

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**OBJECT MANIPULATION THROUGH INTUITIVE HAND GESTURES IN
VIRTUAL REALITY: THE RELATION BETWEEN OBJECTS' SPATIAL
PROPERTIES AND GESTURES**



M.Sc. THESIS

Erkut SİRDAŞ

Department of Game and Interaction Technologies

Program of Game and Interaction Technologies

JANUARY 2022

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**OBJECT MANIPULATION THROUGH INTUITIVE HAND GESTURES IN
VIRTUAL REALITY: THE RELATION BETWEEN OBJECTS' SPATIAL
PROPERTIES AND GESTURES**



M.Sc. THESIS

**Erkut SIRDAŞ
(529171006)**

Department of Game and Interaction Technologies

Program of Game and Interaction Technologies

**Thesis Advisor: Prof. Dr. Hatice KÖSE
Thesis Co-Advisor: Assoc. Prof. Dr. Sema ALAÇAM**

JANUARY 2022

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**SANAL GERÇEKLİKTE SEZGİSEL EL JESTLERİYLE NESNE
MANİPÜLASYONU: NESNELERİN UZAMSAL ÖZELLİKLERİYLE
JESTLERİN İLİŞKİSİ**

YÜKSEK LİSANS TEZİ

**Erkut SİRDAŞ
(529171006)**

Oyun ve Etkileşim Teknolojileri Anabilim Dalı

Oyun ve Etkileşim Teknolojileri Programı

**Tez Danışmanı: Prof. Dr. Hatice KÖSE
Eş Danışman: Doç. Dr. Sema ALAÇAM**

OCAK 2022

Erkut SİRDAŞ, a M.Sc. student of ITU Graduate School student ID 529171006, successfully defended the thesis entitled “**OBJECT MANIPULATION THROUGH INTUITIVE HAND GESTURES IN VIRTUAL REALITY: THE RELATION BETWEEN OBJECTS’ SPATIAL PROPERTIES AND GESTURES**”, which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Prof. Dr. Hatice KÖSE**
Istanbul Technical University

Co-advisor : **Assoc. Prof. Dr. Sema ALAÇAM**
Istanbul Technical University

Jury Members : **Prof. Dr. Leman Figen GÜL**
Istanbul Technical University

Assoc. Prof. Dr. Güzden VARİNLİOĞLU.....
Izmir University of Economics

Assoc. Prof. Dr. Çetin TÜKER
Mimar Sinan Fine Arts University

Date of Submission : 25 October 2021

Date of Defense : 7 January 2022





To new breakthroughs,



FOREWORD

This thesis entitled “Object Manipulation Through Intuitive Hand Gestures in Virtual Reality: The Relation Between Objects’ Spatial Properties and Gestures” is an applied study that aims to research the variables that influence mid-air hand gestures used to manipulate 3D objects in virtual reality. This thesis is written as a graduation requirement for the Master of Science program Game and Interaction Technologies at Istanbul Technical University.

The greatest thanks go to my advisor Prof. Dr. Hatice KÖSE and my co-advisor Assoc. Prof. Dr. Sema ALAÇAM for their absolute interest and constant assistance. Their insight and guidance provided me with the perspective I needed to write this thesis. I want to specially thank my mother, father, and sister for their never-ending encouragement and my grandmother, whom I lost recently, for always believing in me. I wish to thank all participants, without whom the analysis couldn’t have been possible. Also, I would like to express my gratitude to Teracity Software Technologies, the company I have been a part of during my master’s studies, for supporting my education. Last but not least, I wish to thank my cousin for warmly welcoming me into her home during my visits to Istanbul.

January 2022

Erkut SİRDAŞ
(Multidisciplinary Designer)



TABLE OF CONTENTS

	<u>Page</u>
FOREWORD	ix
TABLE OF CONTENTS	xi
ABBREVIATIONS	xv
LIST OF TABLES	xvii
LIST OF FIGURES	xix
SUMMARY	xxi
ÖZET	xxv
1. INTRODUCTION	1
1.1 Purpose of Thesis	3
1.2 Research Questions	4
2. LITERATURE REVIEW	5
2.1 Gesture Classifications	5
2.1.1 Classification based on number of hands.....	5
2.1.2 Classification based on form.....	5
2.1.3 Classification based on nature.....	6
2.1.4 Classification based on flow	7
2.1.5 Classification based on opposition.....	7
2.2 Gesture Interaction and Virtual Reality	8
2.3 Gesture Elicitation Studies	9
3. GESTURE INTERACTION PROPOSAL FOR A VR GAME	11
3.1 Game Concept	11
3.2 Target Platform	11
3.3 Gesture Interaction	11
3.3.1 Object manipulation.....	12
3.3.1.1 Design criteria of object manipulation gestures.....	12
3.3.1.2 Proposed gesture classifications.....	14
3.3.1.3 Proposed object manipulation gestures.....	18
3.3.2 System navigation.....	25
3.3.2.1 Design criteria of system navigation gestures.....	25
3.3.2.2 Proposed system navigation gestures.....	26
3.4 User-Elicitation Study	28
3.4.1 Referents	28
3.4.2 Animation.....	29
3.4.3 Objects	29
3.4.3.1 Form	30
3.4.3.2 Size.....	31
3.4.4 Placement.....	31
3.4.4.1 Position.....	33
3.4.4.2 Orientation.....	33
3.5 Choice-Based Elicitation Study	33
4. EXPERIMENT & RESULTS	35
4.1 Environment	35

4.2 User Interface	35
4.2.1 Information screens	36
4.2.2 Countdown timers	37
4.2.3 Stand here sign	38
4.3 Development.....	38
4.3.1 Development tools.....	38
4.3.1.1 Unity.....	39
4.3.1.2 Xbox Game Bar.....	39
4.3.1.3 Adobe After Effects & Adobe Media Encoder	39
4.3.1.4 Adobe Illustrator	39
4.3.2 Frameworks.....	39
4.3.2.1 XR Plugin Management	39
4.3.2.2 Oculus Integration package.....	40
4.3.2.3 ProBuilder	40
4.3.3 Project setup	40
4.3.4 Oculus hand tracking setup	40
4.3.5 Object animations.....	41
4.3.6 Direct interaction.....	41
4.3.7 Mid-air hand gesture detection.....	43
4.3.8 Scene and timer management.....	46
4.4 Participants	46
4.5 Constraints & Limitations	46
4.5.1 Temporal constraints	46
4.5.2 Participant constraints	47
4.5.3 Technological limitations.....	47
4.6 Experiment Flow	48
4.7 Data Collection.....	49
4.7.1 Screen recording.....	49
4.7.2 Video & audio recording.....	50
4.7.3 Concurrent think-aloud protocol	50
4.7.4 In-person survey.....	50
4.7.4.1 Participant profile section.....	51
4.7.4.2 Gesture evaluation section	51
4.8 Results	53
4.8.1 Survey results	53
4.8.1.1 General profile.....	53
4.8.1.2 Player profile	56
4.8.1.3 Participant attitudes towards predefined gestures	58
4.8.2 User-elicited gestures and gesture classifications	64
4.8.2.1 Gesture classifications used on elicited gestures.....	65
4.8.2.2 Classification of user-elicited gestures.....	66
5. DISCUSSIONS	73
5.1 Gamers Versus Non-Gamers	73
5.2 Discussions on Gesture Classifications	73
5.3 The Similarities Between User-Elicited Gestures	74
5.4 The Effects of Objects' Spatial Properties on Elicited Gestures.....	75
5.4.1 The effects of objects' size on elicited gestures.....	75
5.4.2 The effects of objects' distance on elicited gestures.....	76
5.4.3 The effects of objects' orientation on elicited gestures.....	76
5.5 The Impact of Kinesthetic Feedback on User Interaction Satisfaction	77

5.6 User-Elicited Versus Author-Designed Gestures.....	78
5.7 The Qualities of a Preferable Gesture in HCI	80
6. CONCLUSION.....	83
REFERENCES.....	87
APPENDICES	93
APPENDIX A.1	95
APPENDIX A.2	97
APPENDIX A.3	98
APPENDIX A.4	99
APPENDIX A.5	100
APPENDIX A.6	101
APPENDIX A.7	102
APPENDIX A.8	103
APPENDIX A.9	104
APPENDIX A.10	105
APPENDIX A.11	106
APPENDIX A.12	107
APPENDIX A.13	108
APPENDIX A.14	109
APPENDIX A.15	110
CURRICULUM VITAE.....	111



ABBREVIATIONS

2D	: Two-Dimensional
3D	: Three-Dimensional
API	: Application Programming Interface
COVID-19	: Coronavirus disease of 2019
CTA	: Concurrent Think-Aloud protocol
DIP	: Distal Interphalangeal
IP	: Interphalangeal
ITU	: Istanbul Technical University
FOV	: Field of View
GUI	: Graphical User Interface
HCI	: Human-Computer Interaction
HMD	: Head-Mounted Display
MCP	: Metacarpophalangeal
PC	: Personal Computer
PIP	: Proximal Interphalangeal
POV	: Point of View
SDK	: Software Development Kit
UI	: User Interface
UX	: User Experience
VR	: Virtual Reality
XR	: Extended Reality



LIST OF TABLES

	<u>Page</u>
Table 3.1 : The scenarios in the user-elicitation study. Orientations are around the viewing axis.	28





LIST OF FIGURES

	<u>Page</u>
Figure 3.1 : Hand fingers & bones.....	15
Figure 3.2 : Hand-object intersections.....	16
Figure 3.3 : Proposed taxonomies of kinesthetic feedback.....	18
Figure 3.4 : Punch gesture.....	19
Figure 3.5 : Middle Finger Flick gesture.....	20
Figure 3.6 : Four Finger Flick gesture.....	21
Figure 3.7 : Clap gesture.....	22
Figure 3.8 : Pinch gesture.....	23
Figure 3.9 : Open Hand to Fist gesture.....	24
Figure 3.10 : Thumbs-up gesture.....	26
Figure 3.11 : Pinky-up gesture.....	27
Figure 3.12 : Explosion and implosion animations.....	29
Figure 3.13 : Objects in each scenario.....	30
Figure 3.14 : Dimensions of the objects in each scenario.....	31
Figure 3.15 : The pose of the objects and the user’s upper limbs for each scenario.....	32
Figure 4.1 : Environment from participant’s point of view.....	35
Figure 4.2 : The first information screen, welcoming and informing participants.....	36
Figure 4.3 : The second information screen, showing the next steps to be taken.....	36
Figure 4.4 : Countdown timer showing remaining seconds before the next scenario began.....	37
Figure 4.5 : Countdown timer showing remaining seconds to produce an intuitive gesture.....	37
Figure 4.6 : Stand here sign from participant’s view looking down.....	38
Figure 4.7 : List of tracked points on each hand and their boneIDs in OVRSkeleton script.....	42
Figure 4.8 : Interaction spheres on each hand.....	43
Figure 4.9 : GestureDetector component of right hand for scenario 3.....	44
Figure 4.10 : Flowchart of FourFingerFlickDetector script’s logic.....	45
Figure 4.11 : Punch gesture performed in space.....	51
Figure 4.12 : Punch gesture performed on the object.....	52
Figure 4.13 : Age distribution of the participants.....	53
Figure 4.14 : Dominant hand distribution of the participants.....	54
Figure 4.15 : Self-declared health conditions of the participants.....	54
Figure 4.16 : Education distribution of the participants.....	55
Figure 4.17 : Occupation distribution of the participants.....	55
Figure 4.18 : Gaming frequency of the participants.....	56
Figure 4.19 : VR experience distribution of the participants.....	56
Figure 4.20 : Gesture technology experience distribution of the participants.....	57
Figure 4.21 : Gesture-controlled VR experience distribution of the participants.....	57
Figure 4.22 : VR experience frequency of the participants.....	58
Figure 4.23 : Attitude distribution of Punch gesture.....	59
Figure 4.24 : Attitude distribution of Middle Finger Flick gesture.....	60

Figure 4.25 : Attitude distribution of Four Finger Flick gesture.....	61
Figure 4.26 : Attitude distribution of Clap gesture.	62
Figure 4.27 : Attitude distribution of Pinch gesture.....	63
Figure 4.28 : Attitude distribution of Open Hand to Fist gesture.	64
Figure 4.29 : Number of gestures in each taxonomy category for scenario 1.	67
Figure 4.30 : Number of gestures in each taxonomy category for scenario 2.	68
Figure 4.31 : Number of gestures in each taxonomy category for scenario 3.	69
Figure 4.32 : Number of gestures in each taxonomy category for scenario 4.	70
Figure 4.33 : Number of gestures in each taxonomy category for scenario 5.	71
Figure 4.34 : Number of gestures in each taxonomy category for scenario 6.	72
Figure 5.1 : Agreement distribution of interaction satisfaction increasing due to kinesthetic feedback.....	77
Figure 5.2 : Distribution of preferred gestures for each scenario.	79
Figure 5.3 : Selection frequency of factors that determined the preferred gestures for each scenario.....	81
Figure A.1.1 : Analysis of author-designed gestures and scenarios.	95
Figure A.2.1 : Participant information sheet.....	97
Figure A.3.1 : Experiment walkthrough.	98
Figure A.4.1 : Experiment flow.	99
Figure A.5.1 : General profile part of the survey.....	100
Figure A.6.1 : Player profile part of the survey.	101
Figure A.7.1 : Gesture evaluation section of the survey for the 1 st scenario.	102
Figure A.8.1 : Gesture evaluation section of the survey for the 2 nd scenario.	103
Figure A.9.1 : Gesture evaluation section of the survey for the 3 rd scenario.....	104
Figure A.10.1 : Gesture evaluation section of the survey for the 4 th scenario.	105
Figure A.11.1 : Gesture evaluation section of the survey for the 5 th scenario.....	106
Figure A.12.1 : Gesture evaluation section of the survey for the 6 th scenario.....	107
Figure A.13.1 : Participant profiles.....	108
Figure A.14.1 : User-elicited gesture examples for each scenario.....	109
Figure A.15.1 : Participant preferences.....	110

OBJECT MANIPULATION THROUGH INTUITIVE HAND GESTURES IN VIRTUAL REALITY: THE RELATION BETWEEN OBJECTS' SPATIAL PROPERTIES AND GESTURES

SUMMARY

This thesis aims to investigate the variables that affect mid-air hand gestures used to manipulate 3D objects in virtual reality. Spatial properties of virtual objects (form, size, position, and orientation); intuitiveness and kinesthetic feedback of gestures are the focus points of this study.

VR medium is selected because of its ability to display objects in simulated 3D. While an object displayed on a 2D screen cannot provide spatial information such as the object's volume and distance, an object displayed in a VR headset can. To examine the design implications of this unique aspect on gesture interaction, a gesture-controlled VR game concept is developed. The main purpose of the game is to destroy objects via mid-air hand gestures. The player is surrounded by objects that differ in spatial properties (big-small, within-beyond arm's reach, various orientations). There are two types of objects and to destroy them players need to use one of the two destruction methods: implode or explode.

This game concept is used as a basis to develop a VR experiment using the Unity game engine. The experiment has a scenario-based setup. There are a total of six scenarios. Each scenario includes one of the two animated objects. The first three scenarios have the same exploding object whereas some spatial properties of the object change between scenarios. The first scenario includes a relatively big object that is placed within arm's reach. In the second scenario, the object is smaller. In the third scenario, the object is big again but is placed beyond arm's reach. The fourth, fifth, and sixth scenarios have another object that implodes. The objects in these scenarios are set as in the first three scenarios, but this time the orientation of the object changes between scenarios as well.

For each scenario, a unique hand gesture was proposed by the author to create a gesture vocabulary. This vocabulary was created based on multiple criteria such as ergonomics, ease of performance, being metaphorically and iconically logical, producing kinesthetic feedback, etc. Then, the form, scale, and placement of the objects were fleshed out for each scenario to make proposed gestures more discoverable.

After that, the experiment was carried out in two trials. The first trial was a user-elicitation study and aimed to collect mid-air hand gestures from the general public. Therefore, 15 participants with varying occupations and gaming frequencies were recruited for the study. The gestures were aimed to be intuitively generated. For that purpose, some temporal constraints were set and participants were not informed about the context of the study before the experiment. Next, using a VR headset, participants observed each scenario and proposed a hand gesture that would cause the observed animation. After each scenario, they repeated the gesture they proposed three more

times while explaining why they used that particular gesture. For each scenario, they were given 5 seconds to propose a gesture. This process was repeated for all six scenarios. Thumbs-up and Pinky-up gestures, proposed by the author, were used to navigate between scenarios.

During the experiment, participants were restricted to upper limb movements, allowing them to use their fingers, hands, forearms, arms, etc. They could also decide whether to use one or both hands. Furthermore, they were allowed to touch or grasp the object as they saw fit. Lastly, if deemed necessary, they were allowed to use the same gesture for different scenarios. The objects that were within arm's reach were programmed to play their animations when touched. The objects that were beyond arm's reach, on the other hand, were programmed to react to specific hand gestures.

Following the first study, a novel choice-based elicitation study was conducted. First, participants were informed about the context of the study and were taught author-designed gestures. After that, the second trial of the experiment began. This time, participants tried author-designed gestures for each scenario. Next, an in-person survey was conducted. The survey was used to profile participants and investigate their perception of author-designed gestures. Furthermore, participants were prompted to choose the gesture they deemed more appropriate for each scenario (their user-elicited gesture vs the author-designed gesture). They were also prompted to check up to three factors out of seven (intuitive, natural, logical, simple, effortless, fun, and satisfactory) that strongly influenced their preferences. The experiment was video recorded for further analysis. The research questions were answered through the analysis of the video recordings and survey responses.

Do different users generate similar types of gestures for the same scenarios (RQ1)? There were many similarities between the participants in each scenario. Most of the participants produced gestures using both hands for explode operation. In addition, participants showed a very strong tendency to trace a motion path with their hands to explode objects. Most participants used iconic gestures to explode beyond arm's reach objects while using direct manipulation gestures to explode within arm's reach objects. The number of hands used by the participants also showed strong similarities for implode operation. While the participants mostly produced two-handed gestures in the 4th and 6th scenarios, they mostly produced one-handed gestures in the 5th scenario. The nature of the gestures in the imploding object scenarios was also highly consistent among the participants. The majority of the participants used direct manipulation gestures for imploding objects that were within arm's reach while pantomimic gestures were mostly used for the imploding object that was beyond arm's reach.

Do spatial properties (size, position, orientation) of virtual objects affect the users' gestures (RQ2)? The results showed that the participants produced highly different gestures depending on the change in object size. In scenarios 2 and 5, where the objects were smaller, the number of one-handed gestures increased considerably compared to the previous scenarios. These findings suggested that the size of the object greatly affected the number of hands used. The object size also produced significantly different results in the hand parts used to touch virtual objects. In scenario 5, participants used their fingertips instead of using their palms, unlike in scenario 4. Direct manipulation gestures were most commonly used in all scenarios, where the objects were positioned within arm's reach. This indicated that most of the participants tended to control within arm's reach objects through contact. This finding showed parallelism with author-designed gestures. In the scenarios, where the object was

positioned beyond arm's reach, however, the nature of the most common gestures was different from each other. Although the movements of the gestures were approximately the opposite of each other, the nature of the gestures differed. While the participants produced mostly iconic gestures for the remote explode operation, they produced mostly pantomimic gestures for the remote implode operation.

How does the presence of kinesthetic feedback impact user interaction satisfaction (RQ3)? Results showed that interaction satisfaction due to kinesthetic feedback was highest in direct manipulation gestures. Additionally, gestures that offer an absolute down position acquired the highest ratings. Finally, the results indicated that participants were not adversely affected by the absence of kinesthetic feedback throughout the interaction.

Do users prefer user-elicited or author-designed gestures (RQ4)? All gestures designed by the author produced positive results and it was observed that at least more than half of the participants preferred these gestures to other gestures.

What is the relationship between preferable gestures and user interaction qualities (logical, intuitive, natural, effortless, simple, fun, satisfactory) (RQ5)? In the survey, participants were asked to choose up to three of the seven factors that were influential in their gesture preference for each scenario. The results showed that the most important factors in all scenarios were logical, intuitive, natural, effortless, simple, fun, and satisfactory, respectively. The frequency of responses indicated that especially logic and intuitiveness tended to produce more preferred gestures compared to other factors. In addition, the results of the first scenario suggested that participants tended to choose the more logical gesture over the more intuitive one.

This thesis proposed two new gesture classifications, namely hand-object intersections and repetition per participant. In addition, two novel classifications were proposed for kinesthetic feedback according to its source and temporal state. Furthermore, definitions for the terms 'easy to perform', 'iconically logical', and 'metaphorically logical' (Nielsen et al., 2003) were made. Finally, the scope of pantomimic gestures was expanded to include new use cases observed during the experiments.

Some participants misinterpreted the implosion animation as the object moving away from them, indicating that this animation was one of the weaknesses of the study. It was assumed that this was due to the constant acceleration of the shrink animation used. Initially, a more complex animation, in which some parts moved faster and broke apart as the object collapsed in on itself, was aimed but later abandoned due to time constraints. This animation can be tested in the future.

Thumb positioning was found to be more diverse than expected. It was observed that the participants positioned their thumbs above, below, and to the side of the other fingers while making a fist. Therefore, the success of the system in recognizing certain gestures decreased. In the future, gestures with different thumb positioning can be added to the gesture vocabulary as variations of the same gesture, allowing each variation to be recognized.

Finally, this thesis study elicited an affect display gesture and pointed out that these gestures could potentially be used in the field of human-computer interaction. For example, in a visual novel type game, the user can project their emotions to an in-game avatar using these gestures and change the direction of the game. This game setup can also be used for educational purposes so that children can observe the results of their different attitudes.



SANAL GERÇEKLİKTE SEZGİSEL EL JESTLERİYLE NESNE MANİPÜLASYONU: NESNELERİN UZAMSAL ÖZELLİKLERİYLE JESTLERİNİN İLİŞKİSİ

ÖZET

Bu tezin amacı, sanal gerçeklikte 3 boyutlu nesnelere manipüle etmek için kullanılan el jestlerini etkileyen değişkenleri araştırmaktır. Nesnelerin uzamsal özellikleri (biçim, boyut, konum ve yönelim); jestlerin sezgiselliği ve kinestetik geribildirim bu çalışmanın odak noktalarıdır.

VR ortamı, nesnelere 3 boyutlu olarak görüntüleme olanağı sağladığı için seçilmiştir. 2 boyutlu bir ekranda görüntülenen bir nesnenin hacmi ve mesafesi gibi uzamsal bilgileri sağlanamazken, aynı nesne VR ortamında görüntülendiğinde bu bilgiler sağlanabilir. Bu benzersiz özelliğin jest etkileşimi üzerindeki tasarım etkilerini incelemek için, el jestleri ile kontrol edilen bir VR oyun konsepti geliştirilmiştir. Oyunun temel amacı, nesnelere el jestleriyle dokunmaktır. Oyuncu, uzamsal özellikleri farklı olan nesnelere dokunur (büyük-küçük, kol mesafesinde-kol mesafesinin ötesinde, çeşitli yönelimler). Oyunda iki tür nesne bulunmaktadır ve bu nesnelere dokunmak için oyuncuların iki imha yönteminden birini kullanması gerekmektedir. Bunlar, dışarı ve içeri doğru patlatma operasyonlarıdır.

Bu oyun konsepti bir VR deneyi geliştirmek için temel olarak kullanılmıştır. Deney Unity oyun motoru kullanılarak geliştirilmiştir ve senaryo tabanlı bir düzeneğe sahiptir. Toplamda altı senaryo bulunmaktadır. Her senaryo, iki hareketli nesneden birini içerir. İlk üç senaryo dışarı doğru patlayan bir nesneyi içerirken, nesnenin bazı uzamsal özellikleri senaryolar arasında değişmektedir. İlk senaryo, erişilebilecek bir konuma yerleştirilmiş nispeten büyük bir nesne içermektedir. İkinci senaryoda, nesne daha küçüktür. Üçüncü senaryoda, nesne tekrar büyük olup bu sefer erişilemeyecek bir konuma yerleştirilmiştir. Dördüncü, beşinci ve altıncı senaryolar da ise, içe doğru patlayan başka bir nesne bulunmaktadır. Bu senaryolardaki nesnelere dokunmak ilk üç senaryodaki gibi ayarlanmıştır, ancak bu kez nesnenin yönü de her bir senaryoda değişmiştir.

Yazar tarafından her senaryo için benzersiz el jestleri önerilerek bir jest seti oluşturulmuştur. Bu setteki jestler, ergonomi, uygulama kolaylığı, metaforik ve ikonik olarak mantıklı olma, kinestetik geri bildirim üretme vb. birçok kritere dayalı olarak üretilmiştir. Ardından, önerilen jestleri daha keşfedilebilir kılmak için nesnelerin formu, ölçeği ve yerleşimi her bir senaryo için detaylandırılmıştır.

Deney iki denemeli olarak gerçekleştirilmiştir. İlk deneme, bir jest açığa çıkarma çalışması olup genel halktan el jestleri toplamayı amaçlamıştır. Bu doğrultuda, farklı mesleklere ve oyun oynama sıklığına sahip 15 katılımcı çalışmaya dahil edilmiştir. Jestlerin sezgisel olarak üretilmesi hedeflenmiştir. Bu nedenle, deneye bazı zaman kısıtlamaları getirilmiş ve katılımcılara deney öncesinde çalışmanın içeriği hakkında bilgi verilmemiştir. Ardından, katılımcılar bir VR başlığı kullanarak her bir senaryoyu gözlemlemiş ve gözlemledikleri animasyona neden olacak bir el jesti üretmiştir.

Katılımcılar her bir senaryodan sonra ürettikleri jesti üç kez daha tekrarlayarak bu jesti neden kullandıklarını sesli olarak gerekçelendirmiştir. Her bir senaryoda katılımcılara jest üretmeleri için 5 saniye süre tanınmıştır. Bu süreç altı senaryonun tümü için tekrarlanmıştır. Senaryolar arasında ileri ve geri geçiş sağlamak için yazar tarafından önerilen başparmak ve serçe parmak kaldırma jestleri kullanılmıştır.

Deney sırasında katılımcılar üst ekstremiteler (parmaklar, eller, önkollar, kollar vb.) hareketleriyle sınırlandırılmıştır. Buna karşın, katılımcılar bir veya iki elini kullanmakta özgür bırakılmıştır. Ayrıca, katılımcıların nesneye dokunmalarına veya nesneyi kavramalarına izin verilmiştir. Son olarak, katılımcılara, gerekli gördükleri takdirde aynı jesti birden fazla senaryo için kullanma şansı tanınmıştır. Kol mesafesindeki nesnelere animasyonları dokunma ile etkinleştirecek şekilde programlanmıştır. Kol mesafesinin ötesindeki nesnelere animasyonları ise belirli el jestlerinin tespitiyle etkinleştirecek şekilde programlanmıştır.

İlk çalışmayı takiben, bir seçime dayalı açığa çıkarma çalışması yapılmıştır. Öncelikle, katılımcılara çalışmanın içeriği hakkında bilgi verilmiş ve yazar tarafından tasarlanan jestler öğretilmiştir. Devamında, deneyin ikinci denemesine başlanmıştır. Bu sefer katılımcılar, her bir senaryo için yazar tarafından tasarlanan jestleri test etmiştir. Ardından da karşılıklı görüşme anketi düzenlenmiştir. Anket aracılığıyla, katılımcıların profili oluşturulmuş ve yazar tarafından tasarlanan jestlere ilişkin katılımcı algıları araştırılmıştır. Ayrıca, katılımcılardan her bir senaryo için daha uygun buldukları jesti seçmeleri istenmiştir (kullanıcı tarafından üretilen jeste karşı yazar tarafından tasarlanan jest). Son olarak, katılımcılara hangi faktörlerin (sezgisel, doğal, mantıklı, basit, zahmetsiz, eğlenceli ve tatmin edici) tercih ettikleri jesti belirlemelerinde daha etkili olduğu sorulmuş ve bu faktörlerden en fazla üç tanesini seçmeleri istenmiştir. Deney, daha sonra analiz edilmek amacıyla videoya alınmıştır. Araştırma soruları video kayıtlar ve anket sonuçlarının analizi yoluyla cevaplanmıştır.

Farklı kullanıcılar aynı senaryolar için benzer türde jestler üretir mi (RQ1)? Her senaryoda katılımcılar arasında birçok benzerlik görülmüştür. Katılımcıların çoğu dışarı doğru patlatma operasyonu için iki elini kullanarak jest üretmiştir. Ek olarak, katılımcılar bir nesneyi dışarı doğru patlatmak için elleriyle bir yol takip etme konusunda çok güçlü bir eğilim göstermiştir. Katılımcıların çoğu, kol mesafesindeki nesnelere dışarı doğru patlatmak için doğrudan manipülasyon jestlerini kullanırken, erişilemeyen nesnelere dışarı doğru patlatmak için çoğunlukla ikonik jestler kullanmıştır. Katılımcıların kullandığı el sayısı içeri doğru patlatma senaryolarında da güçlü benzerlikler göstermiştir. Katılımcılar 4. ve 6. senaryolarda büyük çoğunlukla iki elini kullanarak jest üretirken, 5. senaryoda büyük çoğunlukla tek elini kullanarak jest üretmiştir. İçeri doğru patlatma senaryolarındaki jestlerin doğası da katılımcılar arasında oldukça tutarlılık göstermiştir. Katılımcıların çoğunluğu erişilebilen ve içeri doğru patlayan nesnelere için doğrudan manipülasyon jestleri kullanırken, erişilemeyen ve içeri doğru patlayan nesne için genellikle pantomimik jestler kullanmıştır.

Sanal nesnelere uzamsal özellikleri (boyut, konum, yön) kullanıcıların jestlerini etkiler mi (RQ2)? Sonuçlar, katılımcıların nesne boyutundaki değişime bağlı olarak büyük oranda farklı jestler ürettiğini göstermiştir. Nesnelere daha küçük olduğu 2. ve 5. senaryolarda, önceki senaryolara oranla tek eli jestlerin sayısı oldukça artmıştır. Bu bulgular, nesnenin boyutunun kullanılan el sayısını büyük ölçüde etkilediğini ileri sürmüştür. Nesne boyutu, sanal nesnelere temas etmekte kullanılan el kısımlarında da önemli ölçüde farklı sonuçlar üretmiştir. Katılımcılar 5. senaryoda, 4. senaryodan farklı olarak avuçlarını kullanmak yerine parmak uçlarını kullanmıştır. Nesnelere

yakında konumlandırıldığı tüm senaryolarda en yaygın olarak doğrudan manipülasyon jestlerinin kullanıldığı görülmüştür. Bu da katılımcıların çoğunun kol mesafesindeki nesnelere temas yoluyla kontrol etme eğiliminde olduğuna işaret etmektedir. Bu bulgu yazar tarafından önerilen jestlerin tasarımıyla da paralellik göstermiştir. Nesnenin erişilemediği senaryolarda ise en yaygın olarak görülen jestlerin doğalarının birbirinden farklı olduğu tespit edilmiştir. 3. ve 6. senaryolardaki jestlerin hareketleri yaklaşık olarak birbirinin zıttı olsa da jestlerin doğası birbirinden farklı çıkmıştır. Katılımcılar uzaktan dışarı doğru patlatma operasyonu için daha çok ikonik jestler üretirken, uzaktan içeri doğru patlatma operasyonu için daha çok pantomimik jestler üretmiştir.

Kinestetik geribildirim varlığı kullanıcı etkileşimi memnuniyetini nasıl etkiler (RQ3)? Sonuçlar, kinestetik geribildirim sayesinde etkileşimden elde edilen memnuniyetin doğrudan manipülasyon jestlerinde en yüksek seviyede olduğunu göstermiştir. Ayrıca, mutlak bir bitiş konumu sunan jestlerin en olumlu geri dönüşlere sahip olduğu tespit edilmiştir. Ek olarak, sonuçlar, kinestetik geribildirim etkileşim boyunca devam etmemesinden katılımcıların olumsuz etkilenmediğini işaret etmiştir.

Kullanıcılar, kendileri tarafından açığa çıkarılan jestleri mi yoksa yazar tarafından tasarlanan jestleri mi tercih eder (RQ4)? Yazar tarafından tasarlanan tüm jestler olumlu sonuç vermiş ve katılımcıların en az yarısından fazlası bu jestleri diğer jestlere tercih etmiştir.

Tercih edilen jestler ile kullanıcı etkileşimi nitelikleri (mantıklı, sezgisel, doğal, zahmetsiz, basit, eğlenceli, tatmin edici) arasındaki ilişki nedir (RQ5)? Anketle, katılımcılardan her senaryo için bir jesti diğerine tercih etmelerinde etkili olan yedi faktörden en fazla üçünü seçmeleri istenmiştir. Tüm senaryolarda en önemli faktörlerin sırasıyla mantıklı, sezgisel, doğal, zahmetsiz, basit, eğlenceli ve tatmin edici olduğu ortaya çıkmıştır. Cevapların sıklığı, özellikle mantık ve sezgiselliğin, diğer faktörlere kıyasla daha çok tercih edilen jestler üretme eğiliminde olduğunu göstermiştir. Ek olarak, ilk senaryonun sonuçları, katılımcıların daha sezgisel olandansa daha mantıklı olan jesti seçme eğiliminde olduklarını ileri sürmüştür.

Bu tez çalışması, el-nesne kesişimleri ve katılımcı bazında tekrar olarak adlandırılan iki yeni jest sınıflandırması önermiştir. Ayrıca, kaynağına ve zamansal durumuna göre kinestetik geribildirim için iki yeni sınıflandırma önerilmiştir. Ek olarak, bu tez çalışmasında 'gerçekleştirilmesi kolay', 'ikonik olarak mantıklı' ve 'metaforik olarak mantıklı' (Nielsen ve meslektaşları, 2003) terimleri için tanımlamalar yapılmıştır. Son olarak, pantomimik jestlerin kapsamı, deneyler sırasında gözlemlenen yeni kullanım örneklerini içerecek şekilde genişletilmiştir.

Bazı katılımcıların içeri doğru patlama animasyonunu nesne kendilerinden uzaklaşıyor muş olarak hatalı yorumlaması, bu animasyonun çalışmanın zafiyetlerinden biri olduğunu göstermiştir. Bu durumun, kullanılan küçülme animasyonunun sabit ivmesinden kaynaklandığı varsayılmıştır. Başlangıçta, nesnenin kendi içine çökmesi esnasında bazı kısımların daha hızlı hareket ederek parçalandığı daha karmaşık bir animasyon hedeflenmiştir, ancak zaman kısıtlamaları nedeniyle bu yaklaşımdan vazgeçilmiştir. Gelecekte bu animasyon test edilebilir.

Başparmak konumlandırmasının beklenenden daha çeşitli olduğu saptanmıştır. Katılımcıların yumruk yaparken başparmaklarını diğer parmakların üstüne, altına ve yanına konumlandığı görülmüştür. Bu nedenle sistemin belirli jestleri tanıma başarısı düşmüştür. Gelecekte, farklı başparmak konumlandırmalarına sahip jestler,

aynı jestin varyasyonları olarak jest sözlüğüne eklenerek her bir varyasyonun tanınması sağlanabilir.

Son olarak, bu tez çalışması, bir duygu ifade etme jesti açığa çıkarmış ve bu jestlerin potansiyel olarak insan-bilgisayar etkileşimi alanında da kullanılabilmesine işaret etmiştir. Örneğin, görsel roman türü bir oyunda, kullanıcı duygularını oyun içi bir avatara bu jestleri kullanarak yansıtırak oyunun yönünü değiştirebilir. Bu oyun kurgusu, eğitim amaçlı kullanılarak çocukların sergilediği farklı tutumların olası sonuçlarını gözlemlenmelerini sağlayabilir.



1. INTRODUCTION

This thesis explores the variables that impact mid-air hand gestures used for object manipulation purposes in virtual reality (VR). For that purpose, we initially decided to use an Oculus Rift VR headset coupled with a Leap Motion Controller, a hand tracking module, as both of these devices were readily available at Istanbul Technical University (ITU). However, due to the lack of resources on creating a custom gesture vocabulary with Leap Motion Controller, we later switched to the newly-released Oculus Quest 2, a VR headset with built-in hand tracking. Oculus Quest 2 headset quickly became a mainstream device following its release thanks to its capabilities and competitive price. Thus, educational materials were abundant on how to build apps for this device, which eased the development side of this study.

In the literature review process, gesture elicitation methodology was found suitable for the study. This methodology helps distinguish whether a gesture is appropriate or not for a given application domain. Even though the method doesn't have strict rules, researchers typically create a list of operations, named 'referents' (Wobbrock et al., 2009). The referents are then verbally or visually presented to target end-users, who either propose (user-elicitation) or select (choice-based elicitation) gestures they regard appropriate, thus, creating a gesture set.

Choice-based elicitation studies (Dim et al., 2016; Morris et al., 2010; Silpasuwanchai & Ren, 2015) have a distinct advantage as they allow integrating designed gestures into the choices along with user-elicited gestures. However, Wobbrock et al. (2009) hypothesized the end-users would generate a better gesture set than human-computer interaction (HCI) experts. Their results showed that none of the participants elicited 19.1% of the gestures designed by the experts. Thus, Wobbrock et al. (2009) remarked the experts might be 'hopelessly lost'. In contrast, Stern et al. (2008) found little agreement amongst participants about the relationship between gestures and the anticipated outcome. Therefore, Aigner et al. (2012) speculated a gesture vocabulary should be designed instead of observed as designed languages can be constructed to prevent disputes while yet being high-bandwidth and ergonomic. As a result, we

decided to incorporate both user- and choice-based elicitation methods in our study and aimed to investigate whether user-elicited or author-designed gestures would be preferable by the users and why.

Vuletic et al. (2018), who researched intuitive gesture interaction for manipulation of three-dimensional (3D) objects in conceptual design, set some temporal restrictions to collect instant and intuitive gestures from the users. The method of Vuletic et al. (2018) is adopted in this thesis study as it is aimed for user-elicited gestures to be intuitive.

A vast number of studies reviewed in literature propose independent operations as referents (Arefin et al., 2016; Kray et al., 2010; Ruiz et al., 2011; Vatavu, 2012; Wobbrock et al., 2009). In contrast, we decide to create referents that have the same operation but differ on other levels. This approach allows us to investigate the variations in the elicited gestures caused by an isolated variable such as the size of an object. Most of the studies reviewed present their referents on a two-dimensional (2D) screen (Silpasuwanchai & Ren, 2015; Vuletic et al., 2018). Therefore, users are unable to infer some spatial properties of the objects in question such as the object's actual size and distance. It is possible to speculate that providing this information may alter the gestures elicited by the users. For that purpose, virtual reality is selected as the main medium to conduct the gesture elicitation study in the scope of this study.

Virtual Reality (VR) medium offers some of the most immersive experiences in today's technology. It is assumed that one of the critical advantages of a VR system is to provide a sense of spatial awareness. Hence, the VR platform allows us to investigate the spatial properties of 3D objects for the gesture elicitation study. Spatial properties refer to the location, shape, size, orientation, and sphere influence of an object (Eldrandaly et al., 2019). Even though the objects might have constant spatial properties, our perception of them may vary. In a broader sense, when we see an object, we perceive it from a certain point of view. As a result, we may perceive the object differently based on our perspective. For example, a pipe can seem circular, rectangular, or cylindrical when observed from various angles. Similarly, two streetlights of the same size positioned at different distances from an observer, appear to differ in size, however, the observer can still perceive them to be the same size (Cohen, 2010; Peacocke, 1985; Schellenberg, 2008).

According to Wang and MacKenzie (2000), manipulating virtual objects using mid-air hand gestures diminishes performance outcomes significantly due to the lack of physical touch. Therefore, a few studies in the literature (Grossman et al., 2004; Vogel & Balakrishnan, 2005) propose a gesture in which the thumb presses on the side of the index finger as a way of providing kinesthetic feedback for selection operation. In this thesis study, Vogel and Balakrishnan's (2005) method of providing kinesthetic feedback through the touch sensation between the hand surfaces is utilized while designing object manipulation gestures.

Moreover, in the scope of this study, classification and taxonomy terms are used interchangeably. Predefined gesture and author-designed gesture terms as well as gesture set and gesture vocabulary terms are also used in place of one another. Lastly, "near-field" and "far-field" terms are used to represent "within arm's reach" and "beyond arm's reach" terms respectively.

1.1 Purpose of Thesis

The purpose of this thesis is to explore the factors that influence mid-air hand gestures used for the manipulation of 3D objects in virtual reality. Spatial properties of virtual objects (form, size, position, and orientation); intuitiveness and kinesthetic feedback of gestures are the focus points of this study.

For this purpose, the elements of mid-air gesture interaction in VR: hands, virtual objects, and space perception are investigated. It is hypothesized that the gestures used to manipulate virtual objects will differentiate based on the spatial properties of the virtual objects, and kinesthetic feedback will improve the user satisfaction. These initial hypotheses are tested through the inspection of experiment sets with a systematic view. A gesture vocabulary is formed by collecting gestures from the general public to gain more insight into the user responses in terms of intuitive gesture behavior. It is explored whether we can create a taxonomy from the acquired gestures and how this taxonomy relates to the other taxonomies available in the literature. Moreover, we propose a secondary gesture vocabulary using a human based approach. Then, both gesture vocabularies are compared to reveal the interaction qualities that make gestures preferable.

VR medium is selected because of its ability to display objects in simulated 3D. While an object displayed on a 2D screen cannot provide spatial information such as the object's volume and distance, an object displayed in a VR headset can. This thesis examines the design implications of this unique aspect on mid-air gesture interaction.

1.2 Research Questions

This thesis study states 5 research questions:

Research Question 1 (RQ1): Do different users generate similar types of gestures for the same scenarios?

Research Question 2 (RQ2): Do spatial properties (size, position, orientation) of virtual objects affect the users' gestures?

Research Question 3 (RQ3): How does the presence of kinesthetic feedback impact user interaction satisfaction?

Research Question 4 (RQ4): Do users prefer user-elicited or author-designed gestures?

Research Question 5 (RQ5): What is the relationship between preferable gestures and user interaction qualities (logical, intuitive, natural, effortless, simple, fun, satisfactory)?

2. LITERATURE REVIEW

2.1 Gesture Classifications

Mitra and Acharya (2007) define gestures as “expressive, meaningful body motions involving physical movements of the fingers, hands, arms, head, face, or body with the intent of: 1) conveying meaningful information or 2) interacting with the environment”. Even though ‘gesture’ encompasses not only hand movements but also face expressions and gaze shifts in a wider context (Quek et al., 2002) within the context of this study, gesture refers to a posture or movement of a person’s upper limbs. Throughout decades, numerous classifications have been made for human gestures. The following sections explore some of the most prevalent classifications formed within the domain of human-computer interaction (HCI).

2.1.1 Classification based on number of hands

Number of hands dimension, also referred to as hand usage, is one of the most commonly used classifications in literature (Aigner et al., 2012; Arora et al., 2019). This classification indicates how many hands are utilized while performing a gesture and consists of two categories: unimanual for one hand and bimanual for two hands.

2.1.2 Classification based on form

Form dimension classifies gestures based on their spatiotemporal status. It is proposed by Wobbrock et al. (2009), who studied tabletop gestures. This classification includes 6 categories; static pose, dynamic pose, static pose and path, dynamic pose and path, one-point touch, and one-point path. Form is static pose if the posture of the hand is held still without a change in its spatial position. An example is showing the palm of your hand to someone to stop them. If the posture of the hand changes over time while preserving its location, the gesture classifies as dynamic pose. In static pose and path, the posture of the hand is held still while the hand changes location. An example is swiping your hand from left to right. If the posture of the hand changes over time as the hand changes location, the gesture classifies as dynamic pose and path. One-point

touch is a special version of static pose that utilizes one finger. Similarly, one-point path also utilizes one finger and is a special version of static pose and path. Nevertheless, Wobbrock et al. (2009) argue one-point touch and one-point path should be distinguished due to their resemblance to mouse actions.

2.1.3 Classification based on nature

According to Dim et al. (2016), nature dimension classifies gestures according to the relationship between the gesture and the meaning.

Quek et al. (2002) explore hand gestures accompanying speech and propose definitions for gesture types like semaphores and manipulative gestures. They define semaphores as a learned set of static or dynamic gestures that use arm and hand. By their definition, manipulative gestures are used to maneuver an entity. Moreover, the movements of the gesture and the manipulated object demonstrate a strong connection. Quek et al. (2002) along with many others (Eisenstein & Davis, 2007; Kettebekov, 2004) consider gesticulation accompanying speech to be the most natural way of gesturing.

Karam and Schraefel (2005) classify gestures according to their nature in the scope of HCI. Their model proposes deictic, manipulation, semaphores, gesticulation, and sign language. Deictic gestures indicate direction as to pinpoint an entity or a certain location in space. Even though some other types of gestures inherently involve pointing such as targeting an entity for identification in order to manipulate it, Karam and Schraefel (2005) remark that a lot of interactions utilize solely deictic gestures. Gesticulation is used to illustrate the outline of entities as they are communicated through speech. Sign language is based on linguistics and is usually evaluated separately from other gesture types. Sign language gestures produce grammatical structures using a combination of distinct hand poses. Karam and Schraefel (2005) adopt the definitions of semaphores and manipulation from Quek et al. (2002).

Aigner et al. (2012) mainly build on Karam and Schraefel's (2005) model along with other HCI-focused work. They propose 5 categories: pointing, semaphoric, pantomimic, iconic, and manipulation. Pointing is a variation of deictic. They rename this category since there isn't any speech involved in their work. Hence, there wasn't any deixis by their setup's nature. Aigner et al. (2012) speculate the gesticulation category fails to distinguish between pantomimic and iconic gestures. Therefore, divide gesticulation into two categories: pantomimic and iconic. Pantomimic gestures

imitate a particular operation. It is stated these gestures are generally executed in the absence of an actual object in the user's immediate vicinity such as holding an imaginary glass with one hand and tilting an imaginary carafe with the other hand. Pantomimic gestures often contain several more fundamental gestures such as grasping, controlling, and releasing. The model of Aigner et al. (2012) regards all of these gestures as one pantomimic gesture.

Arora et al. (2019) study mid-air gestures for animation composing purposes in VR. They classify gestures based on the nature of action. Their model includes direct manipulation, demonstrative, semaphoric and widget use. Arora et al. (2019) define direct manipulation as interacting with an object in the scene directly such as sculpting to modify the shape of an object. Demonstrative is another term for iconic gestures that mimics the shape and/or the motion path of an entity. Arora et al. (2019) adopt semaphoric category from Aigner et al. (2012). Widget use, inherited from Wobbrock et al. (2009), refers to the use of buttons and sliders for certain operations, which are deemed unfit by users for gestural control.

2.1.4 Classification based on flow

Flow dimension, proposed by Wobbrock et al. (2009), classifies gestures in relation to the system's response and consists of discrete and continuous. Flow is discrete if the response occurs after the gesture is fully performed. An example is drawing 'x' to reject a dialog. A gesture's flow is continuous if the response occurs while the user acts, therefore, requires ongoing recognition. An example is dragging an object to move it. Although Wobbrock et al. (2009) bring discrete and continuous terms together under flow dimension, these terms have been mentioned in literature before (Wu et al., 2006).

2.1.5 Classification based on opposition

Feix et al. (2015) propose a taxonomy of human grasps based on opposition. The categories in their work vary depending on the direction of the applied force between the surfaces of the hand and consist of pad opposition, palm opposition, and side opposition. Pad opposition occurs when the direction is parallel to the palm. It uses the tip of the fingers such as holding a needle between the pad of the thumb and index finger. Palm opposition occurs when the direction is perpendicular to the palm. This grasp utilizes the palm as well as fingers such as holding a hammer. Side opposition,

defined by Feix et al. (2015), occurs when the direction is transverse to the palm. It uses the shaft of the fingerbones with or without the pad of the thumb such as holding a cigarette between the index and middle fingers or holding a key between the pad of the thumb and the side of the index finger.

2.2 Gesture Interaction and Virtual Reality

Human-computer interaction (HCI) is the study of how humans conceive, employ, and implement computer systems. It is a field of study used to research humans, computers, and their interactions.

As a crucial supporting technology for VR, human-computer interaction offers a wide range of interactive modes depending on distinct functions and goals, allowing individuals to enjoy immersive experiences in 3D virtual environments. VR is an interactive human-computer interaction interface that is still in its development stage. As a result, the Post-WIMP (Van Dam, 1997) and Non-WIMP (Green & Jacob, 1991) interface paradigms have been proposed, and the interaction expands beyond a 2D graphical user interface based on keyboard and mouse to a natural user interface. Unlike typical WIMP, in a natural user interface scenario, the user can interact in a 3D world through interactive scenario design or VR simulation alone. Thus, avoiding resource waste and lowering production costs (Yang et al., 2019).

Direct manipulation of a 3D object with users' hands, rather than using a keyboard or mouse, may enhance the user experience (Dardas & Alhaj, 2011). The users can manipulate a 3D object's location and orientation with their hands using this method, similar to how they manipulate objects in the real world. For VR applications, this method more accurately approximates the operations that users will execute.

Mid-air gesture interaction is the manipulation of digital content or remote devices without touching them, using non-invasive sensors (mostly based on computer vision) that track body motions, postures, and gestures (Vogiatzidakis & Koutsabasis, 2018). Mid-air gesture interaction has grown into a distinct kind of HCI in the last several years. The beginnings of mid-air gesture interaction may be traced back to the MIT Media Room and the "Put-That-There" demo in the late 1970s (Bolt, 1980).

There are a vast number of classifications on the use of gestures in VR such as a meta-analysis provided by Yang et al. (2019), a specific gesture definition for punch, move

forward, grab, and throw functions (Rautaray & Agrawal, 2011), animation authoring (Arora et al., 2019), pedestrian and automated vehicle interaction (Gruenefeld et al., 2019), online shopping systems (Wu, Wang, et al., 2019).

2.3 Gesture Elicitation Studies

Although mid-air hand gestures have gained widespread prominence in recent years, the majority of gesture-controlled applications are designed by experts in HCI, and end-users are seldom given the chance to contribute to the gesture design process. However, system designers may neglect the usability of gestures while trying to conform to the limitations of image processing and gesture recognition algorithms (Chan et al., 2016; Choi, 2016). Thus, the gesture disagreement problem (Wu, Zhang, et al., 2019) may arise based on the difference of opinion on what is deemed as a suitable gesture between the system designers and end-users.

Gesture elicitation studies, which originated in the domain of participatory design, are now commonly used to assist designers in selecting the most suitable gesture vocabulary for their respective applications. To reveal and design suitable gestures that are discoverable, easy to learn, remember and perform in the context of HCI, a gesture elicitation methodology has been proposed by researchers. This methodology enables the target end-users to partake in the gesture design process of gesture-based applications. Even though there isn't a strict procedure for the methodology, a typical gesture elicitation study consists of presenting the starting and ending states of target operations to end-users, who then propose a gesture they deem suitable for the observed operations. Generally, researchers also explore the users' mental model through rich qualitative data collected by think-aloud protocol and video analysis. Finally, after analyzing the gathered data, the researcher devises a gesture set.

Gesture elicitation studies have been used in a variety of application domains in literature such as mobile devices (Di Geronimo et al., 2017; Pyryeskin et al., 2012; Ruiz et al., 2011), smartwatches (Malu et al., 2018), smart home environments (Gheran et al., 2018), gaming (Silpasuwanchai & Ren, 2015), human-drone interaction (Obaid et al., 2016), virtual reality (Arora et al., 2019; Yan et al., 2018), augmented reality (Pham et al., 2018; Piumsomboon et al., 2013), operating rooms (Jurewicz et al., 2018).

A variant of this methodology is the choice-based elicitation approach (Dim et al., 2016; Morris et al., 2010; Silpasuwanchai & Ren, 2015), in which intended users pick their preferred gesture from a list of predefined alternatives. The choices in the list can include previously elicited gestures as well as expert-designed gestures. Incorporating expert-designed gestures can serve as a non-threatening technique to introduce designed alternatives into the process. This is especially useful if the users have little to no awareness of the spectrum of possibilities for appropriate gestures. Furthermore, as Morris et al. (2010) imply, allowing users to choose their preferred options reduces the risk of designers creating excessively complex gestures. This approach also raises the likelihood of deciding on a gesture that is appropriate from both the designer's and the users' perspectives.

The user evaluation of elicited gestures based on a variety of criteria, scales, and qualitative remarks is a critical part of an elicitation study. In general, user evaluations might occur while executing gestures (concurrently), after performing a single gesture (post-task), or after all the gestures are performed (post-test). A considerable number of studies reviewed in literature use concurrent user evaluation methodology. The think-aloud protocol, in which the participants verbalize their thoughts is the most common approach (Vogiatzidakis & Koutsabasis, 2018).

A gesture elicitation study's principal objective is to extract suitable gestures for mid-air interactions. For that purpose, the majority of the reviewed studies inspect a gesture vocabulary that fits the intended operations better (without providing insight on what this entails). On the other hand, some studies investigated elicited gestures' appropriateness on specific levels such as discoverability (Obaid et al., 2016), intuitiveness (Bostan et al., 2017), easy to perform gestures (Ruiz et al., 2011), gesture simplicity (Löcken et al., 2012), perceived fatigue (Bostan et al., 2017), comfort (Arefin et al., 2016; Bostan et al., 2017; Jurewicz et al., 2018).

3. GESTURE INTERACTION PROPOSAL FOR A VR GAME

3.1 Game Concept

For this study, a conceptual virtual reality game that utilizes mid-air hand gestures as the main gameplay element is designed. The game spawn objects vertically from the ground at varying locations all around the player. The objects leave the ground and suspend in mid-air at various heights. The player stands on a fixed spot for the duration of the game and uses mid-air hand gestures to destroy these objects. A real-time visual representation of the player's hands can be seen in virtual reality (VR). The right and left hands are colored differently, green and purple respectively. Each object has green-and/or purple-colored surfaces on them as well indicating which hand should be used to destroy the objects. Objects can be destroyed either by exploding or imploding. The form of the object dictates the destruction method.

Furthermore, objects require unique gestures to be destroyed based on their size and distance. Proper destruction of an object awards the player a score. Orientation of the gesture is another important factor as it determines the awarded score. The better aligned the gesture, the higher the score. Lastly, objects should be destroyed as fast as possible as they will stay afloat for only a set duration of time before disappearing, and the longer the objects stay afloat, the lesser the score they will award.

3.2 Target Platform

The game requires a head-mounted display (HMD) and a gesture recognition solution. Therefore, Oculus Quest 2 headset is selected as it is an affordable and widely available HMD with built-in hand tracking.

3.3 Gesture Interaction

There is not a set of mid-air gestures that are universally regarded as the most appropriate for all the applications (Nielsen et al., 2003). Even though there are some design guidelines and best practices, they are heavily dependent on their context and

tracking technology. Moreover, these guidelines are mainly created based on interactions that are used the most in a given application domain. In literature, there are limited studies on object manipulation through deformation (Arora et al., 2019), and even fewer studies on object manipulation through destruction. In that context, this study aims to contribute to the literature.

3.3.1 Object manipulation

According to Quek et al. (2002), manipulative gestures refer to the act of controlling an entity. Grasp, move, and rotate gestures can be considered as one of the most common gestures which can be encountered in daily life (Yang et al., 2019). These common gestures are also studied extensively in literature (Piumsomboon et al., 2013), and while affecting objects' position and orientation, these gestures don't cause a change in manipulated objects' form or size. In this thesis study, however, dichotomous implode and explode operations are selected as the concept of the game is to destroy objects. Therefore, objects deform in such ways that they can't preserve their form and/or size. Based on these two operations, a gesture vocabulary is designed.

3.3.1.1 Design criteria of object manipulation gestures

Six different gestures are proposed for explode and implode operations regarding multiple criteria. Nielsen et al. (2003) propose using a human based approach in the creation of gesture vocabulary. Their approach is built on usability principles and consists of the following: intuitive, easy to perform & remember, ergonomic, iconically & metaphorically logical. This thesis study aims to compare author-designed gestures with user-elicited gestures. User-elicited gestures are considered to be generated intuitively, whereas intuitiveness is neglected in the creation of author-designed gestures to allow creative freedom. Besides the criteria of Nielsen et al. (2003), kinesthetic feedback is also taken into account in the gesture design process.

- **Easy to perform:** This criterion can be considered as gestures requiring minimal time and effort (mental and physical) to execute. Using both hands at the same time requires using both hemispheres of the brain. Thus, increases the mental effort needed to gesture. Consequently, all of the gestures in the gesture set are designed to be executed with one hand except for one gesture, which is added to compare its perceived level of ease with unimanual gestures.

Wobbrock et al. (2009) classify gestures by their form as detailed in Section 2.1.2. This classification consists of six categories: static pose, dynamic pose, static pose and path, dynamic pose and path, one-point touch, and one-point path. Gestures in dynamic pose and path category change their pose over time while also moving in 3D space. Therefore, conceptually it is the most complex category. Hence, all of the gestures in the gesture set are designed to be categorized other than dynamic pose and path to reduce their complexity.

- Iconically & metaphorically logical: This category suggests that the pose and/or motion path of the gesture should present some sensible connections with the intended operation on a visual and semantic level. Therefore, each gesture in the set is either associated with real-world gestures or mimics the target operation visually.
- Ergonomic: Nielsen et al. (2003) define ergonomic as not causing physical stress after being used several times. All of the author-designed gestures in this study are designed to be recognized regardless of the hand's orientation at least around one axis, which allows the users to position their wrist in a way that is the most comfortable for them.
- Kinesthetic feedback: Because virtual objects lack corporeality, it is not possible to get any tactile feedback from the objects when using mid-air hand gestures in VR. In contrast, interactions such as a mouse click, button press, or screen tap provide immediate tactile feedback indicating that the input has been applied. Wang and MacKenzie (2000) present that using hand gestures without touching a physical surface to manipulate virtual objects reduces performance results considerably. Therefore, this thesis study aims to compensate for the lack of tactile feedback by proposing gestures that produce kinesthetic feedback via contact between hands or hand parts. This approach has been used in literature before (Grossman et al., 2004; Vogel & Balakrishnan, 2005). Both of these studies propose a gesture that has the thumb press the side of the index finger for select function. Even though they speculate this gesture has a distinct advantage due to providing kinesthetic feedback as well as an absolute down position, users find this gesture to be uncomfortable and tiring. We speculate that this is due to thumb movement not conforming to ergonomic standards. Our thesis study investigates whether gestures that generate kinesthetic

feedback would produce more satisfactory interaction for object manipulation purposes.

3.3.1.2 Proposed gesture classifications

This thesis study focuses solely on mid-air hand gestures in the context of human-computer interaction and classifies gestures accompanying speech and gestures that are used to manipulate objects under one unified model. In this thesis study, we adopt some of the classifications in literature such as number of hands, nature, flow, and form, all of which are detailed in Section 2.1. Apart from these, two novel classifications named hand-object intersections and repetition per participant are proposed. Moreover, kinesthetic feedback is broken down into novel classes and categories as well.

- Hand-object intersections: This is a novel classification proposed in this thesis study. Silpasuwanchai and Ren (2015) classify gestures based on body parts' suitability. Their research is focused on full-body gestures. Hence, they investigate which parts of the body such as the head, eyes, hands, torso, legs, etc. are the most suitable for gesturing. Hand-object intersections category proposed in this study has a similar approach. It explores which parts of the hand are the most appropriate for manipulating virtual objects by direct interaction.

This novel classification only applies to direct manipulation gestures as it requires the hand to coincide with the virtual objects. Through this class, this thesis study aims to investigate if there are meaningful changes in used hand parts between scenarios based on the change of the manipulated object's size and orientation. The intersections defined by this class are also translated into affordances and signifiers in the object design process to make proposed gestures more discoverable.

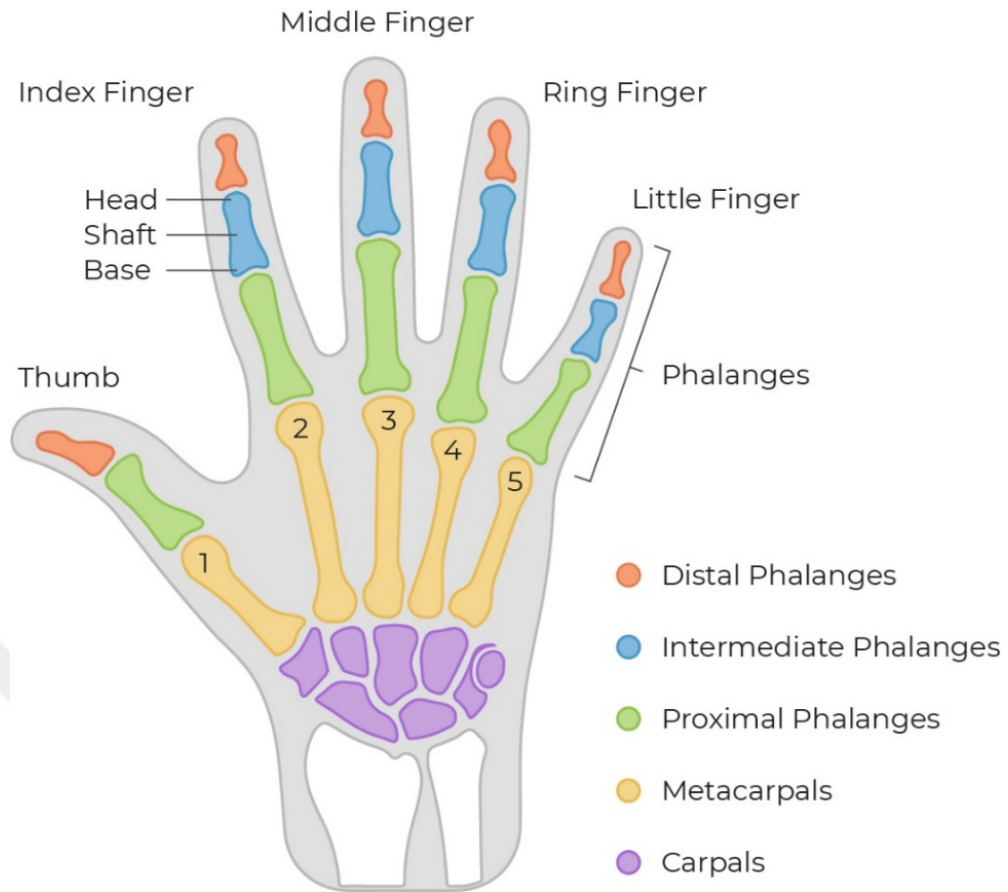


Figure 3.1 : Hand fingers & bones.

Feix et al. (2015) classify human grasps based on opposition. Even though their taxonomy is based on the direction of the force between the opposing hand surfaces, hand parts responsible for each grasp also diverge. This is evident by the example grasps they mention as well as the naming they use for each category.

Similar to the hand parts pointed out in the classification of Feix et al. (2015), detailed in Section 2.1.5, hand-object intersections classification categorizes human touch in VR medium based on hand parts used to touch the virtual object and consists of four categories: palmar, phalangeal, distal, and no intersection. Palmar refers to the use of the palm with or without the fingers. The finger bones in the hand are called phalanges. Figure 3.1 shows the names and parts of the bones in the hand. Each finger except for the thumb has three phalanges proximal, intermediate, and distal. Thumb doesn't have an intermediate phalanx. Phalangeal category in this study is a subcategory of palmar and refers to using any combination of finger bones.

Metacarpophalangeal (MCP), interphalangeal (IP), proximal interphalangeal (PIP), and distal interphalangeal (DIP) joints are part of this category as well. Distal is a subcategory of phalangeal, and in this thesis study, refers to using any number of distal phalanges. Any hand surface around the distal phalange such as the nail, pad, and tip of the finger is categorized as distal in this thesis study. From palmar to distal, the intersection of the hand and the virtual object tends to get smaller as shown in Figure 3.2. If the hand and the virtual object don't intersect, it is classified as no intersection.





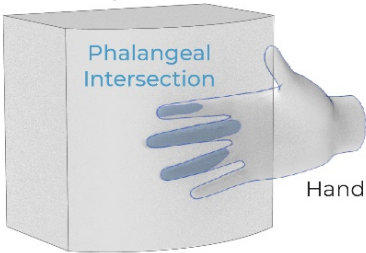
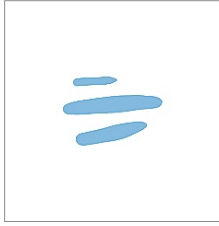

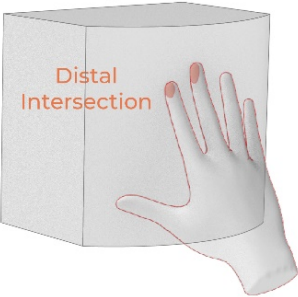
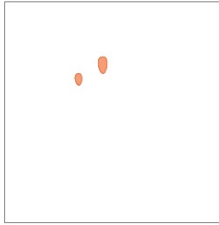
INTERACTING HAND PARTS	HAND-OBJECT INTERSECTIONS	INTERSECTION TYPE
 <p>● Palm & Fingers</p>	<p>Virtual Object</p>  <p>Hand</p>	 <p>Palmar Intersection</p>
 <p>● Fingers</p>	<p>Virtual Object</p>  <p>Hand</p>	 <p>Phalangeal Intersection</p>
 <p>● Fingertips</p>	<p>Virtual Object</p>  <p>Hand</p>	 <p>Distal Intersection</p>

Figure 3.2 : Hand-object intersections.

- Repetition per participant: At the heart of this thesis is the investigation of gestural differences based on virtual objects' spatial properties. Therefore, a novel classification named repetition per participant is also added to identify duplicate gestures within participants. It is proposed specifically for this study. It has three categories: identical, oriented, and unique. The gestures are categorized based on visual similarities instead of semantics as the Oculus Quest 2 headset recognizes the finger bones and their movements but not the user's intention. Therefore, classifications such as number of hands and flow are crucial whereas the nature of the gesture is neglected in the process of forming this classification.

Identical is when the gesture of a participant is the same as a previously used gesture by the same participant. It requires the hand pose, motion path, and direction of two gestures to match exactly in relation to world coordinates.

If the hand pose, motion path, and direction of a gesture are the same as a previous gesture relative to acting hands' wrists, but different relative to world coordinates, the gesture classifies as oriented. In this thesis study, all oriented gestures are generated solely by positioning upper limbs differently before gesturing due to the constraints of the experiment.

Unique is when the gesture of a participant is different from all the other gestures they generated before. Gestures that don't meet the requirements of the other two categories classify as unique. Thus, any gesture that has either a different hand pose, motion path, or direction relative to acting hands' wrists is categorized as unique in this thesis study.

- Kinesthetic feedback: As mentioned in Section 3.3.1.1, kinesthetic feedback is previously mentioned in literature (Grossman et al., 2004; Vogel & Balakrishnan, 2005). However, this thesis study further explores and divides kinesthetic feedback into different categories based on the source as self-kinesthetic and pair-kinesthetic. Self-kinesthetic feedback occurs between the hand surfaces of one hand. Pair-kinesthetic feedback occurs between the surfaces of right and left hands.

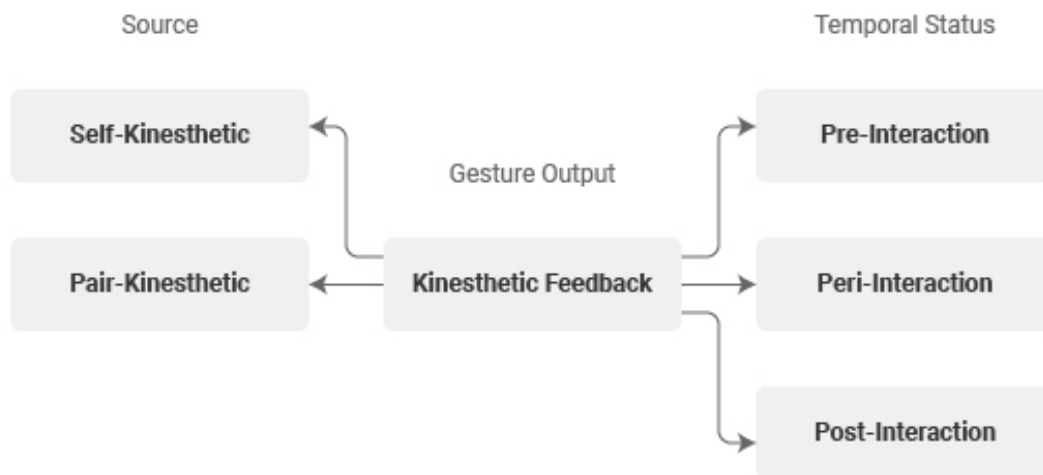


Figure 3.3 : Proposed taxonomies of kinesthetic feedback.

Additionally, as shown in Figure 3.3, kinesthetic feedback is divided into three other categories based on its temporal status in relation to interaction as pre-interaction, peri-interaction, and post-interaction. In the pre-interaction category, kinesthetic feedback occurs and ends before the interaction is recognized or fully responded to. If kinesthetic feedback and the system's response to the interaction occur simultaneously (disregarding the processing time needed) it is classified as peri-interaction. Lastly, post-interaction kinesthetic feedback occurs after the interaction is recognized and partially responded to. This thesis study investigates whether source or temporal status affects the level of satisfaction acquired through kinesthetic feedback.

3.3.1.3 Proposed object manipulation gestures

Based on the criteria explored in Section 3.3.1.1, a gesture vocabulary is proposed. There are six gestures in the gesture vocabulary, and moving forward they will be referred to as predefined gestures or author-designed gestures. All predefined gestures and their attributes can be seen in Figure A.1.1. These gestures are designed to represent explode and implode operations in different scenarios. There are six scenarios in total.

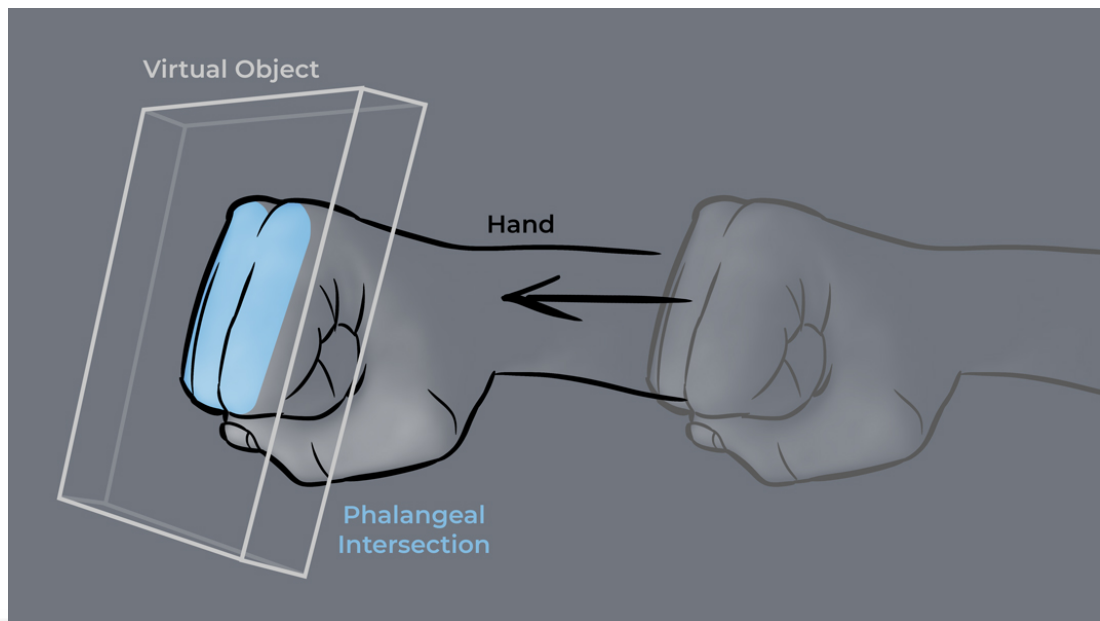


Figure 3.4 : Punch gesture.

- Punch gesture: This is a unimanual gesture, which requires making a fist and using the proximal phalanges of the second to fifth fingers to punch the virtual object. Thus, the hand-object intersection is phalangeal. This gesture shown in Figure 3.4, is proposed to explode relatively big (medicine ball-sized) objects that are within arm's reach. The form of the gesture is static pose & path. The flow of the gesture is discrete. The nature of the gesture classifies as direct manipulation as the gesture requires making contact with the virtual object. In the real world, when punched, a glass panel shatters and scatters. This reaction can be considered as an appropriate representation of an explosion. Punch gesture is the only gesture that doesn't provide kinesthetic feedback within the predefined gesture vocabulary. A unique part of this gesture, however, depending on the object's distance, the arm may have to be fully extended to initiate contact, in which case the elbow reaches an absolute down position producing strong musculoskeletal feedback.

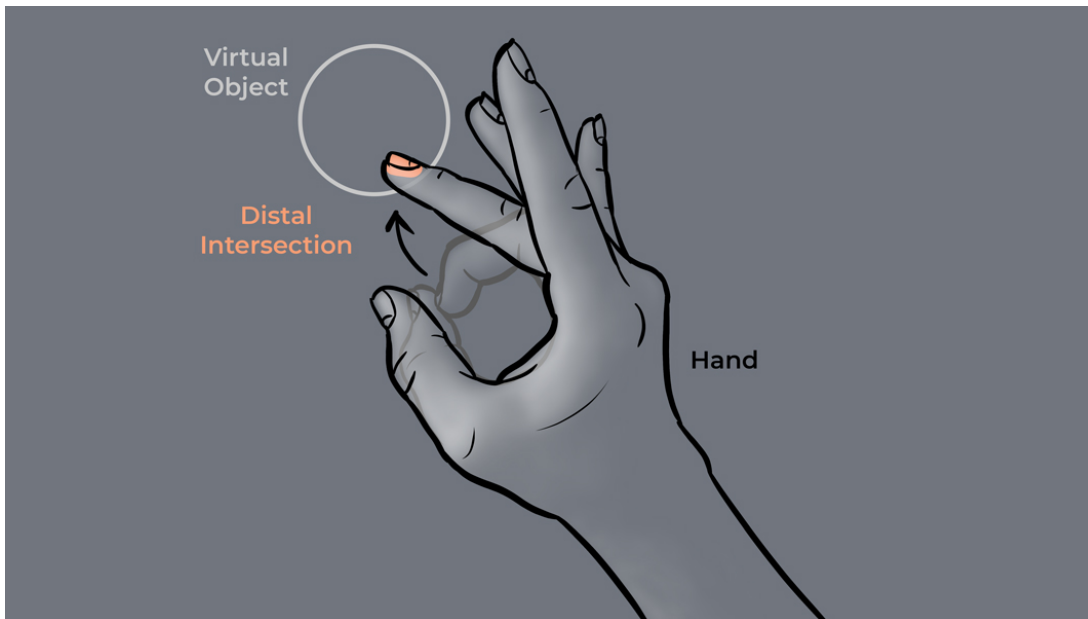


Figure 3.5 : Middle Finger Flick gesture.

- Middle Finger Flick gesture: This gesture, shown in Figure 3.5, requires building up tension between the nail of the middle finger and the pad of the thumb, then by releasing the middle finger, one can translate this tension into a high-speed motion to hit the virtual object with the middle finger's nail. Hence, the hand-object intersection is distal. This gesture is proposed to explode relatively small (baseball-sized) objects that are within arm's reach. This is a unimanual, direct manipulation gesture. The form of the gesture is dynamic pose. The flow of the gesture is discrete. It also provides pre-interaction self-kinesthetic feedback. However, considering this is a very short-distance and fast-paced gesture, it is assumed this temporal difference would be negligible and wouldn't reduce this gesture's effectiveness at providing increased interaction satisfaction through tactility. This gesture can be used to shoot some small objects or poke someone on the ear or forehead in the real world. Therefore, it has a playful nature, which is assumed to improve this gesture's fun factor and appropriateness as it will be used in a game.

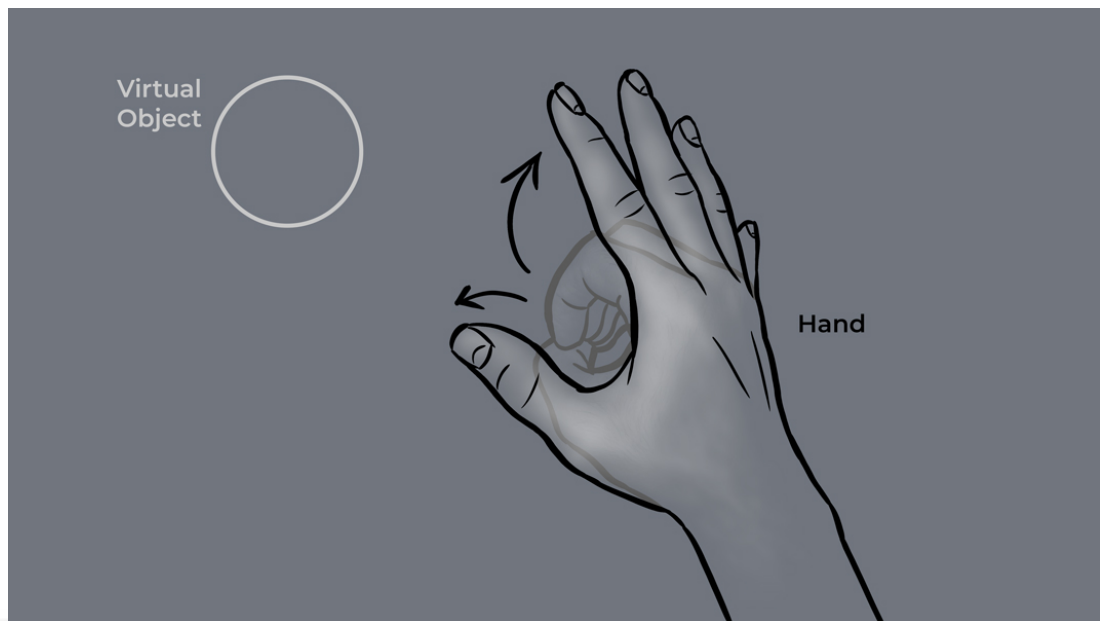


Figure 3.6 : Four Finger Flick gesture.

- Four Finger Flick gesture: This gesture, shown in Figure 3.6, builds up tension between the nails of the four fingers (second to fifth fingers) and the side of the thumb, then releases this tension in an explosive manner creating a high-speed flicking motion. This is another fast-paced unimanual gesture. The form of the gesture is dynamic pose. It is a discrete gesture providing self-kinesthetic feedback that occurs pre-interaction. It is proposed to explode relatively big (medicine ball-sized) objects that are beyond arm's reach. In the real world, this gesture is mostly used to flick water at someone, who is most likely out of the user's reach as well. It is playful by nature, which is assumed to make this gesture a good fit as it will be used in a game. The motion path of the gesture also mimics the explosion iconically as the fingers splay from a central position. Therefore, based on the user's intention, the same gesture can be classified differently. Mimicking the way people flick water at someone makes it pantomimic while mimicking the motion path of the explosion makes it iconic. As the gesture does not require making contact with the object, it is designed to be recognized regardless of its orientation around the axis of view. This allows the gesture to be performed in varying orientations, from which the user can prefer the most ergonomic one for themselves. Nevertheless, as the user should perform this gesture towards an object, any other rotation about different axes is not possible in order not to disrupt the user's aim.

A similar gesture is observed in the user-defined gesture set of Wobbrock et al. (2009). In their study, the tips of all the fingers are brought together to form a downwards pointing cone, then the fingertips are placed on an interactive tabletop and splayed for enlarge operation. Due to the difference in thumb's placement, there isn't any built-up tension in their version though.

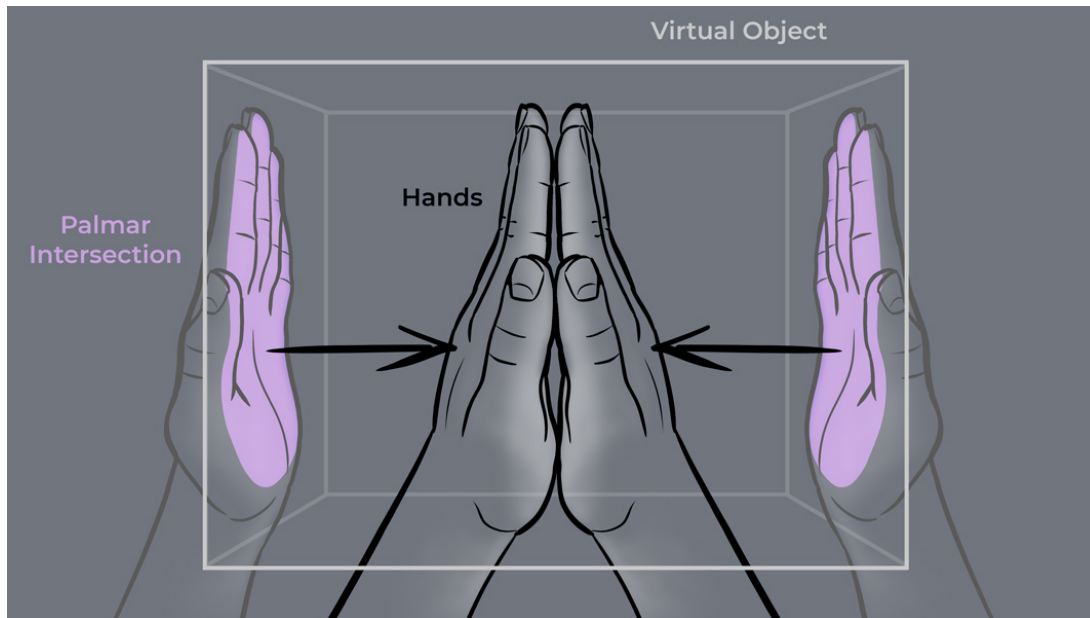


Figure 3.7 : Clap gesture.

- Clap gesture: This gesture, shown in Figure 3.7, requires squashing the virtual object between the palmar side of the hands symmetrically. Therefore, the hand-object intersection is palmar. It is proposed to implode relatively big (medicine ball-sized) objects that are within arm's reach. This is the only bimanual gesture in the gesture set. It also provides post-interaction pair-kinesthetic feedback. The form of the gesture is static pose & path for both hands. The flow of the gesture is continuous. To reduce the gesture's complexity, a simple hand pose and motion path that moves symmetrically is selected. This is a direct manipulation gesture. It is mostly used to swat mosquitoes in mid-air in the real world. It is designed to be performed as fast as its real-world counterpart. Therefore, it produces a loud clapping sound, which is unique in the gesture set. Moreover, when the two hands unite, the gesture introduces an absolute down position. Consequently, this gesture is assumed to be regarded as highly satisfying due to its tactility and sound.

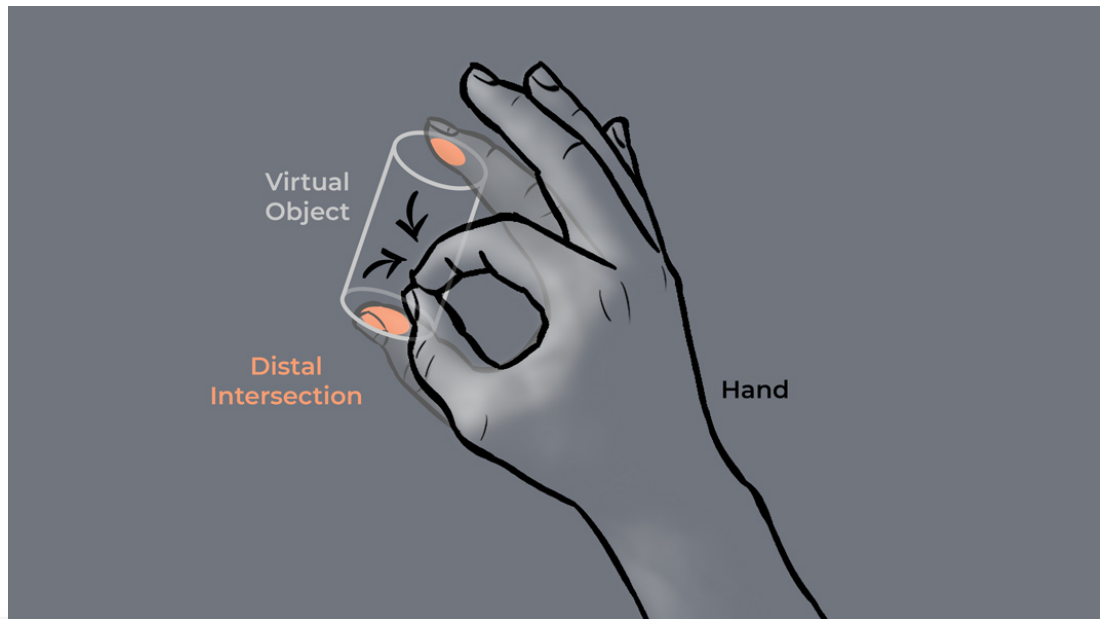


Figure 3.8 : Pinch gesture.

- Pinch gesture: This gesture, shown in Figure 3.8, requires connecting the pad of the thumb and index finger to squeeze the virtual object. Consequently, the hand-object intersection is distal. It is designed to implode objects that are relatively small (baseball-sized) and within arm's reach. This is a unimanual continuous gesture. The form of the gesture is dynamic pose. The nature of the gesture is direct manipulation. This gesture is mostly used to snuff out candles in the real world. Furthermore, it is commonly used to zoom out on touch screen devices. Pinch gesture provides post-interaction self-kinesthetic feedback and an absolute down position. Depending on how small the object is, kinesthetic feedback can occur almost simultaneously with the interaction.

A similar gesture is observed in the user-defined gesture set of Wobbrock et al. (2009). In their study, the gesture is used on an interactive tabletop to shrink objects. Instead of being a mid-air gesture, their version required the tips of the fingers to perform on a physical surface though.

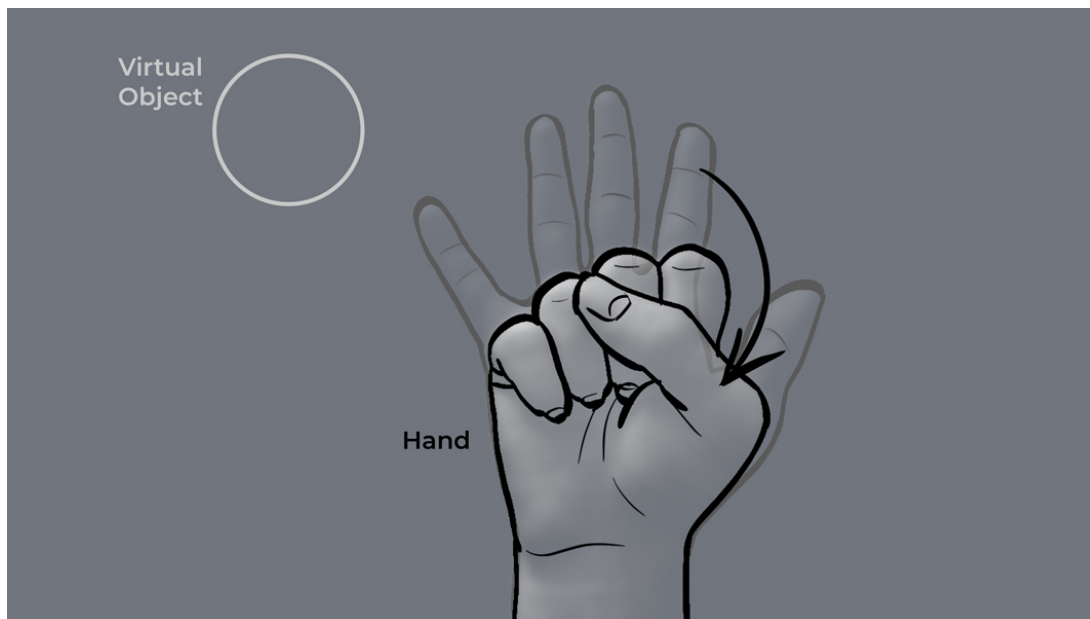


Figure 3.9 : Open Hand to Fist gesture.

- **Open Hand to Fist gesture:** This gesture, shown in Figure 3.9, starts with an open hand and requires making a fist. It is a unimanual discrete gesture. The form of the gesture is dynamic pose. It is proposed to implode relatively big (medicine ball-sized) objects that are beyond arm's reach. This gesture is mostly used to squeeze sponges in the real world. The gesture's motion path also imitates the implosion in an iconic manner as the splayed fingers are brought together towards the center of the hand. Hence, depending on the user's intention, the same gesture can be classified differently. Gesturing as if squeezing a sponge makes it pantomimic while mimicking the motion path of the implosion makes it iconic. Open Hand to Fist gesture produces peri-interaction self-kinesthetic feedback and provides an absolute down position. As the gesture does not require making contact with the object it is designed to be recognized regardless of its orientation around the viewing axis. This allows the gesture to be performed in varying orientations, from which the user can prefer the most ergonomic one for them.

According to Wobbrock et al. (2009), users tend to perform reversible gestures for dichotomous operations. This reversible nature also reduces users' cognitive load and makes the gestures easier to remember. Therefore, this gesture is also designed to be the reverse of the gesture that is used to explode objects that are beyond arm's reach.

3.3.2 System navigation

Navigation design entails developing, assessing, and implementing methods enabling users to navigate across an app. The navigation-based interactions in this thesis study are designed to be activated by mid-air hand gestures instead of graphical user interface (GUI) elements like floating buttons, menus, etc.

3.3.2.1 Design criteria of system navigation gestures

To navigate the system, semaphoric gestures that have static pose are selected because it is assumed that participants would mostly generate manipulative gestures, which are inherently motion-based, as they try to control virtual objects. Therefore, utilizing fundamentally different gestures, static semaphores, to navigate the system would allow the headset to easily differentiate between the object manipulation and system navigation gestures. Thus, effectively reducing the chances of unintentional triggers of system controls while participants generate gestures to manipulate objects.

However, semaphoric gestures have some shortcomings for gesture-controlled systems such as:

- Amongst all gesture types, semaphoric gestures are the most dependent on users' background and if previously unbeknownst, they can be hard to learn and remember,
- If they are arbitrarily mapped to tasks, they can be hard to learn because of the strain they put on user's memory,
- If they are overly generic, they may be performed and registered accidentally,
- They might be deemed vulgar for certain cultures,
- If they are extremely fluent, the system may fail to distinguish when and where they start/end.

Hence, a few decisions are made to compensate for these shortcomings. For instance:

- Gestures that are considered common knowledge are selected,
- For each task, a gesture is selected based on or in accordance with reason or logic,

- Gestures that require flexion or extension of some sort and don't position the hand in a relaxed state are selected to avoid unintentional triggers,
- Gestures that might be considered inappropriate in Turkish culture is avoided,
- Lastly, static gestures are chosen as dynamic gestures are harder to recognize for the current technology due to the added level of complexity introduced by the motion. Therefore, using static gestures improved the overall responsiveness of the system.

3.3.2.2 Proposed system navigation gestures



Figure 3.10 : Thumbs-up gesture.

- Thumbs-up gesture: This gesture, shown in Figure 3.10, is proposed to start the experiment. This is a unimanual semaphoric gesture. The form of the gesture is static pose. The flow of the gesture is discrete. It represents concepts such as requesting permission to act, approval, etc. This gesture is preferred because it is widely known and used in many countries around the world. Moreover, the concept of approval aligns with the intended interaction, which indicates being ready to start.

As this gesture doesn't require targeting a particular object and the gesture vocabulary does not include a different orientation of the same gesture, it is designed to be recognized in varying positions and orientations. Thus, it affords to be used in the most ergonomic way for each user. Nevertheless, the gesture

is proposed to be used as the palm facing sagittal plane, which puts the least amount of stress on the wrist.

The same gesture is also used for proceeding to the next scenario. The system decides which action to trigger based on the active scene. Using the same gesture for coherent interactions conditionally has some distinct advantages. As the user needs to learn one less gesture, the mental effort needed to operate the system reduces.

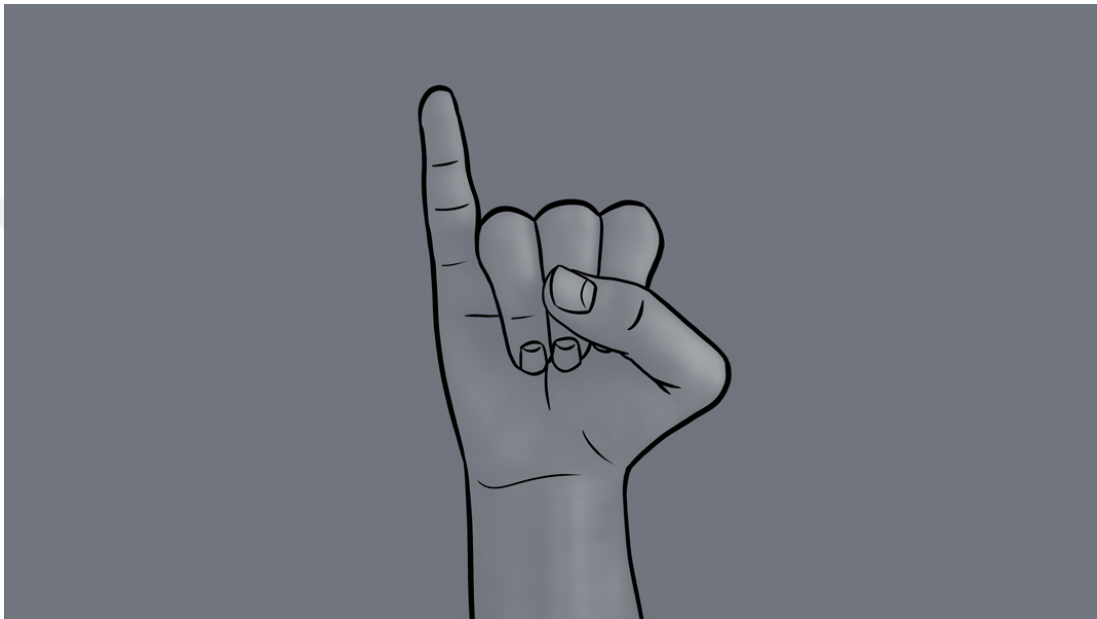


Figure 3.11 : Pinky-up gesture.

- Pinky-up gesture: In case a participant skips one scenario unintentionally, Pinky-up gesture, shown in Figure 3.11, is proposed for returning back to the previous scenario. Initially, Thumbs-up gesture positioned above the shoulder showing the user's back with the thumb (Thumbs-back) is considered for this interaction. It is assumed to be a good fit because not only does the meaning of the gesture match the task, but also it relates to the Thumbs-up gesture that is used to proceed to the next scenario. Both of these gestures are used to navigate through scenarios, semantically match with the task, and last but not least are the same gesture oriented differently, all of which would reduce the cognitive load. Therefore, making the gestures easier to learn and remember. However, the Thumbs-back gesture is positioned on the border of HMD's tracking volume, and the angular difference between the Thumbs-up and Thumbs-back gestures is not great enough, therefore failing to work reliably.

Consequently, a different approach is used. It is hypothesized that using the finger on the opposite end of the hand to achieve the opposite effect might prove useful at creating a coherent mental model for participants. Therefore, Pinky-up gesture is proposed for returning back to the previous scenario. This decision also enables Thumbs-up gesture to work more reliably by allowing it to be recognized regardless of its orientation. Pinky-up gesture is proposed to be used as the palm facing the sagittal plane to reduce the stress on the wrist.

3.4 User-Elicitation Study

Following the completion of the author-designed gesture vocabulary, a user-elicitation study is conducted to gather intuitively generated gestures that are assumed to improve the user experience (UX).

3.4.1 Referents

In general, the first step of a user-elicitation study is to define operations, named referents, that are going to be performed via gestures. A considerable amount of previously conducted user-elicitation studies (Arefin et al., 2016; Kray et al., 2010; Ruiz et al., 2011; Vatavu, 2012; Wobbrock et al., 2009) create a list of frequently used operations as referents. Yan et al. (2018) use objects instead of operations as all of the scenarios in their study explore the grasping operation of varying objects. This thesis study, however, offers only two operations (explode and implode) and focuses on exploring the variations in user-elicited gestures caused by objects' spatial properties such as position, orientation, and size. Therefore, as seen in Table 3.1, the referents of this study are generated based on multi-dimensional scenarios instead of operations alone. There are six scenarios in the user-elicitation study.

Table 3.1 : The scenarios in the user-elicitation study. Orientations are around the viewing axis.

Scenario	Operation	Object	Size	Target Distance	Orientation
1	Explode	Cylindrical	Big	Within arm's reach	0°
2	Explode	Cylindrical	Small	Within arm's reach	0°
3	Explode	Cylindrical	Big	Beyond arm's reach	0°
4	Implode	Double Conic	Big	Within arm's reach	0°
5	Implode	Double Conic	Small	Within arm's reach	45°
6	Implode	Double Conic	Big	Beyond arm's reach	90°

3.4.2 Animation

Harrison and Hummell (2010) define animation as “The rapid display of a sequence of images to create an illusion of movement”. For the user-elicitation study, two animations are designed to represent dichotomous explode and implode operations. As shown in Figure 3.12, the explosion animation is visualized with an object breaking down to cube-shaped parts that scatter around whereas the implosion animation is visualized with an object shrinking towards its center.

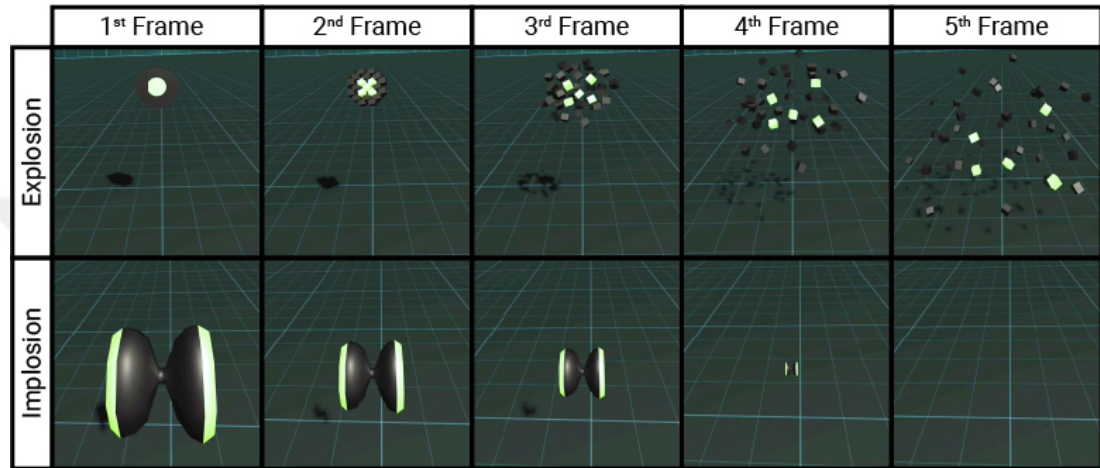


Figure 3.12 : Explosion and implosion animations.

3.4.3 Objects

For each operation, implode and explode, a unique object is designed. The objects are designed to be geometrically simple to reduce the time needed for participants to perceive the objects’ form upon seeing (Vuletic et al., 2018). Both of the objects have some design elements that provide cues on how to interact with them. For example, virtual hands and object surfaces intended to intersect with hands are painted in the same green color. Spatial properties of these two objects such as form, size, and placement are altered between scenarios in a way that would afford users to execute author-designed gestures. Figure 3.13 shows the objects in all six scenarios. The objects are designed to be as basic as possible to avoid distracting users and to make the objects more effective at leading users’ gaze through color, shapes, and angles to areas of interest.

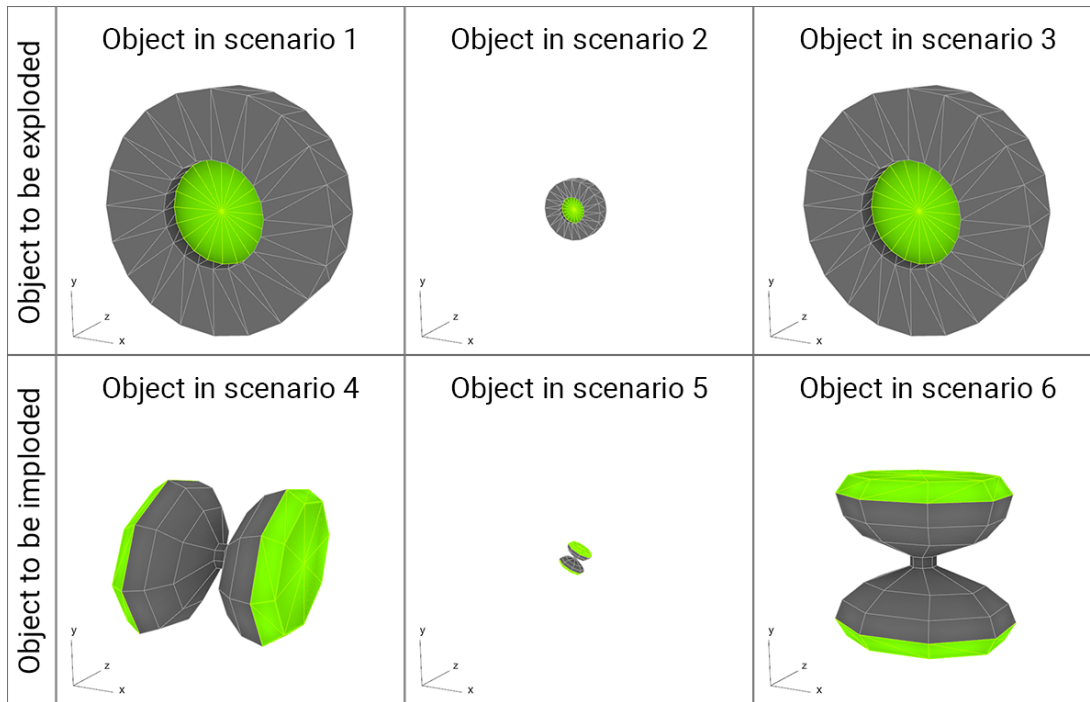


Figure 3.13 : Objects in each scenario.

3.4.3.1 Form

Form, also called shape, is a 3D configuration of mass (Who discovered 3D shapes, 2021). The hand-object intersections of the author-designed gestures are utilized as signifiers to make predefined gestures more discoverable. For instance, the green areas on the objects are shaped and sized to conform to the hand-object intersections of the author-designed gestures.

Object to be exploded is designed to be hit at a certain area when placed within arm's reach. This area is colored green and centered on a cylindrical object as seen in scenario 1 in Figure 3.13. Central placement and the extrude cut feature intend to focus users' gaze towards the center of the object. Z-axis shown in Figure 3.13 represents the viewing axis of the user for each scenario.

Object to be imploded is designed to be squished from certain areas when placed within arm's reach. These areas are colored green as seen in scenario 4 in Figure 3.13. Moreover, the center of the object is considerably narrower compared to the other areas. This form factor aims to guide users in bringing their hands together towards the object center.

3.4.3.2 Size

Size is the geometric quantity of an object such as volume, area, and length (Eldrandaly et al., 2019). It is hypothesized that the size of the object would affect user-elicited gestures. To test this hypothesis, objects' size is altered between scenarios. For instance, the object in scenario 2 is the same as the object in scenario 1 but smaller. Likewise, the object in scenario 5 is the same as the object in scenario 4 but smaller. The dimensions of all the objects in each scenario can be seen in Figure 3.14.

Additionally, the objects are sized to afford author-designed gestures. For example, a fist neatly fits into the green area of the object in scenario 1 as shown in Figure 3.15. Likewise, the tip of a finger can fit into the green area of the object in scenario 2.

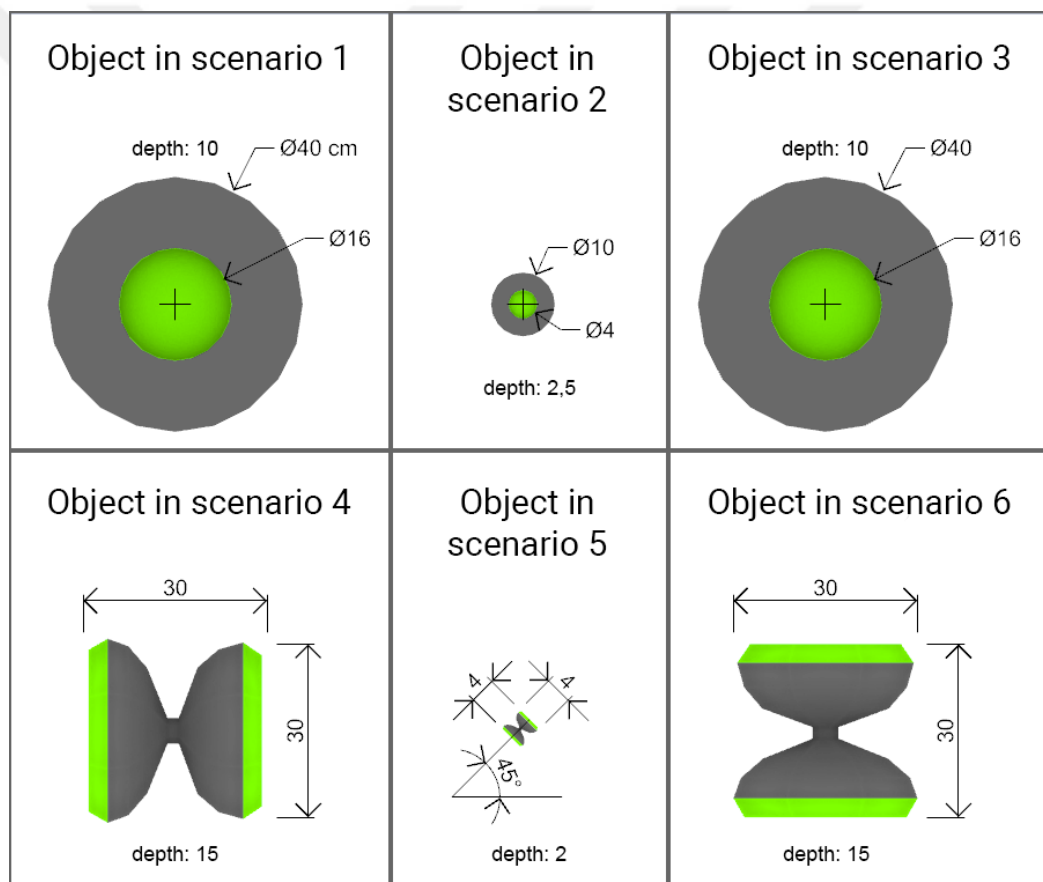


Figure 3.14 : Dimensions of the objects in each scenario.

3.4.4 Placement

Placement, also named pose, refers to the combination of position and orientation of an object. Placement of the object is hypothesized to affect the user-elicited gestures as well. To test this hypothesis, some scenarios introduce the same objects posed

differently. For instance, the object in scenario 3 is the same as the object in scenario 1 but placed further away from the user as seen in Figure 3.15. Likewise, the object in scenario 6 is the same as the object in scenario 4 but placed further away from the user as well as rotated 90° around the viewing axis.

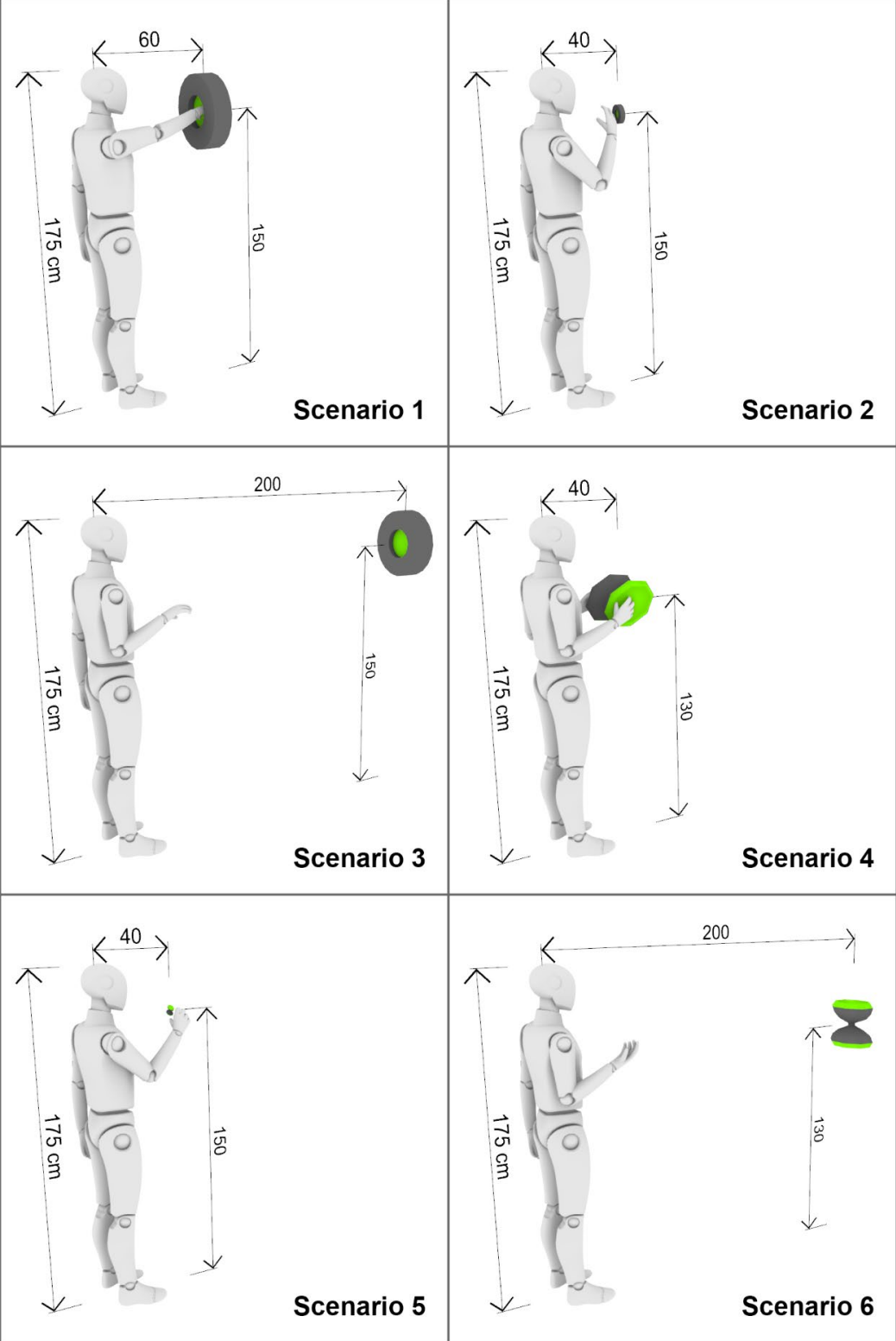


Figure 3.15 : The pose of the objects and the user’s upper limbs for each scenario.

3.4.4.1 Position

Position, also referred to as location, is where an object is within a space (Wentz, 2000). All of the objects are placed directly in front of the user for each scenario. On the other hand, target distance varies between scenarios. In this thesis, the distance that is reachable with an upper limb flexion while standing still is referred to as within arm's reach whereas anywhere beyond this distance is referred to as beyond arm's reach. As shown in Figure 3.15, the objects in scenarios 1, 2, 4, and 5 are placed within arm's reach whereas the objects in scenarios 3 and 6 are placed beyond arm's reach.

3.4.4.2 Orientation

Orientation is the direction or the rotation of an object (Eldrandaly et al., 2019). The orientation of the objects in the first three scenarios doesn't change as the purpose of these scenarios is to examine the gestural differences caused by the change in size and target distance. In contrast, for the remaining three scenarios the orientation of the object changes. In regards to the object in scenario 4, the object in scenario 5 is rotated 45° around the axis of view, and the object in scenario 6 is rotated 90° around the viewing axis as seen in Figure 3.14.

3.5 Choice-Based Elicitation Study

Following the user-elicitation study, a novel choice-based elicitation study is conducted. This methodology provides the end-users with a list of options from which the users pick the most suitable. The list can include both elicited and expert-designed gestures. Thus, the possibility of deciding on a gesture that is suitable from both the designer's and the users' standpoint increases. Because of this possible benefit, we decided to use a choice-based elicitation method for this thesis study. In conclusion, while developing appropriate gestures for any application is a difficult task, our study takes advantage of both user- and choice-based elicitation methodologies to find the most appropriate mid-air gestures.

In this thesis study, each participant is presented with two choices, a gesture produced by the participant themselves (user-elicited) and a predefined gesture (author-designed). Previous studies in literature used choice-based elicitation methodology before (Dim et al., 2016; Morris et al., 2010; Silpasuwanchai & Ren, 2015). However, all of these studies present either all user-elicited gestures that are collected from

different participants in a previous user-elicitation study and/or all author-designed gestures.



4. EXPERIMENT & RESULTS

4.1 Environment

For the experiment, a scenario-based setup was created. Each scenario consisted of 3-second animations. Due to the short duration of scenarios, even a glance might have participants miss the animation, which would affect the results. Therefore, a basic static environment was created to avoid distractions that could cause participants to shift their focus from the subject. The environment, shown in Figure 4.1, was a cubic room. The floor had a grid pattern, whose lines converged at the horizon to enhance perspective and depth perception. The walls and the ceiling also had the same grid pattern with a larger scale, which increased the consistency of the environment while hinting at the distance as well. To reduce eyestrain a dark environment was created.

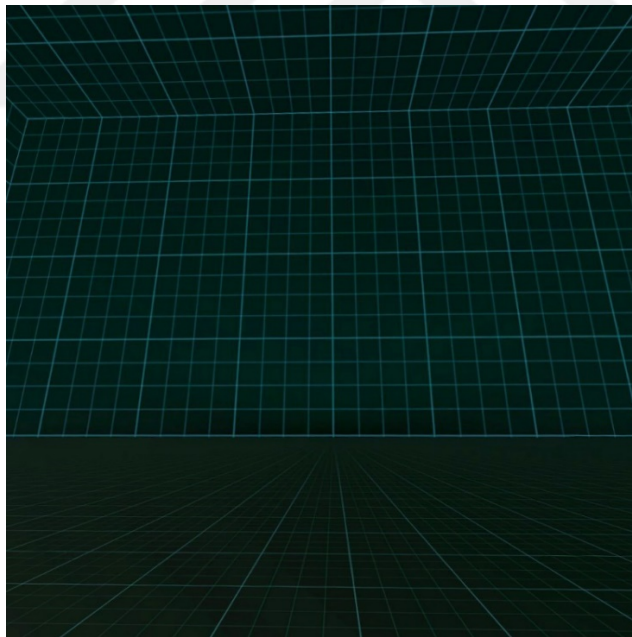


Figure 4.1 : Environment from participant's point of view.

4.2 User Interface

For the experiment, a minimal user interface (UI) was aimed. Graphical user interface (GUI) components that were created and optimized for 2D screens such as menus, buttons, sliders, etc., or their 3D counterparts, most of which were extruded in the 3rd

dimension and explored relatively extensively (Alger, 2015) over the past few years were avoided in this thesis study. To provide a more natural experience that was inspired by gestures accompanying speech, the system-level interactions were coded to be triggered by hand poses rather than virtual components.

4.2.1 Information screens

The starting scene of the experiment displayed an information screen, shown in Figure 4.2, that welcomed participants and overviewed the experiment one last time before the experiment began.

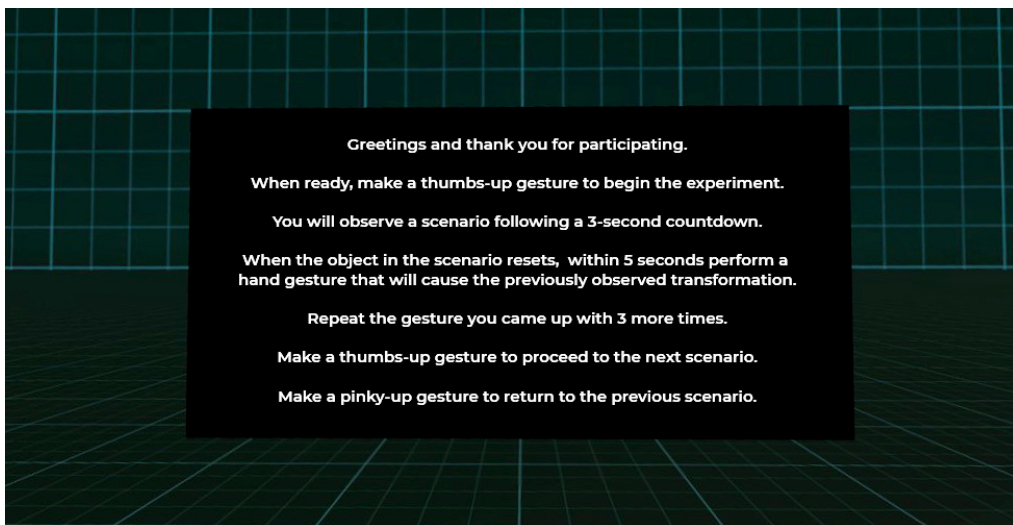


Figure 4.2 : The first information screen, welcoming and informing participants.

The second information screen, shown in Figure 4.3, appeared after the sixth scenario, stating the completion of the experiment's first trial. It also informed participants about the next steps such as learning predefined gestures and testing them.

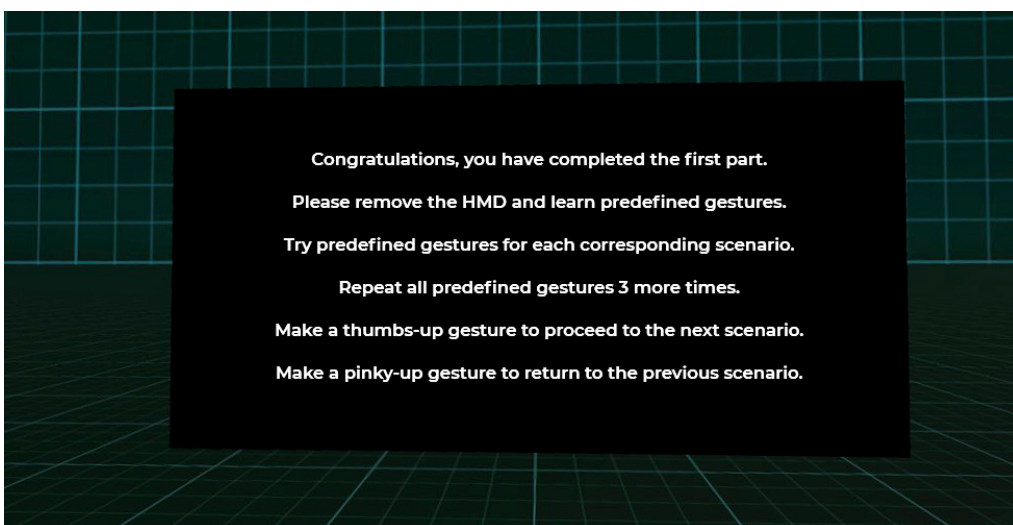


Figure 4.3 : The second information screen, showing the next steps to be taken.

4.2.2 Countdown timers

As part of the experiment, the scenarios could appear at different heights and distances. Due to short animation times (3 seconds), if the participants were gazing in the wrong direction when a scenario started, they could miss the animation. Therefore, at the beginning of each scenario a 3-second timer was shown to draw participants' gaze to the correct location, where the object would appear. The color of the text was set as white to increase the contrast with the dark background. At the end of the timer, a custom phrase 'OBSERVE' was shown instead of '0' to further emphasize what to do next. This sequence can be seen in Figure 4.4.

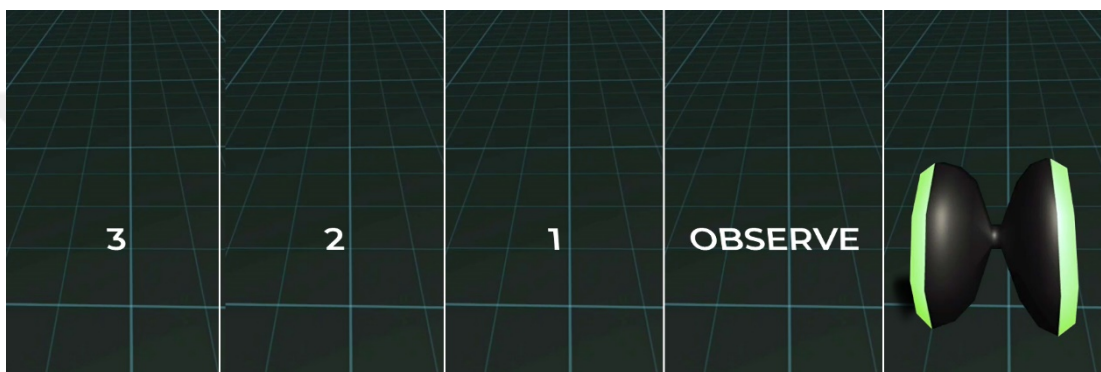


Figure 4.4 : Countdown timer showing remaining seconds before the next scenario began.

A secondary countdown timer, which indicated the remaining duration until the participants should produce a gesture was set to 5 seconds and located above the objects. This timer was smaller and had lower contrast than the other timer in order not to distract participants. An outlined circle, encasing the remaining seconds gradually shortened and eventually disappeared as the seconds went by as shown in Figure 4.5.

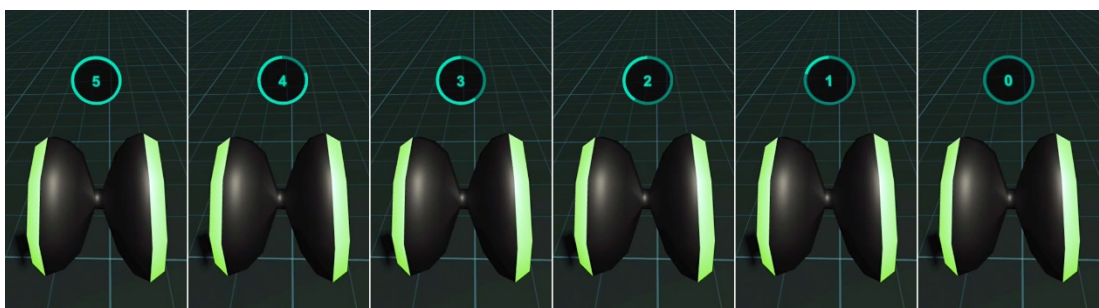


Figure 4.5 : Countdown timer showing remaining seconds to produce an intuitive gesture.

4.2.3 Stand here sign

To properly position and orient the participants in the virtual world, a stand here sign shown in Figure 4.6, was designed and placed at the origin of the scene. When participants positioned themselves comfortably while facing forward in the real world, the HMD's screen reset function was used to match their position and orientation to the origin of the virtual world. Between trials, while in standby mode, the headset could reposition the origin, causing the scene to displace. Hence, the stand here sign was utilized to point out this error.



Figure 4.6 : Stand here sign from participant's view looking down.

4.3 Development

4.3.1 Development tools

Development process is carried out using several software. Unity game engine is used to develop the app. 3D models are generated using the ProBuilder framework. 2D art assets are created using Adobe Illustrator. Predefined gestures are performed and recorded while using the VR experiment to generate gif animations, which are used to teach predefined gestures to participants. The footage is recorded using Xbox Game Bar, edited using Adobe After Effects, and rendered using Adobe Media Encoder.

4.3.1.1 Unity

Unity is a video game engine that allows 2D or 3D game development across multiple platforms such as mobile, desktop, web, and extended reality (XR). This engine is selected as the main development platform for its VR support and the extensive number of educational resources such as learning projects, video tutorials, etc. These resources eased and enhanced the development process. Unity version 2020.1.15f1 is used for the project.

4.3.1.2 Xbox Game Bar

Xbox Game Bar is a built-in windows 10 utility that enables screen capture along with a few other gaming-related functionality. In this study, it is used to record the VR experiment.

4.3.1.3 Adobe After Effects & Adobe Media Encoder

Adobe After Effects is a visual effects and motion graphics software that enables timeline-based video compositing. Adobe Media Encoder is a rendering solution and operates as an encoding engine for Adobe After Effects. In this study, predefined gestures are performed and recorded while using the developed app. This footage is then edited in Adobe After Effects to isolate each gesture. Subsequently, Adobe Media Encoder is used to export these clips in gif format. The gif animations are first used to teach predefined gestures to participants and later used in the survey to remind the predefined gestures.

4.3.1.4 Adobe Illustrator

Adobe Illustrator is a vector-based graphic design program. 2D UI elements such as ground pattern and stand here sign are created using this software.

4.3.2 Frameworks

Even though Unity game engine is used for core development, a few other software tools such as XR Plugin Management, Oculus Integration package, and ProBuilder are also used in the development process.

4.3.2.1 XR Plugin Management

This plugin is required to configure the unity project for XR development.

4.3.2.2 **Oculus Integration package**

Oculus Integration package is an all-in-one source to ease developing apps for Oculus devices in Unity. The package includes multiple essential software development kits (SDKs) such as scripts, core VR features, components, and plugins.

4.3.2.3 **ProBuilder**

ProBuilder is a software consisting of level design and 3D modeling tools, enabling quick prototyping inside Unity game engine. The objects in scenarios are modeled using this software. ProBuilder is chosen because of its capability to work natively.

4.3.3 **Project setup**

First, a Unity project was created and the build environment was set to Android as Oculus Quest 2 headset utilizes Qualcomm Snapdragon XR2 system-on-chip, which is designed specifically for VR headsets and is a variant of Snapdragon 865. Following, a few player settings were set such as company name and product name. Under the Other Settings tab, the minimum Application Programming Interface (API) level was set to Android 6.0 'Marshmallow' (API level 23). Later, XR plugin Management was installed to enable VR development in Unity. Under the Android tab, Oculus was selected as Plug-in Provider. Finally, the Oculus Integration package was downloaded from the asset store and imported into the project.

4.3.4 **Oculus hand tracking setup**

To enable head and hand tracking in the app, the main camera in the scene was removed and OVRCameraRig was added instead. Tracking Origin Type of the OVRCameraRig was set to Floor Level, which used the height of the player to position the camera. Hand Tracking Support was set to Controllers and Hands to enable hand tracking while maintaining controller support.

To visualize virtual hands in the project, OVRHandPrefab was added as a child of LeftHandAnchor and RightHandAnchor gameobjects under OVRCameraRig, and the fields of each prefab were populated to configure them as left and right hands respectively. A custom material named LimeHandMaterial was created and set for both OVRHandPrefabs. Subsequently, to manage left and right hands, HandsManager prefab was added to the scene. Customized OVRHandPrefab gameObjects were added

to Left Hand and Right Hand fields of HandsManager and Visual Mode of the virtual hands was set to Mesh.

4.3.5 Object animations

The explosion and implosion animations were coded in Unity as well. Object to be exploded (cylindrical object) was modeled based on the form factors detailed in Section 3.4.3. A broken version of the cylindrical object was constructed using 42 cubes. Each cube was given a box collider and rigid body. These cubes were placed in varying orientations and slightly overlapped each other. This arrangement coupled with an appropriate mass value allowed the cubes to scatter around when instantiated. The cylindrical object had a custom script named Exploder. This script managed the explosion effect. Likewise, the object to be imploded (double conic object) had a custom Imploder script. This script managed the implosion effect. It scaled the object down uniformly in all three dimensions until it became invisible to the naked eye.

4.3.6 Direct interaction

To directly interact with the objects in the scene via virtual hands, InteractableToolsSDKDriver prefab was added to the scene. This prefab came populated with FingerTipPokeToolIndex prefab for each hand. The InteractableToolsSDKDriver prefab essentially puts a little sphere (FingerTipPokeToolIndex prefab) to the tip of the index finger of each hand. These spheres enabled direct interaction with other objects in the scene. The default configuration had a shortcoming though as the interaction would take place if and only if the tip of the index finger collided with the objects.

For real-time visualization of virtual hands, Oculus SDK tracks all 14 phalanges of each hand, 1st and 5th metacarpals, the root of the wrist, and a few other points. As seen in Figure 4.7, OVRSkeleton script has a public enum containing the full list of boneIDs for all these points. To enable interaction with other parts and surfaces of the hands beyond the tip of the index finger, these boneIDs were utilized to target particular bones. However, putting interaction spheres on too many hand bones reduced the app performance significantly. Therefore, based on the novel hand-object intersections classification proposed in this thesis study, a few key bones were selected to place interaction spheres. Thumb metacarpal bone named Hand_Thumb1 was targeted to enable palm first interactions. The middle proximal phalange named Hand_Middle1

was selected for phalange first interactions. Lastly, the middle distal phalange named Hand_Middle3 was targeted to enable fingertip interaction for the middle finger. These boneIDs afforded all direct manipulation gestures in the predefined gesture set and many more to be recognized.

```

46 47 bayuru
public enum BoneId
47
48 {
49     Invalid = OVRPlugin.BoneId.Invalid,
50
51     Hand_Start = OVRPlugin.BoneId.Hand_Start,
52     Hand_WristRoot = OVRPlugin.BoneId.Hand_WristRoot, // root frame of the hand, where the wrist is located
53     Hand_ForearmStub = OVRPlugin.BoneId.Hand_ForearmStub, // frame for user's forearm
54     Hand_Thumb0 = OVRPlugin.BoneId.Hand_Thumb0, // thumb trapezium bone
55     Hand_Thumb1 = OVRPlugin.BoneId.Hand_Thumb1, // thumb metacarpal bone
56     Hand_Thumb2 = OVRPlugin.BoneId.Hand_Thumb2, // thumb proximal phalange bone
57     Hand_Thumb3 = OVRPlugin.BoneId.Hand_Thumb3, // thumb distal phalange bone
58     Hand_Index1 = OVRPlugin.BoneId.Hand_Index1, // index proximal phalange bone
59     Hand_Index2 = OVRPlugin.BoneId.Hand_Index2, // index intermediate phalange bone
60     Hand_Index3 = OVRPlugin.BoneId.Hand_Index3, // index distal phalange bone
61     Hand_Middle1 = OVRPlugin.BoneId.Hand_Middle1, // middle proximal phalange bone
62     Hand_Middle2 = OVRPlugin.BoneId.Hand_Middle2, // middle intermediate phalange bone
63     Hand_Middle3 = OVRPlugin.BoneId.Hand_Middle3, // middle distal phalange bone
64     Hand_Ring1 = OVRPlugin.BoneId.Hand_Ring1, // ring proximal phalange bone
65     Hand_Ring2 = OVRPlugin.BoneId.Hand_Ring2, // ring intermediate phalange bone
66     Hand_Ring3 = OVRPlugin.BoneId.Hand_Ring3, // ring distal phalange bone
67     Hand_Pinky0 = OVRPlugin.BoneId.Hand_Pinky0, // pinky metacarpal bone
68     Hand_Pinky1 = OVRPlugin.BoneId.Hand_Pinky1, // pinky proximal phalange bone
69     Hand_Pinky2 = OVRPlugin.BoneId.Hand_Pinky2, // pinky intermediate phalange bone
70     Hand_Pinky3 = OVRPlugin.BoneId.Hand_Pinky3, // pinky distal phalange bone
71     Hand_MaxSkinnable = OVRPlugin.BoneId.Hand_MaxSkinnable,
72     // Bone tips are position only. They are not used for skinning but are useful for hit-testing.
73     // NOTE: Hand_ThumbTip == Hand_MaxSkinnable since the extended tips need to be contiguous
74     Hand_ThumbTip = OVRPlugin.BoneId.Hand_ThumbTip, // tip of the thumb
75     Hand_IndexTip = OVRPlugin.BoneId.Hand_IndexTip, // tip of the index finger
76     Hand_MiddleTip = OVRPlugin.BoneId.Hand_MiddleTip, // tip of the middle finger
77     Hand_RingTip = OVRPlugin.BoneId.Hand_RingTip, // tip of the ring finger
78     Hand_PinkyTip = OVRPlugin.BoneId.Hand_PinkyTip, // tip of the pinky
79     Hand_End = OVRPlugin.BoneId.Hand_End,
80
81     // add new bones here
82
83     Max = OVRPlugin.BoneId.Max
84 }

```

Figure 4.7 : List of tracked points on each hand and their boneIDs in OVRSkeleton script.

To place an interaction sphere on these bones, OVRPlugin and FingerTipPokeTool scripts were customized. In the OVRPlugin script, two new variables were added to the HandFinger enum named ThumbMetacarpal and MiddleProximal. Then, in FingerTipPokeTool script, these variables were associated with corresponding boneIDs. Lastly, two new prefabs named PokeToolThumbMetacarpal and PokeToolMiddleProximal were created for the newly designated bones, and were added to fields in InteractableToolsSDKDriver gameobject as additional elements. The tip of the middle finger didn't require custom coding or prefab creation as Oculus already offered defined variables and prefabs for each fingertip. Therefore, FingerTipPokeToolMiddle prefab was also added as yet another element. As seen in Figure 4.8, in the end, each hand had four interaction spheres on them. The mesh of interaction spheres was hidden in the final build to avoid distraction.

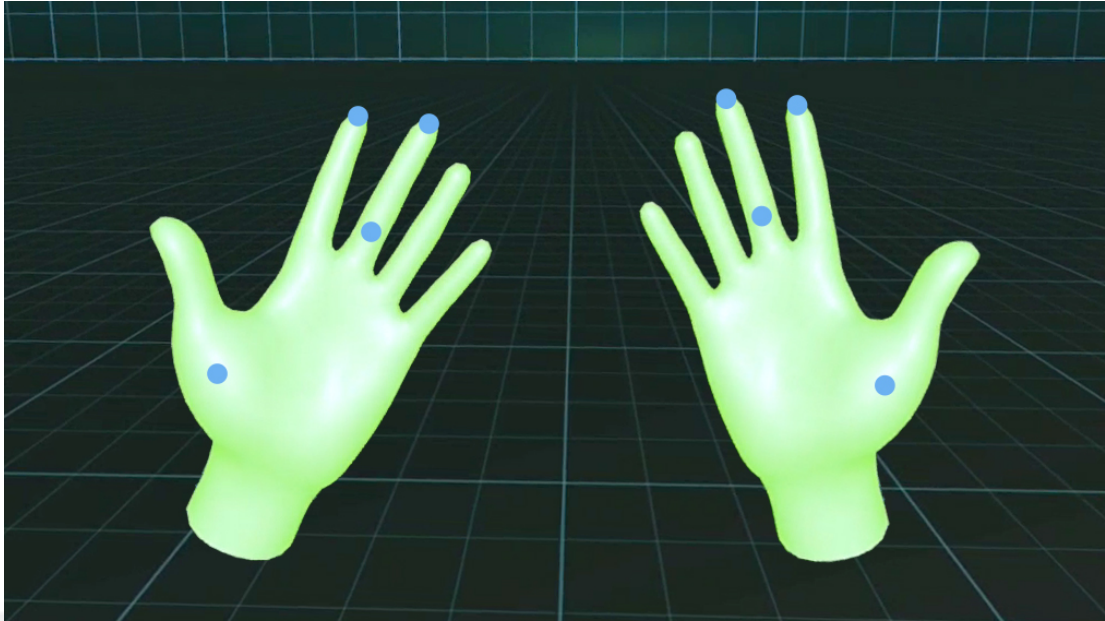


Figure 4.8 : Interaction spheres on each hand.

Next, objects to interact with were modeled and added to the scene. Then, a collider and ButtonController script were added to the objects. The colliders were used to populate the Action Zone field in the ButtonController script. This script was responsible for detecting the collision between the interaction spheres on virtual hands and the object's collider. Lastly, a custom script named ButtonListener was created. It serialized an event handler for the action state of the gameobject. This allowed triggering any public function of a gameobject from the inspector. The event listener was used to invoke implode and explode functions for the corresponding objects.

4.3.7 Mid-air hand gesture detection

To detect mid-air hand gestures such as Four Finger Flick and Thumbs-up, a custom script named GestureDetector was created and added to an empty gameobject with the same name (Valem, 2020). First, all the tracked points on the left and right hands were called and stored in a list. Then, a keypress was coded to record the spatial data of all the bones in both hands from the editor. Using Oculus Link, HMD was connected to Unity. This connection enabled live app testing without building the app. In play mode, all predefined gestures were posed one by one with the right hand, and the spatial data of all the bones for each gesture were collected using the GestureDetector component (Valem, 2020). This process was then repeated for the left hand. The spatial data were used to create the gesture vocabulary. As shown in Figure 4.9, this vocabulary consisted of ThumbsUp, PinkyUp, OpenHandPalmUp, OpenHandPalmDown,

FistPalmUp, and FistPalmDown. Both PalmUp and PalmDown versions were included for Fist and OpenHand gestures to increase their recognition rate in different orientations. Additionally, a threshold number was set to define acceptable deviation on the spatial data for gesture recognition. Lastly, an event handler was serialized for each gesture to enable calling any gameobject's public functions from the inspector when a particular gesture is recognized.

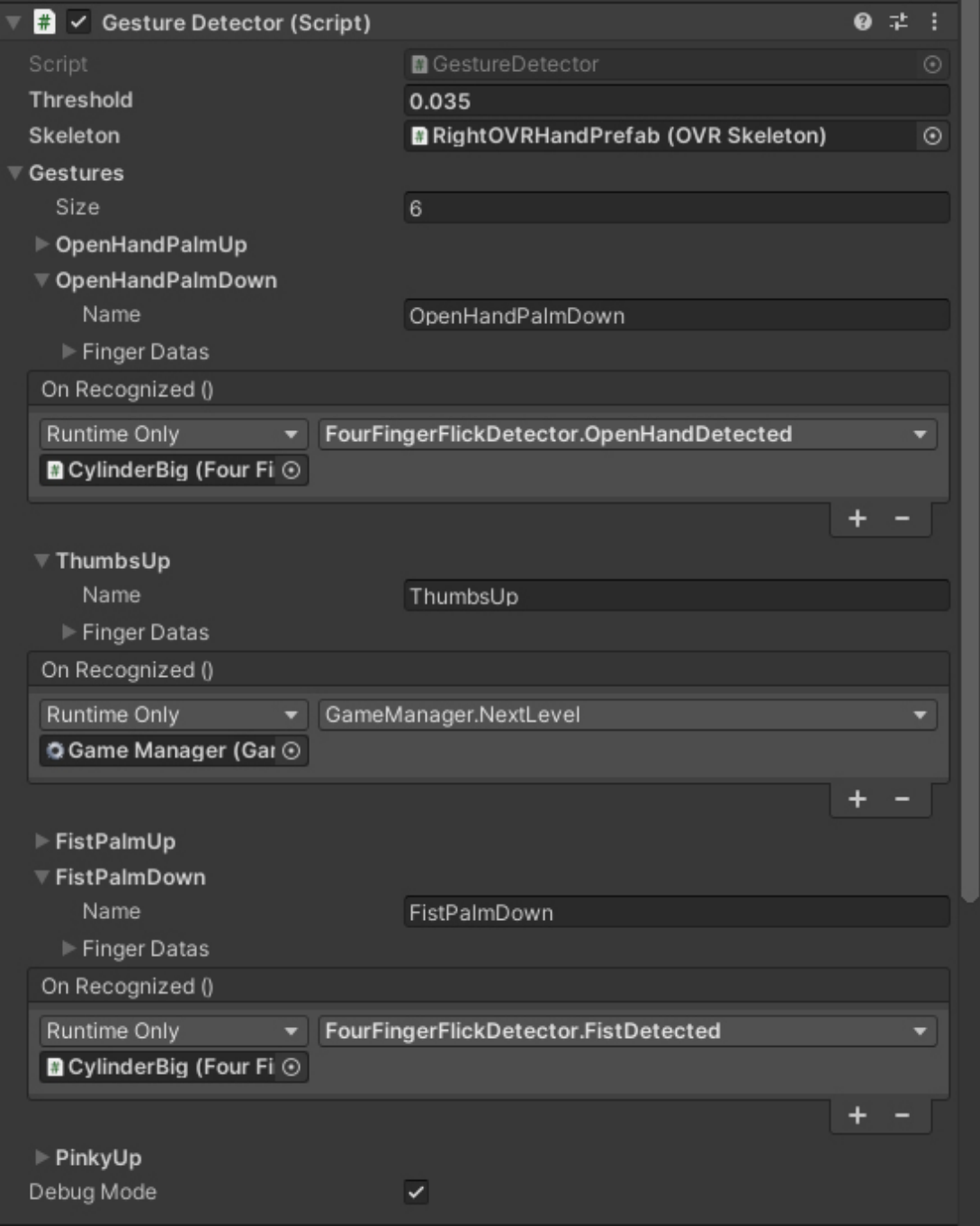


Figure 4.9 : GestureDetector component of right hand for scenario 3.

Thumbs-up and Pinky-up gestures, whose form was static pose, were easily coded to load next and previous scenes respectively when detected. Open Hand to Fist and Four Finger Flick gestures' form, however, was dynamic pose. Initially, the ending poses of these gestures were used to invoke their animations, but the ending pose of the Four Finger Flick gesture was nearly identical to the relaxed state of the hand. Therefore, the animation kept triggering unintentionally. Hence, to invoke these dynamic gestures, both the starting and ending poses were used. For Four Finger Flick gesture, a custom script named FourFingerFlickDetector was created and assigned to the object to be remotely exploded. The flowchart of this script's logic was shown in Figure 4.10. This script had two booleans named fist and openhand that were set to false at the beginning. Both FistPalmUp and FistPalmDown gestures (two variants of the starting pose) toggled fist boolean to true. The same gestures also invoked a 2-second timer that toggled fist boolean back to false if it was still true when the timer ended. Both OpenHandPalmUp and OpenHandPalmDown gestures (two variants of the ending pose) toggled openHand boolean to true if fist boolean had been true when the ending pose was detected. Therefore, using both gestures in succession within 2 seconds set both booleans to true. Finally, update function checking for both booleans' true condition, played the explosion animation, and then set both booleans back to false, therefore, resetting the process.

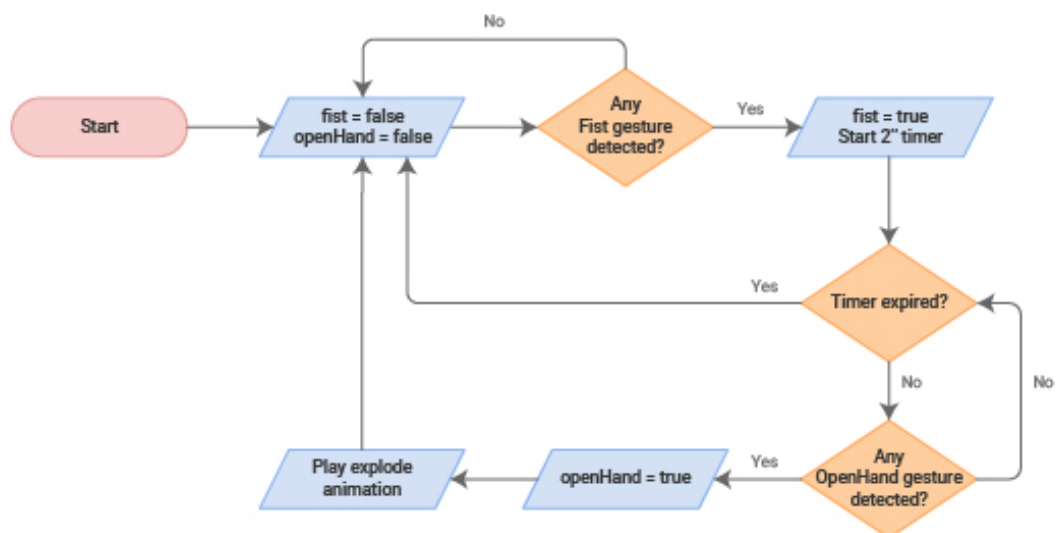


Figure 4.10 : Flowchart of FourFingerFlickDetector script's logic.

4.3.8 Scene and timer management

Each scenario in the experiment was set as a different scene in Unity. To manage scene transitions and countdown timers, an empty gameobject named GameManager was created and a custom script with the same name was added to it. GameManager script serialized two fields that were responsible for loading previous and next levels when invoked. The script also managed the timers appearing at the beginning of each scene. It set the duration of the timer and displayed a custom message (observe) at the end. Moreover, the script managed a secondary timer, which showed the remaining seconds left to gesture for each scenario.

4.4 Participants

15 participants (8 female & 7 male), with ages ranging from 18 to 54 years, were chosen for the experiment. All of the participants were selected amongst Turkish citizens. To investigate the effects of the dominant hand in gestural behaviors: right-handed, left-handed, and ambidextrous people were included in the experiment. To prevent technical bias, participants from a wide variety of occupational groups were selected. Moreover, both gamer and non-gamer participants were recruited to compare their results. VR experience was another important factor that was considered in the participant recruitment process. Thus, people with and without VR experience were recruited for the experiment. Only one participant owned an Oculus Quest 2 headset. Therefore, all participants except this participant, who used their own headset, used the author's HMD for the experiment.

4.5 Constraints & Limitations

The experiment had the following constraints and limitations.

4.5.1 Temporal constraints

- In order to collect intuitive gestures, participants were only allowed 5 seconds to initiate a gesture. Any gesture made after 5 seconds was not included in the study.

4.5.2 Participant constraints

- Participants were asked to stand on a specific spot without moving their lower limbs during the experiment. However, they were given complete freedom of upper limb motion for gesturing.
- Participants were also informed to gaze forward for the duration of the experiment.

4.5.3 Technological limitations

- The environment must be well lit in order for the cameras on the HMD to recognize participant's hands. Without proper lighting levels, virtual hands can't be visualized.
- There are four inside-out cameras on the HMD. To increase the tracking volume, all of these cameras record the hands of the user individually. The recordings are then stitched together using an algorithm to determine the position of the real hands. Because of these stitches, there can be occasional jittering of fingers.
- If the participant's hands touch or occlude each other, the virtual hands disappear instead of showing an approximation of the real hands. This can break the immersion.
- The hands cannot be moved outside of the HMD's tracking volume when gesturing. This means they cannot be moved to the back of the head or stretched all the way to the sides otherwise any gestural input performed would not be recognized.
- The Oculus touch controllers and hands cannot be used at the same time. Using controllers disables virtual hands.

Participants weren't informed about the technological limitations in order not to interfere with the intuitiveness of their gestures. To compensate for the limitations of the cameras on the HMD, a secondary camera was placed as a failsafe mechanism in front of the participants to record their gestures from another viewpoint.

4.6 Experiment Flow

Before the experiment, participants were not informed of the purpose of the experiment in order not to create any bias. Nevertheless, they were informed about how they will be involved through a participant information sheet as seen in Figure A.2.1. A walkthrough of the experiment was also provided to inform participants on how to operate as shown in Figure A.3.1.

During the experiment, the aforementioned 6 scenarios were tested in two trials. The first trial (user-elicitation study) aimed to collect participants' intuitive gestures, while the second trial (testing predefined gestures) aimed to collect participants' perception of predefined gestures. Lastly, a survey was conducted to compare and evaluate both gesture vocabulary. The flow of the experiment is shown in Figure A.4.1.

To test these scenarios, participants were asked to wear an Oculus Quest 2 VR headset, stand up, look forward, and give a Thumbs-up gesture with either one of their hands when ready. Then, they were informed to lower their hands to a natural relaxed position by their sides. Their real-world position was mapped to an in-game stand here sign to make sure they were positioned and oriented correctly in VR.

Following their Thumbs-up gesture, a 3-second countdown timer appeared in front of them. When the countdown was over, a virtual object appeared, went through a transformation, and disappeared. Participants were asked to observe the transformation while staying still and cause the previously observed transformation using hand gestures when the object reappeared. The animation lasted 3 seconds before disappearing, and 1 second later the object reappeared. Participants were given 5 seconds to initiate a gesture following the reappearance of the object.

Short animation time and fast response requirements were designed to capture participants' intuitive reactions. Temporal restrictions were placed to prevent participants from having enough time to think about alternative ways of achieving the same effect, creating different mental models, and applying the best solution they believe they came up with. Given enough time, they might even become biased, starting to think how they would achieve the same effect using existing interfaces and come up with a solution based on the way other interfaces work (Vuletic et al., 2018).

Throughout the experiment, participants were only allowed upper limb movements, so they could use their arms, forearms, hands, fingers, etc. They were also free to utilize

either one hand or both hands. Moreover, they were free to touch or grab the object whether they deemed necessary. The near-field objects, which were strongly assumed to be touched by participants, were coded to react to participants' touch, and their animations were played at the time of contact. The far-field objects, however, were coded to react to certain hand poses. After gesturing for the first scenario, participants were asked to repeat the gesture they came up with three more times while voicing their reasoning as to why they used that specific gesture.

Participants were instructed to give another Thumbs-up gesture with either one of their hands indicating they are ready to move forward, and after another 3-second countdown the next scenario began. They were allowed to use the same gesture for multiple scenarios if deemed necessary. The first trial was completed after the participants gestured for all 6 scenarios.

Following the first trial, participants were instructed to take off the HMD by an in-game information screen. Next, they were informed about the context of the experiment and were taught predefined gestures via gif animations and oral instructions. Then, participants returned to the VR experiment and the second trial began. This time, instead of coming up with gestures, participants tested predefined gestures. They were again instructed to repeat each gesture three more times while voicing their thoughts. This process was repeated for all six scenarios.

Lastly, an in-person survey was conducted to gather data regarding the experiment.

4.7 Data Collection

The portable nature of the required technologies allowed the experiment to be set up at varying test locations for different participants. We collected various qualitative and quantitative data through screen recording, video & audio recording, concurrent think-aloud protocol (CTA), and in-person survey.

4.7.1 Screen recording

To capture the gestures of the participants in relation to the objects in VR, HMD's screen was streamed to a personal computer (PC) using the headset's native casting functionality. The whole experience of the participant was then recorded via Xbox

Game Bar on PC. This method was particularly useful in identifying gestures' directness as it recorded hand gestures and the VR environment at the same time.

4.7.2 Video & audio recording

A camera positioned in front of the participants was used to record gestures from another point of view (POV). This method was especially effective at recording finer details of the gestures that might be missed by the HMD. As mentioned in Section 4.5.3, the HMD might fail to visualize virtual hands due to occlusion, lack of light, or gesturing outside the tracking volume of the HMD's cameras. Hence, the secondary camera was utilized as a failsafe mechanism against these probabilities.

4.7.3 Concurrent think-aloud protocol

During the experiment, participants were asked to voice their chain of thoughts on why they used the particular gesture they came up with, what they aimed to achieve using this gesture, etc. Participants were asked to present their mental models during the repetition of the gesture, rather than when the gesture was first executed, in order not to reduce task performance. Knowing the mental model of the participants was crucial in classifying their gestures according to the nature dimension, since the nature of a gesture can differ depending on the participant's intention.

4.7.4 In-person survey

To profile participants and collect their views on a series of subjects a survey was created. The survey was conducted face-to-face to help participants remember their intuitive gestures by showing their recordings if needed. Throughout the survey; five-point Likert scale, five-point semantic differential scale, multiple choice single response checklist, multiple choice multiple response checklist question types were used. To avoid primacy and sequencing effects, the order of choices was randomized. The survey is created using Google Forms. To avoid ambiguity on whether a participant skipped a question or couldn't find an appropriate answer, every question was marked as required and included 'none of the above' response when applicable. The survey included two main sections.

4.7.4.1 Participant profile section

This section was intended to profile testers on multiple levels. It consisted of two parts: general profile and player profile.

- General profile part: The questions in the first part aimed to create a general profile of the participants and inquired participants' age, dominant hand, self-declared health condition, education, and occupation. Self-declared health condition question was especially important considering the condition, as well as its severity, might affect the performance and gathered data. To be able to identify the severity of health conditions, the five-point Likert scale question type was used. The points were none, minor, moderate, major, and extreme. The general profile part of the survey was shown in Figure A.5.1.
- Player profile part: As shown in Figure A.6.1, the second part was focused on player characteristics of the participants and information such as gaming frequency, VR experience, gesture tech experience, gesture-controlled VR game experience, and VR gaming frequency in the past 12 months were collected.

4.7.4.2 Gesture evaluation section

The second section had participants review predefined gestures based on multiple criteria. This section was presented in 6 parts, one for each scenario.

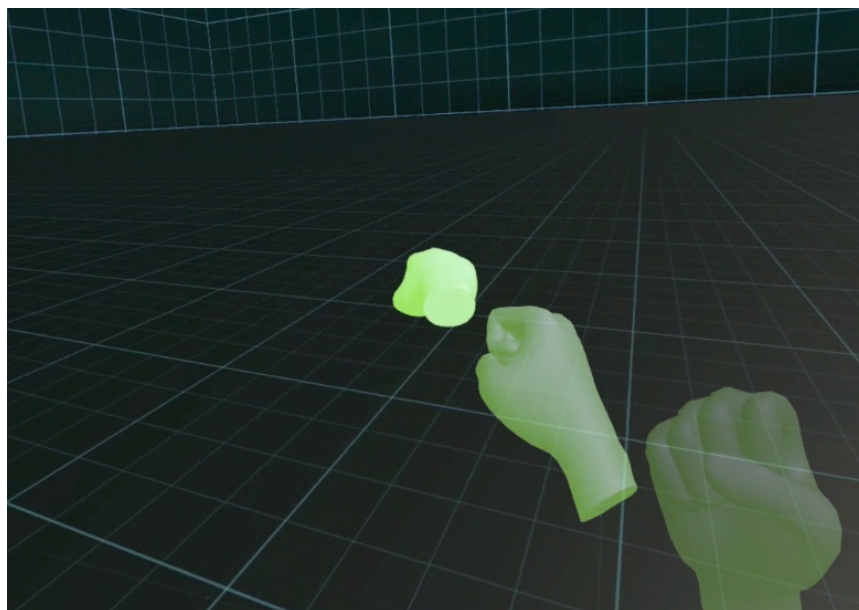


Figure 4.11 : Punch gesture performed in space.

First, participants' perception of predefined gestures' familiarity, complexity, and effortlessness was collected using a five-point semantic differential scale. The polar adjectives were familiar-unfamiliar, simple-complex, and effortless-fatiguing. As seen in Figure 4.11, the gesture itself being performed in VR was presented as a looping gif animation above the corresponding question to help participants remember the gesture before rating it.

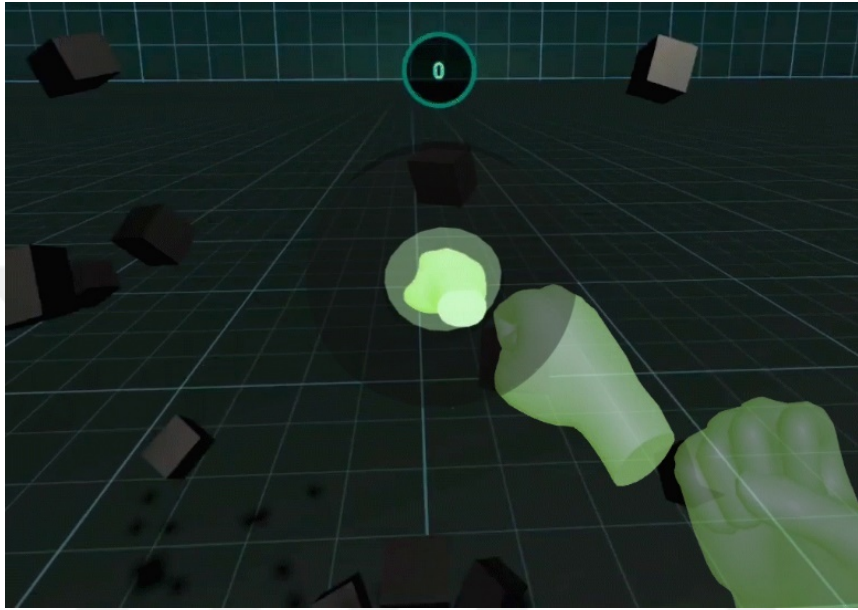


Figure 4.12 : Punch gesture performed on the object.

Second, the same gesture is reviewed based on the scenario it was used. Here, another five-point semantic differential scale reviewed participants' perception of how logical, natural, and fun the predefined gesture was for the scenario, in which it was used. The polar adjectives were logical-arbitrary, natural-pretended, and fun-boring. As seen in Figure 4.12, the question included a looping gif animation of the gesture performed on the object, which reacted as observed in the VR experiment.

Third, a five-point Likert scale, ranging from strongly disagree to strongly agree, 1 to 5 respectively, was presented to evaluate whether the interaction was more satisfying due to gesture producing kinesthetic feedback. This question was not used in the 1st scenario as the Punch gesture didn't produce any kinesthetic feedback when performed.

Subsequently, participants were asked whether they consider their intuitively generated gesture or predefined gesture to be more appropriate for this scenario. To

remind them of their intuitively generated gestures, the screen recording of the VR experiment was shown.

Lastly, participants were prompted to select up to three factors (user interaction qualities) out of seven, which made them consider one gesture more appropriate than the other. The choices were the gesture being simpler, more intuitive, natural, fun, effortless, logical, or satisfactory due to kinesthetic feedback.

All of the questions mentioned in this section were asked for each scenario as shown in Figure A.7.1, Figure A.8.1, Figure A.9.1, Figure A.10.1, Figure A.11.1, and Figure A.12.1.

4.8 Results

4.8.1 Survey results

All the answers gathered in the participant profile section are shown in Figure A.13.1.

4.8.1.1 General profile

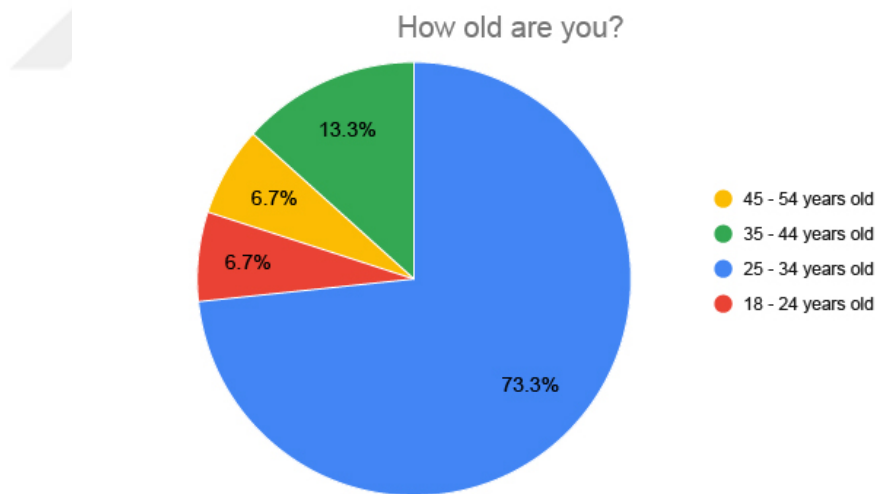


Figure 4.13 : Age distribution of the participants.

15 participants (7 male & 8 female) were recruited for the experiment. All of them were Turkish citizens. As shown in Figure 4.13, 11 of the participants were aged between 25-34. There was 1 person each in 18-24 and 45-54 age groups, the remaining 2 participants were aged between 35-44.

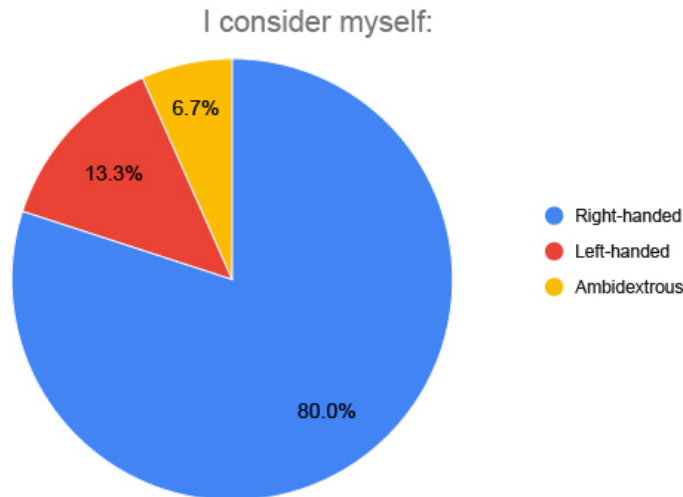


Figure 4.14 : Dominant hand distribution of the participants.

As seen in Figure 4.14, 12 participants were right-handed, 2 participants were left-handed, and 1 participant was ambidextrous.

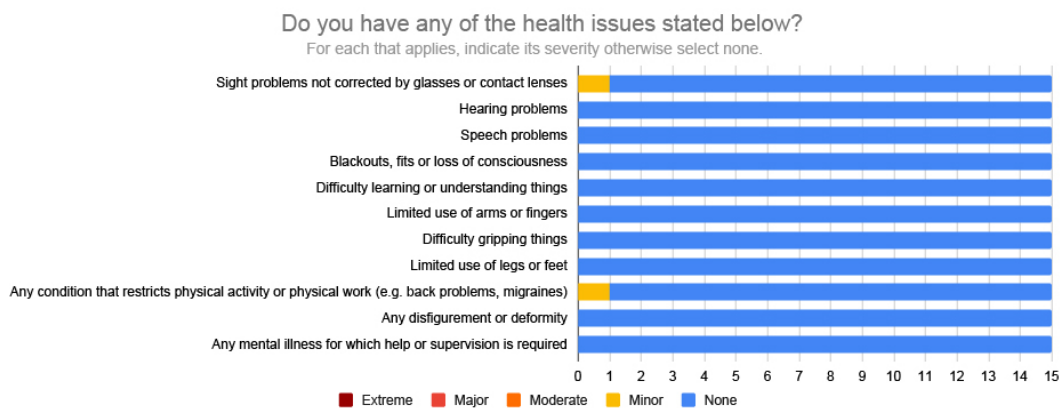


Figure 4.15 : Self-declared health conditions of the participants.

2 people reported health problems in the self-declared health condition question. Participant 8 disclosed minor sight problems not remedied by eyewear (5% sight loss in their right eye due to a childhood accident), and participant 12 declared minor back problems that restrict physical activity as shown in Figure 4.15. While Participant 8 couldn't generate a gesture for the 6th scenario within 5 seconds, both of the participants were able to complete the experiment.

What is the highest level of education you have completed?

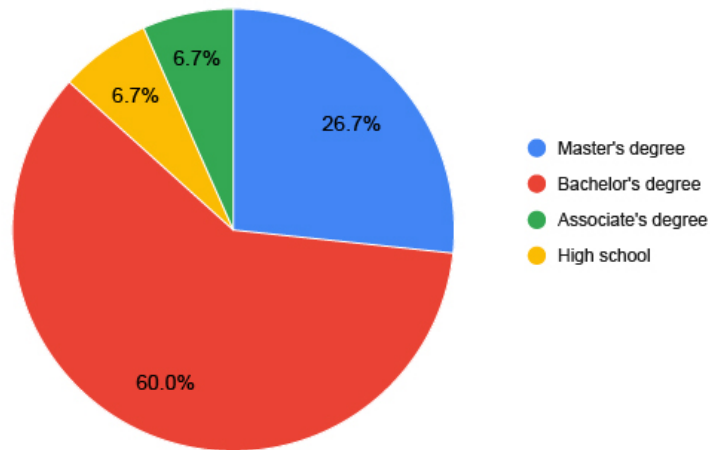


Figure 4.16 : Education distribution of the participants.

Education level of the participants was also collected. As shown in Figure 4.16, 4 participants had master's degree, 9 participants had bachelor's degree, 1 participant had associate's degree and 1 participant was a university student.

What is your occupation?

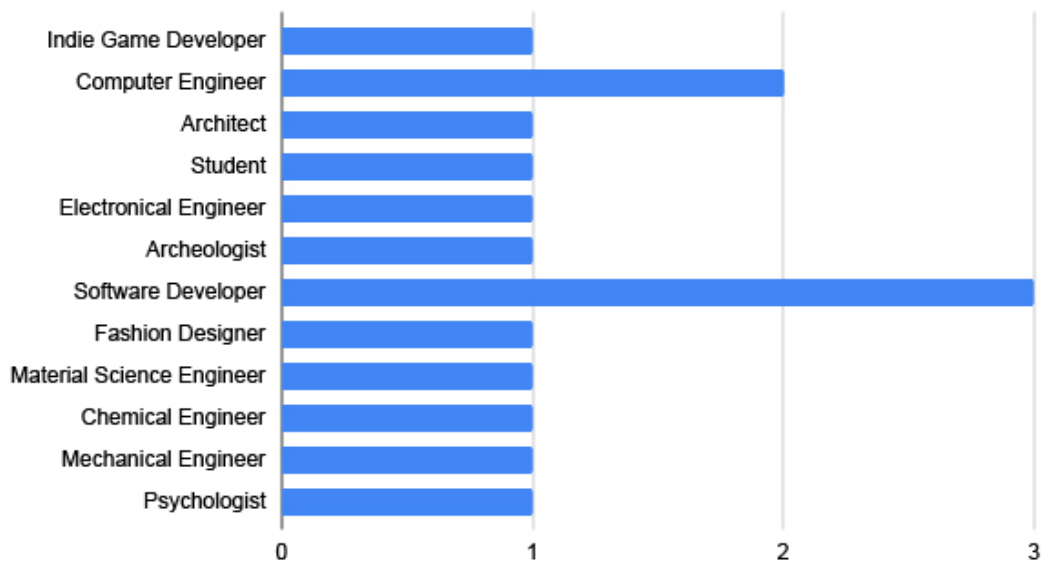


Figure 4.17 : Occupation distribution of the participants.

To avoid technical bias, participants were selected from varying professions such as psychologist, archeologist, fashion designer, and software developer. As seen in Figure 4.17, 12 different occupations were reported.

4.8.1.2 Player profile

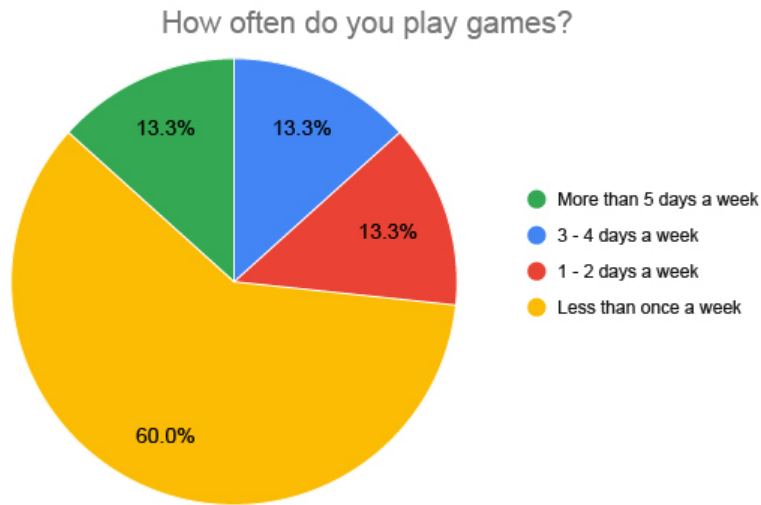


Figure 4.18 : Gaming frequency of the participants.

As seen in Figure 4.18, 60% of the participants reported playing games less than weekly while the remaining 40% declared playing games at least once a week. These groups were classified as non-gamers and gamers respectively.

Have you ever experienced virtual reality?

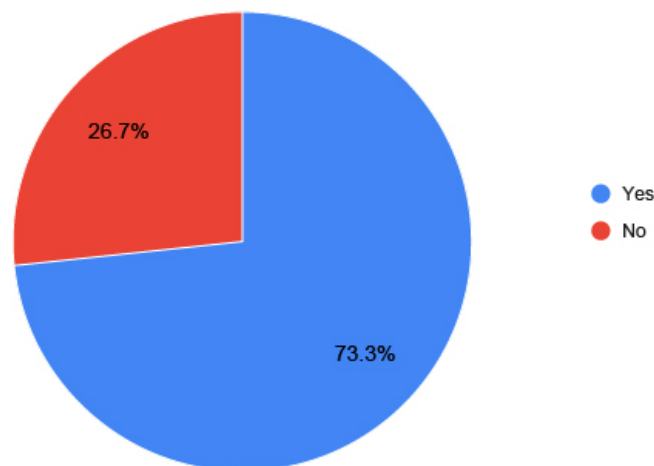


Figure 4.19 : VR experience distribution of the participants.

As shown in Figure 4.19, 73% of the participants have experienced virtual reality before whereas 27% of the participants experienced virtual reality for the first time during the experiment.

Please check all gesture controlled technologies you have experienced.

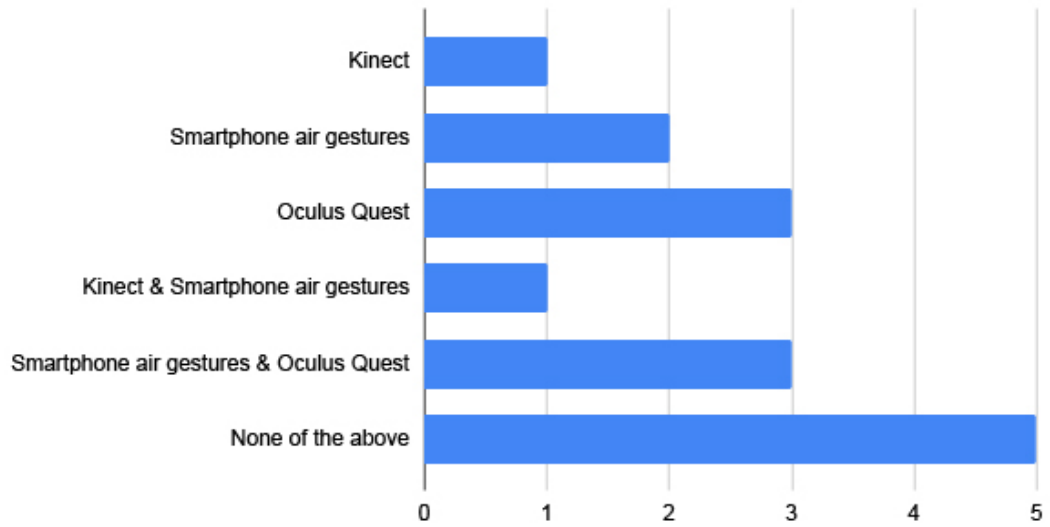


Figure 4.20 : Gesture technology experience distribution of the participants.

As seen in Figure 4.20, 33% of the participants haven't tried any gesture-controlled technology before whereas the remaining 67% of the participants have experienced some form of gesture-controlled technology such as Kinect, smartphone air gestures, and Oculus Quest.

Have you ever tried a hand gesture controlled experience in virtual reality?

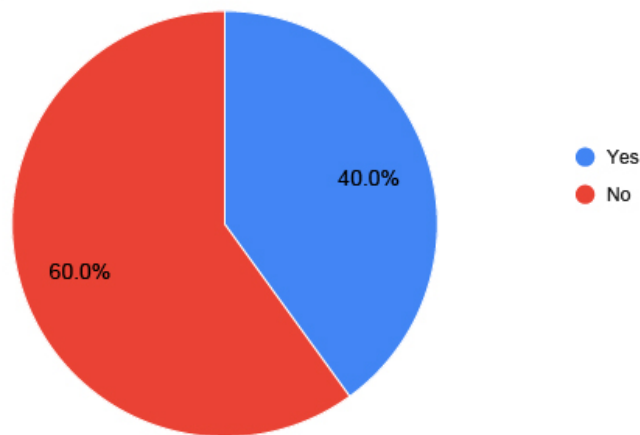


Figure 4.21 : Gesture-controlled VR experience distribution of the participants.

As shown in Figure 4.21, 40% of the participants have experienced a hand gesture-controlled virtual reality application before whereas 60% of the participants

experienced hand gesture interaction in virtual reality for the first time during the experiment.

How often did you experience virtual reality in the past 12 months?

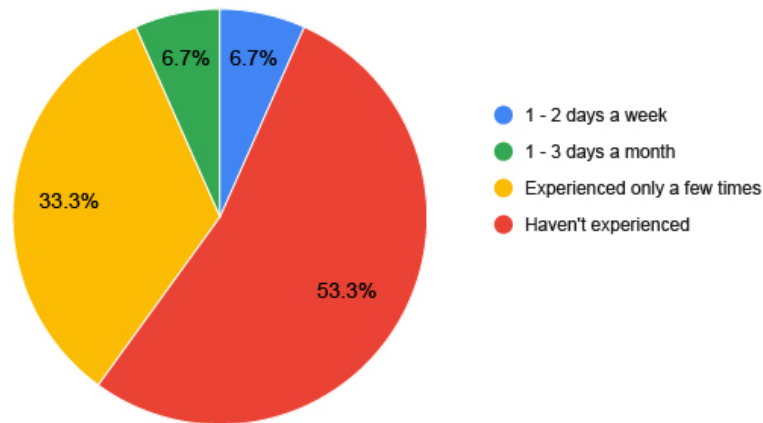


Figure 4.22 : VR experience frequency of the participants.

Even though 73% of the participants reported experiencing virtual reality before, as seen in Figure 4.22, 53% of the participants didn't experience VR in the past 12 months. The remaining 47% of the participants, however, experienced VR at least a few times in the past 12 months.

4.8.1.3 Participant attitudes towards predefined gestures

To collect participants' attitudes towards predefined gestures, semantic differential scale question type was used. To visualize gathered data, diverging stacked bar chart and semantic differential chart methods were utilized and overlapped on top of each other. Figure 4.23 shows the distribution of participants' attitudes towards the Punch gesture and the average ratings. The gesture was explored in two parts. The first part discussed the gesture itself regardless of its scenario whereas the second part discussed the gesture interaction.

The stacks on the chart represent the frequencies of each reply. The vertical line in the middle of the polar adjectives represents the baseline and the negative responses were stacked to the left of this baseline while the positive responses were stacked to the right of this baseline. Neutral responses were split evenly by the baseline. Thus, half of the neutral responses were on the negative side, while the other half were on the positive side. The pink and green diamonds on each bar represent the average ratings.

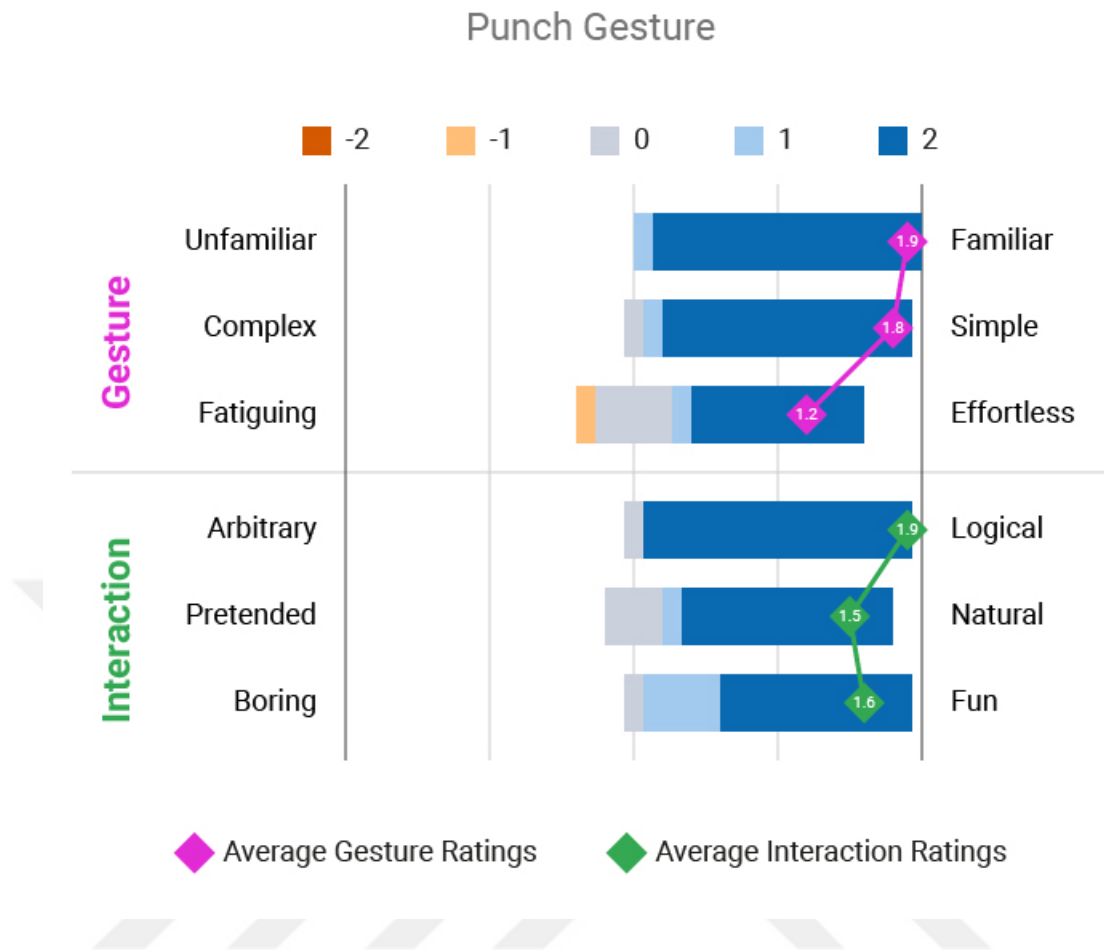


Figure 4.23 : Attitude distribution of Punch gesture.

As seen in Figure 4.23, Punch gesture obtained overwhelmingly positive ratings. The percentages of positive ratings (the sum of light and dark blue stacks) were: 100% for familiarity, 93% for simplicity, 67% for effortlessness, 93% for being logical for its intended use, 80% for naturalness, and 93% for fun factor. Punch gesture acquired the highest average scores in familiarity and logic items. Many participants remarked this gesture is more logical than the gesture they came up with. In contrast, it acquired the lowest average score in effortlessness. This result was expected as the Punch gesture required the highest amount of upper limb flexion amongst the predefined gestures. The greater the distance between the limbs and the body centerline, the greater the physical effort required to maintain the position of the limbs.

Middle Finger Flick Gesture

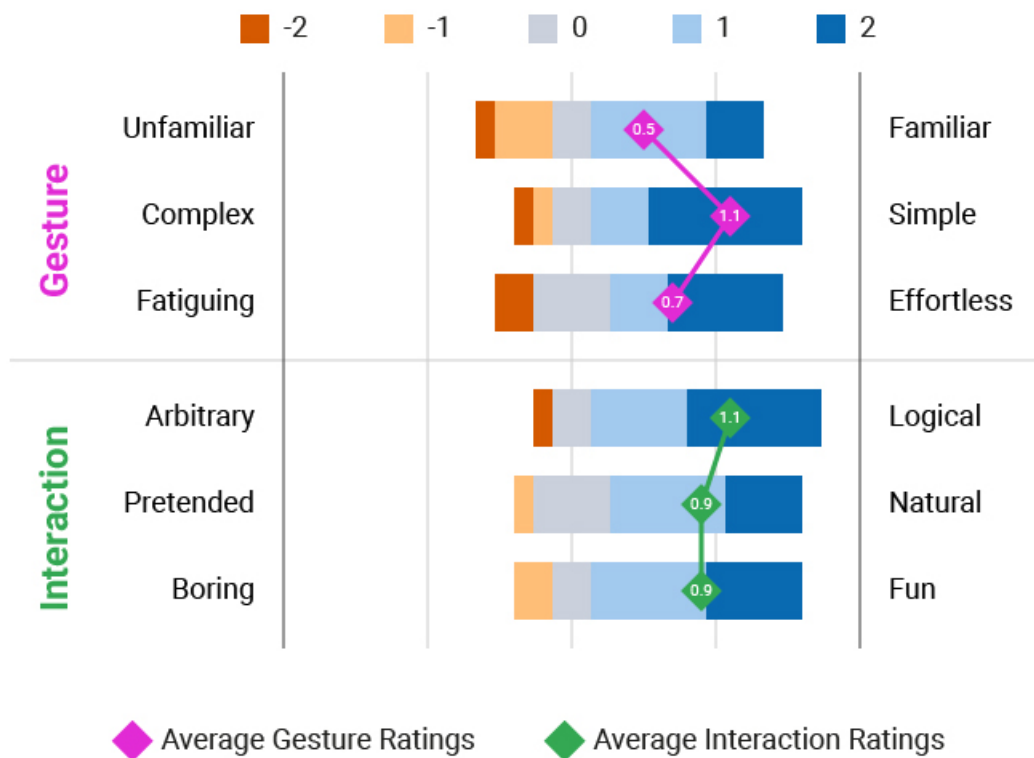


Figure 4.24 : Attitude distribution of Middle Finger Flick gesture.

As shown in Figure 4.24, Middle Finger Flick gesture acquired mostly positive ratings along with a few neutral and negative ratings. The percentages of positive ratings (the sum of light and dark blue stacks) were: 60% for familiarity, 73% for simplicity, 60% for effortlessness, 80% for being logical for its intended use, 67% for naturalness, and 73% for fun factor. Middle Finger Flick gesture acquired the lowest average score in familiarity. This gesture is proven to be the least familiar gesture in the predefined gesture set. However, the same gesture acquired the highest average scores in simplicity and logic items. Participant 3 quoted, “Even though I am not familiar with this gesture, it is very cleverly designed. When I first learned this gesture, I thought it didn't make much sense for its intended use. However, when I tried the gesture, something clicked in my mind, and it suddenly made perfect sense. I also really enjoyed performing it.”

Four Finger Flick Gesture

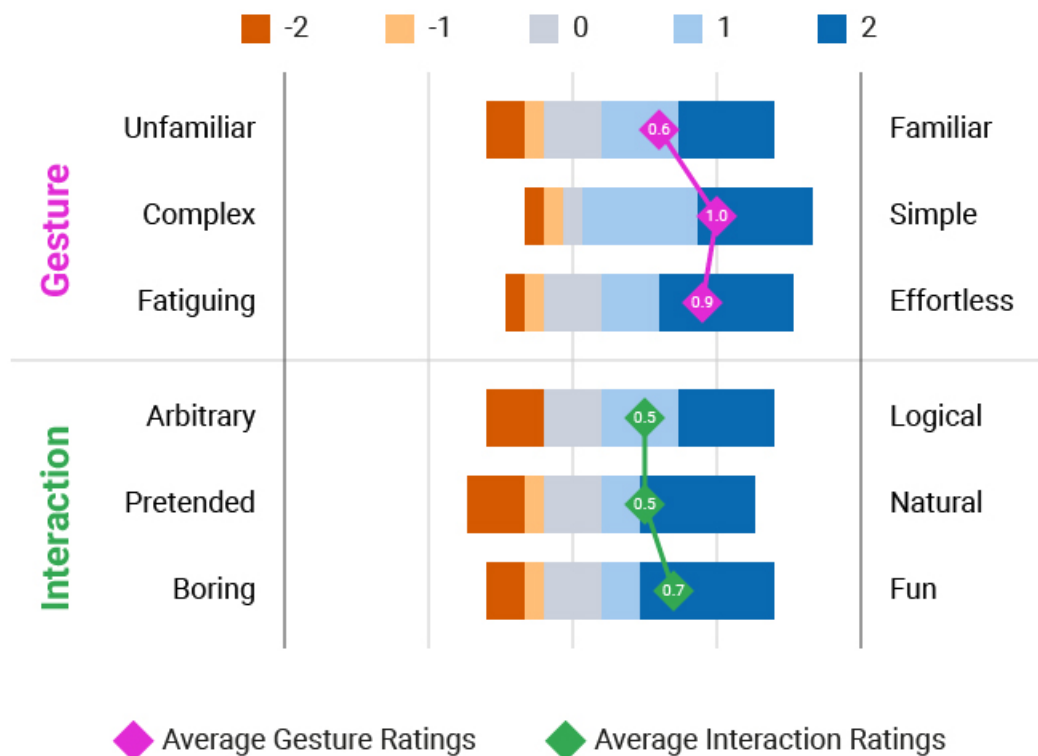


Figure 4.25 : Attitude distribution of Four Finger Flick gesture.

As seen in Figure 4.25, Four Finger Flick gesture acquired mostly positive ratings besides a few neutral and negative ratings. The percentages of positive ratings (the sum of light and dark blue stacks) were: 60% for familiarity, 80% for simplicity, 67% for effortlessness, 60% for being logical for its intended use, 53% for naturalness, and 60% for fun factor. Four Finger Flick gesture acquired the highest average score in simplicity and lowest average scores in naturalness and logic items. Participant 14 stated, “This gesture (Four Finger Flick gesture, which was explained to imitate flicking water) didn't feel very natural to me. The gesture would have felt more natural if there were a visual element (water) that came out of my hand and moved towards the object as you explained.” Even though having all around positive results, this gesture acquired the lowest ratings amongst the predefined gesture set.

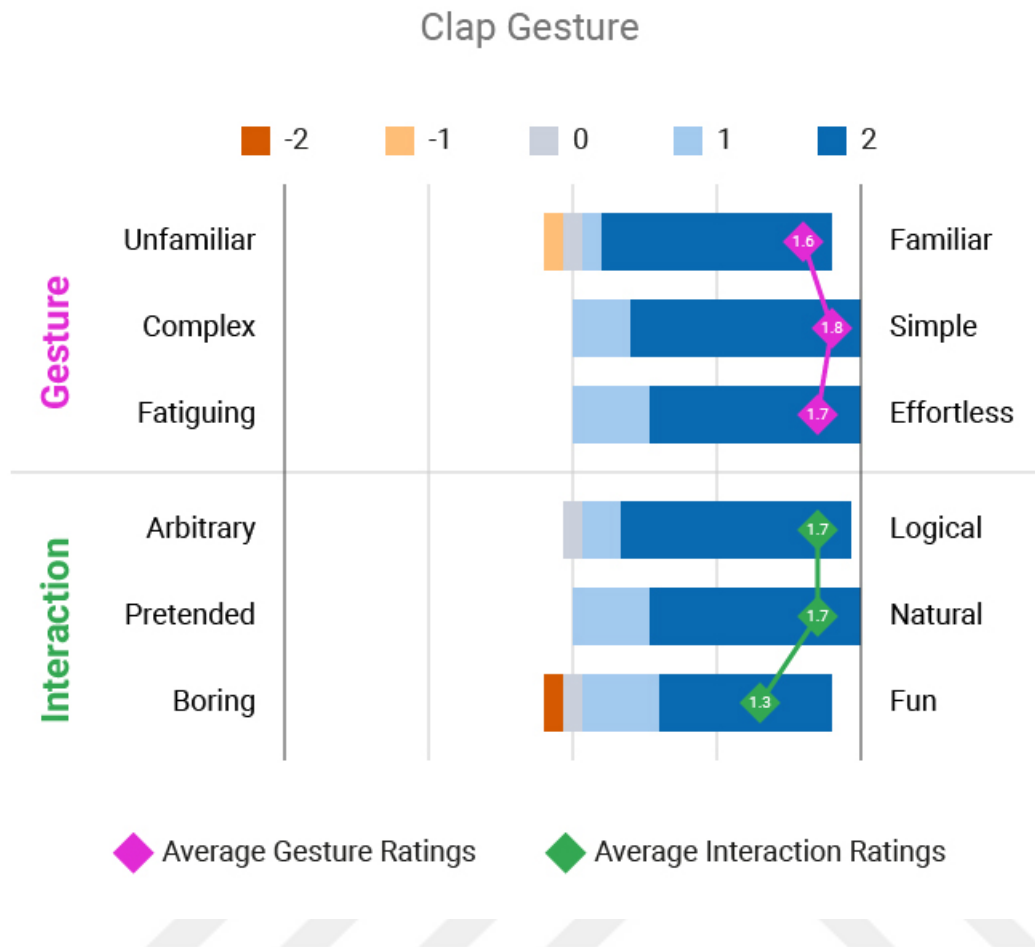


Figure 4.26 : Attitude distribution of Clap gesture.

As shown in Figure 4.26, Clap gesture acquired overwhelmingly positive ratings as well. The percentages of positive ratings (the sum of light and dark blue stacks) were: 87% for familiarity, 100% for simplicity, 100% for effortlessness, 93% for being logical for its intended use, 100% for naturalness, and 87% for fun factor. Clap gesture acquired the highest average score in simplicity. Even though this is a bimanual gesture, the perceived level of simplicity was considerably high. This finding suggested bimanual gestures could be perceived as highly simple when they utilized simple hand poses and motion paths. In contrast, Clap gesture acquired the lowest average score in fun factor. Participant 13 stated, “When my hands touched each other, they disappeared. This reaction made the gesture less enjoyable for me.” The disappearance of hands was due to the technical limitations of the hardware and can be prevented with technological advancements in the future.

Pinch Gesture

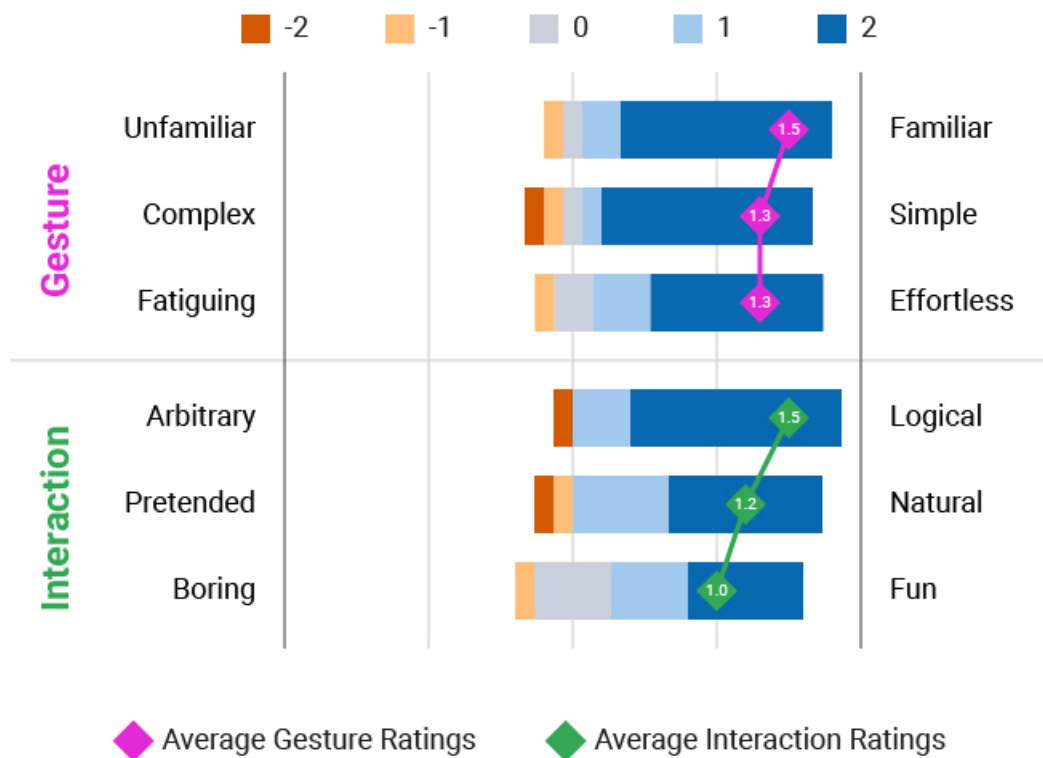


Figure 4.27 : Attitude distribution of Pinch gesture.

As seen in Figure 4.27, Pinch gesture acquired highly positive ratings. The percentages of positive ratings (the sum of light and dark blue stacks) were: 87% for familiarity, 80% for simplicity, 80% for effortlessness, 93% for being logical for its intended use, 87% for naturalness, and 67% for fun factor. Pinch gesture acquired the highest average scores in familiarity and logic items. Participant 15 quoted, “When I was in kindergarten, we used to perform the same gesture in the star collecting game, and each finger would perceive the other as the star.” In contrast, Pinch gesture acquired the lowest average score in fun factor.

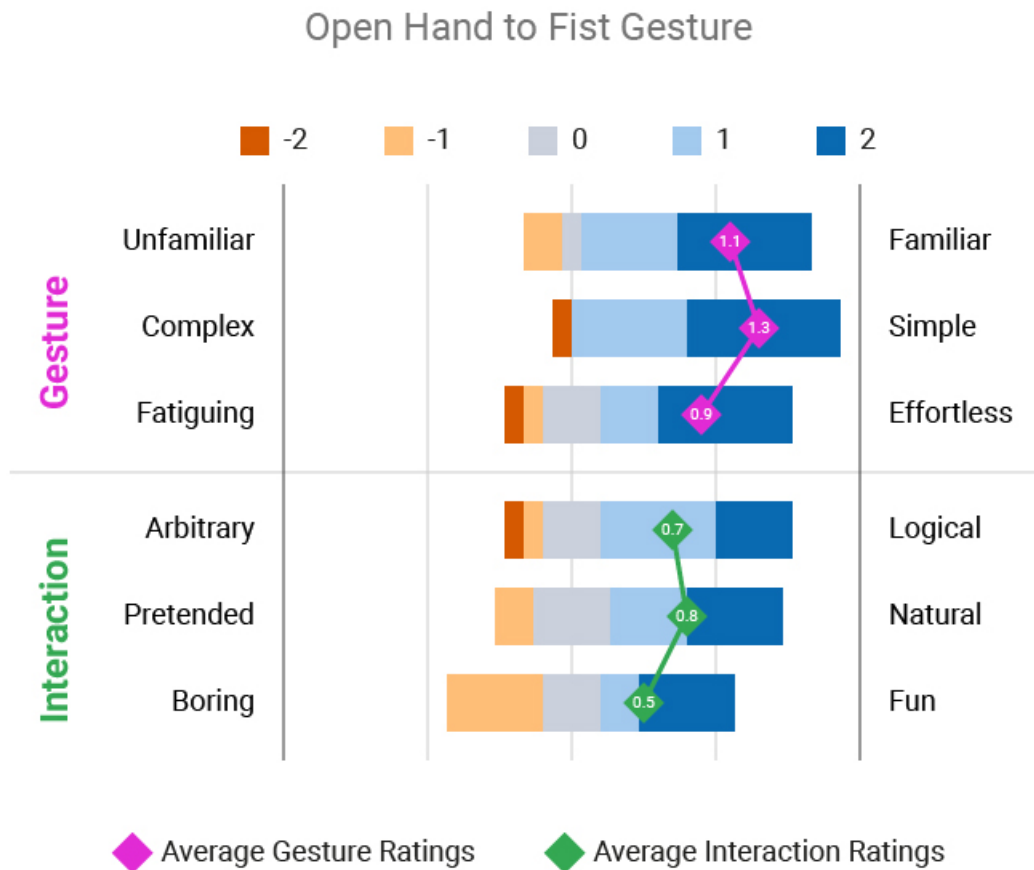


Figure 4.28 : Attitude distribution of Open Hand to Fist gesture.

As shown in Figure 4.28, Open Hand to Fist gesture acquired mostly positive ratings with a few neutral and negative ratings. The percentages of positive ratings (the sum of light and dark blue stacks) were: 80% for familiarity, 93% for simplicity, 67% for effortlessness, 67% for being logical for its intended use, 60% for naturalness, and 47% for fun factor. Open Hand to Fist gesture acquired the highest average score in simplicity and the lowest average score in fun factor. While performing this gesture, most participants held their hands on head level instead of the proposed chest level. This hand position, coupled with the variations observed in thumb placement, reduced the recognition success of the gesture, which frustrated the participants. The relatively low fun ratings were assumed to be due to this frustration.

4.8.2 User-elicited gestures and gesture classifications

Through the first trial of the VR experiment, 89 intuitively generated gestures were collected from 15 participants across six scenarios. Some of the user-elicited gestures that were different from the author-designed gestures were shown in Figure A.14.1.

Participant 8 couldn't generate a gesture for scenario 6 in the allocated 5 seconds. All gestures were coded and classified into five dimensions: number of hands, form, nature, hand-object intersections, and repetition per participant.

4.8.2.1 Gesture classifications used on elicited gestures

- Number of hands: This class includes two categories unimanual and bimanual. Gestures utilizing one hand are classified as unimanual whereas gestures utilizing both hands are classified as bimanual.
- Form: This classification, detailed in Section 2.1.2, consists of six categories in the original study (Wobbrock et al., 2009). 'Dynamic pose', 'static pose and path', and 'dynamic pose and path' categories of form classification are adopted for this thesis study. In contrast, as 'one point touch' and 'one point path' are associated with mouse click and, therefore, used to select and drag components, they weren't observed in our user-elicitation study due to the context of the experiment. Additionally, 'static pose' wasn't observed either as expected due to the manipulative nature of the scenarios. Hence, 'static pose', 'one point touch', and 'one point path' categories were omitted from form class in this thesis study.
- Nature: This thesis study classifies gestures based on their nature as well. This classification consists of five dimensions: direct manipulation, pantomimic, iconic, semaphoric, and affect display.
 - Direct manipulation: Direct manipulation occurs when users maneuver the virtual object by placing their hand on it. Thus, it requires reaching out to the object until an intersection takes place. This category is adopted from Arora et al. (2019).
 - Pantomimic: Pantomimic gestures mimic the way users interact with an object (Aigner et al., 2012). The difference between direct manipulation and pantomimic gestures is that users performing pantomimic gestures do not intend to control the virtual object by contact.
 - Iconic: Imitating the form or the motion path of an entity is classified as iconic (Aigner et al., 2012).

- Semaphoric: This category is adopted from Quek et al. (2002), who defines semaphoric gestures as a learned set of static or dynamic gestures that use arm and hand.
- Affect Display: Gestures that convey an emotion or intention are classified as affect display (Ottenheimer, 2005).
- Hand-object intersections: As detailed in Section 3.3.1.2, this novel classification categorizes gestures by the hand parts intersecting with the virtual object and consists of palmar, phalangeal, distal, and no intersection. Palmar intersection occurs when the palm of the hand is used to contact the object. Phalangeal intersection occurs when only the fingers are used to contact the object. Distal intersection occurs when only the fingertips are used to contact the object. If an intersection does not take place, it is classified as no intersection.
- Repetition per participant: As detailed in Section 3.3.1.2, this novel classification investigates reused gestures within each participant and consists of identical, oriented, and unique categories.

In this thesis, the identical category is further explored for each scenario to produce more comprehensive results. Therefore, terms as 1st Gest, 2nd Gest, 3rd Gest, 4th Gest, and 5th Gest are used next to identical term in the figures of Section 4.8.2.2. 1st Gest means the gesture in question is identical to the gesture used in the first scenario, 2nd Gest means the gesture in question is identical to the gesture used in the second scenario, and so on.

Oriented category is further explored just like the identical category. It is formatted as Nth Gest Rn°. N is the number of the scenario, R represents rotation, and n is the degree of the rotation around the viewing axis. For example, 4th Gest R90° means the gesture in question is the same as the gesture in the 4th scenario, except the gesture in question is rotated 90° around the viewing axis.

4.8.2.2 Classification of user-elicited gestures

Figure 4.29 shows the taxonomy breakdown of user-elicited gestures in scenario 1. The vertical axis of the stacked column chart shows the number of participants, while

the horizontal axis shows the different taxonomies. The first column (number of hands) includes unimanual and bimanual categories. The second column (form) includes dynamic pose, static pose & path, dynamic pose & path. The third column (nature) includes direct manipulation, pantomimic, iconic, semaphoric, and affect display. The fourth column (hand-object intersections) includes palmar, phalangeal, distal, and no intersection. The fifth column (repetition per participant) includes unique, oriented, and identical. Unobserved categories are omitted from the charts for each scenario. Nature and hand-object intersections columns are placed side-by-side as they are highly associated. Numbers at the bottom of each stack show the number of participants in that stack. Figure 3.15 can be referenced for the configuration of each scenario while reviewing the figures in this section.

Taxonomy Breakdown of User-Elicited Gestures in Scenario 1

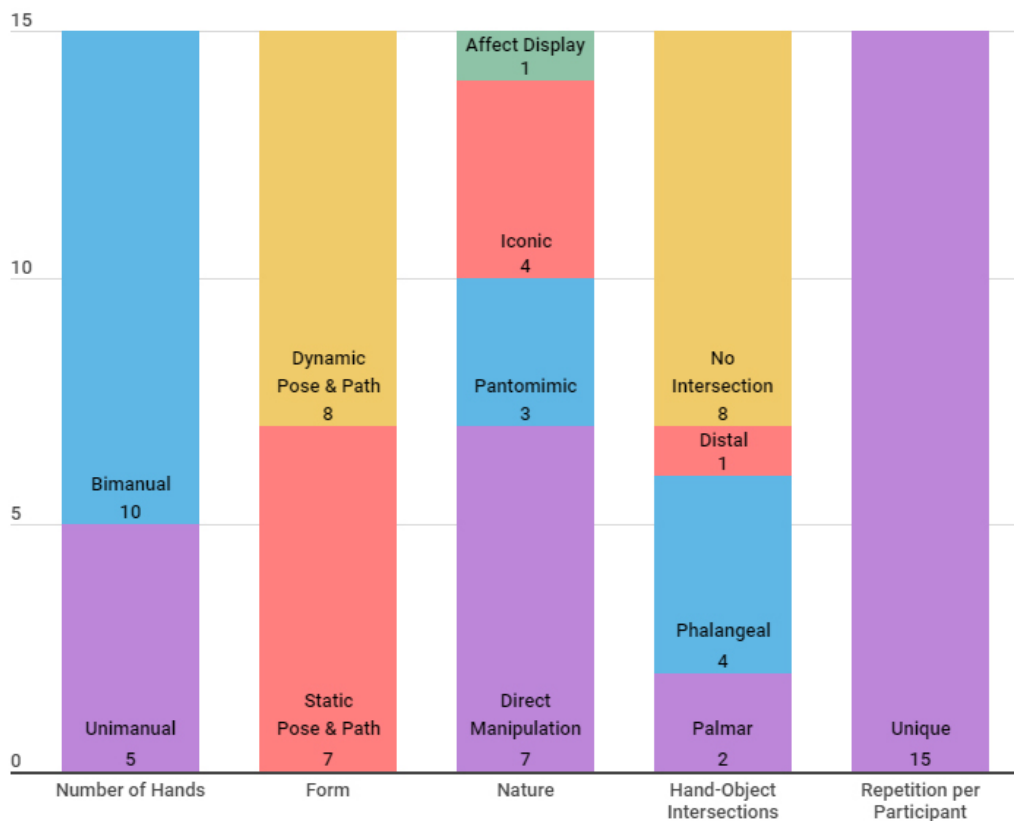


Figure 4.29 : Number of gestures in each taxonomy category for scenario 1.

In the 1st scenario, where a relatively big, near-field object exploded, 67% of the participants used bimanual gestures, while the remaining 33% used unimanual gestures, as shown in Figure 4.29. The forms of the gestures were split almost evenly

between dynamic pose & path (53%) and static pose & path (47%). In the nature column, the majority of the participants used direct manipulation gestures (47%). The most common hand-object intersection amongst direct manipulation gestures was phalangeal intersection. As each participant gestured for the first time in the 1st scenario, all of the gestures were unique within participants.

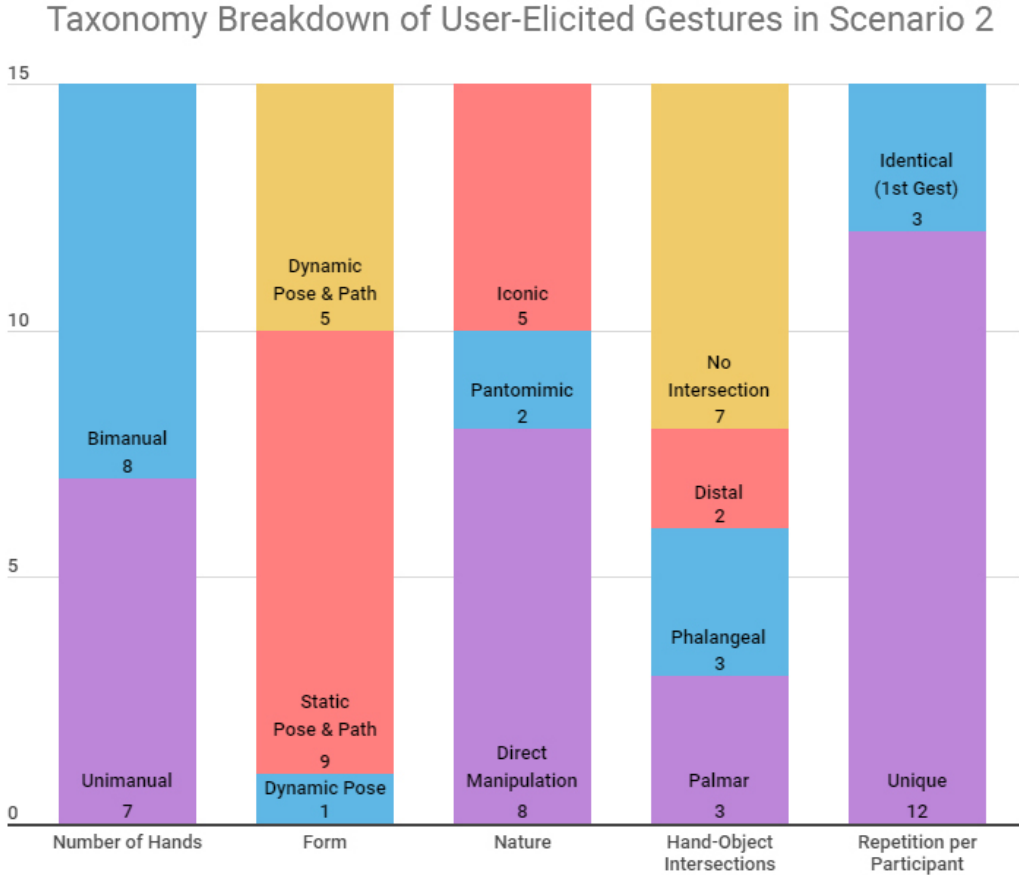


Figure 4.30 : Number of gestures in each taxonomy category for scenario 2.

In the 2nd scenario, where the object in the 1st scenario was smaller, unimanual gestures rose from 33% to 47%, yet bimanual gestures were still more frequent with 53%. The majority of the gestures’ form was static pose & path (60%), as shown in Figure 4.30. The nature of the gestures was mostly direct manipulation (53%). Hand-object intersections were split almost evenly and didn’t produce notable results. As seen in repetition per participant column, only 3 people out of 15 used a gesture that was identical to their first one. Hence, 80% of the participants gestured differently based on the exploding object’s smaller size.

Taxonomy Breakdown of User-Elicited Gestures in Scenario 3

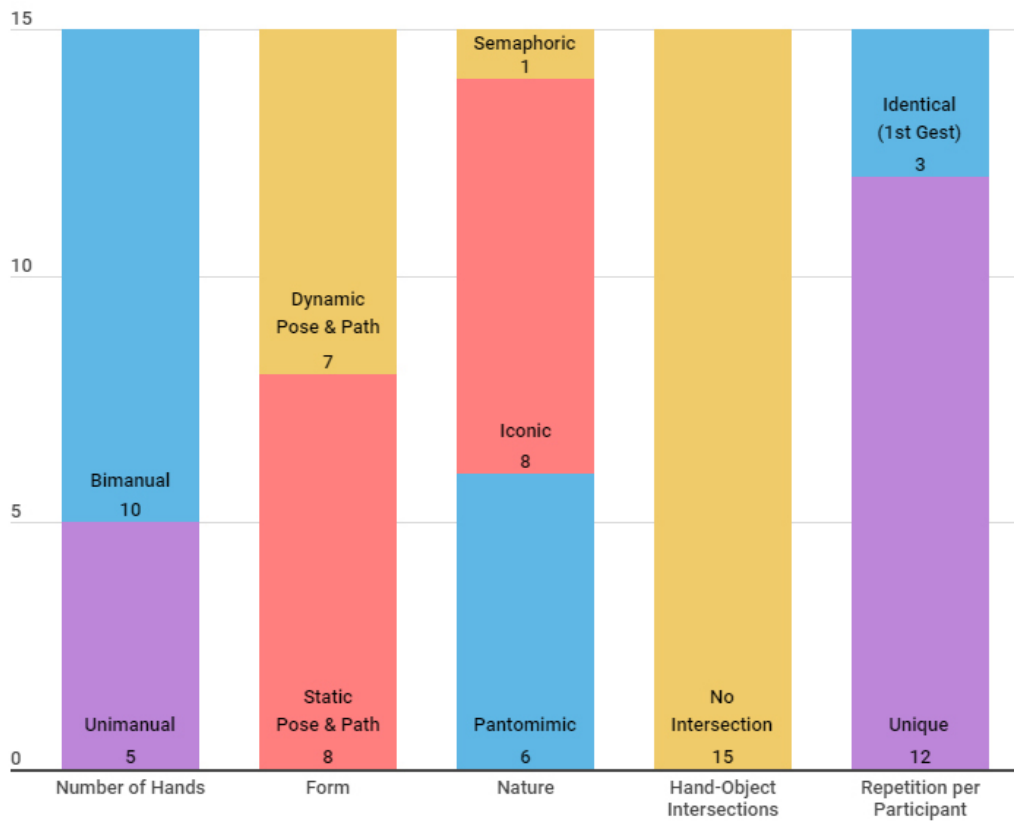


Figure 4.31 : Number of gestures in each taxonomy category for scenario 3.

In the 3rd scenario, where all variables were the same as the 1st scenario except the object was further away, 67% of the participants used bimanual gestures, as shown in Figure 4.31. Participant 9 stated, “Since the object is far away, I thought I should provide more power by using both hands.” after producing a bimanual gesture as opposed to their unimanual gestures in the 1st and 2nd scenarios, where the same object was within arm’s reach. In the form column, 53% of the gestures were static pose & path. Most of the gestures’ nature was iconic (53%). As the virtual object was beyond arm’s reach in this scenario, direct manipulation wasn’t possible. Hence, all 15 gestures were categorized as no intersection in the hand-object intersections dimension. As shown in repetition per participant column, there were only 3 people out of 15 who used a gesture that was identical to their first gesture. Thus, 80% of the participants gestured differently based on the exploding object’s increased distance.

Taxonomy Breakdown of User-Elicited Gestures in Scenario 4

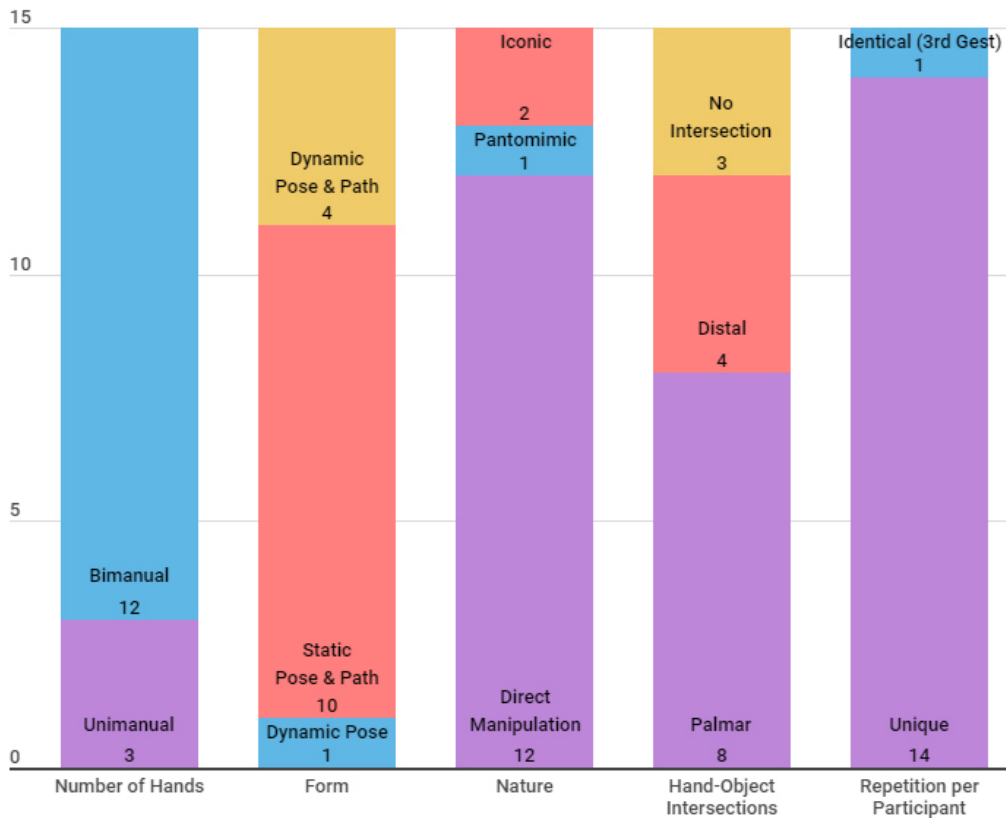


Figure 4.32 : Number of gestures in each taxonomy category for scenario 4.

The 4th scenario introduced a new object and a new animation; an imploding double conic object. In the number of hands column, 80% of the participants used bimanual gestures, as seen in Figure 4.32. In the form column, 67% of the gestures were static pose & path. The majority of the gestures' nature was direct manipulation (80%). Palmar intersection was the most observed hand-object intersection with 53%. As this scenario had a different object and animation than the ones before, almost everyone used a unique gesture as shown in repetition per participant column. Participant 9, who used an identical gesture (his gesture in the 3rd scenario), explained that he misread the animation as if the object was moving away from him rather than scaling down. Therefore, he commented he pushed the virtual object away.

Taxonomy Breakdown of User-Elicited Gestures in Scenario 5

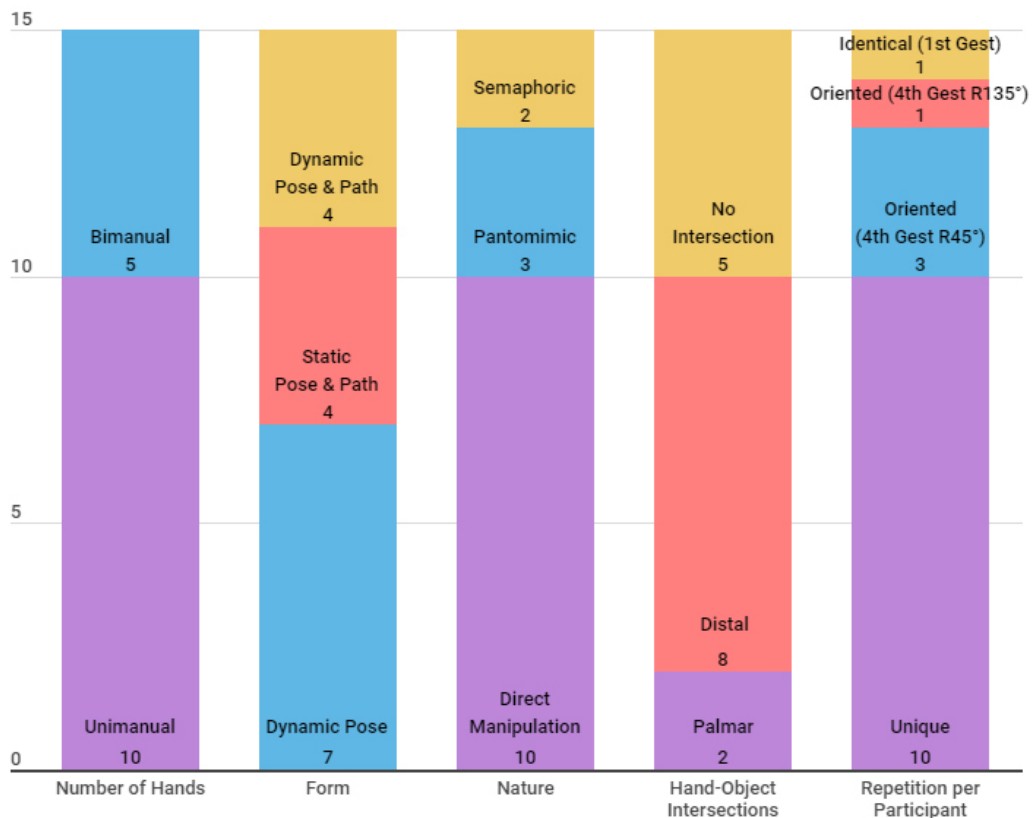


Figure 4.33 : Number of gestures in each taxonomy category for scenario 5.

In the 5th scenario, where the object in the 4th scenario was smaller and rotated 45° around the viewing axis, 67% of the participants used bimanual gestures, as seen in Figure 4.33. In this scenario, most of the gestures' form was dynamic pose (47%). The majority of the gestures' nature was direct manipulation (67%). Distal intersection was the most observed hand-object intersection with 53%. As seen in repetition per participant column, 1 person used an identical gesture, while 4 people used oriented gestures. Participant 9, who misread the animation again, used another pushing gesture, which was identical to his gesture in the 1st scenario. Amongst 4 people who used oriented gestures, 3 of them rotated their 4th gesture 45° around the viewing axis, while 1 person rotated his 4th gesture 135° around the viewing axis. These 4 people, who used oriented gestures, oriented their previously used gestures to match the new orientation of the object. This behavior suggested that the participants were highly aware of the object's orientation and gestured accordingly. It also implied that they would have gestured identically had the object not been rotated. Thus, 67% of the participants used a unique gesture according to the imploding object's smaller size.

Taxonomy Breakdown of User-Elicited Gestures in Scenario 6

15

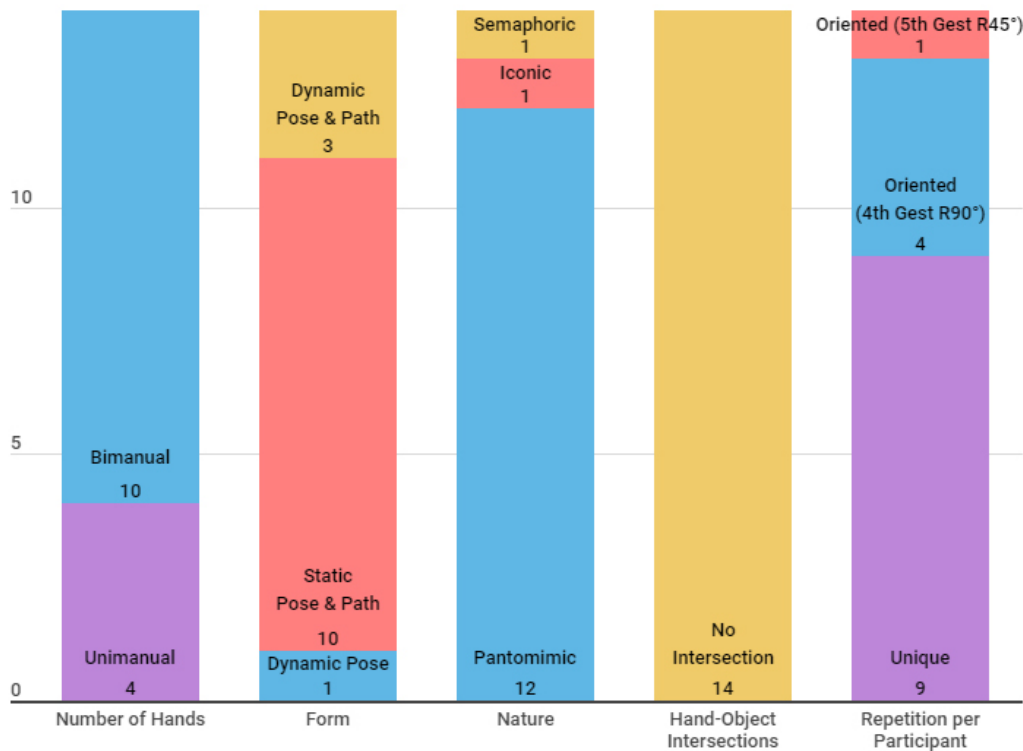


Figure 4.34 : Number of gestures in each taxonomy category for scenario 6.

In the 6th scenario, where the object in the 4th scenario was further away and rotated 90° around the viewing axis, 1 person couldn't gesture within the allocated 5 seconds. Hence, their data was omitted from this scenario, as seen in Figure 4.34. In the number of hands column, 71% of the participants used bimanual gestures. In the form column, 71% of the gestures were static pose & path. 86% of the gestures' nature was pantomimic. As the object was beyond arm's reach, direct manipulation was not possible. Therefore, all 14 gestures were categorized as no intersection in the hand-object intersections dimension. In repetition per participant column, 5 people used oriented gestures. 4 of these people rotated their 4th gesture 90° around the viewing axis, and 1 person rotated his 5th gesture 45° around the viewing axis. All of these gestures were oriented to match the object's new orientation. Thus, 64% of the participants used a unique gesture according to imploding objects' increased distance.

5. DISCUSSIONS

5.1 Gamers Versus Non-Gamers

40% of the participants (gamers) disclosed playing games at least once a week, whereas the remaining 60% (non-gamers) declared playing games less than weekly, shown in Figure 4.18.

Gamers tended to find gestures that produced physics-defying results (Four Finger Flick and Open Hand to Fist gestures) to be more logical than non-gamers. None of the gamers gave negative scores to these gestures for being logical for their intended use. Participant 13 stated, “It (Four Finger Flick gesture) seemed familiar and logical to me as this is a gesture performed by wizards in video games.”

Furthermore, gamers incorporated imagined elements such as magical orbs and bullets into their gestures. In the 3rd scenario, where the object exploded afar, two gamers (Participants 1 and 10) mimicked the shape of a gun with their hands and flicked their hands vertically as if they were shooting the object with a bullet. Participant 4 (another gamer) pushed her hands forward and quoted, “I imagined a magic power coming out of my hands and going towards the object.” In contrast, Participant 14 (a non-gamer) said, “This gesture (Four Finger Flick gesture, which was explained to imitate flicking water) didn't feel very natural to me. The gesture would have felt more natural if there were a visual element (water) that came out of my hand and moved towards the object as you explained.” While gamers naturally incorporated imagined elements into their gestures, the same approach felt unnatural to non-gamers.

5.2 Discussions on Gesture Classifications

During the experiment, all of the participants performed pantomimic gestures towards the virtual objects as a way of ‘immediate spatial reference’ (Quek et al., 2002). Moreover, some participants even reached for the virtual objects as far as they could in far-field scenarios. Participant 11 stated, “Even though I knew I couldn't touch the

object, I thought I had to reach out to interact with it.” Therefore, this behavior was assumed to be an extension of spatial referencing.

In their study, Aigner et al. (2012), who proposed pantomimic category, prompted a person to arrange some objects by observing the gestures of a participant, who could only see these objects from a 2D screen. Aigner et al. (2012) stated that participants generating pantomimic gestures imagined these objects as the actual objects were not present within the participant’s vicinity. In this thesis study, however, participants were gazing and acting directly towards the virtual objects rather than completely imagining the objects. Therefore, within the scope of this study, the pantomimic category slightly diverged from its precedents.

Participant 2, who has never experienced VR before, remarked feeling startled upon viewing the exploding object in scenario 1. For this scenario, Participant 2 proposed the hand gesture (fingers splayed, palms facing forward, pulling both hands back to either side of the face) that they performed when startled in the real world. It’s worth noting that Participant 2, who let out a barely audible scream right after observing the scenario, didn’t execute the gesture alongside the scream. Participant 2 proposed this gesture after a while, within 5 seconds. Therefore, the gesture was executed deliberately rather than unconsciously. Although Participant 2 did not make a statement beyond "it just felt right" while justifying the gesture, it was speculated that Participant 2 might have expected the object to explode in a cartoonish fashion when startled. Therefore, Participant 2 might have proposed this gesture to direct the object into reacting as startled. This gesture is an affect display gesture (Ottenheimer, 2005), which is a classification used in fields such as anthropology, linguistics, and cognitive science. None of the HCI-focused studies we reviewed investigated this classification.

5.3 The Similarities Between User-Elicited Gestures

Do different users generate similar types of gestures for the same scenarios (RQ1)?

There were many similarities between participants within each scenario. Most of the participants used bimanual gestures for the explode operation. The percentages of bimanual gestures were: 67% in the 1st scenario, 53% in the 2nd scenario, and 67% in the 3rd scenario. Additionally, in the first three scenarios, where the object explodes, except for one participant’s gesture in scenario 2, all of the elicited gestures’ forms

were either static pose & path or dynamic pose & path. Therefore, people showed a very strong tendency to trace a motion path when gesturing to explode an object. The majority of the participants used direct manipulation gestures to explode near-field objects (47% in the 1st scenario and 53% in the 2nd scenario) and iconic gestures to explode far-field objects (53% in the 3rd scenario).

The number of hands category showed strong similarities between participants for implode operation within each scenario as well. 80% of the participants in the 4th scenario and 67% of the participants in the 6th scenario used bimanual gestures, whereas 67% of the participants used unimanual gestures in the 5th scenario. The nature of gestures for the implode operation was very consistent between participants for each scenario too. The majority of the participants used direct manipulation gestures in near-field implosion scenarios (80% in the 4th scenario and 67% in the 5th scenario), whereas 80% of the participants used pantomimic gestures in the far-field implosion scenario (6th scenario).

5.4 The Effects of Objects' Spatial Properties on Elicited Gestures

Do spatial properties (size, position, orientation) of virtual objects affect the users' gestures (RQ2)?

Out of the 89 gestures collected during the experiment, 17 were recurring gestures. Eight of these gestures were identical, and the remaining nine were oriented. Therefore, the vast majority of gestures collected were unique, and virtual objects' spatial properties proved to affect users' gestures.

5.4.1 The effects of objects' size on elicited gestures

73% of the participants gestured differently according to the change in manipulated object's size. This percentage was calculated based on the total number of unique gestures in the 2nd and 5th scenarios, where the objects were smaller than the rest.

In the 2nd scenario, the number of unimanual gestures rose from 33% to 47% compared to the 1st scenario. Likewise, in the 5th scenario, it rose from 20% to 67% compared to the 4th scenario. These findings showed parallelism with the results of previous studies (Pham et al., 2018; Piumsomboon et al., 2013) and heavily suggested the size of the object affected the number of hands used for otherwise same scenarios.

Hand-object intersections class produced notable results based on the change in object size, too. In the 2nd scenario, Participant 1 said, “Since the object was smaller, I thought touching with the tip of my finger would be enough this time.” after producing a distal intersection compared to her phalangeal intersection in the 1st scenario. Between the 4th and 5th scenarios, palmar intersections decreased from 53% to 13%, whereas distal intersections increased from 27% to 53%. Consequently, implode operation produced significantly different results in hand-object intersections class based on the change in manipulated object’s size, and the most frequent intersections switched from palmar to distal.

5.4.2 The effects of objects’ distance on elicited gestures

66% of the participants gestured differently based on the change in manipulated object’s distance. This percentage was calculated based on the total number of unique gestures in the 3rd and 6th scenarios, where objects were further away than the rest.

Direct manipulation was the most common nature in all near-field scenarios. The percentages of direct manipulation gestures were 47% in the 1st scenario, 53% in the 2nd scenario, 80% in the 4th scenario, and 67% in the 5th scenario. Hence, most of the participants’ immediate response to near-field objects was to control them by contact. This result showed parallels with the nature of predefined gestures, all of which were direct manipulation in near-field scenarios. Unlike near-field scenarios, the most common natures in far-field scenarios differed from one another. In the 3rd scenario, most of the gestures were iconic (53%), whereas, in the 6th scenario, most of the gestures were pantomimic (80%). Even though the motion paths of the elicited gestures in the 3rd and 6th scenarios were very much alike and approximately the reverse of each other, the nature of the gestures diverged. While participants mostly used iconic gestures for remote explode operation, they mostly used pantomimic gestures for remote implode operation.

5.4.3 The effects of objects’ orientation on elicited gestures

79% of the participants gestured according to the manipulated object’s orientation. This percentage was calculated based on the total number of gestures that utilized an arm or forearm movement to orient itself to the object in the 5th and 6th scenarios compared to previous scenarios with the same object. The gestures of Participants 2, 6, and 9 were seemingly unrelated to the object’s orientation both in the 5th and 6th

scenarios. As seen in Figure 4.33, 27% of the gestures in the 5th scenario were classified as oriented. Likewise, as shown in Figure 4.34, 36% of the gestures in the 6th scenario were classified as oriented. 40% of the participants elicited Pinch gesture (predefined gesture) in the 5th scenario. Therefore, their gestures matched the orientation of the object as well. Although the remaining participants produced unique gestures, they still applied some form of arm/forearm movement that aligned their hands concerning the object's orientation.

5.5 The Impact of Kinesthetic Feedback on User Interaction Satisfaction

How does the presence of kinesthetic feedback impact user interaction satisfaction (RQ3)?

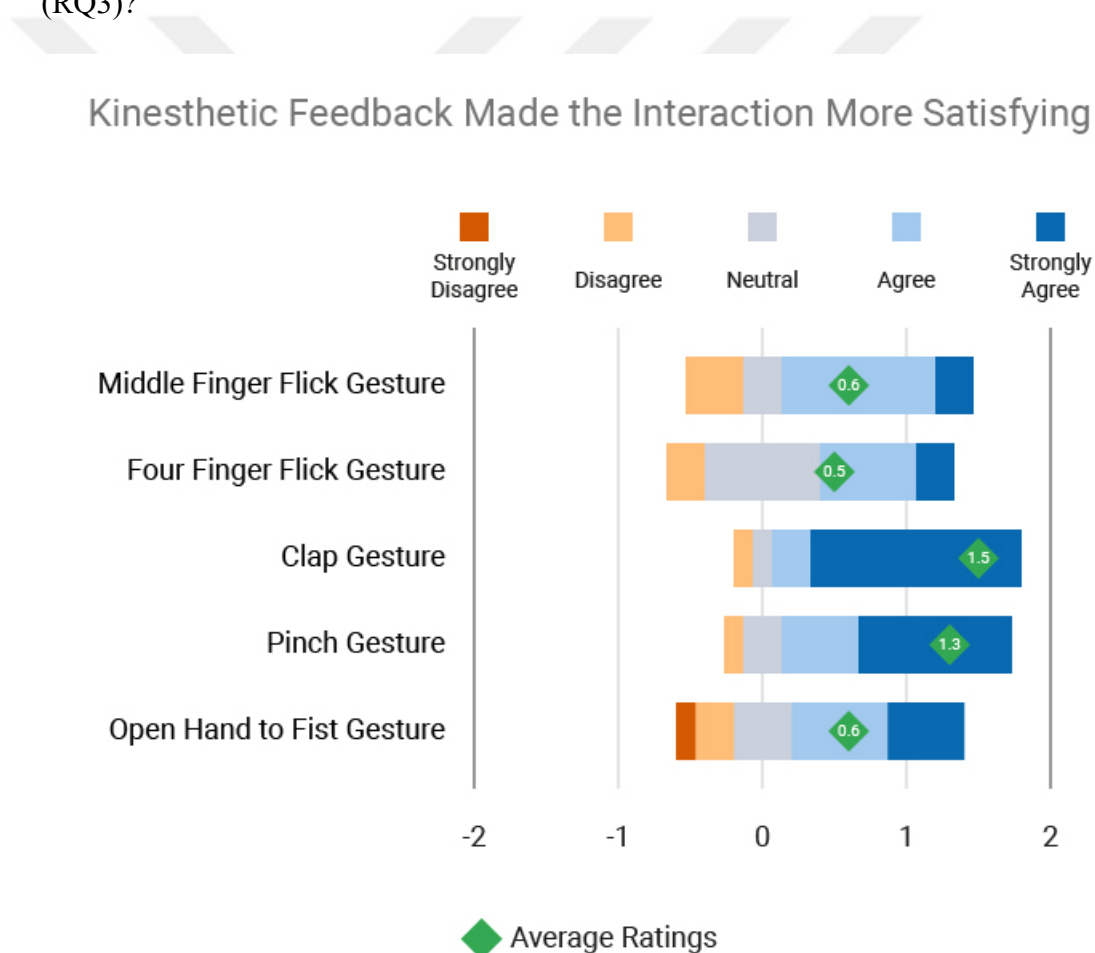


Figure 5.1 : Agreement distribution of interaction satisfaction increasing due to kinesthetic feedback.

As seen in Figure 5.1, all predefined gestures acquired highly positive ratings on the perception that kinesthetic feedback made interaction more satisfying except the

Punch gesture, which didn't produce kinesthetic feedback. The percentage of positive ratings (the sum of strongly agree and agree answers) were 67% for Middle Finger Flick gesture, 47% for Four Finger Flick gesture, 87% for Clap gesture, 80% for Pinch gesture, and 60% for Open Hand to Fist gesture. Nevertheless, the percentage of positive ratings were lowest in far-field scenarios (the 3rd and 6th scenarios) and decreased around 20% compared to near-field alternatives. Additionally, all predefined gestures in near-field scenarios were direct manipulation gestures, whereas the predefined gestures in far-field scenarios were pantomimic or iconic. These findings suggested that the satisfaction gained from the interaction due to kinesthetic feedback increased when the nature of the gesture was direct manipulation compared to other alternatives. Nevertheless, further testing is required for more definitive results as all predefined gestures were different from one another in this thesis study. Testing direct manipulation and pantomimic alternatives of the same gesture on an object positioned within and beyond arm's reach, respectively, would yield more definite results.

Moreover, gestures that produce an absolute down position acquired the highest satisfactory interaction ratings due to kinesthetic feedback. Clap gesture obtained 87% positive ratings, and Pinch gesture received 80% positive ratings. Another factor increasing Clap gesture's positive ratings was the presence of the sound. Participant 10 remarked, "The sound of clapping my hands together and the sense of rhythm created by the successive use of this gesture made the interaction more satisfying and immersive for me."

Even though Clap and Pinch gestures produced post-interaction kinesthetic feedback, they scored the highest ratings. This result suggested the participants weren't affected negatively by the absence of kinesthetic feedback at the beginning of the interaction. Therefore, we inferred that for kinesthetic feedback to occur at the end of the interaction was sufficient to make the interaction more satisfying.

5.6 User-Elicited Versus Author-Designed Gestures

Do users prefer user-elicited or author-designed gestures (RQ4)?

In contrast to findings on previous research in literature (Morris et al., 2010), our results showed that participants mostly preferred author-designed gestures over user-

elicited ones. The participants tended to elicit gestures that are physically and conceptually more complex than the gestures in the author-designed set. For example, the user-elicited gestures were more likely to use bimanual gestures compared to unimanual author-designed gestures for the explode operation. The preferred gestures of all the participants for each scenario can be seen in Figure A.15.1.

The distribution of participants' preferred gestures for each scenario is shown in Figure 5.2. The bars on the chart represent the frequencies of each reply. The red bars are the percentage of participants who preferred their own (user-elicited) gestures, the blue bars are the percentage of participants who preferred predefined (author-designed) gestures, and the gray bars are the percentage of participants who elicited predefined gestures. The vertical line in the middle of the figure represents the baseline. Participants who preferred gestures that were different from predefined gestures were stacked to the left of the baseline, whereas participants who preferred predefined gestures, whether they elicited or not, were stacked to the right of the baseline.

User-Elicited Versus Author-Designed Gestures

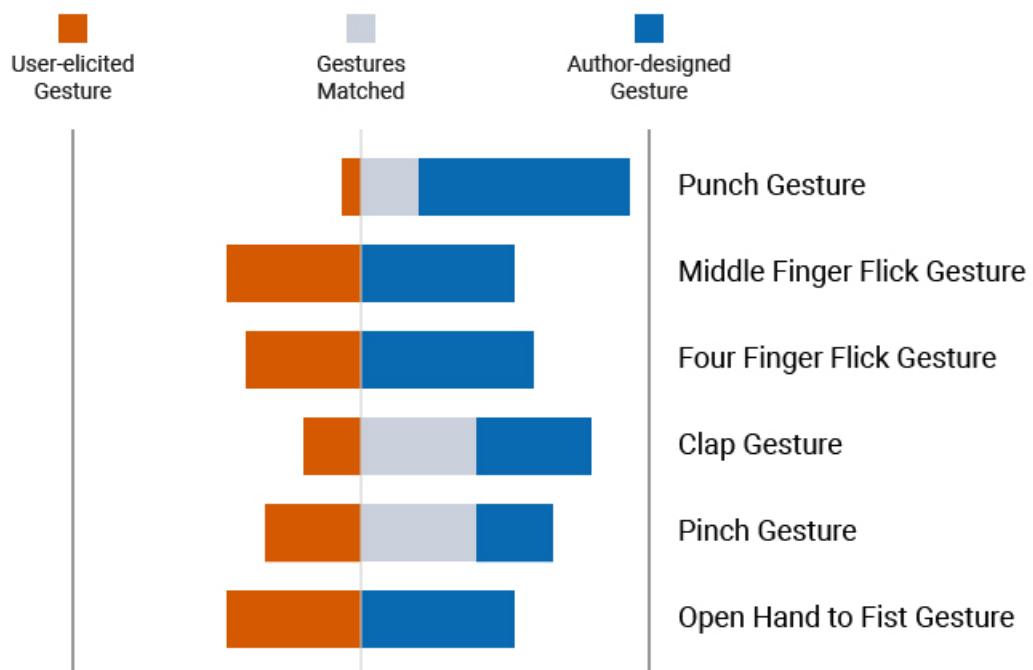


Figure 5.2 : Distribution of preferred gestures for each scenario.

As seen in Figure 5.2, the most intuitive gestures in the predefined gesture set were Clap and Pinch gestures. 40% of the participants elicited these gestures in the user-

elicitation study (the first trial of the experiment). These two gestures were followed by the Punch gesture, which 20% of the participants elicited. The remaining predefined gestures didn't appear in the user-elicitation study. Therefore, they could be considered unintuitive. However, the purpose of the predefined gestures was to be the most suitable gestures for their respective scenarios regarding multiple criteria except intuitiveness, which was neglected purposefully to allow creative freedom. In contrast, the scenarios were designed to make predefined gestures more discoverable. In that regard, there is room for improvement for Middle Finger Flick, Four Finger Flick, and Open Hand to Fist gestures.

As the purpose of the predefined gestures was to be most suitable, their success lies in the percentage of them being preferred for their respective scenarios. Percentage of total participants who either elicited or preferred predefined gestures were: 93% for Punch gesture, 53% for Middle Finger Flick gesture, 60% for Four Finger Flick gesture, 80% for Clap gesture, 67% for Pinch gesture, and 53% for Open Hand to Fist gesture. All predefined gestures produced positive results, and at least more than half of the participants either elicited or preferred predefined gestures over other user-elicited gestures.

5.7 The Qualities of a Preferable Gesture in HCI

What is the relationship between preferable gestures and user interaction qualities (logical, intuitive, natural, effortless, simple, fun, satisfactory) (RQ5)?

In the survey, participants are prompted to select up to three factors (user interaction qualities) out of seven that made them prefer one gesture over the other for each scenario. Figure 5.3 shows the selection frequency of factors that determined the preferred gestures for each scenario. The horizontal axis of the stacked column chart shows the different factors that affected preferred gestures. The vertical axis displays the number of times each user interaction quality is selected as a determining factor. The stacks in each column represent a different scenario. Numbers at the bottom of each stack show how many times the specified factor was selected. The frequency of selected factors decreases from left to right. The factors selected by each participant are shown in Figure A.15.1.

User Interaction Qualities That Determined the Preferred Gestures

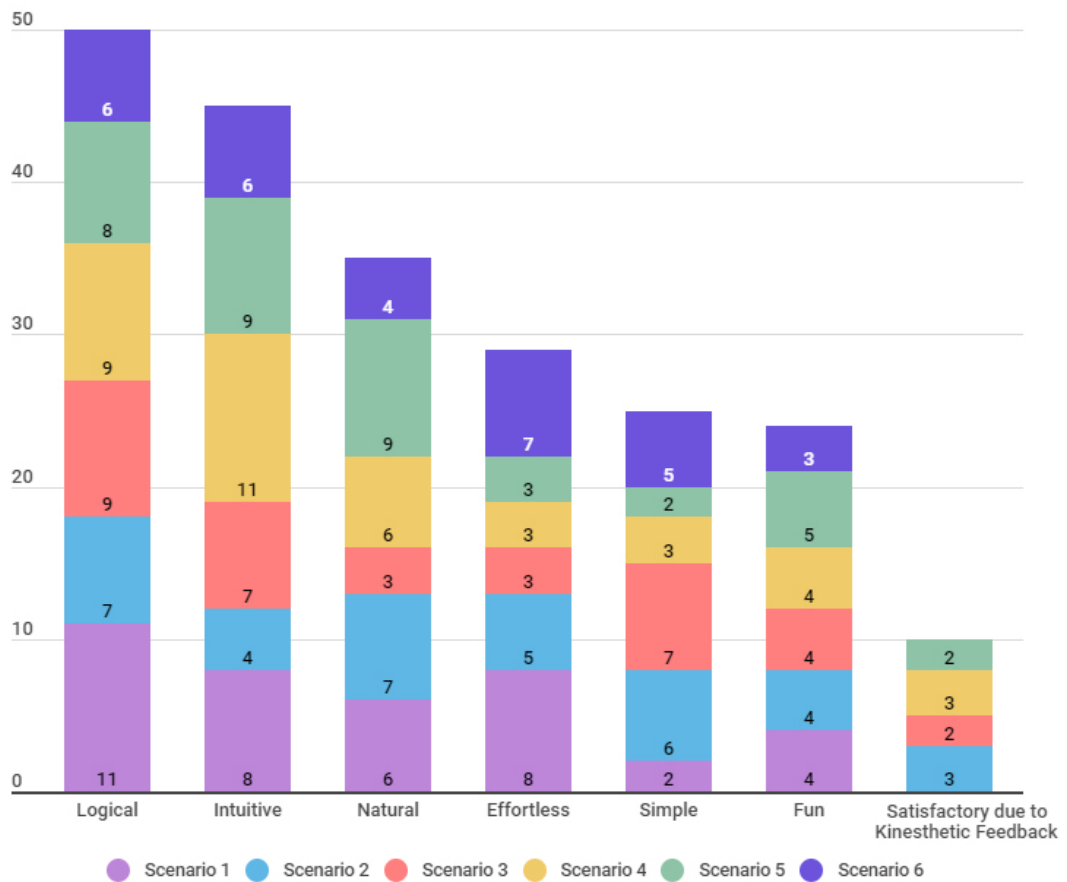


Figure 5.3 : Selection frequency of factors that determined the preferred gestures for each scenario.

The number of times each user interaction quality was selected as a determining factor were 50 times for logical, 45 times for intuitive, 35 times for natural, 29 times for effortless, 25 times for simple, 24 times for fun, and 10 times for satisfactory due to kinesthetic feedback. Therefore, the three most important factors across all scenarios were logical, intuitive, and natural, respectively. These results implied that logic and intuition often generated preferable gestures compared to other factors. Additionally, as shown in Figure 5.2, 93% of the participants preferred the Punch gesture for the 1st scenario. This gesture also had very positive ratings across all items, as seen in Figure 4.23. Moreover, as shown in Figure 5.3, most participants (11 out of 15) reported they preferred this gesture over their intuitively generated gesture due to being more logical. These findings suggested that when participants' intuition and logic conflicted, they tended to choose the logical gesture over the intuitive one.



6. CONCLUSION

In this thesis, we aimed to investigate the variables that affect mid-air hand gestures used to manipulate objects in virtual reality. Spatial properties of virtual objects (form, size, position, and orientation); intuitiveness and kinesthetic feedback of gestures were the focus points of this study.

The experiment setup required participants to wear a head-mounted display (HMD), whereas the timing of the study coincided with the Coronavirus disease of 2019 (COVID-19) global pandemic, which hindered the experimentation. In addition, it was required for the author to be present during the experiments due to the experiment setup. Furthermore, none of the participants, except for Participant 1, had an Oculus Quest 2 headset. Therefore, these participants used the author's HMD for the experiment. To ensure sanitary conditions, the HMD was disinfected with 80% alcohol and was rested after each experiment. Moreover, only people vaccinated against COVID-19 were recruited as participants. Lastly, the number of participants was limited to 15.

This thesis study proposed two novel gesture classifications named hand-object intersections and repetition per participant. Furthermore, two new classifications were proposed for kinesthetic feedback based on its source and temporal status. Additionally, the terms 'easy to perform,' 'iconically logical,' and 'metaphorically logical' (Nielsen et al., 2003) were defined in this thesis study. Lastly, the extent of pantomimic gestures was expanded to include the novel use cases observed during the experimentation.

Participants' field of view (FOV) affected the gestures for the explosion scenarios. The blast radius of the explosion animation encompassed different regions for the 1st, 2nd, and 3rd scenarios in VR. For example, in the 1st scenario, where the object was relatively big and located near the participant, some fragments were scattered outside the participant's FOV. Here, participants mostly gestured according to the initial seconds of the explosion and moved their hands symmetrically up and apart. However, in the 2nd scenario, where the object was still near the participant but was smaller, the

fragments didn't scatter around as much as they did in the 1st scenario. This time, participants mostly gestured according to the final seconds of the animation and moved their hands directly downwards without moving them apart. In the 3rd scenario, where the object was relatively big again but located further away, all of the fragments of the explosion remained within sight. Here, participants mostly moved their hands down and apart.

The color-coding of the objects and the virtual hands proved to be effective. Participant 11 quoted, "I felt like my green hands should interact with the green surfaces on the object." Participants 13 and 14 made similar remarks as well.

System navigation gestures received highly positive feedback, too. Thumbs-up gesture was deemed considerably fun by participants. Participant 2 quoted, "Oh, this is so much fun!" after effectively using the Thumbs-up gesture for the first time. In contrast, there were some rare cases of unintended triggers of system navigation gestures. These cases predominantly occurred when the virtual fingers misbehaved due to the lack of light or the occlusion.

The implosion animation proved to be a weakness of this thesis study since some of the participants misread this animation as if the object was moving away from them. This was assumed to be due to the uniform nature of the scale-down animation. Initially, a more complex animation, where the object deformed unevenly with some parts moving faster and breaking apart as the object collapsed in on itself, was aimed, but later discarded due to time limitations. In the future, this animation can be tested.

Thumb positioning proved to be more diverse than anticipated. While making a fist, participants placed their thumbs on top of, below, and next to the other fingers. The Open Hand to Fist gesture was designed to end with the thumb resting on top of the other fingers, as shown in Figure 3.9. However, due to the variations of the thumb placement, the system failed to recognize the gesture. This failure frustrated some participants and was assumed to be the cause of relatively low fun ratings of Open Hand to Fist gesture. In the future, gestures with different thumb placements can be added to the gesture vocabulary as variations of the same gesture, allowing recognition for each case.

In the future, the experiment can be conducted again by recruiting participants from different nationalities, allowing us to investigate how both the elicited gestures and the

attitudes towards author-designed gestures change as the culture plays a significant role in shaping peoples' gestures. For example, the Thumbs-up gesture used to start the experiment is a sign of approval in Turkish culture, whereas the same gesture has a very offensive meaning in numerous countries in the Middle East and West Africa.

Apart from the role of the culture, other criteria such as education level, occupation, and VR experience can be further explored in the future by recruiting participants conforming to specific groups within each category. As this thesis study recruited participants from the general public, the role of the aforementioned criteria on elicited gestures and gesture preferences was inconclusive.

Lastly, this thesis study elicited an affect display gesture (Ottenheimer, 2005), which was not observed in any other HCI-focused study we reviewed in the literature. Hence, these gestures showed potential to be used in this field as well. In the future, for example, in a visual novel type of game, affect display gestures can be used to project the player's emotions to an in-game avatar at specific points, altering the direction of the game. This game setting can even be used for educational purposes enabling children to explore the consequences of their various emotional responses.



REFERENCES

- Aigner, R., Wigdor, D., Benko, H., Haller, M., Lindbauer, D., Ion, A., Zhao, S. & Koh, J. T. K. V.** (2012). Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for HCI. *Microsoft Research TechReport MSR-TR-2012-111*, 2, 30.
- Alger, M.** (2015). Visual design methods for virtual reality. *Ravensbourne*. http://aperturesciencellc.com/vr/VisualDesignMethodsforVR_MikeAlger.pdf.
- Arefin Shimon, S. S., Lutton, C., Xu, Z., Morrison-Smith, S., Boucher, C., & Ruiz, J.** (2016, May). Exploring non-touchscreen gestures for smartwatches. In *Proceedings of the 2016 chi conference on human factors in computing systems* (pp. 3822-3833).
- Arora, R., Kazi, R. H., Kaufman, D. M., Li, W., & Singh, K.** (2019, October). Magicalhands: Mid-air hand gestures for animating in VR. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (pp. 463-477).
- Bolt, R. A.** (1980, July). "Put-that-there" Voice and gesture at the graphics interface. In *Proceedings of the 7th annual conference on Computer graphics and interactive techniques* (pp. 262-270).
- Bostan, I., Buruk, O. T., Canat, M., Tezcan, M. O., Yurdakul, C., Göksun, T., & Özcan, O.** (2017, June). Hands as a controller: User preferences for hand specific on-skin gestures. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (pp. 1123-1134).
- Chan, E., Seyed, T., Stuerzlinger, W., Yang, X. D., & Maurer, F.** (2016, May). User elicitation on single-hand microgestures. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 3403-3414).
- Choi, S.** (2016). Understanding people with human activities and social interactions for human-centered computing. *Human-centric Computing and Information Sciences*, 6(1), 1-10.
- Cohen, J.** (2010). Perception and computation. *Philosophical Issues*, 20, 96-124.
- Dardas, N. H., & Alhaj, M.** (2011). Hand gesture interaction with a 3D virtual environment. *The Research Bulletin of Jordan ACM*, 2(3), 86-94.
- Di Geronimo, L., Bertarini, M., Badertscher, J., Husmann, M., & Norrie, M. C.** (2017, September). Exploiting mid-air gestures to share data among devices. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 1-11).

- Dim, N. K., Silpasuwanchai, C., Sarcar, S., & Ren, X.** (2016, June). Designing mid-air TV gestures for blind people using user-and choice-based elicitation approaches. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (pp. 204-214).
- Eisenstein, J., & Davis, R.** (2007). Visual and linguistic information in gesture classification. In *ACM SIGGRAPH 2007 courses* (pp. 15-es).
- Eldrandaly, K. A., Abdel-Basset, M., & Shawky, L. A.** (2019). Internet of spatial things: A new reference model with insight analysis. *IEEE Access*, 7, 19653-19669.
- Feix, T., Romero, J., Schmiedmayer, H. B., Dollar, A. M., & Kragic, D.** (2015). The grasp taxonomy of human grasp types. *IEEE Transactions on human-machine systems*, 46(1), 66-77.
- Gheran, B. F., Vanderdonckt, J., & Vatavu, R. D.** (2018, June). Gestures for smart rings: empirical results, insights, and design implications. In *Proceedings of the 2018 Designing Interactive Systems Conference* (pp. 623-635).
- Green, M., & Jacob, R.** (1991). SIGGRAPH'90 Workshop report: software architectures and metaphors for non-WIMP user interfaces. *ACM SIGGRAPH Computer Graphics*, 25(3), 229-235.
- Grossman, T., Wigdor, D., & Balakrishnan, R.** (2004, October). Multi-finger gestural interaction with 3d volumetric displays. In *Proceedings of the 17th annual ACM symposium on User interface software and technology* (pp. 61-70).
- Gruenefeld, U., Weiß, S., Löcken, A., Virgilio, I., Kun, A. L., & Boll, S.** (2019, September). VRoad: Gesture-based interaction between pedestrians and automated vehicles in virtual reality. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings* (pp. 399-404).
- Harrison, H. L. H., & Hummell, L. J.** (2010). Incorporating animation concepts and principles in STEM education. *Technology and Engineering Teacher*, 69(8), 20.
- Jurewicz, K. A., Neyens, D. M., Catchpole, K., & Reeves, S. T.** (2018). Developing a 3D gestural interface for anesthesia-related human-computer interaction tasks using both experts and novices. *Human factors*, 60(7), 992-1007.
- Karam, M. & Schraefel, M. C.** (2005) A Taxonomy of Gestures in Human Computer Interactions. Technical Report ECSTR-IAM05-009, Electronics and Computer Science, University of Southampton.
- Kettebekov, S.** (2004, October). Exploiting prosodic structuring of coverbal gesticulation. In *Proceedings of the 6th International Conference on Multimodal interfaces* (pp. 105-112).
- Kray, C., Nesbitt, D., Dawson, J., & Rohs, M.** (2010, September). User-defined gestures for connecting mobile phones, public displays, and tabletops.

- In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services* (pp. 239-248).
- Löcken, A., Hesselmann, T., Pielot, M., Henze, N., & Boll, S.** (2012). User-centred process for the definition of free-hand gestures applied to controlling music playback. *Multimedia systems*, 18(1), 15-31.
- Malu, M., Chundury, P., & Findlater, L.** (2018, April). Exploring accessible smartwatch interactions for people with upper body motor impairments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).
- Mitra, S., & Acharya, T.** (2007). Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(3), 311-324.
- Morris, M. R., Wobbrock, J. O., & Wilson, A. D.** (2010). Understanding users' preferences for surface gestures. In *Proceedings of graphics interface 2010* (pp. 261-268).
- Nielsen, M., Störing, M., Moeslund, T. B., & Granum, E.** (2003, April). A procedure for developing intuitive and ergonomic gesture interfaces for HCI. In *International gesture workshop* (pp. 409-420). Springer, Berlin, Heidelberg.
- Obaid, M., Kistler, F., Kasparavičiūtė, G., Yantaç, A. E., & Fjeld, M.** (2016, October). How would you gesture navigate a drone? a user-centered approach to control a drone. In *Proceedings of the 20th International Academic Mindtrek Conference* (pp. 113-121).
- Ottenheimer, H. J.** (2005). *The anthropology of language: an introduction to linguistic anthropology*. Wadsworth Publishing.
- Peacocke, C.** (1985). Sense and Content: Experience, Thought & Their Relations. *Mind*, 94(375).
- Pham, T., Vermeulen, J., Tang, A., & MacDonald Vermeulen, L.** (2018, June). Scale impacts elicited gestures for manipulating holograms: Implications for AR gesture design. In *Proceedings of the 2018 Designing Interactive Systems Conference* (pp. 227-240).
- Piumsomboon, T., Clark, A., Billinghamurst, M., & Cockburn, A.** (2013, September). User-defined gestures for augmented reality. In *IFIP Conference on Human-Computer Interaction* (pp. 282-299). Springer, Berlin, Heidelberg.
- Pyryeskin, D., Hancock, M., & Hoey, J.** (2012, November). Comparing elicited gestures to designer-created gestures for selection above a multitouch surface. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces* (pp. 1-10).
- Quek, F., McNeill, D., Bryll, R., Duncan, S., Ma, X. F., Kirbas, C., McCullough, K. E. & Ansari, R.** (2002). Multimodal human discourse: gesture and speech. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 9(3), 171-193.
- Rautaray, S. S., & Agrawal, A.** (2011, December). Interaction with virtual game through hand gesture recognition. In *2011 International Conference on*

Multimedia, Signal Processing and Communication Technologies (pp. 244-247). IEEE.

- Ruiz, J., Li, Y., & Lank, E.** (2011, May). User-defined motion gestures for mobile interaction. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 197-206).
- Schellenberg, S.** (2008). The situation-dependency of perception. *The Journal of Philosophy*, 105(2), 55-84.
- Silpasuwanchai, C., & Ren, X.** (2015). Designing concurrent full-body gestures for intense gameplay. *International Journal of Human-Computer Studies*, 80, 1-13.
- Stern, H. I., Wachs, J. P., & Edan, Y.** (2008). Designing hand gesture vocabularies for natural interaction by combining psycho-physiological and recognition factors. *International Journal of Semantic Computing*, 2(01), 137-160.
- Valem.** (2020, March 2). *Hand Tracking Gesture Detection - Unity Oculus Quest Tutorial* [Video]. YouTube. <https://www.youtube.com/watch?v=1BzwUKQ3tbw>
- Van Dam, A.** (1997). Post-WIMP user interfaces. *Communications of the ACM*, 40(2), 63-67.
- Vatavu, R. D.** (2012, July). User-defined gestures for free-hand TV control. In *Proceedings of the 10th European conference on Interactive tv and video* (pp. 45-48).
- Vogel, D., & Balakrishnan, R.** (2005, October). Distant freehand pointing and clicking on very large, high resolution displays. In *Proceedings of the 18th annual ACM symposium on User interface software and technology* (pp. 33-42).
- Vogiatzidakis, P., & Koutsabasis, P.** (2018). Gesture elicitation studies for mid-air interaction: A review. *Multimodal Technologies and Interaction*, 2(4), 65.
- Vuletic, T., Duffy, A., Hay, L., McTeague, C. P., Campbell, G., Choo, P. L., & Grealy, M.** (2018). Natural and intuitive gesture interaction for 3D object manipulation in conceptual design. In *DS 92: Proceedings of the DESIGN 2018 15th International Design Conference* (pp. 103-114).
- Wang, Y., & MacKenzie, C. L.** (2000, April). The role of contextual haptic and visual constraints on object manipulation in virtual environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 532-539).
- Wentz, E. A.** (2000). A shape definition for geographic applications based on edge, elongation, and perforation. *Geographical Analysis*, 32(2), 95-112.
- Who discovered 3D shapes?** (2021). Retrieved October 15, 2021, from <https://notes-of-history.com/library/details/read/189044-who-discovered-3d-shapes>
- Wobbrock, J. O., Morris, M. R., & Wilson, A. D.** (2009, April). User-defined gestures for surface computing. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1083-1092).

- Wu, H., Wang, Y., Qiu, J., Liu, J., & Zhang, X.** (2019). User-defined gesture interaction for immersive VR shopping applications. *Behaviour & Information Technology*, 38(7), 726-741.
- Wu, H., Zhang, S., Liu, J., Qiu, J., & Zhang, X.** (2019). The gesture disagreement problem in free-hand gesture interaction. *International Journal of Human-Computer Interaction*, 35(12), 1102-1114.
- Wu, M., Shen, C., Ryall, K., Forlines, C., & Balakrishnan, R.** (2006, January). Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces. In *First IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP'06)* (pp. 8-pp). IEEE.
- Yan, Y., Yu, C., Ma, X., Yi, X., Sun, K., & Shi, Y.** (2018, April). Virtualgrasp: Leveraging experience of interacting with physical objects to facilitate digital object retrieval. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).
- Yang, L. I., Huang, J., Feng, T. I. A. N., Hong-An, W. A. N. G., & Guo-Zhong, D. A. I.** (2019). Gesture interaction in virtual reality. *Virtual Reality & Intelligent Hardware*, 1(1), 84-112.



APPENDICES

APPENDIX A.1: Analysis of author-designed gestures and scenarios

APPENDIX A.2: Participant information sheet

APPENDIX A.3: Experiment walkthrough

APPENDIX A.4: Experiment flow

APPENDIX A.5: General profile part of the survey

APPENDIX A.6: Player profile part of the survey

APPENDIX A.7: Gesture evaluation section of the survey for the 1st scenario

APPENDIX A.8: Gesture evaluation section of the survey for the 2nd scenario

APPENDIX A.9: Gesture evaluation section of the survey for the 3rd scenario

APPENDIX A.10: Gesture evaluation section of the survey for the 4th scenario

APPENDIX A.11: Gesture evaluation section of the survey for the 5th scenario

APPENDIX A.12: Gesture evaluation section of the survey for the 6th scenario

APPENDIX A.13: Participant profiles

APPENDIX A.14: User-elicited gesture examples for each scenario

APPENDIX A.15: Participant preferences



APPENDIX A.1

The attributes and classification of author-designed gestures and their corresponding scenarios are shown in Figure A.1.1.

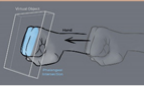


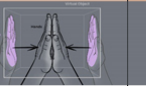


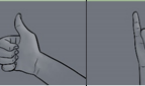

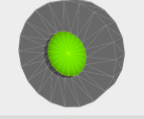
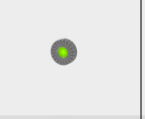
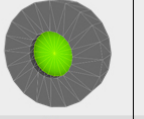
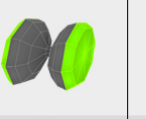
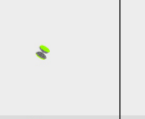
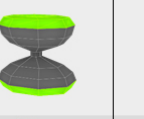
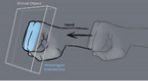


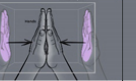




				Object Manipulation Gestures						System Navigation Gestures			
				punch	middle finger flick	four finger flick	clap	pinch	open hand to fist	thumbs-up	pinky-up		
													
affordances	object	spatial properties	form										
			size	width	∅40 cm	∅10 cm	∅40 cm	30 cm	4 cm	30 cm			
				height	∅40 cm	∅10 cm	∅40 cm	30 cm	4 cm	30 cm			
				depth	10 cm	2,5 cm	10 cm	15 cm	2 cm	15 cm			
			position	x-axis	directly in front	directly in front	directly in front	directly in front	directly in front	directly in front	directly in front		
				y-axis	eye height	chest height	eye height	elbow height	chest height	elbow height	elbow height		
				z-axis	within arm's reach	within arm's reach	beyond arm's reach	within arm's reach	within arm's reach	beyond arm's reach	beyond arm's reach		
			orientation	form					45° around view axis	90° around view axis			
				color	green	green	green	green	green	green			
				animation	explosion	explosion	explosion	implosion	implosion	implosion			
virtual hand	signifier	color	green	green	green	green	green	green					
gesture input	upper-limb motion	elbow flexion	elbow flexion	elbow flexion	elbow flexion	elbow flexion	elbow flexion	elbow flexion	elbow flexion	elbow flexion	elbow flexion		
		shoulder flexion	shoulder flexion	shoulder flexion	shoulder flexion	shoulder flexion	shoulder flexion	shoulder flexion	shoulder flexion	elbow flexion	elbow flexion		
	gesture taxonomy	number of hands	unimanual	unimanual	unimanual	unimanual	unimanual	unimanual	unimanual	unimanual	unimanual	unimanual	
			bimanual				bimanual						
			static pose								static pose	static pose	
		form	dynamic pose										
			static pose & path										
			dynamic pose & path										
		nature	direct manipulation	direct manipulation	direct manipulation	direct manipulation	direct manipulation	direct manipulation	direct manipulation	direct manipulation			
			pantomimic										
			iconic										
		hand-object intersections*	semaphoric										
			palmar*								semaphoric	semaphoric	
			phalangeal*	phalangeal									
		flow	distal*										
discrete	discrete		discrete	discrete	discrete	discrete	discrete	discrete	discrete	discrete			
		continuous					continuous	continuous					

Figure A.1.1 : Analysis of author-designed gestures and scenarios.

				Object Manipulation Gestures					System Navigation Gestures				
				punch	middle finger flick	four finger flick	clap	pinch	open hand to fist	thumbs-up	pinky-up		
													
gesture input	upper-limb motion	elbow flexion		elbow flexion	elbow flexion	elbow flexion	elbow flexion	elbow flexion	elbow flexion	elbow flexion	elbow flexion		
		shoulder flexion		shoulder flexion	shoulder flexion		shoulder flexion		shoulder flexion		shoulder flexion		
	gesture taxonomy	number of hands	unimanual		unimanual		unimanual		unimanual		unimanual	unimanual	
			bimanual				bimanual						
		form	static pose									static pose	static pose
			dynamic pose										
			static pose & path										
		nature	dynamic pose & path										
	direct manipulation			direct manipulation	direct manipulation		direct manipulation	direct manipulation					
	pantomimic						pantomimic				pantomimic		
	iconic						iconic				iconic		
	hand-object intersections*	semaphoric									semaphoric	semaphoric	
palmar*						palmar							
phalangeal*			phalangeal										
flow	distal*									distal			
	discrete		discrete	discrete	discrete			discrete	discrete	discrete	discrete		
gesture output	kinesthetic feedback	source*	self-kinesthetic*		self-kinesthetic	self-kinesthetic		self-kinesthetic	self-kinesthetic	self-kinesthetic			
		temporal status*	pair-kinesthetic*				pair-kinesthetic						
			pre-interaction*		pre-interaction	pre-interaction							
			peri-interaction*							peri-interaction			
	sound	post-interaction*					post-interaction sound		post-interaction				
		explode object		explode object	explode object	explode object							
	system response	implode object					implode object	implode object	implode object				
		start/next scenario								start/next scenario			
		previous scenario									previous scenario		

* Proposed in this dissertation.

Figure A.1.1 (continued) : Analysis of author-designed gestures and scenarios.

APPENDIX A.2

To inform participants about how they will be involved, the participant information sheet shown in Figure A.2.1 is used.

Participant Information Sheet

1. What this study is about?
This study aims to analyze intuitively generated hand gestures of the participants as well as the factors that make a gesture more preferable as part of my master's thesis.
2. Do I have to take part?
Participation is entirely voluntary and refusal or withdrawal will involve no penalty or loss, now or in the future.
3. What will taking part involve?
 - Testing a virtual reality experience while presenting your thoughts in a concurrent think aloud session, approximately 8 minutes,
 - Learning predefined gestures, approximately 2 minutes,
 - Testing the experiment again with predefined gestures, approximately 5 minutes,
 - Reviewing the gestures you generated, and answering an in-person survey, approximately 15 minutes.
4. What are the possible risks and benefits of taking part?
The lack of experience with VR hardware may cause VR sickness. Some of the symptoms are dizziness, eyestrain and disorientation. The results differ based on your susceptibility.
5. Will taking part be confidential?
Any personal data you provide during the course of the experiment will be kept strictly confidential.
6. Why and how will information you provide be recorded?
 - The experiment will be recorded by a front-facing camera in order for me to analyze the gestures you generate.
 - Everything you see inside VR will be recorded via headset's screen record function so following the experiment, we can watch and comment on your gestures.
 - The camera will also record audio to capture CTA session, which will be reviewed by me later on.
 - Survey answers will be collected via Google Forms to interpret the data.
7. What will happen to the results of the study?
The results of the study will be presented in my master's thesis.
8. What do I give consent to?
 - I confirm that I have read and understand the Participant Information Sheet.
 - I understand that all personal information will remain confidential and that all efforts will be made to ensure I cannot be identified.
 - I agree that data gathered in this study may be stored anonymously and securely, and may be used for future research.
 - I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason.
 - I agree to take part in this study.

Figure A.2.1 : Participant information sheet.

APPENDIX A.3

To inform the participants about the flow of the experiment, the experiment walkthrough shown in Figure A.3.1 is used.

Experiment Walkthrough

- Wear Oculus Quest 2 headset, stand still on the virtual “stand here sign” and look forward for the duration of the experiment.
- Inside the VR, you will see real-time visual representation of your hands. Feel free to explore them. When ready, make a thumbs-up gesture to begin the experiment.
- First, you will see a 3-second countdown. When the countdown is over, an object will appear right where the countdown was and will go through a transformation.
- The object will disappear once it completes its transformation, and reappear in its untransformed state.
- Following that, perform a hand gesture that you believe would cause the previously observed transformation. You will have 5 seconds to initiate a gesture. There will be another countdown indicating how much time you have left on top of the object.
- While gesturing you are free to move your upper limbs such as fingers, arms, shoulders etc. You may use either one hand or both hands.
- The object may or may not react to your gesture.
- After you gesture, the scene will reload. Please perform the gesture you came up with 3 more times while elaborating on the gesture. You can mention which aspects of the scenario triggered you to use this gesture, what your intention was etc.
- To proceed to the next scenario, make a thumbs-up gesture again.
- If you accidentally skip a scenario you can return to it with pinky-up gesture.
- There are a total of 6 scenarios.
- You are free to use the same gesture more than once if you deem necessary.
- After you complete the experiment once, I will present you a set of predefined gestures and request you to try them for each corresponding scenario.
- Please repeat these gestures 3 more times as well while comparing them out loud with the ones you came up with.
- Once you tried all 6 predefined gestures, the VR experiment will end.
- Lastly, you will answer an in-person survey.

Figure A.3.1 : Experiment walkthrough.

APPENDIX A.4

The flowchart of the experiment is shown in Figure A.4.1.

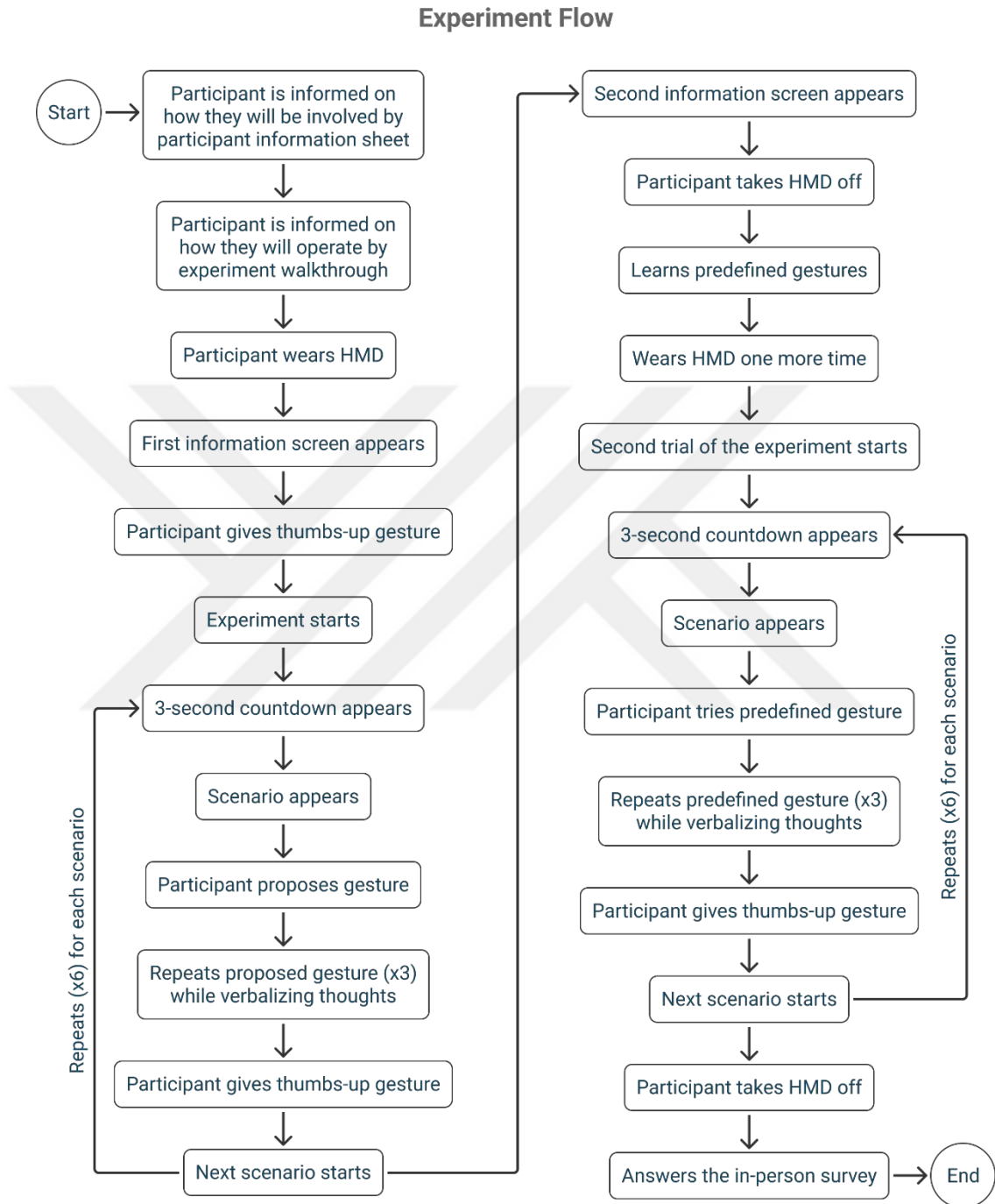


Figure A.4.1 : Experiment flow.

APPENDIX A.5

General profile part of the survey is shown in Figure A.5.1.

How old are you? *

Choose

I consider myself: *

Right-handed

Left-handed

Ambidextrous

Do you have any of the health issues stated below? *

For each that applies, indicate its severity otherwise select none.

	None	Minor	Moderate	Major	Extreme
Sight problems not corrected by glasses or contact lenses	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hearing problems	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Speech problems	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Blackouts, fits or loss of consciousness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Difficulty learning or understanding things	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Limited use of arms or fingers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Difficulty gripping things	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Limited use of legs or feet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Any condition that restricts physical activity or physical work (e.g. back problems, migraines)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Any disfigurement or deformity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Any mental illness for which help or supervision is required	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

What is the highest level of education you have completed? *

Choose

What is your occupation? *

Your answer

Figure A.5.1 : General profile part of the survey.

APPENDIX A.6

Player profile part of the survey is shown in Figure A.6.1.

How often do you play games? *

More than 5 days a week

3 - 4 days a week

1 - 2 days a week

Less than once a week

Have you ever experienced virtual reality? *

Yes

No

Please check all gesture controlled technologies you have experienced. *

Kinect

Smartphone air gestures

Leap Motion

Oculus Quest

HoloLens

Myo armband

None of the above

Other: _____

Have you ever tried a hand gesture controlled experience in virtual reality? *

Yes

No

How often did you experience virtual reality in the past 12 months?? *

More than 5 days a week

3 - 4 days a week

1 - 2 days a week

1 - 3 days a month

Experienced only a few times

Haven't experienced


Figure A.6.1 : Player profile part of the survey.

APPENDIX A.7

Gesture evaluation section of the survey for the 1st scenario is shown in Figure A.7.1.

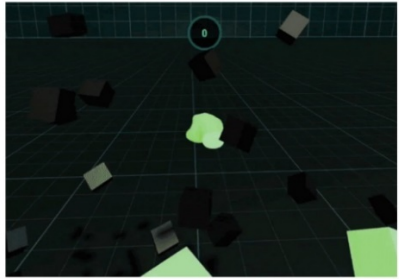
Scenario 1 - Punch gesture

Can you rate the below gesture based on the following criteria? *



	1	2	3	4	5
[1]Familiar ... [5]Unfamiliar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Simple ... [5]Complex	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Effortless ... [5]Fatiguing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Can you rate the same gesture for the below scenario based on the following criteria? *



	1	2	3	4	5
[1]Logical ... [5]Arbitrary	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Natural ... [5]Pretended	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Fun ... [5]Boring	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Which gesture do you prefer for this scenario? *

Both gestures were the same

Predefined gesture

Gesture I came up with

What made you choose this gesture? *

Check 3 items at most. Kinesthetic feedback refers to the touch sensation you feel through physical contact between parts of the hand/hands such as the tip of your thumb and index finger.

It felt more natural

It was more effortless

It was more logical

It was simpler

It was more fun

It was more intuitive

Kinesthetic feedback made it more satisfactory

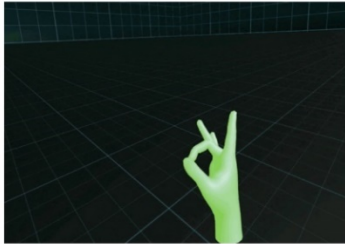
Figure A.7.1 : Gesture evaluation section of the survey for the 1st scenario.

APPENDIX A.8

Gesture evaluation section of the survey for the 2nd scenario is shown in Figure A.8.1.

Scenario 2 - Middle finger flick gesture

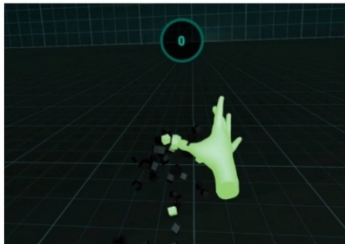
Can you rate the below gesture based on the following criteria? *



1 2 3 4 5

[1]Familiar ... [5]Unfamiliar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Simple ... [5]Complex	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Effortless ... [5]Fatiguing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Can you rate the same gesture for the below scenario based on the following criteria? *



1 2 3 4 5

[1]Logical ... [5]Arbitrary	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Natural ... [5]Pretended	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Fun ... [5]Boring	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Share your opinion regarding the following statement. *

Kinesthetic feedback refers to the touch sensation you feel through physical contact between parts of the hand/hands such as the tip of your thumb and index finger.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Kinesthetic feedback made this gesture more satisfactory	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Which gesture do you prefer for this scenario? *

Gesture I came up with

Predefined gesture

Both gestures were the same

What made you choose this gesture? *

Check 3 items at most

- It was more fun
- It was more intuitive
- It was simpler
- Kinesthetic feedback made it more satisfactory
- It was more effortless
- It was more logical
- It felt more natural

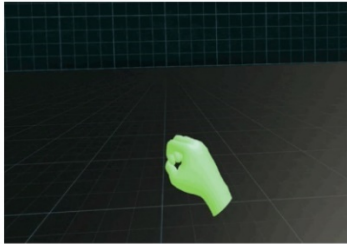
Figure A.8.1 : Gesture evaluation section of the survey for the 2nd scenario.

APPENDIX A.9

Gesture evaluation section of the survey for the 3rd scenario is shown in Figure A.9.1.

Scenario 3 - Four finger flick gesture

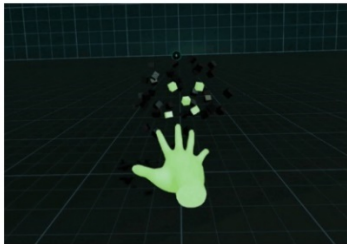
Can you rate the below gesture based on the following criteria? *



1 2 3 4 5

[1]Familiar ... [5]Unfamiliar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Simple ... [5]Complex	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Effortless ... [5]Fatiguing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Can you rate the same gesture for the below scenario based on the following criteria? *



1 2 3 4 5

[1]Logical ... [5]Arbitrary	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Natural ... [5]Pretended	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Fun ... [5]Boring	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Share your opinion regarding the following statement. *

Kinesthetic feedback refers to the touch sensation you feel through physical contact between parts of the hand/hands such as the tip of your thumb and index finger.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Kinesthetic feedback made this gesture more satisfactory	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Which gesture do you prefer for this scenario? *

Predefined gesture

Both gestures were the same

Gesture I came up with

What made you choose this gesture? *

Check 3 items at most:

It was more logical

Kinesthetic feedback made it more satisfactory

It was more fun

It felt more natural

It was simpler

It was more effortless

It was more intuitive

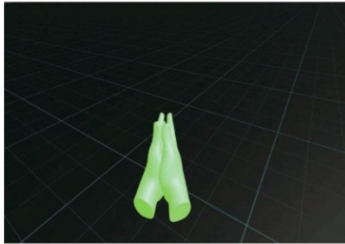
Figure A.9.1 : Gesture evaluation section of the survey for the 3rd scenario.

APPENDIX A.10

Gesture evaluation section of the survey for the 4th scenario is shown in Figure A.10.1.

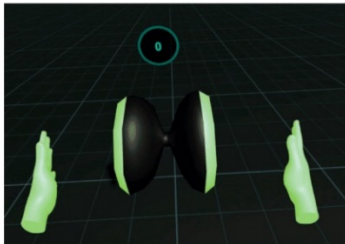
Scenario 4- Clap gesture

Can you rate the below gesture based on the following criteria? *



	1	2	3	4	5
[1]Familiar ... [5]Unfamiliar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Simple ... [5]Complex	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Effortless ... [5]Fatiguing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Can you rate the same gesture for the below scenario based on the following criteria? *



	1	2	3	4	5
[1]Logical ... [5]Arbitrary	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Natural ... [5]Pretended	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Fun ... [5]Boring	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Share your opinion regarding the following statement. *

Kinesthetic feedback refers to the touch sensation you feel through physical contact between parts of the hand/hands such as the tip of your thumb and index finger.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Kinesthetic feedback made this gesture more satisfactory	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Which gesture do you prefer for this scenario? *

Both gestures were the same

Gesture I came up with

Predefined gesture

What made you choose this gesture? *

Check 3 items at most

- It was more intuitive
- It was simpler
- It was more effortless
- It was more logical
- It felt more natural
- Kinesthetic feedback made it more satisfactory
- It was more fun

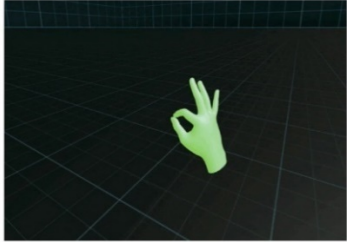
Figure A.10.1 : Gesture evaluation section of the survey for the 4th scenario.

APPENDIX A.11

Gesture evaluation section of the survey for the 5th scenario is shown in Figure A.11.1.

Scenario 5 - Pinch gesture

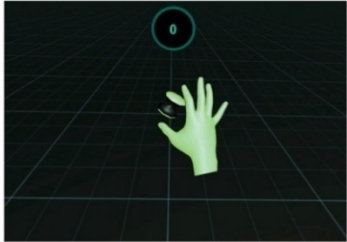
Can you rate the below gesture based on the following criteria? *



1 2 3 4 5

[1]Familiar ... [5]Unfamiliar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Simple ... [5]Complex	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Effortless ... [5]Fatiguing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Can you rate the same gesture for the below scenario based on the following criteria? *



1 2 3 4 5

[1]Logical ... [5]Arbitrary	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Natural ... [5]Pretended	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Fun ... [5]Boring	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Share your opinion regarding the following statement. *

Kinesthetic feedback refers to the touch sensation you feel through physical contact between parts of the hand/hands such as the tip of your thumb and index finger.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Kinesthetic feedback made this gesture more satisfactory	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Which gesture do you prefer for this scenario? *

Gesture I came up with

Both gestures were the same

Predefined gesture

What made you choose this gesture? *

Check 3 items at most

It was more fun

It was simpler

It was more logical

It was more effortless

It felt more natural

Kinesthetic feedback made it more satisfactory

It was more intuitive


Figure A.11.1 : Gesture evaluation section of the survey for the 5th scenario.

APPENDIX A.12

Gesture evaluation section of the survey for the 6th scenario is shown in Figure A.12.1.

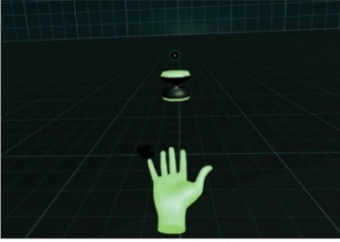
Scenario 6 - Open hand to fist gesture

Can you rate the below gesture based on the following criteria? *



	1	2	3	4	5
[1]Familiar ... [5]Unfamiliar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Simple ... [5]Complex	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Effortless ... [5]Fatiguing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Can you rate the same gesture for the below scenario based on the following criteria? *



	1	2	3	4	5
[1]Logical ... [5]Arbitrary	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Natural ... [5]Pretended	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[1]Fun ... [5]Boring	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Share your opinion regarding the following statement. *

Kinesthetic feedback refers to the touch sensation you feel through physical contact between parts of the hand/hands such as the tip of your thumb and index finger.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Kinesthetic feedback made this gesture more satisfactory	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Which gesture do you prefer for this scenario? *

Both gestures were the same

Gesture I came up with

Predefined gesture

What made you choose this gesture? *

Check 3 items at most

It was more fun

It was more effortless

Kinesthetic feedback made it more satisfactory

It was simpler

It was more logical

It felt more natural

It was more intuitive

Figure A.12.1 : Gesture evaluation section of the survey for the 6th scenario.

APPENDIX A.14

Still frames of some of the user-elicited gestures for each scenario are shown in Figure A.14.1.



Figure A.14.1 : User-elicited gesture examples for each scenario.

APPENDIX A.15

The preferred gestures of all the participants and their reasoning are shown in Figure A.15.1.

	Scenario 1		Scenario 2		Scenario 3		Scenario 4		Scenario 5		Scenario 6	
	PG	Preference Factors	PG	Preference Factors	PG	Preference Factors	PG	Preference Factors	PG	Preference Factors	PG	Preference Factors
Participant 1	M	I E	P	N SK L	P	F SK	M	I N L	M	I N F	U	I N F
Participant 2	P	I N	P	N F	P	N F S	U	I N L	U	I N S	P	I E
Participant 3	M	I S L	P	S E L	P	S E L	U	I	P	N L	P	E
Participant 4	P	N E L	U	N E L	P	I S E	P	I F	M	F SK L	U	F S
Participant 5	P	I E L	P	S SK L	P	S SK L	M	I S SK	P	N L	U	N S E
Participant 6	P	N F L	U	N F	P	N F L	P	E L	U	N	P	S L
Participant 7	P	I N L	P	N L	P	L	M	I N L	P	I N L	P	N S L
Participant 8	P	F L	P	N L	P	I S L	P	I N F	P	F E L	P	E
Participant 9	P	L	P	L	U	S L	U	I L	U	I S	U	I
Participant 10	M	I N L	U	I F SK	U	I L	M	F SK L	M	F SK L	U	I F L
Participant 11	P	I N E	U	I S E	U	I S E	P	S E L	U	I	P	E L
Participant 12	P	E L	U	N S	U	I	M	I F E	M	I N E	P	E L
Participant 13	P	F E L	P	I S E	P	F L	M	I N L	M	I N L	U	I S E
Participant 14	P	S E L	U	S E	U	I N L	P	I N L	M	I N L	U	I N
Participant 15	U	I F E	U	I F	U	I	P	S SK	U	I F E	P	L

PG = Preferred Gesture	U = User-elicited Gesture
	P = Predefined Gesture
	M = Gestures Matched

Preference Factors	I = Intuitive
	N = Natural
	F = Fun
	S = Simple
	E = Effortless
	SK = Satisfactory due to Kinesthetic Feedback
	L = Logical

Figure A.15.1 : Participant preferences.

CURRICULUM VITAE



Name Surname : Erkut Sırdaş



EDUCATION :

- **B.Sc.** : 2016, Anadolu University, School of Industrial Arts, Industrial Design
- **B.Sc. Erasmus Exchange:** 2013, L'École de design Nantes Atlantique, International Class, Industrial Products

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2018-Present Teracity Software Technologies
- 2017-2018 Efes Play
- 2016-2017 GLR Metal
- 2015-2016 TMS Stand
- 2014-2015 Dekorex Architecture
- 2013 Hexagon Studio
- 2013 Farba Automotive
- 2012 Arçelik