distribution of participants formations among robot formations are described above. The total amount of H and V formations are higher than the other formations and I formation does not have a single instance. In order to make the comparison between two phases of the experiment easier to follow, the distribution of the dominant F-formations formed by the participants is listed according to the experiment phases they occurred. (see Table 10)

Table 10: Table describes the dominant F-formation shapes formed by participants in the eye gaze and lip movement phases of the experiment.

Robot formations	Eye gaze	Lip movement
C	V,	H
H	L	V
I	V, N	V, N
L	V, H, C	H
N	L	C
V	V	H

In the table 10, lip movement and eye gaze stand for the kinemorphs and the F-formation shapes stand for the proxemics features. When robots are having an eye contact with the participants, V (%31) and L (%21,3) shaped F-formations are most frequently used among six F-formation shapes; when robots are talking to each other, V(%26.5) and H(%27) are more frequently used than the other F-formation shapes (see table 8 And 9). However, there are also other shapes occurred in a frequency which should be taken into consideration. When their occurrence is analyzed with robot formations as in the Table 10, a pattern reveals itself. For example, in the eye gaze phase, if robots are standing in N formation, participant tend to form L formation with the robot they choose to interact, while in lip movement phase they tend to form C formation. Except for the state in which robots are standing in I formation, all robot formations seem to be coupled with different participant formations in different kinesics states (experiment phases). However, the problem is that although robot formations are coupled with different participant formations 'between' the kinesics states, some robot formations coupled with the same participant formations 'within' the kinesics states. It means that when robot formation is C, participants tend to form V formation in eye gaze phase, while they tend to form H formation in the lip movement phase; however, participants tend to form V formation, while robots are in C formation and N formation in the eye gaze phase. It means that there must be another pattern which can explain the relationship between the robot formations and participant formations. In that sense, analyzing the relationship between symmetricity of the robot states and the participant formations with respect to the kinesics states seem to be an acceptable strategy. The reason is that in symmetrical cases, participants tend to decide the vaguely on which robot to interact with and this decision becomes much more precise in asymmetrical cases. (see 4.1) Symmetricity of the robot states might also have an effect on the F-formation shapes. 4.2.1.1 and 4.2.1.2 describes the relationship between symmetricity of the robot states and the distribution of participants' F-formation shapes on F-formation shapes of the robots by focusing on the difference between the experiment phases.

4.2.1.1 Eye Gaze Phase

Table 11: Table describes the distribution of the participant formations according to the robot states in C formation of eye gaze phase.

First Character	Second Character	C	N	H	V	L	I
-	-	%6	%10	%2	%8	%0	%0
+	+	%6	%6	%4	%6	%2	%0
-	+	%0	%0	%10	%14	%2	%0
+	-	%2	%4	%16	%4	%2	%0

The results that were reported in Table 10 (previous section) shows that V formation is formed by the participants, while the robots are standing in C formation. However, the reason why V formation looks like the dominant F-formation may be that it has a great deal of occurrences in the symmetrical cases (-/-, +/+). When it comes to the asymmetrical cases which lead participants to interact with one of the robots (see Section 4.1), C is not the dominant F-formation, instead H and V are the dominant F-formations. As it can be seen from the table 11, in symmetrical cases, participants are distributed widely on the F-formations; however, in asymmetrical cases H and V formations constitute %81 of the cases.

Table 12: Table describes the distribution of the participant formations according to the robot states in the H formation of eye gaze phase

First Character	Second Character	C	N	H	V	L	I
-	-	%1.9	%3.8	%0	%3.8	%13.4	%0
+	+	%3.8	%0	%0	%5.7	%13.4	%0
-	+	%9.6	%1.9	%0	%3.8	%11.5	%0
+	-	%5.7	%0	%0	%9.6	%11.5	%0

The results that were reported in Table 10, the dominant F-formation shape is L formation. Results do not seem to be changed, even if the frequency of C and V formation increases and gets closer to L formation in the asymmetrical cases. The reason why H formation was not formed by the participants might be that it is physically unfeasible. If two robots are standing in H formation, participants should be able to move one of the robots and take the place of it. Since the design of the experiment did not allow participants to move the robots from their places, the frequency of the H-formation formed by the participants is equal to zero.

Table 13: Table describes the distribution of the participant formations according to the robot states in the I formation of eye gaze phase

First Character	Second Character	C	N	H	V	L	I
-	-	%1.8	%9.1	%7.2	%7.2	%0	%0
+	+	%0	%10.9	%7.2	%5.4	%0	%0
-	+	%0	%7.2	%9.1	%9.1	%0	%0
+	-	%0	%3.6	%5.4	%16.3	%0	%0

In table 10, the most frequently formed F-formations are V and N, when robots are standing in I formation. On the contrary, table 13 successfully illustrates why N formation seems like a dominant F-formation for participants. In the symmetrical states (-/- and +/+), N formation has frequent occurrences; however, the frequency of it decreases in the asymmetrical states. Instead of N formation, H and V formations increases in the asymmetrical states.

Table 14: Table describes the distribution of the participant formations according to the robot states in the L formation of eye gaze phase

First Character	Second Character	C	N	H	V	L	I
-	-	%9.2	%1.8	%7.4	%5.5	%1.8	%0
+	+	%7.4	%0	%0	%11.1	%5.5	%0
-	+	%3.7	%3.7	%3.7	%7.4	%7.4	%0
+	-	%0	%1.8	%12.9	%7.4	%1.8	%0

Table 10 shows that V, H and C formations are the most frequently formed F-formations (although none of them was able to achieve %30 of occurrence). In asymmetrical cases, the frequency of C formation decreases. Instead of C formation, H formation increases. Again, the most dominant F-formations are H and V formations.

Table 15: Table describes the distribution of the participant formations according to the robot states in the N formation of eye gaze phase

First Character	Second Character	C	N	H	V	L	I
-	-	%3.8	%1.9	%3.8	%1.9	%11.5	%0
+	+	%7.7	%0	%0	%3.8	%15.4	%0
-	+	%5.7	%0	%1.9	%9.6	%9.6	%0
+	-	%7.7	%0	%0	%5.7	%9.6	%0

According to Table 10, participants tend to form L formation with the robot they choose to interact, while the robots are standing in N formation. As it can be followed from table 15, the reason why L formation looks much more dominant than the others is that it has many occurrences in symmetrical cases. However, in asymmetrical cases, C and V formations seem to increase, while L formation decreases. Although the frequency of L formation decreases, still C, V and L formation seem to be the dominant F-formations.

Table 16: Table describes the distribution of the participant formations according to the robot states in the V formation of eye gaze phase

First Character	Second Character	C	N	H	V	L	I
-	-	%9.4	%1.9	%1.9	%5.8	%5.8	%0
+	+	%5.8	%0	%1.9	%9.4	%5.8	%0
-	+	%1.9	%0	%1.9	%13.2	%9.4	%0
+	-	%1.9	%0	%7.5	%13.2	%3.8	%0

According to table 10, participants tend to form V formation with the robot they choose to interact, while robots are standing in V formation. Table 16 confirms its result.

4.2.1.1.2 Interim Summary of the Gaze Phase Results

Although H and V formations are the dominant F-formation shapes among others when the robots have asymmetrical gaze states, L and C formation has also dominancy in two of the cases. When the robots were standing in H formation, the participants tended to form L and V formation with the robot they focused on. Also, when the robots were standing in N formation, the participants tended to form V, L and C formations with the robot they focused on. Other than these two cases, there is no difference between the distribution of the participants' F-formations on robots F-formations. However, all six cases revealed that in symmetrical robot states, the participants were distributed widely on the F-formations; however, in asymmetrical states, F-formation they tended to form became more specific.

4.2.1.3 Lip Movement Phase

Table 17: Table describes the distribution of the participant formations according to the robot states in the C formation of lip movement phase

First Character	Second Character	C	N	H	V	L	I
-	-	%3.6	%7.2	%9.1	%3.6	%1.8	%0
+	+	%5.4	%5.4	%7.2	%5.4	%0	%0
-	+	%1.8	%3.6	%12.7	%7.2	%0	%0
+	-	%3.6	%3.6	%14.5	%3.6	%0	%0

Table 17 confirms the result of Table 10. H formation is the most dominant F-formation shape in the asymmetrical robot states.

Table 18: Table describes the distribution of the participant formations according to the robot states in the H formation of lip movement phase

First Character	Second Character	C	N	H	V	L	I
-	-	%5.7	%0	%5.7	%9.6	%7.7	%0
+	+	%11.5	%1.9	%0	%5.7	%3.8	%0
-	+	%3.8	%0	%0	%9.6	%9.6	%0
+	-	%3.8	%0	%0	%15.4	%5.7	%0

Table 18 confirms the result of table 10 in the sense that the most dominant F-formation formed by the participants in the asymmetrical robot states were V formation; however, in addition to it, L formation constituted %32 of the all F-formation occurrences in the asymmetrical robot states.

Table 19: Table describes the distribution of the participant formations according to the robot states in the I formation of lip movement phase

First Character	Second Character	C	N	H	V	L	I
-	-	%0	%12.7	%5.4	%7.3	%0	%0
+	+	%1.8	%7.3	%7.3	%7.3	%0	%0
-	+	%1.8	%5.4	%5.4	%12.7	%0	%0
+	-	%0	%7.3	%5.4	%12.7	%0	%0

According to table 10, N formation were one of the most frequently formed F-formations along with the V formation; however, Table 19 shows that the dominancy of N formation is coming from the fact that it has many occurrences in symmetrical states. In asymmetrical states of the robots, N formation does not seem to achieve the quantity of V formation. On the other hand, the frequency of N formation decreases in asymmetrical states.

Table 20: Table describes the distribution of the participant formations according to the robot states in the L formation of lip movement phase

First Character	Second Character	C	N	H	V	L	I
-	-	%10.9	%1.8	%7.3	%5.4	%0	%0
+	+	%9.1	%3.6	%5.4	%3.6	%1.8	%0
-	+	%3.6	%3.6	%9.1	%7.3	%1.8	%0
+	-	%3.6	%1.8	%9.1	%9.1	%1.8	%0

Table 10 shows that the most frequent F-formation shapes of the participants were H formation, while the robots were standing in L formation. According to Table 20, there is a widely distributed F-formation shapes in symmetrical states. In the asymmetrical states, H and V becomes more frequent than the others.

Table 21: Table describes the distribution of the participant formations according to the robot states in the N formation of lip movement phase

First Character	Second Character	C	N	H	V	L	I
-	-	%11.1	%1.8	%1.8	%3.7	%7.4	%0
+	+	%11.1	%0	%3.7	%5.5	%5.5	%0
-	+	%3.7	%0	%9.3	%1.8	%7.4	%0
+	-	%7.4	%0	%7.4	%3.7	%7.4	%0

While the robots were standing in N formation, the participants tended to form C formation with the robot they wanted to interact, according to Table 10. As it is described many times above, the abundance of C formation is caused by the symmetrical states. Conversely, in asymmetrical states, this abundance disappears. Conversely, H and L formations seem to increase in asymmetrical states.

Table 22: Table describes the distribution of the participant formations according to the robot states in the V formation of lip movement phase

First Character	Second Character	C	N	H	V	L	I
-	-	%9.1	%0	%7.3	%3.6	%3.6	%0
+	+	%9.1	%1.8	%5.4	%7.3	%1.8	%0
-	+	%3.6	%0	%9.1	%9.1	%3.6	%0
+	-	%5.4	%1.8	%14.5	%3.6	%0	%0

According to the results described in Table 10, H formation is the most frequently occurred F-formation shape, while robots are standing in V formation. The results in Table 22, confirms this result since there is an increase in asymmetrical states of H formation and it seems that it is still the most frequent F-formation among them.

4.2.1.4 Interim Summary of the Lip Movement Phase Results

The results of the lip movement phase show similar patterns with the gaze phase. Again, H and V formations were the dominant F-formation shapes among others, when the robots had asymmetrical gaze states. L formation had also a dominancy, when the robots were standing in H and N formations. All of the cases demonstrate that in symmetrical robot states, the participants were distributed widely on the F-formations; however, in asymmetrical states, F-formation they tended to form became more specific.

4.2.1.5 Evaluation of the Two-Robots Condition Results

All the tables starting from table 11 to Table 22 describes a pattern in which the distribution of the F-formation shapes formed by participants changed according to the symmetricity of the robot states. Participants' distribution on the F-formations has a wide range in symmetrical states; conversely, its range gets smaller in the asymmetrical states. The narrowed range of the asymmetrical states are shown in the Table 23 below.

Table 23: Table describes the dominant F-formation shapes formed by participants in the asymmetrical states.

Robot formation	Eye gaze	Lip movement
C	H, V	H
H	L	V, L
I	H, V	V
L	H, V	H, V
N	C, V, L	H, L
V	V	H

The findings presented in Table 23 suggest that both in eye gaze and lip movement phases, there is no much variety in F-formations. Mostly H and V formations, few of the times C and L formations occurred. Moreover, it seems hard to say that a particular robot formation leads participants to form a particular F-formation shape. Three questions arise from the results so far:

1. While there are six F-formation shapes in real-life one-to-one interactions, why only five of them occurred in the virtual environment?

2. Concerning the frequency of occurrences, why specifically two of the F-formations have the dominance over other F-formations? Does it make sense that people tend to use H and V formations more frequently than other formations in virtual environment?

3. Do people tend to use a narrower list of F-formations, while they are trying to join a dyadic interaction as a third interactant?

As an answer to the first question concerning the absence of I formation in the virtual environment, it can be claimed that the absence of I formation is caused by the kinemorphs that is chosen to be examined. The reason is that I formation requires a position in which two interactants stand near each other and looking the same direction (forward). Therefore, in this F-formation, interactants are not able to see each other. This situation makes it hard to use I-formation in an interaction consists of visual cues. If olfaction, thermal or vocal codes (Agnus, 1968) were to be examined in the research, the occurrence of I formation would be meaningful; however, it is not a suitable way to interact through visual codes.

As an answer to the second and third questions concerning the frequency of the occurrences of the F-formations, it can be claimed that both of the possibilities seem plausible. The high frequency in the occurrence of H and V formations might be because there is a difference between dyads and triads and also it might be because there is a difference between human-robot interaction in a virtual environment and human-human interaction in real life. In order to understand whether there is a difference between dyadic and triadic interactions, dyadic interactions in a virtual environment should be examined. (see 4.2.2) If there is difference between dyadic and triadic interactions concerning F-formations and all formations occur in dyads, then it is hard to say that the dominance of H and V formations is caused by the difference between human-human and human-robot interaction. Conversely, if there is no difference between dyads and triads, it becomes meaningful to claim that the narrowed list of F-formations is used in human-robot interaction in the virtual environment. Therefore, two findings are certain:

1. H and V formations are the most frequently formed F-formations by the participants. V formation is much more frequent than H formation in eye gaze phase, while H formation has only one more occurrence than V formation in the lip formation phase.

2. If the robot states are symmetrical, the range of the F-formation distribution is wide; if robot states are asymmetrical, the range of the F-formation distribution gets smaller.

In this section, eye gaze phase and lip movement phase of the two-robots condition were compared by focusing on the relationship between the symmetricity of the robot states and the distribution of the participants' F-formation shapes on the robots' F-formation shapes. In the following section, almost the same thing is done; however, there is a single difference. Section 4.2.2 describes the results of the one-robot condition.

4.2.2 One-robot Condition

4.2.2.1 Eye Gaze Phase

According to Table 24, there are V, H and N formations in the eye gaze phase of the experiment, while there is only one robot in the scene and participants join the interaction as the second interactant. In one-to-one human-robot interaction in the virtual environment, most frequently formed F-formation shape (%78) is V formation for the eye gaze kinemorph

Table 24: Table describes the eye gaze phase of the experiment. There are two states in the one-robot condition; robot is not looking at the participant and robot is looking at the participant. Each line refers to a participant. There are 14 lines.

Looking	Not Looking
H	V
V	V
V	V
V	V
V	V
V	N
H	V
V	V
V	V
H	V
H	H
V	V
N	V
V	V

4.2.2.2 Lip Movement Phase

In Table 25, there are V, H and C formations in the lip movement phase of the experiment, while there is only one robot in the scene and participants join the interaction as the second interactant. In one-to-one human-robot interaction in the virtual environment, most frequently formed F-formation shape (%68) is H formation for the lip movement kinemorph.

Table 25: Table describes the lip movement phase of the experiment. There are two states in the one-robot condition; robot is not speaking and robot is speaking. Each line refers to a participant. Therefore, there are 14 lines.

Speaking	Not Speaking
H	H
H	H
V	V
H	H
H	V
H	C
H	H
H	V
H	H
V	H
V	V
H	H
V	H
H	H

Although there is a difference between lip movement and eye gaze, there are no major differences between the levels of these two kinemorphs. In the eye gaze phase, V formation seems to be the dominant F-formation shape, while in the lip movement phase, H formation seems to be the dominant F-formation shape. However, there are no differences between the conditions in which the robot is looking at the participants and it is not looking at the participants. Also, there are no differences between the conditions in which the robot is speaking and the robot is not speaking. This part requires a further research; however, results of one-robot condition reveal that H and V formations are the dominant F-formation shapes in the virtual environment. On the other hand, L and I formations do not even have a single instance. The lack of L formation might be because of the small sample size (14 participants) or there might be a difference between the length of the F-formation list used in triads and dyads. Both possibilities require further researches; however, two findings are certain:

1. In the eye gaze phase, V formation is the most frequently formed one by the participant and the robot agent.

2. In the lip movement phase, H formation is the most frequently formed by the participant and the robot agent.

4.3. Deviant Shapes

In the present study, some shapes that are not found in the human-human interaction occurred; however, they constituted only %2,3 of the all trials. It means that in 15 among 657 sessions, deviant formations occurred. Fig 17 (below) demonstrates these abnormal shapes. The states of the robots in Fig 17 is described in the following bulleted list.

- In the first, fifth and the last images, first robot is looking at the participant and the second robot is not.

- In the second, third, fourth and sixth images, no robots are looking at the participant.

- In the seventh image, both robots are looking at the participant.

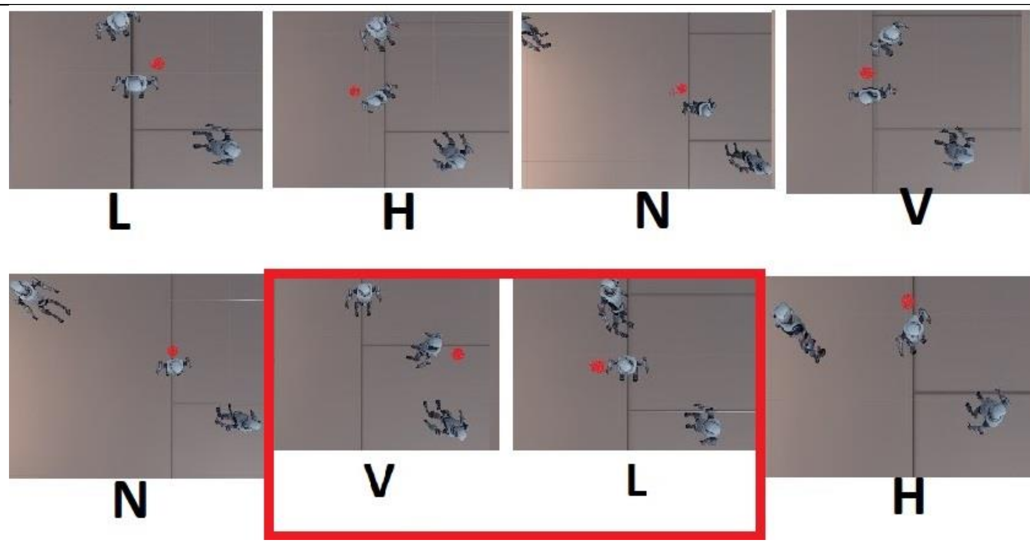

Figure 19: Two images in the borders of the red square are the F-formation shapes occurred in the lip movement phase and the rest of the images belong to the eye gaze phase. Red dots in each image indicates the robot which is focused by the participant

The choice of “collider” in the Unity Engine might be the cause of these deviant formations. In order to create an interactable game object in Unity, a component called “collider” must be added to the game object. The collider can be in various shapes; however, “box collider” is easier to render when it is compared with the others because it renders a more generic shape. The outcome of this ease is that rays coming from participants’ gaze might hit the collider of the wrong robot and might not find its actual target. As it can be followed from Fig 18, the magnitude of the box collider (see Fig. 18) prepares a convenient condition for such a problem. Images in the Fig 17 are also candidates for this problem because of the robot’s distances from each other. All of the abnormal shapes might be because collider of the wrong robot prevents the target robot from getting the ray hit. The reason is that if the rays had hit the other robot in these abnormal cases, participants would be forming one of the six defined F-formation shapes.

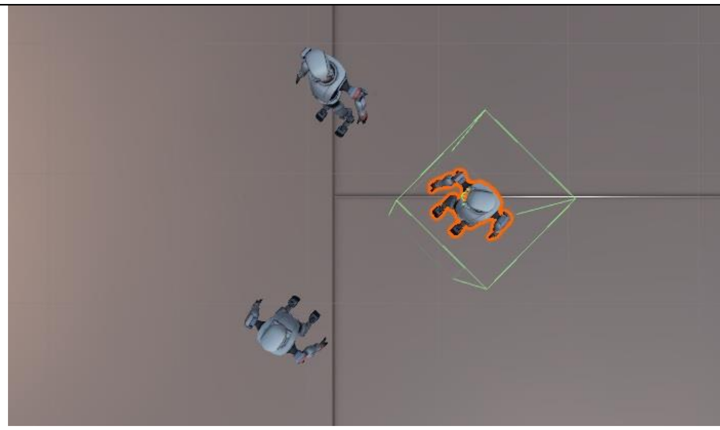

Figure 20: The yellow lines describes the magnitude of the box collider

Accordingly, the deviant formations can be conceived as outliers in the present study. The reason is that the number of the deviant formations is 15 among 657 formations, as it is mentioned in the first sentences of the present section. In further studies, instead of a "box collider", a "mesh collider" can be used in order to define the border of the robot bodies better.

4.4. Interrater reliability

Since the F-formation shape data is self-annotated in the present study, a volunteer researcher also annotated the %25 of the visuals in order to conduct an interrater reliability test. Cohen's Kappa analysis is conducted between the self-annotated data and the data annotated by the volunteer. There is a substantial level of agreement between the self-annotated data and the data annotated by the volunteer, $\kappa = .715$, $p < .0005$.

CHAPTER 5

CONCLUSION

This is a study to investigate how kinemorphs influence F-formation shapes in a virtual environment by focusing on the occurrences of one-to-one F-formation shapes in a social interaction setting. Whether the shape of a social group's spatial arrangement is influenced by kinemorphs was the first research question. In order to answer this question, two facial kinemorphs (head rotation, lip movement) on robot faces were introduced to participants in a virtual environment. The results revealed that participants tend to look at the robots with the kinemorphs. Moreover, the number of people who looked at the robot with head rotation was higher than the number of people who looked at the robot with the lip movement. Therefore, it can be claimed that head rotation (gaze) is more salient than lip movement in a virtual interaction.

Whether there is an influence of one-to-one F-formation shapes on people's joining behavior was the second research question. More precisely, in a joining-a-dyadic-interaction-as-a-third-interactant-in-virtual-environment scenario, which F-formations people would form was the question. In order to address likely postures for a virtual robot agent in a triadic interaction setting, the participants F-formations with the robot they looked is comparatively analyzed by focusing on the difference between real life human-human interaction and the interaction between the participants and the virtual robots. The results revealed that H and V formations are formed more frequently than other F-formation shapes in the virtual environment. (further findings are described in 5.1)

Cognitive science is such an interdisciplinary department that links a lot of study area. It investigates not only how the natural cognitive systems work but also how to build artificial cognitive systems that natural systems can interact with. In that sense, what is the contribution of this study to the literature of cognitive science is that it aims to describe the F-formation in a way that it can be translated to the artificial cognitive systems. As it is explained in Chapter 2, the main goal is to achieve a notion like "F-formation as computation". Correspondingly, what influences the F-formation and how it influences F-formation were described in the present study. In the following section (5.1), the key findings of the study are demonstrated.

5.1. Key Findings

Both kinemorphs (gaze and lip movement) seem to determine the focus of the participants in a social interaction with robots; however, the effect of gaze is higher than lip movement.

In one-robot condition, there were two phases of the experiment separating head rotation (gaze) kinemorph from the lip movement kinemorph for a comparative study. In eye gaze phase, V formation was the most frequently formed by the participant and the robot agent. In the lip movement phase, H formation was the most frequently formed by the participant and the robot agent. In both of the phases, L and I formations do not even have a single instance.

In two-robots condition, there were, again, two phases because of the same reasons; however, it seems hard to talk about a difference between these two phases when there were two robots in the scene. In both phases, H and V formations are the most frequently formed F-formations by the participants. V formation is much more frequent than H formation in eye gaze phase, while H formation has only one more occurrence than V formation in the lip formation phase. Moreover, if robot states are symmetrical, the range of the F-formation distribution is wide; if robot states are asymmetrical, the range of the F-formation distribution gets smaller. H and V formations becomes extremely dominant when robot states are asymmetrical. On the other hand, I formation does not even have a single instance.

With the given kinesmorphs (lip movement and eye gaze), there is no concrete difference between the dyadic F-formation shapes in terms of their occurrence in a triad. There is no such pattern that might reveal which F-formation shape one would form with the robot that s/he focus on in a triad.

5.2. Limitations of the Study

Among kinemorphs, gaze's effect is higher on the focus of the participants than the lip movement. The reason why gaze attracts the attention of the participants more than lip movement might be that lip movement is a regulator behavior. As it is mentioned in the chapter 2, lip movement mostly carries information about who's turn in the conversation. Gaze may be more salient than lip movement. When a participant's gaze intersects a robot's gaze, mutual gaze occurs. Mutual gaze is brought up as an important part of perception of other's emotion states and social communication by Rogers (2013). When participants are told that they are going to interact with the robot they choose, they might tend to choose the robot that provides more information.

In the one-robot conditions, the frequency of H and V formations are much higher than the other formations. Moreover, V is dominant when there is a robot with the gaze kinemorph; H is dominant when there is a robot with the lip movement kinemorph. The reason why H is dominant when there is a lip movement kinemorph might be caused by the experiment design. In the lip movement phase, robot heads do not have any rotation ability. Heads are stable but lips are moving as if robots are talking. Therefore, participants need to place themselves right in front of the robot faces in order to feel that they are in an interaction. If there were other means of signals like vocal codes, they would not need to do it. However, with such a restricting visual signal, they need to place themselves where robot faces can be seen. The formation that one can see other interactant's face in a direct way is H formation. Therefore, choosing H formation when there is a lip movement kinemorph in a conversation is plausible. Conversely, in the gaze phase of the experiment, robots are following the participants wherever they go in the scene by their head movement. Therefore, participants do not need much effort in order to have an interaction with these following robots. Even if they stand behind the robots, robots are looking at them. Therefore, they do not have to stand right in front of the robots. The environment which they can use is expanded by the following behavior of the robots. However, although it explains the frequency of H formation, it doesn't explain why V formation is used much more frequently than remaining four formations, while two of these four formations do not even have a single instance.

The results of the two-robots condition are supporting the results of the one-robot condition in the sense that H and V formations are the dominant formations. While robot states are symmetrical, there is no domination of any formation and range of F-formation used is very wide; however, in asymmetrical robot states, H and V becomes dominant. The total number of V formations are higher in gaze phase, while total number of H and V formations are almost the same in the lip movement phase. It means that while one robot is looking at the participant and the other robot is not; participants tend to form V formation with the robot they focus on. On the other hand, while one robot is speaking and the other robot is not, participants tend to form H and V formation with the robot they focus on. The key point is that this does not mean that people tend to form H and V formation with the speaking robot or V formation with the following robot. Even though people tend to focus on the speaking or following robots, these dominant formations are not always formed with the speaking or following robots. There are a lot of cases in which the robot that is not speaking, however, gets the attention of the participant. In these cases, H and V formations are continued to be formed by the participants. Therefore, it can be claimed that when robots have asymmetrical gaze states, V formation occurs more frequently than the other formations. However, it can't be claimed that when robots have asymmetrical gaze states, participants tend to form V formation with the following robot. The reason is that loglinear analysis did not give a significant result for the 3-way interaction between the robot states, participant formations and the robot that is looked by the participants. This situation might be caused by the small number of subjects.

Another topic which has an importance is that the I formation does not even have a single instance in the two-robots condition, while both L and I formations do not have instances in the one-robot condition. Firstly, the reason why it does not even have a single instance might be, again, the structure of the experiment. As it is mentioned in Chapter 2, this is a visual code experiment and I formation is such a spatial organization that two interactants are not able to see each other's face. It means that in the I formation, information is gathered through other codes than visual code. If there were vocal codes, there would be a chance that I formation occurs. However, in an experiment on visual codes on faces, participants need to see robots' faces. Secondly, the lack of L formation in one-robot condition might also be caused by the small subject number. The reason is that of 56 one-robot cases, H and V constitutes 53. It means that there are only three cases shared by C and N formations. Therefore, the possibility of L formation to occur in the next 56 cases is not much lower than C and N formations.

About the conceptual discussion on kinesics and proxemics which is brought in 2.3, it is hard to come up with a claim. The I formation does not seem to be used when only visual codes are used in the virtual environment. However, it is not certain that I formation does not exist because this is a virtual environment or because of the lack of other codes than visual codes. If it was known that I formation has instances when other codes are used, it would have an important meaning. It would mean that variables of proxemics described by Hall are not only effective on the distance of the participants but also the shape of the F-formation. However, in this present study, if the effect is caused by the difference between the virtual environment and the real life or it is caused by the kinemorphs is not clear.

About the discussion in page 49, the results showed that neither there is a difference between dyads and triads in terms of the dominant F-formation nor there is enough evidence to say that a narrowed list of F-formations is used in human-robot interaction in the virtual environment.

About the possible influences of the distributed cognition on F-formation, the results did not show any effect. While designing the experiment, robots are given an ability to rotate their head in 360 degrees. It could have an effect on the F-formation shapes; however, it does not have any effect. Robots are following the participants with their head, even though participants are going behind the robots. Yet, not even one participant chooses to stand behind of the robots. Even though there is a new body ability in robots, participants' use the same F-formations. It is a surprising result; however, it is also understandable because of three possibilities. Firstly, new body abilities require time to become new active representations in the human mind. Therefore, it might be implausible to expect such quick response to such a change. Secondly, while robots have 360 degrees rotation ability, participants do not have it. Therefore, embodiment of this rotation ability is not recognized by the participants. This situation might have led the participants to use the F-formations they always use. Thirdly, head rotations might not be very effective in the use of F-formations, when it is compared with the body rotations.

5.3. Future Work

- There is an ambiguity on whether there is a difference between the virtual environment and the real life in terms of the occurrence of F-formation shapes. (I formation is missing in the virtual environment) This ambiguity is caused by using only visual codes. Therefore, experiments with other codes of Hall (1968) would be very explanatory. On the other hand, it would also help to decide whether F-formation shapes are the topic of kinesics or proxemics.

- The frequency of F-formations under the effect of robots with kinemorphs should be investigated ‘in the wild’, if one wants to contrast the differences between the virtual environment and the real life.

- The frequency of F-Formations under the effect of kinemorphs should be investigated with embodied conversational agents which has human-like appearance. The reason is that whether the effect is caused by the virtual environment or human-robot interaction is not clear in the present study.

- The 3-way interaction between robot states, participants’ formations and the robot that is looked by the participants is needed to be investigated again with more participants.

- Effects of new body abilities other than 360 degrees of head rotation can be introduced to people in order to see their effect on F-formations.

REFERENCES

1. Bibliography

Agnus, O. M. (2012). Proxemics: The Study of Space.

Ciolek, T. M., & Kendon, A. (1980). Environment and the Spatial Arrangement of Conversational Encounters. In *Language and Social Interaction.*

Ciolek, T. M. (1977). A Study of Spatial Patterns in Social Encounters (Published PhD thesis). Australian National University.

Edmonson, B. & Han, S.S. (1983). EVects of socialization game on proximity and prosocial behavior of aggressive mentally retarded institutionalized women. *American Journal of Mental Deficiency*, 87, 435–550.

Ekman, P. & Friesen, W.V. (1969a). The repertoire of nonverbal behavior: categories, origins, usage, and coding. Semiotica, 1, 49–98

Frischen A., Bayliss A. P., Tipper S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. Psychol Bull 133(4):694

Goffman, E. (1963). *Behavior in Public Places: Notes on the Social Organization of Gathering*. New York: Free Press.

Hall, E. T. (1968). Proxemics, University of Chicago, Chicago.

Harris, T. E., & Sherblom, J. (2011). *Small group and team communication*. (p.119)

Hollan, J., Storneta, S. (1992). Beyond being there. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI ’92, Monterey, CA, May 3-7), P. Bauersfeld, J.Bennet and G. Lynch, Eds. ACM Press, New York, NY, 119-125.

Hollan, J., Hutchins, E., & Kirsh D. (2000). Distributed cognition: toward a new foundation for human-computer interaction research. ACM Transactions on Computer-Human Interaction, 7(2), 174-196.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge: MIT Press.

Hutchins, E. (2000). Distributed Cognition.

JASP Team (2019). JASP (Version 0.10.2) [Computer software].

Kendon, A. (2010). Spacing and orientation in co-present interaction. In: Proceedings of the second international conference on development of multimodal interfaces: active listening and synchrony, COST'09. Springer, Berlin, Heidelberg, pp 1–15

Knapp, M. L. (2013), *Nonverbal Communication in Human Interaction,* Rinehart & Winston, Holt, New York.

Koay, K. L., Syrdal, D. S., Walters, M. L., Dautenhahn, K. (2007) 'Living with Robots: Investigating the Habituation Effect in Participants' Preferences During a Longitudinal Human-Robot Interaction Study', Proceedings of the 16th IEEE International Workshop on Robot and Human Interactive Communication (ROMAN 2007), South Korea, 564-569,

Lakoff, G., & Johnson, M. (2001). Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought.

Louwerse, M. M., Graesser, A. C., McNamara, D. S., & Lu, S. (2009). Embodied Conversational Agents as Conversational Partners. APPLIED COGNITIVE PSYCHOLOGY, 1244-1255.

Lusk, M. M., & Atkinson, R. (2007). Animated pedagogical agents: Does their degree of embodiment impact learning from static or animated worked examples? Applied Cognitive Psychology, 21(6), 747-764.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, W.H.Freeman.

McCall, C. (2016). Mapping Social Interactions: The Science of Proxemics.

Microsoft Corporation [Computer Software]. (2016)

Nomura T., Shintani T., Fujii K. & Hokabe K., (2007). 'Experimental Investigations of Relationships Between Anxiety, Negative Attitudes, and Allowable Distance of Robots (HCI 07)', Proceedings of the Second International Conference on Human Computer Interaction (HCI 07), Beijing, P.R. China, 13-18.

Rogers S.J. (2013) Mutual Gaze. In: Volkmar F.R. (eds) Encyclopedia of Autism Spectrum Disorders. Springer, New York, NY

Ruttkay, Z., & Pelachaud, C. (2004). From Brows to Trust: Evaluating Embodied Conversational Agents. doi:10.1007/1-4020-2730-3

Silver J. (Producer), & Wachowski L. (Director), & Wachowski L. (Director), (2003). *Matrix Reloaded* (Motion Picture).

Sommer, R. (1969). *Personal space: The behavioral basis of design.* Englewood Cliffs, NJ: Prentice-Hall.

Vine, I. (1975). Territoriality and the Spatial Regulation of Interaction. In Organization of Behavior in Face-to-Face Interaction (pp. 357-388).

Yılmaz, E. C. (2018). Deictic Gaze in Virtual Environments (Unpublished master's thesis). Middle East Technical University.

Yumak, Z., & Thalmann, N. M. (2016). Multimodal and Multi-party Social Interactions. In *Human-Computer Interaction Series*

APPENDIX

Katılım ve Bilgi Formu

Deneyimize hoşgeldiniz. Bu çalışma kapsamında, üç boyutlu sanal bir ortamda karşılaştığınız karakterlerle etkileşime girmeniz istenecektir. Etkileşim sadece görsel özellikler taşımakta ve klavye ile mouse kullanmak dışında herhangi bir şey yapmanız beklenmemektedir.

Bu çalışma, Doç. Dr. Cengiz Acartürk'ün yönetiminde, Gizem Özen'in MSc tezi kapsamında, insanların sanal ortamda alan kullanımını araştırmak için yapılan bir çalışmadır. Çalışmada gözbakış datası toplanmaktadır. Çalışmaya katılım tamamiyle gönüllülük temelinde olmalıdır. Kişisel bilgileriniz kimseyle paylaşılmayacak, deney sırasında toplanacak veriler ise isim verilmeden katılımcı kodu ile yayımlanacaktır. Deney sırasında göz takip kaydı ve ekrandan data kaydı alınacaktır. Elde edilecek bilgiler bilimsel yayımlarda kullanılacaktır.

Çalışmamız kişisel rahatsızlık verecek sorunlar veya eylemler içermemektedir. Ancak, katılım sırasında sorulardan ya da herhangi bir nedenden ötürü kendinizi rahatsız hissederseniz çalışmayı yarıda bırakıp çıkmakta serbest olacaksınız. Böyle bir durumda size eşlik eden profesyonele, çalışmayı tamamlamadığınızı söylemeniz yeterli olacaktır. Çalışma sonunda, ilgili mevcut sorularınız cevaplanacaktır.

Bu çalışmaya katıldığınız için şimdiden teşekkür ederiz.

Çalışma hakkında daha fazla bilgi almak için Bilişsel Bilimler bölümü öğrencilerinden Gizem Özen (E-posta: gizem.ozen@metu.edu.tr) ile iletişim kurabilirsiniz.

Bu çalışmaya tamamen gönüllü olarak katılıyorum ve istediğim zaman yarıda kesip çıkabileceğimi biliyorum. Verdiğim bilgilerin bilimsel amaçlı yayımlarda isimsiz olarak kullanılmasını kabul ediyorum.

İsim Soyisim	Katılımcı No	Tarih	İmza
		___/___/_____	