Researching virtual reality and augmented reality devices to enhance human-computer interaction

DOI:10.18136/PE.2020.761

PhD Thesis

Tibor Guzsvinecz

Supervisor: Cecilia Sik-Lanyi
associate professor

Doctoral School of Information Science and Technology
Faculty of Information Technology
University of Pannonia

2020

Researching virtual reality and augmented reality devices to enhance human-computer interaction

Thesis for obtaining a PhD degree in the Doctoral School of Information Science and Technology of the University of Pannonia

in the branch of Information Sciences

Written by: Tibor Guzsvinecz
Supervisor(s): Cecilia Sik-Lanyi

Propose acceptance (yes / no)

....
Supervisor(s)

As reviewer, I propose acceptance of the thesis:

Name of Reviewer: .... (yes / no)

....
Reviewer

Name of Reviewer: .... (yes / no)

....
Reviewer

The PhD-candidate has achieved ....% at the public discussion,

Veszprém/Keszthely, ...

....
Chairman of the Committee

The grade of the PhD Diploma .... (........%)

Veszprém/Keszthely, ...

....
Chairman of the UDHC

Abstract

In this PhD thesis research about human-computer interaction using virtual reality and augmented reality devices are presented. Human-computer interaction not only differs from device to device, but from application to application as well. Thus, with the increasing number of both virtual reality and augmented reality devices and their applications, it is important to assess human-computer interaction and to propose new design choices with the goal to make the interaction easier for the users. Therefore, two different parts of human-computer interaction were investigated.

Spatial ability is essential when using virtual reality and augmented reality applications as the user is placed in a four-dimensional space. While the spatial skills of the users can be negatively affected by a poorly designed virtual environment, they can actually be enhanced by a well-designed one. Therefore, the first part of this research is about assessing and enhancing the spatial skills of the users in virtual environments in case of two different display devices: a desktop display device and the Gear VR head-mounted display device. For this, a spatial ability measuring application using three spatial ability test types has been developed and the skills of 240 and 61 students using the desktop display and the Gear VR were measured, respectively. In this research, three different aspects have been investigated based on their results on the tests: the first is about which attributes of the virtual environment and the display device used enhance the spatial ability of the users by increasing their probabilities of correct answers. The second is about which skills of the users and the display device used influence the completion times on the spatial ability tests and the third is to find which device used correlates with which human skill.

According to the results of this research of spatial skills, the optimal preference for the virtual environments to positively influence the correct answers on the spatial ability tests by affecting the human–computer interaction is a perspective camera type, a camera rotation of -45° or 0° or 45°, a contrast ratio of 1.5:1 or 3:1, and the Gear VR display device. Also, the probabilities of the correct answers and test completion times are not independent and the latter is significantly affected by the used display device, the test type and the gender of the user. Lastly, compared to the desktop display, the results of female or left-handed or older students are significantly improved with the use of the Gear VR. With it, the Purdue Spatial Visualization Test is also made significantly easier.

The second part of this research focuses on the use of the Kinect sensor in medical applications, mainly in physical rehabilitation. In this part of the research two various aspects have been investigated: the first aspect is about whether more expensive sensors can be substituted by the Kinect. This is a crucial aspect in telerehabilita-

tion as the hospitals have became overcrowded nowadays and rehabilitation can be more convenient in the homes of the patients. Low-cost sensors such as the Kinect can also be affordable for the families of the patients. To make physical rehabilitation easier in home environments, the Asynchronous Prediction-Based Movement Recognition algorithm has been developed. The gesture descriptors of four groups of people were measured and the results were evaluated both in real-time and from a file. All evaluations were done in three acceptance domains using two computers and six mean techniques.

Based on the evaluation of the algorithm, the prediction-based gesture recognition method is viable and usable in a home environment with the Kinect v1. By comparing to the previous algorithm it was based on, it can be concluded that the prediction-based algorithm has an increased average gesture acceptance rate by 358.2%-535.3% in the ± 0.05 m acceptance domain depending on the used mean technique. The increase in the ± 0.10 m and the ± 0.15 m acceptance domains is 87.8%-125.4% and 22.7%-47.3%, respectively.

With this research the aim of the author is to make the human-computer interaction easier, therefore the output of this research is dual: the first is to form a new recommendation in the design of virtual environments which can enhance the spatial skills of the users and the second is to present a new, easy-to-use gesture recognition algorithm which can help the physical rehabilitation of people with movement disabilities in their homes.

Tartalmi kivonat

A doktori értekezésemben azt vizsgálom, hogy virtuális és kiterjesztett valóság eszközök használata során miképpen zajlik az ember és gép közti interakció (human-computer interaction). Az interakció fajtája változik eszközről eszközre, alkalmazásról alkalmazásra. Nem beszélve arról, hogy a virtuális valóság és kiterjesztett valóság alapú eszközök és azok alkalmazásai napról napra nőnek. Ezek miatt fontosnak tartom az ember és gép közti interakció vizsgálatát, aminek kimeneteképpen új tervezési ajánlásokat fogalmazhatok meg, hogy könnyebbé tegyem az ember és gép közti interakciót. A doktori értekezésemben ezen interakció két különböző részét vizsgáltam.

Mivel a felhasználót egy négydimenziós térbe helyezzük a virtuális- és kiterjesztett valóság használata során, ezért a térérzékelés képessége létfontosságú. Egy rosszul megtervezett virtuális környezet ronthatja a felhasználó térérzékelési képességeit, de egy jól megtervezett pedig fokozhatja azt. Utóbbit vizsgálom a kutatásom első felében. Ehhez egy asztali monitort és a Gear VR fejre helyezhető virtuális valóság kijelzőt használtam. Fejlesztettem egy alkalmazást, ami képes felmérni a felhasználó térérzékelési képességét. Ezt három különböző térérzékelést vizsgáló teszt típussal teszi meg. Monitorral 240, a Gear VR használatával pedig 61 egyetemi hallgató térérzékelési képességeit mértem fel. Kutatásaim során három különböző aspektust vizsgáltam az eredményeik alapján: az első, hogy melyik eszköz és a virtuális környezet melyik paramétere fokozza a térérzékelést azzal, hogy növeli a felhasználók jó válaszainak valószínűségét. A második, hogy melyik emberi tulajdonság és megjelenítési eszköz befolyásolja a térérzékelési teszt teljesítési idejét és a harmadik, hogy milyen kölcsönhatás van az eszközök és az emberi tulajdonságok között.

A térérzékeléssel kapcsolatos kutatási eredményeim azt bizonyítják, hogy a virtuális környezet optimális felépítése pozitívan befolyásolja a térérzékelési teszteken elért jó válaszokat és ezáltal az ember és gép közti interakciót is. Ez az optimális felépítés nem más, mint egy perspektivikus kamera, ami vagy -45°-kal, vagy 0°-kal, vagy 45°-kal kerül elforgatásra, illetve egy 1,5:1 vagy egy 3:1 kontrasztarány, és a Gear VR megjelenítő használata. Ezen felül, a jó válaszok valószínűsége és a teszt teljesítési idők nem függetlenek egymástól és az utóbbit szignifikánsan befolyásolja a használt megjelenítő, a teszt típusa és a felhasználó neme. Végül, a Gear VR használata szignifikánsan javítja a jó válaszok arányát női vagy balkezes vagy idősebb hallgatók esetén, illetve könnyebbé teszi a Purdue Spatial Visualization Test nevű teszt típust.

A kutatásom második része a Kinect szenzor egészségügyi – főként mozgásrehabilitációs – felhasználására fókuszál. Ezt a részt két különböző nézőpontból vizsgál-

tam, amiből az első, hogy a Kinect helyettesíteni tud-e nála drágább szenzorokat. Ez egy kritikus része a telerehabilitációnak, hiszen manapság túlzsúfoltak a kórházak, aminek következtében a rehabilitáció hatásosabb lehet a páciensek otthonában. A Kinecthez hasonló olcsó szenzorok könnyen elérhetőek a családok számára. Hogy könnyebbé tegyem a rehabilitációt otthoni környezetben, létrehoztam az Aszinkron Predikció Alapú Mozgásfelismerő (Asynchronous Prediction-Based Movement Recognition) algoritmust. Négy felhasználói csoport gesztusleíróit mértem és értékeltem ki valós időben és fáljból is. Minden kiértékelés három elfogadási tartományban történt, két számítógéppel és hat átlagolási technikával.

A kiértékelést követően arra a konklúzióra jutottam, hogy a predikció-alapú mozgásfelismerő módszer hasznos és használható otthoni környezetben a Kinect v1 szenzorral. Mivel egy korábbi algoritmus továbbfejlesztése ez a módszer, így összehasonlításra került vele. Az eredmények azt mutatják, hogy a predikció-alapú algoritmus esetében – a használt átlagolási technikától függően – 358,2%-535,3%-kal nőtt az átlagos gesztus elfogadási ráta a $\pm 0,05$ méteres elfogadási tartományban. Ez a növekedés a $\pm 0,10$ méteres elfogadási tartományban 87,8%-125,4%, míg a $\pm 0,15$ méteres elfogadási tartományban 22,7%-47,3%.

A célom, hogy könnyebbé tegyem az ember és gép közti interakciót, tehát a kutatásom kimenete kettős: egy olyan ajánlás megfogalmazása, amivel térérzékelési képességet fokozó virtuális környezetek tervezhetőek, illetve egy egyszerű, otthoni környezetben használható, mozgásrehabilitációs célú, gesztusfelismerő algoritmus bemutatása.

Estratto di contenuto

Nelle mie tesi di dottorato sto analizzando, come funziona la interazione tra umano e computer (human-computer interaction) durante uso degli strumenti della realtà virtuale e della realtà aumentata. Il tipo di interazione si sta modificando per ogni tipo di strumento e per ogni tipo di applicazione. Per non parlare di quello che stanno crescendo gli strumenti in base della realtà virtuale e della realtà aumentata ed anche le loro applicazioni. Per questo tengo importante analizzara le interazioni tra umano e computer, e dopo posso proporre raccomandazioni di progettazione per facilitare la interazione tra umano e computer. Nelle mie tesi di dottorato ho esaminato le due parti di queste interazioni.

Abbiamo posizionato l'utente in uno spazio quadridimensionale durante uso la realtà virtuale ed aumentata. Un'ambiente progettata male potrebbe potrebbe peggiorare la capacità di percezione dello spazio dell'utente, mentre un'ambiente progettata giustamente potrebbe migliorarla. Sto esaminando quest'ultimo nella prima parte della mia ricerca. Uso uno schermo del computer e un display Gear VR della realtà visiva montato sulla testa. Ho sviluppato un'applicazione che controlla la capacità della percezione dello spazio dell'utente, che lo fa con tre diversi tipi di test della percezione dello spazio. Ho controllato le capacità della percezione dello spazio di 240 studenti con monitor e 61 studenti con Gear VR. Nelle mie ricerche ho esaminato tre diversi aspetti in base delle risulte: per primo: quale parametro di quale strumento e di quale ambiente virtuale aumenta percezione e di conseguenze si aumenta la probabilità delle risposte buone. Per sedondo: quale proprietà umano e strumeno influenza la durata del test della percezione, e per terzo: che tipo di interazione esiste tra le proprietà umano e strumenti.

I risultati delle mie ricerche delle percezioni dello spazio dimostrano, che la contruzione ottimo dell'ambiente virtuale influenza positivamente le risposte giuste nei test della percezione dello spazio, e così anche l'interazione tra umano e computer. La costruzione ottimo nientemeno che una camera prospettiva, che si gira -45° o 0° oppure 45°, e rapporto di contraso 1,5:1 oppure 3:1, ed uso di Gear VR. Inoltra la probabilità delle risposte e tempi di test non sono indifferenti, e questi sono stati influenzati dal dispositivo di visualizzazione, il tipo di test e il sesso dell'utente. Infine l'uso di Gear VR migliora significamente il rapporto delle risposte giuste nei casi delle donne oppure dei mancini, oppure dei più vecchi studenti, intanto migliora anche il tipo di test di Purdue Spatial Visualization Test.

La seconda parte delle mie ricerche messa a fuoco all'uso sanitario – principalmente per la riabilitazione – di sensore Kinect. L'ho esaminato da due punto di vista, nel primo ho controllato se il sensore Kinect possa sostituire i sensori più costosi. Questa parte è la più sensibile del teleriabilitazione, perché oggi come oggi gli ospedali soso pieni, conseguentemente la riabilitazione potrebbe essere più efficace nelle case dei pazienti. I sensori simili al sensore Kinect sono più accessibili per le famiglie. Per facilitare la riabilitazione nell'ambito domestico ho creato un algoritmo: Asincrono Base di Predizione Riconoscimento di Movimento (Asynchrnous Prediction-Based Movement Recognition). Ho misurato descrittori di gesti di quattro gruppi ed ho analizzato in tempo reale e dei file. Tutti tre valutazioni sono stati fatti in tre gamme di accettazione, con due computer e con sei tecniche della media.

Dopo l'analisi sono arrivato alla conclusione, che il metodo di riconoscimento di movimento in base di predizione è molto utile, e può essere utilizzato con sensore Kinect v1 in un ambiente domestico. Siccome questo metodo è un ulteriormente sviluppato di un algoritmo precedente, ed è stato confrontato con quello. I risultati sono in caso di algoritmo di base di predizione – dipende quale tecnica della media usata – aumentato il tasso medio di accettazione dal 358,2% al 535,3%, nella gamma di accettazione di $\pm 0,05$ metro. Quest'aumentazione nella gamma di accettazione di $\pm 0,10$ metro è da 87,8% al 125,4%, mentre nella gamma di accettazione di $\pm 0,15$ metro aumentato da 22,7% al 47,3%.

Il mio scopo è migliorare l'interazione tra umano e computer, quindi l'uscita della mia ricerca è duale: una formulazione di una raccomandazione, con quale si progetta gli ambiti virtuali di intensificatore delle capacità di percezione dello spazio, intanto la presentazione un algoritmo di riconoscimento di movimento, che si utilizza semplicemente nell'ambito domestico per la riabilitazione del movimento.

Acknowledgement

First of all, the author of this PhD work would like to thank the Doctoral School of Information Science and Technology and its head, Professor Katalin Hangos for the support he received during the years of his PhD studies.

The author is grateful to his supervisor, Dr. Cecília Sik-Lányi as she helped him with many suggestions and continuous guidance throughout their meetings. Thanks to her support, the author was able to publish the results of his research in multiple international journals and he was also able to present them at several national and international conferences.

The author would also like to thank Dr. Éva Orbán-Mihálykó for her continuous support during his research of spatial ability in virtual environments by providing the necessary knowledge and literature regarding statistics, and Dr. Erika Perge for introducing him to the paper-based spatial ability tests and organizing the measurements at the University of Debrecen.

The continuous insight and support of Dr. Veronika Szücs made a huge impact on the studies of the author and he would not be who he is without her. The author is very grateful for that. With the knowledge she and Dr. Attila Magyar provided, the author was able to complete the second topic of his PhD thesis.

The author would also like to thank the both the Faculty of Information Technology and the Department of Electrical Engineering and Information Systems for the working environment.

The mother of the author, Erika Slezák supported him with everything and for that, the author is very grateful. The author is also grateful for his significant other and his friends as they mentally helped him through the writing process.

Last, but not least, the author would like to thank financial support of EFOP-3.6.1-16-2016-00015 and NTP-NFTÖ-B-0090. Without their support, this PhD thesis could not be made.

Contents

List of Figures

List of Tables

List of Abbreviations

AC	Advanced Computer. xii, 41, 70–72, 85, 86, XXIII–XXV
AD	Acceptance Domain. xi, xiii, 36, 38–40, 43, 70, 71, 73–75, 85–87, 91, XXVI
AE	Architectural/social Engineering. 68, 69
AGAR	Average Gesture Acceptance Rate. 69–71, 73–76, 85–88, 91, XXII–XXIV, XXVI–XXVIII
AMT	Arithmetic Mean Technique. 69–71, 74, 75, 85, 86, 91, XXII–XXVIII
APBMR	Asynchronous Prediction-Based Movement Recognition. xi, xii, xv, 5, 21, 22, 26–28, 35, 38–41, 43, 44, 69, 71–77, 85–93, 95, 96, XXV, XXVI, XXVIII
AR	Augmented Reality. 1, 2, 5, 7, 9, 11
ASD	Autism Spectrum Disorder. 14
CHMT	Contraharmonic Mean Technique. 69–72, 74, 75, 85–87, 91, XXII–XXVIII
CI	Confidence Interval. xi, xii, 44, 46, 47, 49, 50, 52–57, 59–64
CMT	Cubic Mean Technique. 69, 71, 72, 74, 85–87, 91, XXII–XXVIII
CogInfoCom	Cognitive InfoCommunications. 7
CPU	Central Processing Unit. 41
DD	Desktop Display. xii, xiv, xv, 8, 10, 19–21, 23–26, 29, 32, 33, 46, 48, 49, 52, 53, 56, 57, 60, 61, 63–69, 79, 82–84, 88, 90, III, VI–VIII, X, XI, XIII–XIX, XXVIII
D-P	Douglas-Peucker. 14
ECDF	Empirical Cumulative Distribution Function. xii, 66, 67, XXVIII–XXX
EM	ElectroMagnetic. 14, 15
F	Female. 63, 64, X, XI, XVII
FAAST	Flexible Action and Articulated Skeleton. 14
FoV	Field of View. xi, xiii, xiv, 11, 13–15, 22, 23, 29, 30, 32, 45, 47, 78, 80, 88, 89, I, II, XII
fps	Frames per second. 13, XXI
GC	General Computer. xii, 41, 70, 72, 85, 86, XXII, XXV
GMT	Geometric Mean Technique. 69, 74, 85–87, 91, XXII–XXVIII
GPU	Graphics Processing Unit. 41
GUI	Graphical User Interface. 5, 7
GVR	Gear VR. xii, xiv, xv, 20, 23–26, 29, 31–33, 46, 48, 49, 52, 53, 57, 58, 60–69, 78, 79, 81–84, 88–90, 92, 93, II–IV, VI–XVIII, XX, XXVIII
H	Hypothesis. 22–28, 35, 53, 58, 59, 65–67, 69, 77–87
HCI	Human-Computer Interaction. 2, 3, 5–7, 9, 19, 21, 22, 77, 79, 80, 84, 89, 92
HMD	Head-Mounted Display. 1, 2, 4, 7–10, 19–21, 25, 29, 46, 58, 79, 84, 89, 90, 93
HMT	Harmonic Mean Technique. 69–71, 74, 75, 85–87, 91, XXII–XXVIII
HSVT	Heinrich Spatial Visualization Test. 3, 7
IR	InfraRed. 12, 13, 15, XXI
IT	Information Technology. 1, 31, 69
LH	Left-Handed. 67
LiDAR	Light Detection And Ranging. 14, 15
M	Male. 64, X–XII, XVII
MCT	Mental Cutting Test. xi, 3, 6, 8, 9, 20, 22, 25, 26, 28, 29, 32, 60, 61, 63–65, 81–83, 88, 90, 95, X, XI, XVI–XVIII
ME	Mechanical Engineering. 68, 69
MR	Mixed Reality. 1, 2
MRT	Mental Rotation Test. xi, 3, 6, 8, 9, 20, 22, 24, 28, 32, 33, 60–65, 81, 83, 88–90, 95, IX–XII, XVI–XVIII

MT Mean Technique. xi–xiii, 21, 26, 27, 36, 37, 39–41, 43, 73–75, 85–87, 89, 91, XXII–XXV, XXVII, XXVIII

MTLS Mobile Terrestrial Laser Scanner. 14

NaN Not a Number. 37

PLL Polhemus Liberty Latus. 14, 15

PSVT Purdue Spatial Visualization Test. xi, 3, 6, 8, 20, 22, 24–26, 28, 29, 32, 60, 61, 63–65, 81–84, 88, 90, 92, 95, IX–XII, XVI–XVIII

QMT Quadratic Mean Technique. 69–72, 74, 85–87, 91, XXII–XXVIII

RAM Random Access Memory. 41

RDAMR Reference Distance Based Asynchronous Movement Recognition. xii, xv, 41, 73, 74, 87, 88, 91, 95, XXVI

RDSMR Reference Distance Based Synchronous Movement Recognition. 41, 73

RH Right-Handed. 67

RoM Range of Motion. 18

RQ Research Question. 22–28, 77, 84, 85

SDK Software Development Kit. 17

SDT Surface Development Test. 3, 7

sEMG Surface ElectroMyoGraphy. 14, 15

SMD Specified Measuring Distance. 13

T Thesis. 77–87

TLS Terrestrial Laser Scanner. 14

ToF Time-of-Flight. 12, 13, 15, 16

VE Virtual Environment. 2–4, 6, 7, 9, 19–22, 28, 41, 44, 46, 58, 69, 77–80, 84, 88–90, 92, 93

VMC Video Motion Capture. 18

VR Virtual Reality. xi, 1–11, 19–21, 28, 29, 44, 77–84, 88–90, 92–95

Chapter 1

Introduction

Information technology (IT) has multiple fields which are growing each day, but virtual reality (VR) [1] is one of its most dynamically growing areas. Since then, augmented reality (AR) [2] and mixed reality (MR) [3] also appeared and grown drastically in popularity as well. According to eMarketer [4], the user-base of VR and AR was 22.5 million and 37.6 million in 2016, respectively. These two realities are expected to reach 57.1 million and 85.0 million users by 2021, respectively. According to Business Wire [5], the market of AR and MR was valued at 2.98 billion dollars in 2019. It is expected to reach 27.44 billion dollars by 2025.

Another fact is that as the technology becomes stronger and progresses forward, sensors and applications become easier to develop. The sensors can also use more functionalities as the time progresses. For example, the Valve Index which is available since early-2020 can measure the power of the grip of the users and this feature was not available in the earlier head-mounted displays (HMDs) such as the HTC Vive. Not only the devices and the applications are more popular, but the userbase of these sensors and applications is also increasing every day. This is due to the technological advancements in the sensors as the user interaction becomes easier.

Since VR, AR and MR are widespread in this day and age, new possibilities became available: they are popular in multiple areas such as education [6, 7, 8], medical applications [9, 10], training [11, 12, 13], military [14] and even in entertainment [15, 16]. These realities are mostly simple to use and are located on different parts of a so-called "reality-scale" which is shown in Figure 1.1 and defined in [17].

Figure 1.1: Reality-scale from the real to virtual environment.

VR may be the simplest to use as it is a synthetic reality and is made up of six parts: the environment itself, a computer, a network, the input, the output devices and the users themselves. Due to this definition the objects in VR can be interacted with. AR is different than VR as it expands upon our reality: objects are placed in the real-world by it (through glasses for example), but those objects cannot be interacted with. MR is between VR and AR. With MR, interaction is allowed. Human-computer interaction (HCI) is an essential element in these realities.

The interaction and the behavior of the humans with the computer depend on the tasks, the tools that are available and even the application design [18]. Using the HCI principles, applications can be designed for different purposes, such as learning applications [19, 20], mobile applications [21], helping with assistive technologies [22], serious games with gamification elements [23, 24], entertainment applications [25], interfaces in VR [26], and even the virtual environments (VEs) themselves [27]. It is also indicated by the preceding paper that there is no ideal HCI principle for VR, as this depends greatly on the application type. The opinion of its authors is that user-centric development has proven useful in the past and thus, should be used [28].

There is a possibility that HCI can be made easier with the use of motion tracking sensors besides following a user-centric development. These sensors can either be found inside HMDs which allow the head of the user to be tracked with their inertial accelerometer(s) and gyroscope(s), or can be whole body tracking sensors such as the Microsoft Kinect which uses its depth cameras and microphones.

1.1 Motivation and aims

As can be suspected, human motion tracking and its applications is a vast field of research. It is a quite relevant field as according to Aditya et al. the standard keyboard, mouse and other traditional input devices will become obsolete due to motion tracking sensors [29]. This means that new types of interaction between the human and the computer will emerge. This fact is important in all fields which involve motion analysis, mainly in medical applications [30]. The spatial skills of the users are important as well. Not only because well-developed spatial skills are required by most jobs, but these skills are also important during physical rehabilitation. Sadly, post-stroke patients can have difficulties seeing and they need to have a good spatial sense to successfully complete certain movements during rehabilitation. Since the world is transitioning into a digital age, these facts are needed to be investigated inside VEs.

Therefore, the aim of the author is to make the HCI easier for the user. Thus, this goal is focused on in this PhD dissertation. To achieve this, the previously mentioned two different areas of HCI were investigated during this PhD research. The first area – shown in subsection 1.1.1 – presents a short introduction and research motivation regarding the spatial skills of the users with the goal to prove that HMDs have a positive effect on the spatial skills of the users. To achieve this, three various aspects regarding the spatial skills in VEs were investigated: the influence of display parameters and display devices on the probabilities of correct answers on the tests; how the rates of correct answers are affected by the display devices; and the effect of the skills of the users on the completion times. The second area of research is presented in subsection 1.1.2. In this research, the Kinect sensor is investigated. This research has two outputs: the first is to prove that the Kinect is still useful to this day and can substitute more expensive sensors in different fields of research and the other is to present an easy-to-use "motion predictor" algorithm that can help patients in physical rehabilitation in their homes. As the research of the author is about two areas of HCI, the structure of this PhD work is shown in section 1.2.

1.1.1 The importance of spatial skills in engineering and medical applications

Spatial ability is a cognitive skill [31] which is made of the concept of definitions, originally by three: mental rotation which activates areas of the brain which involves motor stimulation [32]; spatial perception which involves the parietal lobe and the human visual system in the brain [33]; and spatial visualization. According to [34] people who used spatial visualization, the brain activities in the lateral occipital complex, in the right superior parietal, dorsolateral prefrontal cortex and in the right ventrolateral pre-frontal cortex were greater than objects visualizers. Later, Maier expanded the definition of spatial ability and therefore, of spatial intelligence [35]. According to his study, spatial ability is made of the concept of five definitions: spatial perception; visualization; mental rotation; spatial relations and spatial rotations.

A well-developed spatial ability is essential in the life of an engineer, mainly in tasks that require design and creation. Subjects such as descriptive geometry and technical representation are taught by several universities to train the spatial skills of engineering students. Due to these training courses, the spatial skills of engineering students are better than the spatial skills of non-engineering students as the engineering students outperformed the female engineering students and the non-engineering students as well [36]. It is also shown by the studies in [37, 38] that the spatial skills of males are better than of females.

Spatial ability is also essential in physical rehabilitation in the case of post-stroke patients as they can suffer from sensory deficits besides movement disabilities [39]. Due to this, post-stroke patients can lose their spatial skills as well. According to Lee [40], VR-based spatial ability training can return the spatial sense of the patients to some degree.

Therefore, it is possible to train the spatial ability of people with or without disabilities. It can be trained by doing everyday activities that require spatial skills and by solving geometric problems of which many exist. The geometric problems include the Mental Rotation Test (MRT) [41, 42, 43, 44], the Mental Cutting Test (MCT) [45, 42, 46, 47, 48, 49, 50], the Purdue Spatial Visualization Test (PSVT) [51, 42, 52], the Heinrich Spatial Visualization Test (HSVT) [53] and the Surface Development Test (SDT) [54]. It can also be trained by playing video games. According to [55, 56], the spatial skills of the users can also be improved by playing "Tetris" and first-person shooter games, even if these types of games are not developed with the goal to train spatial ability. This is an interesting conclusion, as VR is used by computer games.

This opens up the possibility that VR has the ability to enhance and train spatial ability as there is no good HCI in VR without spatial ability. With this fact, three goals are formed in this PhD dissertation regarding spatial skills in VEs. The first is to find the optimal user-centric preference in VR by using different display parameters and devices. With this, it can be investigated whether the spatial skills of the users can be enhanced by using HMDs. The second is to see how the time of doing tasks in VR is affected by the used display device and the skills of the users. Lastly, to examine and find the influence of certain skills and studies of the users on their spatial skills. Therefore, following are hypothesized:

1. The spatial skills of the users can be affected by the parameters of the VE and the used display devices.

2. The spatial ability test completion times can be affected by the used display devices and the human skills.

3. The ratio of correct answers on the tests can be affected by the used display devices and the human skills.

1.1.2 The use of motion tracking devices in medical applications

As mentioned previously in this introductory section, the aim of the author is dual with the Kinect: the first is to prove that expensive sensors can be substituted by the Kinect and that it can still be used to this day. The second is to present a "motion predictor" algorithm which allows people with movement disabilities to do physical rehabilitation exercises in their homes.

The use of motion sensors is essential in medical applications. For example, due to stroke, patients can lose their skills to move – sometimes partially, sometimes fully. An early physical rehabilitation is needed to prevent muscular dystrophy as the muscles that are not used can stiffen or even shorten. Due to these, the muscles can be locked in a wrong position. Muscular dystrophy can also happen when the muscles are not used in a long time. This condition can be prevented by physical rehabilitation as motor coordination is its focus and thus, the muscles can be strengthened by it. First, the instructions are shown to the patient by the therapist and after that, rehabilitation training can be started in the hospital. Afterwards, the training can be done at home using a (possibly low-cost) motion sensor.

Telemedicine can also be an important factor [57, 58, 59]. Nowadays, it is more important than previously, as the hospitals are overcrowded, mainly in the current situation. Therefore, rehabilitation at the home of the patients is much more convenient and safer for both the patient and the therapist. VR applications that can be used at home can help the work of medical experts as well. For the developers, it is easier to build upon existing professional applications when creating a VR one [60, 61]. Since expensive sensors do not exist in the homes of the users and cannot be taken out of the hospitals, these sensors should be substituted by less expensive ones, such as the Kinect [62, 63, 64]. It looks like the substitution is possible, but a literature review is conducted in section 2.2 in order to find the answers to the first goal [65].

Since most of the stroke patients are of the elderly generation, their needs have to be assessed when developing the applications [66]: the graphical user interface

(GUI) have to be understandable, logical, easy to navigate and the application itself has to be easy to use. There are many neurorehabilitation techniques based on VR technology that promised to help people with phobias or reduce frustration [67, 68, 69, 70, 71, 72, 73, 74], however they did not spread across the field of healthcare. The main reasons were that they were difficult to use and to customize, thereby the patient lost motivation [75].

In the past, multiple algorithms have been developed by the author and his co-authors [76, 77]. These algorithms are easy to use and serve as a basis to this research. Using these algorithms, first the information is presented to the patient by the therapist, then the algorithm is taught the gesture that have to be repeated by the patient. This allows the patient to do the physical rehabilitation exercises at home. In [76, 77] it is concluded that the acceptance domain-based gesture classification can be used in real-time with the Kinect. This could adapt to the current capabilities of the patient, while maintained their motivation in the rehabilitation process. However, it was not always accurate. Therefore, due to these reasons, the Asynchronous Prediction-Based Movement Recognition (APBMR) algorithm is proposed in the PhD dissertation of the author. Its accuracy is evaluated and it is compared to one of the previous algorithms that it is based upon.

The next movement of the user is "predicted" by the APBMR algorithm. It is done by evaluating the previous three and it is decided whether the next user-input gesture can be considered the same movement with the goal to maintain motivation. To make the decision of accepted gestures easier, the position of the user is also followed by it and their speed is also matched. Therefore, the following is hypothesized:

- The average of accepted gestures is larger with the APBMR algorithm than in the case of the previous algorithm it was based upon and it can also be used for telerehabilitation.

1.2 Structure of the thesis

In the PhD dissertation of the author two different parts of HCI were investigated while using VR/AR devices, therefore it is structured as the following: after this short introductory chapter, a literature review is conducted of both parts of the research of the author in chapter 2. The materials and methods that were used during the research are presented in chapter 3. In chapter 4 the results of the measurements are shown. The discussions and the conclusions can be seen in chapter 5 and the application of the new scientific results are presented in chapter 6.

The theses of the author are focused on in chapter 7 as they are summarized in its first section. Not only these theses are summarized, but the respective publications of the author are also assigned to them. In the second section of chapter 7, possible future research in these fields are shown. The publications of the author are presented in the last section of chapter 7, split into three groups: publications that are directly relevant to this PhD dissertation; publications that are indirectly relevant to this PhD dissertation; and publications that are not relevant to this PhD dissertation.

Chapter 2

State of the art

In this chapter the state of the art of human spatial skills in VEs and the literature regarding the Kinect sensors are assessed. Both are important in the case of healthcare, usability and even the accessibility of a VR-based application. After observing the state of the art in both cases and investigating the pros and cons of each, methods are formulated at the end of the chapter not only to solve problems regarding HCI, but to enhance it as well.

2.1 Human spatial skills and their enhancement in virtual environments

A theory is proposed by Gardner in 1983, saying that every human has multiple types of intelligence and spatial intelligence is one of them [78]. This theory was built upon by Maier and it was concluded that spatial intelligence is made up of five different parts [35]: spatial perception, visualization, mental rotation, spatial relations and spatial rotations. According to Miller and Bertoline, this ability is not a biological susceptibility, but can be improved through time [79]: improvement can occur simply by life experiences or by being exposed to certain learning environments. It has been suggested in [80] that spatial ability training should be included in the curriculum of engineering studies. According to Ghiselli, the success in the fields of engineering, mathematics and architecture is related to the spatial skills of the person [81]. However, the spatial skills of non-engineering students were measured by Sorby et al. and the improvement of spatial skills of both engineering and non-engineering students was also investigated in their study [82]. According to their results, the improvements were the same between the two groups when comparing results of the pre-tests and the post-tests. In some cases, the improvements of non-engineering students were better than the improvements of engineering students.

A considerable amount of tests was developed through the years to improve the spatial intelligence and ability of the people. These tests are the following: Cards Rotation Test and the Paper Folding Test [54], Embedded Figures Test [83], Water Level Test [84], Identical Blocks Test [85], Rod-and-frame test [86], Paper Form Board [87]. However, there are more elements in this list: the MRT, MCT, PSVT, HSVT and SDT test types were already mentioned in the introductory section, but as can be suspected, the spatial skills of the people can be improved by these tests. There is a possibility that even more of these tests exist.

Spatial ability can be improved even without solving these tests: according to the study of Hijazi, simply by fencing can increase the spatial perception of the player [88], or according to the study of Romeas and Faubert, simply by playing soccer can also increase spatial perception [89]. Mental rotation skills can also be improved by solving the Rubik's Cube [90], or by playing video games which have gameplay about rotation such as Tetris or first-person shooter video games such as Unreal Tournament [91].

As can be seen, the number of spatial ability improving methods is huge, even if only the tests are taken into account. This is due to it being an important skill to have in the modern day and age. A well-developed spatial ability is required by several jobs [92]. As such, it is an important skill for multiple disciplines, for example the engineering discipline: it is essential during design and development phases. If this ability is well-developed, the spatial relationships between objects and space can more easily be understood by the person.

The methods that improve spatial ability – such as these tests – mostly exist on paper. The Paper Folding Test has been implemented in VR in [93]. This is good as it is shown by the previous studies that the learning skills [94, 95, 96] and even the spatial skills [97, 98, 99, 100, 101, 102, 103] of people can be improved by VR. It is concluded by the last two studies that the mental rotation of males are better than of females in VR and in AR. Also, it is suggested by the last study that AR could be a good tool for improving spatial ability.

Positively affecting learning in VR became easier since the inception of CognitiveInfoCommunications (CogInfoCom) [104, 105, 106]. This is due to new human-computer interfaces [107, 108, 109, 110, 111], virtual laboratories [112] and virtual learning spaces [113]. As the user is placed in VR by the CogInfoCom environment, new possibilities of research regarding spatial ability emerge.

When the users step into VR, new factors must be considered such as the HCI and the display parameters in the VE. To help the users with HCI, a toolkit was developed by Takala which makes it easier to create VR applications using building blocks [114]. Developers using the toolkit can create applications for HMDs and even for the Kinect device. In the study the spatial GUI ideas of the students are presented and the toolkit is evaluated. According to them, the toolkit and the spatial GUIs received positive feedback.

Since most tests that train spatial ability are paper-based, the literature is scarce about VR applications that were developed with the aim to improve the spatial skills of the users. In subsection 2.1.1 existing VR applications that improve the spatial skills of the users are presented, in subsection 2.1.2 it is assessed whether the spatial skills of the users can be enhanced by the design of a VR application and in subsection 2.3.1 the literature review is concluded.

2.1.1 Existing virtual reality applications that improve spatial ability

A web-based VR application was developed by Rafi et al. which can be used to test and to improve the spatial skills of the users [115]. The mental rotation and spatial visualization tasks were used by them. In their study, a pre-test and a post-test were conducted. According to their results, the spatial skills of the users were improved by using their web-based VR application.

According to a study of Chang et al., a perspective-test was developed in VR to measure the spatial skills of the users [116]. Similarly, to the previous study, pre-tests and post-tests were conducted on three groups: users who interacted with the application with motion; users who interacted with a keyboard and a mouse; and users who interacted with motion, but used non-spatial tasks. Their conclusion was that the first two groups improved between the tests. However, significant improvements were only found in the case of the first group.

The results of two groups which done the MRT test type were compared by Jiang and Laidlaw [117]: a desktop environment was used by one group and a VR environment by the other. According to them, low spatial ability participants benefited from learning between the pre-test and the post-test. Their conclusion was that the results on the MRT test were not significantly affected by VR.

In the study of Oman et al. it was found out that the performance of users who used HMDs was slightly better than of those who did not use HMDs [118]. Their conclusion was that VR can be used for training the spatial skills of the users and thus, is excellent for this purpose. The spatial perception of users was compared by Schnabel and Kvan in their study [119]: real 2D environments, a desktop display (DD) and an HMD were used. It was tested whether the users could understand spatial volumes. When rebuilding the volume, the highest accuracy was reached by the engineering students out of all groups. If the HMD was used, the volume and its building blocks could be understood by the students. Therefore, it was concluded that with the use of the HMD, the complexity of the volume would be easier to be understood by the users.

According to Passig and Eden, that the mental rotation of children with hearing difficulties was improved by playing 3D Tetris in VR [120]. It was also mentioned that these types of games require mental rotation. This is true, as according to Johnson, five stages of mental rotation exist [121]: the creation of the mental picture; the mental rotation itself; the mental comparison between the rotated object and the original; the decision whether the two are the same object; and lastly, the acceptation. These stages happen during a game of Tetris, as the falling object has to be mentally rotated by the player.

The PSVT-R test in VR was created and evaluated by Molina et al. [99]. Two groups of users were tested with. A DD was used by one group and an HMD was used by the other one. Pre-tests and post-tests were done by both groups. In the study, the conclusion was that there is improvement in spatial skills with both groups, but the improvement is significant with the HMD.

Studies exist that only present the design of VR spatial ability tests and the research plan of the authors. An MCT test in VR was developed by Hartman et al. with the goal to help other scientist in the creation of future MCT tests [122]. A testing method was outlined by presenting the procedure and the data analysis. As the study is about the creation of VR MCT tests, the results of the tests are not published in this study. A VR MRT test outline was created by Rizzo et al. by presenting their future plan [123]. Later, their preliminary results of the tests were published [124]. According to their preliminary results, the VR MRT test helped the users as their results were improved on the post-tests.

Summarizing the studies mentioned in this subsection, it can be noted that only a few VR-based spatial ability tests were developed during the years. These tests include a web-based application, a perspective-test, MRT tests, a PSVT-R test, a spatial perception test of volumes. A study said that even by playing 3D Tetris the spatial skills of the users can be improved. In most of the mentioned studies pre-tests and post-tests were used to determine the improvements of spatial skills. It is concluded in all but one study [117] that contain pre-tests and post-tests that by using VR the spatial skills of the users can be significantly improved. According to the few tests that used an HMD, the improvements are larger when using an HMD than when using other display devices.

2.1.2 How the design of a virtual reality application can help the spatial skills of the users

The design of the VR application is important, because if the VR application is carefully designed, the spatial skills of the users can be enhanced. This is due to VR having two advantages: the interaction and the spatial visualization inside the VE are different than in reality, therefore HCI and human-computer interfaces can be redefined [125]. The spatial skills of the users can be improved by the existence of these mentioned advantages [126].

To test the importance of design in VR applications, it was investigated by Cutmore et al. whether the spatial skills of the users can be improved by a 2D projection of a room in VR [127]. In the study it was found out that the performance of the users was better in dynamic environments. Similarly, to enhance the spatial skills of the users in VR, the design of three different VEs was researched by Naceri and Chellali [128]. According to them, with the use of an HMD the users can correctly tell the distance of objects up to 55 cm in a rich VE. Similarly, VEs were assessed using three different effects and an HMD in a study of Cidota et al. [129]. These three effects were the blur effect, the fade effect and the standard environment. Although the presence of the user was decreased by the effects, the performance of the users was positively affected. In AR it was affected by the standard environment and in VR it was affected by the blurry and the faded environments.

According to Renner et al. the distances in VR applications are usually underestimated by the users [130]. In their study the conclusion was that binocular disparity should be provided, but high quality graphics in a rich VE should be used with a carefully adjusted virtual camera to enhance the sense of presence of the users. With these, the spatial skills of the users could be increased. It was also concluded by Armbrüster et al. that the distances in VEs can be underestimated by the users [131]. Better estimations are provided in peripersonal space than in extrapersonal space. According to the research, a closed room environment is preferred by the users.

A study on visual and audio cues was conducted by Rébillat et al. with a similar aim to increase the spatial perception in VR. However, according to their results, the depth perception of the users was not influenced by the audio cues as the audio source was overestimated [132]. In another study by Abdullah et al. a system is proposed to help the spatial perception of users by sending haptic feedback [133]. Their conclusion was that when using their system, less spatial mistakes are made by the users. A CAVE system which allows the freedom of movement was studied by Ng et al., but the spatial perception of the users was not enhanced by it [134].

The effect of depth cues both on a DD and an HMD was researched by Gerig et al. [135]. According to them, the impact that was made by the depth cues was minor on the performance of the users and motivational aspects should be used instead. However, it was also concluded that the performance of the users was better when using an HMD and in this case, fewer wrong moves were produced by the users. This also means that the spatial skills of the users can be enhanced by using HMDs.

To summarize, the design of the VR applications is important as the spatial skills of the users can be affected by it. According to the authors of these studies, binocular disparity should be provided with a carefully adjusted camera. Also, human perception can be positively affected by the dynamic, blurry, faded, rich environments. It can even be influenced by motivation. However, the distance of objects can only be correctly perceived up to 55 cm.

2.2 Using the Kinect for physical rehabilitation

According to Zhou and Hu the sensors used in motion tracking can be divided into three classes [136]: "non-visual tracking," "visual-tracking" and "robot-aided tracking". These three classes are defined by them as the main motion tracking sensor classes and all of these main classes have separate subclasses. Also, in their classification study a survey was conducted on the use of sensors in the field of rehabilitation. Several sensors from all three key groups were surveyed with the exception of the Kinect as it did not exist in 2008 yet. It was concluded that they are not only not patient-oriented, but home use is not allowed by them and are expensive. In contrast to the sensors that were surveyed by them, the Kinect is a low-cost sensor and can be used in a home environment if home use is allowed by the therapist.

As mentioned, the Kinect sensor did not exist in the time of writing the mentioned survey, but based on the classes, the Kinect could be classified as a marker-free visual-based sensor. According to the authors of the survey, the sensors that are placed in this class have inefficient computation, high compactness, high precision, cheaper and occlusion is their only drawback.

The Kinect arrived to the consumer market in November 2010 and was well-received by the scientists as according to Hondori and Khademi, the number of papers indexed by PubMed drastically increased from 2011 [137]. Although the Kinect was still young in 2011, its educational usefulness was quickly assessed, yielding skeptical, but still optimistic results [138]. In 2014, an analysis of AR developments was done by Bacca et al. and it was noted that demand exists for educational Kinect applications in AR. It was concluded in the study that its tracking of objects could be enhanced algorithmically [139]. According to the conclusion of Da Gama et al., motor rehabilitation is possible with the Kinect, but the tracking of its skeleton stream should be improved [140]. According to Reis et al., only the rehabilitation of upper limbs is focused on in most studies. Also, according to them, serious games are focused on more in these studies by using the Kinect to entertain and motivate the users during rehabilitation and education [141].

As the reader may suspect, the Kinect is used in many fields of research. It can be used in interactive virtual laboratories for educational purposes [142], children with special educational needs can be supported with it [143], as well as moderate cognitive impairment can be measured and enhanced using it [144] and even motivation can be increased with it [145]. It can also be used for VR therapies [146, 147], exercise gaming [148], creating gesture-controlled systems for people with disabilities [149] and even the performance of people with physical disabilities can be evaluated in medical applications [150, 151].

It has to be noted that since the Kinect is a marker-free visual-based sensor, only the motion that is in front of it can be tracked. This means that the movements have to be done in its field of view (FoV). An indoor localization study was done by Song et al. [152] to increase the FoV of the Kinect. In the study, three different Kinect sensors were connected at various angles. The user stood in front of the sensors. The monitoring of the user was started by one of the three Kinect sensors, based on their position and angle. However, the monitoring was only done by one Kinect at a given time. The Kinect which does the monitoring is selected by two methods: the bivariate gaussian probability density function and the maximum likelihood estimation methods. It was concluded in the study that more expensive sensors can be substituted by the Kinect. Another conclusion was that it is a precise sensor even though it is marker-free visual-based.

Before conclusions are made about the Kinect, it is introduced first and then its specifications are summarized in subsection 2.2.1. Afterward, it is compared to other motion sensors in subsection 2.2.2 and its precision and accuracy are assessed in subsection 2.2.3. The conclusions that are based on these facts are made in subsection 2.3.2.

2.2.1 A brief summary of the Kinect

Two versions of the Kinect exist: the Kinect v1 and the Kinect v2. Both are able to track the whole human body. Two depth cameras are used by each version of the sensors. This means two forms of depth mapping techniques are combined: focused and stereo mapping. Focused mapping means that the objects become more blurry if they move away from the depth camera. To improve the accuracy, astigmatic lenses of different focal lengths on the x and y axes are used by both versions of the Kinect [153]. With their stereo mapping, the depth from the disparity can be calculated [154]. Each of the Kinect sensors has a microphone array consisting of four microphones, in addition to the depth cameras. The microphones have a multichannel echo cancellation feature and the position of sounds can be detected as well. Noise can also be reduced and even suppressed. Apart from these, the hardware of both Kinects differ from each other. The Kinect v1 has a 64 MB DDR2 SDRAM and uses a PrimeSense PS1080-A2 chip that allows data to be processed before it is transmitted [155]. This architecture is presented in Figure 2.1. The infrared (IR) emitter of the Kinect has a laser diode of 60 mW, and works at a wavelength of 830 nm. The Kinect v2 is fitted with a Samsung K4B1G1646 G 128MB DDR3 SDRAM and a Microsoft X871141-001 chip instead of the PrimeSense PS1080-A2 [156].

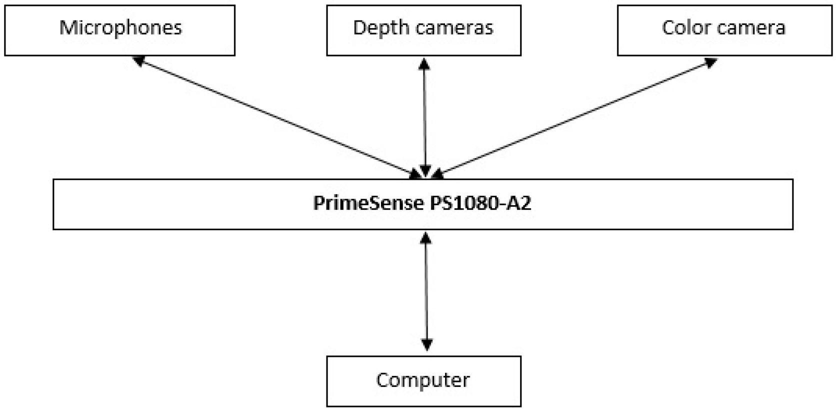

Figure 2.1: The architecture of the Kinect v1.

Both versions of the Kinect devices mainly function as depth sensors. As opposed to their hardware, both are similar in the way they use depth mapping. However, their strategies for mapping the depth are different. In case of the Kinect v1, IR dots are emitted by its IR emitter and the distance is measured based on their distortions [157]. In contrast to the first version, a Time-of-Flight (ToF) method is used by the Kinect v2. This means that distance is calculated by measuring the round trip time of the artificial light signal provided by the sensor [158].

Another way exists to track human motion with both versions of the Kinect: Software Development Kits (SDKs) were published by Microsoft for both Kinects and by using them, a so-called "skeleton stream" can be accessed in real-time. This allows the developers to track the whole body or even just certain "joints" (meaning body parts) in real-time. According to the official specifications, the number of tracked people with the skeleton stream is two in the case of the Kinect v1 and six in the case of the Kinect v2. Also, twenty joints of the user can be tracked with the Kinect v1, while twenty-five can be tracked with the Kinect v2.

Sadly, the Kinect sensors were discontinued by Microsoft in the fall of 2017 and the USB adapter for the Kinect devices was also discontinued on 01.02.2018. This is important as the sensors could not connect to a PC without the adapter. One of the reasons behind the discontinuation was that there were not enough decent games that use the Kinects. Even though the entertainment industry was not satisfied with the use of the Kinect, researchers were and the sensors are still used to this day. Nearly 30 million units have been sold over the lifetime of the Kinect sensors, with 23 million being the Kinect v1 and the remainder being the Kinect v2 [159, 160].

2.2.2 Comparing the Kinect to other motion sensors

Comparing the Kinect to similar price-range motion sensors

The Kinect v1 was compared to the Asus Xtion in a study by Gonzalez-Jorge et al. [161]. The Xtion sensor also features a depth camera [162] and therefore, functions similarly to the Kinect v1: IR dots are projected by both sensors using their depth camera which allow them to calculate the depth. This is not the only similarity between the sensors as both use the PrimeSense IR measuring unit. However, there is one difference: an external power supply is required by the Kinect v1, contrary to the Xtion. In the case of the latter, the power is received through the USB port of the computer. In this study, artefacts were measured in three different angles and the precision and accuracy of both sensors were not affected by the angles. However, an image over 7 m was not generated by the sensors. The two versions of the Kinect, the Xtion Pro Live and the Intel RealSense SR300 were compared to each other in the study of Breedon et al. [163]. The Xtion Pro Live has a better depth camera than the Kinects, although it did not offer many improvements over the two Kinect sensors. The resolution of the depth camera of the Xtion Pro Live is 640×480, while the resolution of the depth camera of the Kinect v2 is 512×424. Similarly, to both Kinects, the Xtion Pro Live was also discontinued.

In 2018, the new version of its RealSense sensor, called the D415 was released by Intel. This was compared to its predecessor and to the Kinect v2 in a study by Carfagni et al. [164]. According to the study, fewer probing form and probing size errors, less sphere spacing errors and less flatness errors are provided by its raw data than by the raw data of the Kinect v2. The conclusion of the study is that the Intel RealSense D415 can be used as a low-cost device in 3D-scanning applications and in motion tracking. Therefore, it can be used for gesture recognition as well.

According to the comparison between the similar price-range motion sensors, a better resolution is provided by the color camera of the Kinect v1 than by the color camera of the Xtion sensor. Its specified measuring distance is better as well. Similarly, a better resolution is provided by the color camera of the Kinect v2 than by the color camera of the Xtion Pro Live, while the resolution of its depth camera is worse. Its specified measuring distance is larger. Another fact is that both RealSense sensors outperform both Kinect sensors. This is because the RealSense sensors are newer than both Kinect sensors, meaning that their hardware is more advanced. A summary of these comparisons is presented in Table 2.1. In the table, SMD stands for specified measuring distance.

Table 2.1: The comparison of similar price-range motion sensors.

	Color camera	**Depth camera**	**Depth technology**	**FoV**	**SMD**	**Connectivity**
Kinect v1	1280×720 at 12 fps, 640×480 at 30 fps	320×240 at 30 fps	IR	57° H, 43° V	0.4 m or 0.8 m – 4 m	USB 2.0 or 3.0
Kinect v2	1920×1080 at 30 fps	512×424 at 30 fps	ToF	70° H, 60° V	0.5 m – 4.5 m	USB 3.0
Xtion	640×480 at 30 fps	320×240 at 30 fps	IR	58° H, 45° V	0.8 m – 3.5 m	USB 2.0
Xtion Pro Live	1280×1024 at 15 fps, 640×480 at 60 fps	640×480 at 30 fps, 320×240 at 60 fps	IR	58° H, 45° V	0.8 m – 3.5 m	USB 2.0
RealSense SR300	1920×1080 at 30 fps, 1280×720 at 60 fps	640×480 at 30 fps	Coded light	73° H, 59° V	0.3 m – 2 m	USB 3.0
RealSense D415	1920×1080 at 60 fps	1280×720 at 90fps	Stereoscopic active IR	69.4° H, 42.5° V	0.16 m – 10 m	USB 3.0 Type-C

Comparing the Kinect to more expensive motion sensors

In a study of Romero et al. [165], it was investigated whether the Polhemus Liberty Latus (PLL) wireless system [166] could be replaced with the Kinect v1 when assessing the motor skills of children with autism spectrum disorder (ASD). It was concluded in the study that the PLL can be replaced by the Kinect v1 in some respects. However, it cannot be replaced in others: with its electromagnetic (EM) field position and rotation mapping, a more accurate measurement is offered by the PLL than by the Kinect v1. Therefore, it is superior to the Kinect v1 sensor in this regard. The measuring skills of the PLL is better suited to small-scale, high-precision activities than the Kinect v1. However, on larger scales, meaning that if the patient moves multiple limbs or whole-body tracking is used, better results are provided by the Kinect v1. Also, the data of the Kinect v1 is easier to use.

According to Sun et al. [167], gesture recognition is possible by using the Kinect with its color camera and surface electromyography (sEMG) together. Both sensors were used simultaneously, fusing the data collected. The Fourier transformation and a characteristic line method were used with the Kinect, and the results were modeled on a histogram. Also, the noise had to be filtered from the video, since the color camera was used. First, a polygonal approximation was used to achieve this, afterward the Douglas-Peucker (D-P) algorithm was used. Fifty training samples and another fifty as samples of silhouettes were collected prior to the study. Two hand, four wrist, and four finger gestures were tested. There were twenty of each gesture in five separate groups. This means that minimum 100 movement descriptors were tested for each gesture. The conclusion is that the Kinect is superior to sEMG. The use of the Kinect resulted in an average of 80-90 gestures, while the use of the sEMG resulted in an average of 60-65 gestures.

A firearms training simulator was developed by Bogatinov et al. with the goal to replace existing simulators on the market that are more expensive [168]. With the use of the Flexible Action and Articulated Skeleton (FAAST) toolkit [169], the movements within the framework have been created. According to them, their firearms training simulator is better and cheaper than the MINT-PD for example, which is another military simulator [170]. They say that their system should be used with nine calibration points and the user has to stand 2.5 m away from the Kinect sensor. Contrary to this, it is more expensive than the MINT-PD as it requires a complete controlled environment to be set-up: it consists of a laser, a laser-tracker and a microphone for speech recognition. A tablet can also be used for special input.

Not only human motion tracking is possible with the Kinect sensor but it can also be used for environmental tracking. The Kinect v2 was used by Rosell-Polo et al. for outdoor agricultural applications [171]. This is interesting as Light Detection and Ranging (LiDAR) sensors that consist of Terrestrial Laser Scanners (TLS) or Mobile Terrestrial Laser Scanners (MTLS) are usually used in this area. Because of their color and depth cameras, the conclusion is that the Kinect v2 sensor is similar to both TLS and MTLS. While they are similar, there are some differences: a shorter range and a smaller FoV were provided by the Kinect v2 than the LiDARs. In the mentioned study the Kinect v2 sensor was combined with a real-time kinematic Global Navigation Satellite System (GNSS). Various FoVs with varying sampling rates were used: 5.15 Hz with a single column FoV, 0.75 Hz with partial FoV and 0.15 Hz with maximum FoV. Naturally, the best performance and output are provided by using the 5.15 Hz option, but errors may occur up to 1.5%. It is concluded that the authors of the study have created a low-cost and efficient replacement for LiDAR sensors. It is shown in research of Keightley and Bawden that the ILRIS 3D LiDAR sensor can be used to monitor the environment [172]. This allows for comparison between the Kinect v2 and a LiDAR.

According to the reviewed comparisons, more expensive sensors can be substituted by the Kinect. The sensors that can be substituted are the PLL, sEMG, MINT-PD, and LiDARs. These comparisons were mostly done during human motion tracking. However, environmental tracking was used in the case of the LiDAR sensors.

Comparing the Kinect to all mentioned motion sensors

A summary of the mentioned motion sensors is provided by Table 2.2. Four columns can be found in the table: the name of the sensor, its method of mapping, its sampling rate and its price. The PLL, the sEMG, and the ILRIS 3D are, as expected, highly expensive. There is no information available on the cost of the MINT-PD as it was designed for the military. It must also be noted that the first four sensors in Table 2.2 are discontinued. Therefore, there are only used ones available on the market, so the price could vary between sellers.

Table 2.2: Comparison of all mentioned motion sensors.

Name	Mapping	Sampling rate	Cost
Kinect v1	Depth (IR)	30 Hz	US$99.95
Kinect v2	Depth (ToF)	30 Hz	US$99.99
Xtion	Depth (IR)	30 Hz	€50
Xtion Pro Live	Depth (IR)	15 Hz	US$140
Intel RealSense SR300	Depth (Coded light)	30 Hz	€68.12
Intel RealSense D415	Depth (Stereo active IR)	90 Hz	US$149
PLL	EM field	188 Hz or 94 Hz	US$12 500–US$600 002
sEMG	Electrodes	800 Hz – 1 kHz	US$250 004
MINT-PD	Laser	No information.	Not available.
ILRIS 3D	Laser	2500 points/s	€16 000

2.2.3 Assessing the accuracy and the precision of the Kinect

The depth sensor

In a study of Wasenmüller and Stricker [173], the depth camera of the Kinect v1 was compared to the Kinect v2. As mentioned earlier, IR dots are projected by the IR emitter that can be found in the Kinect v1. Due to the distortion of the IR dots, the depth is calculated. In contrast, a ToF method is used by the Kinect v2. Similarly, IR light is projected by it into the environment. However, the depth is not measured based on the distortion of the IR dots, instead, the speed of the IR light is measured back and forth. In the mentioned study of Wasenmüller and Stricker, a set of 300 depth images of the same environment was captured with a camera. However, the sensed depth can be affected in case of both Kinects by the current temperature of the sensor. According to them, the mean depth of the Kinect v1 can decrease to less than 2 mm. In the case of the Kinect v2, the seen distance can increase to 20 mm after 16 minutes of use. However, it can decrease to 3 mm when its 5 V DC fan (labeled U40R05MS1A7-57A07A [156]) turns on. Afterward, it can increase slightly when the fan rotates with the same speed. Due to these facts, it is advised to turn on the Kinect v2 16-20 minutes before using it. The accuracy and the precision of the Kinect v1 are less when detecting depth at an increasing distance as the offset increases exponentially: the offset is below 10 mm at 0.5 m away from the sensor, however, it can be more than 40 mm at 1.8 m away. Also, a stripe pattern appears on the depth image of the Kinect v1 with the increasing distance. The number of the stripes increases with the distance as well. The precision of the Kinect v2 decreases in the case of the different distances, but the accuracy remains the same with an offset of -18 mm: the central pixels are the same, while the ones in the corners could be incorrect. If the plane is flat, the precision is greater in the case of the Kinect v1. In the case of the Kinect v2, the precision is less and flying pixels could appear if the plane is not flat or if discontinuities exist. Since the Kinect v1 does not have a ToF camera, flying pixels are not present in its case. The depth estimation of the Kinect v2 can also be affected by the environment color as black colors have 10 mm more depth value. Multipath interference also exist in the case of the Kinect v2. This means that the concave geometry is represented with bulges.

The accuracy and the precision of both Kinect versions were studied by Gonzales-Jorge et al. by measuring an artefact at multiple angles [174]. The measurements were conducted at the angles of 45°, 90° and 135°, but the accuracy and the precision were not affected by the angles. A similar conclusion was made when the Kinect was compared to the Xtion sensors in an earlier study [161]. Multiple distances were also used during the measurements: 1 m was the closest and 6 m was the farthest. According to the conclusions of the measurements, the Kinect v1 can sense up to the range of 6 m and the Kinect v2 can only sense up to 4 m. As can be seen, the range of the Kinect v1 is larger than the range of the Kinect v2. However, the first version is less accurate. In the case of both sensors, the precision worsens as the range increases. Although, the differences in precision are always larger in the case of the first version of the Kinect sensor: with the increasing range, the precision of the Kinect v1 decreases with the second order polynomial. For the decreasing precision, no mathematical model can be found for the Kinect v2, however, similar results can be achieved with an equation. The precision values of both versions of the Kinect sensor can be defined with the following two equations (2.1 and 2.2):

$$y_{Kinect1} = 2.0399Z^2 - 2.016Z + 2.0957 \tag{2.1}$$

$$y_{Kinect2} = 0.5417Z^2 - 0.885Z + 2.7083 \tag{2.2}$$

A study was conducted by Khoshelham and Elberink where it is mentioned that when measuring the depth with the Kinect v1, the error can increase quadratically to 4 cm at the range of 5 m, while the resolution of the depth decreases quadratically [175]. According to their study, the Kinect v1 has a standard deviation of approximately 15 mm in its depth accuracy. It is concluded that the depth measurements with the Kinect v1 should be done between 1-3 m.

In case of the Kinect v2, the aim of Yang et al. was to improve its depth accuracy [176]. According to the authors of the study, until the user is 3 m away from the sensor, the average depth accuracy error is less than 2 mm. The average depth accuracy error is between 2 mm and 4 mm if the user is between 3 m and 4 m away from the sensor. If the user is at least 4 m away from the sensor, the average depth accuracy error is more than 4 mm. However, this is only the case if the user stands directly in front of the Kinect v2. The average depth accuracy error increases when the user steps sideways. A study was conducted by Bragança et al. to assess how precise the Kinect sensor is [177]. A 3D scanner system was made out of four Kinect sensors by them. To assess how precise the sensor is, measuring tapes were used and their manual measurements were compared to their 3D scanner system. According to their comparison, the use of the Kinect sensors is a viable solution for lower levels of precision. Also, their conclusion was that the software is as equally important as the hardware for precision. Similar conclusions were made by Mankoff and Russo [178]. It is mentioned in their study that the actual distance to an object is less than the distance that is sensed by the Kinect.

The volume of eggs was calculated by Chan et al. using the Kinect v2 [179]. The eggs were at least 70 cm away from the sensor. In the study, the Kinect v2 sensor was placed in four different positions in multiple angles of 45° and it always faced the eggs. According to them, the best distance to measure the volumes of eggs is between 70-78 cm, but the best results are with 74 cm. Without shear parameters the deviation of volume estimation is between ± 1.74 ml and ± 3.62 ml, while with shear parameters the volume estimation is between ± 0.05 ml and ± 9.54 ml, but these depend on the size of the egg. It was concluded that the worst accuracy of the Kinect v2 was 84.7% and the best was 97.9% with an average of 93.3%.

The specifications of both Kinect sensors were studied by Kadambi et al. [180]. According to them, the depth of both Kinect sensors can depend on the environmental conditions, although these depth ranges were not tested by them in the study. Their conclusion is that Kinect v1 should be used between 0.4 m - 3.0 m or 0.8 m - 4.0 m and the Kinect v2 should be used between 0.5 m - 4.5 m.

According to the reviewed studies regarding the depth camera of both Kinect sensors, the precision and the accuracy of both decrease or change at different distances. Also, the depth seen by the Kinect sensors can be affected by multiple environmental factors. Therefore, when measuring with the Kinect sensors, these factors should be taken into account. During reviewing these studies, data has also been collected regarding both Kinect versions. These technical specifications can be seen in Table D1 in the appendix.

The skeleton stream

In the case of the Kinect v1, it was concluded by Livingston et al. that the skeleton steam can be acquired between 0.85 m - 4 m [181]. This differs from the range recommended by Microsoft which is 0.5 m - 4 m. According to the authors of the study, this range is similar in the case of the depth sensor. Naturally, data is not returned outside of these bounds. This information contradicts the bound of 6 m where the SDK was not used. However, not only the range of the sensor was measured, but the noise of the skeleton stream as well: at 1.2 m away from the sensor, the noise was 1.3 m with a standard deviation of 0.75 mm and at 3.5 m away from the sensor, the noise was 6.9 mm with a standard deviation of 5.6 mm. In the case of the Kinect v1, the average noise varies from dimension to dimension: $x =$ 4.11 mm, $y = 6.2$ mm and $z = 8.1$ mm. According to the study, the right wrist and the right hand have the most noise. The accuracy of the skeleton stream was tested as well and it averages at 5.6 mm with a standard deviation of 8.1 mm. In case of the accuracy, no difference was found between the dimensions. Another fact is that the noise can change depending on the number of people that use the Kinect v1: if it is used by one person, the error in the accuracy is 1.4 mm, while it is 1.8 mm in the case of two people. Even though officially the Kinect v1 is only able to track two people, when it was tested with three people, the error became 2.4 mm. The mean latency with one skeleton is 146 ms and with two skeletons is 234 ms, but according to the authors of the study this largely depends on the hardware of the computer and on other, simultaneously running applications.

According to Otte et al., the range of the skeleton stream of the Kinect v2 is between 0.5 m - 4.5 m [182]. This range is slightly larger than the range of the skeleton stream of the Kinect v1. While the data of Kinect v2 in the study was not smoothed, their conclusion was that adequate results are yielded by the skeleton stream of the Kinect v2. Also, the according to them, the sensor has a harder time differentiating between the feet of the user and the ground.

Upper extremity movements were measured using the skeleton stream in the study of Reither et al. [183]. Both Kinect versions were evaluated during this study. The results were compared to a Video Motion Capture (VMC) system. A fourth order Butterworth filter was used at 6 Hz to filter the data. The reliability of both Kinect versions was good, although the range of motion (RoM) was underestimated by the Kinect v1 in the case of the reaching type gestures and overestimated for the angular gestures. In the case of the Kinect v2, it performed well for the reaching type gestures, however, its performance for the side gestures was not as good. Similarly, to the Kinect v1, the angular gestures were overestimated by the Kinect v2 as well. While the ROMs that were measured were different than the VMCs, both Kinect versions measured movements very well. According to the authors, by transforming the skeleton data of the Kinect to be more similar to the data of the VMC, it could be used in medical applications.

In a study of Huber et al., the Kinect v1 was compared to a magnetic tracking system and a goniometer where joint angles were tested [184]. Frontal and side views were assessed in the study. The mean difference from the goniometer was between -4.1° and 17.8° during the gestures, and the mean difference from the 3D magnetic tracker was between -24.2° and 20.6°. Their conclusion was that the Kinect can be a reliable sensor when the shoulder joints are not occluded.

Three types of gestures are done by ten subjects in the study of Elgendi et al.: slow, medium and fast [185]. These gestures had to be done with a shift of 45° to the right, as the authors of the study wanted to make sure that the body does not interfere with the hand movements. A first order Butterworth low-pass filter was used with a cutoff frequency of 2 Hz. The goal was to reduce bodily or environmental noise from the skeleton stream. Their conclusion was that even without the low-pass filter the hand is the most reliable for detecting speed. Its lowest error rate is 9.33% without the low-pass filter and 8% with the low-pass filter.

The Extended Body-Angles Algorithm was developed with the goal to increase gesture recognition with the skeleton stream [186]. This algorithm was used in the study of Gutiérrez-López-Franca et al. in 2018 [187]. According to them, the number errors is affected by the number of measured joints: while fewer errors are produced by the global movements, more computational power is required by them. The number of errors is larger in the case of the bounded movements, as the position of the neighbouring joints can affect the ones next to them. However, the required computational power is less than in the case of the global movements.

According to the reviewed studies, the skeleton stream of both Kinect sensors can be used for gesture classification purposes. However, the noise inside the stream has to be taken into account. Different amount of noise exists between axes, the Kinect versions and the number of people in front of the sensor. While it is not an inaccurate stream on its own, multiple methods were created to improve its accuracy and to decrease its noise. However, due to it being easy to use, the skeleton stream is popular with researchers.

2.3 Conclusions on the literature review

After the assessment of the state of the art of two parts of HCI, it can be concluded that both are equally important regarding user interaction. This means that a well-designed application or a well-implemented algorithm can enhance the interaction between the human and the machine. In the following two subsections the findings are concluded and detailed goals of this research are formulated based on them.

2.3.1 Concluding on the importance of spatial skills in virtual environments

Spatial ability is a complex field as it consists of five different concepts and it is a quite new research field in VR. The paper-based tests are still used by many researchers, although some tests are created in VEs. However, according to the little VR spatial ability literature available, the spatial skills of the users can be enhanced with the help of VR and VEs.

According to the earlier, paper-based studies that are mentioned in the previous subsections, the results on the spatial ability tests are in favor of men or right-handed people. This means that on the paper-based tests, the performance of men is significantly better than of women, while the performance of right-handed people is better than of their left-handed counterparts. In case of the VR-based spatial ability tests, these facts are not focused on – yet. Within these VR studies, pre-tests and post-tests are done and the results are compared. According to their authors, there are significant improvements between pre-tests and post-tests. When using a DD or an HMD, these are the improvements that are noticeable in the literature.

Also, based on the previous studies, the design of VEs is also important when assessing the spatial skills of the users. According to the studies, binocular disparity should be provided. This is confirmed by other studies in which HMDs are used. As stereo image is used by HMDs, binocular disparity is provided to the user as well. A carefully calibrated virtual camera, rich environments, blurry or faded images also help the user. Also, motivation can have a positive effect.

Therefore, the VEs have to be carefully designed and used with an HMD to have a positive influence. In this PhD research such positive influences are planned to be found and for this goal, VR versions of the MRT, MCT and the PSVT tests were developed and the spatial skills of 240 and 61 students were measured using a DD and a Gear VR (GVR) HMD, respectively. Analyzing their results, three different types of outputs are planned in the following categories:

1. Which attributes of the VE and which used display device influence the probabilities of correct answers on the test types. With this output it can be seen which attributes and used display device have influence on each other and how they interact as well. The output is to propose new design choices for VEs to enhance the spatial skills of the users [188].

2. Which skills of the users and used device influence the completion times of the test types. With this output it can be seen which used display device and human skill increase or decrease the completion time of each test [189].

3. Which used device correlates with which human skill on all three test types. With this output, it can be concluded that which display device is more suited for VR spatial ability tests. Also, which groups of students experienced the most improvement in their spatial skills using each display device [190].

As can be seen, what is investigated in this PhD dissertation is different than the reviewed studies in the literature. Mostly pre-tests and post-tests are contained in these studies and other parts of spatial perception in VEs are focused on in them. Therefore, the mentioned three categories are investigated in this PhD dissertation by analyzing the results of students who either the used the DD or the GVR while evaluating their spatial skills with the MRT, the MCT and the PSVT tests.

2.3.2 Concluding whether the Kinect can substitute more expensive sensors

When analyzing the studies, it became apparent that both versions of the Kinect are popular. According to the number of studies available and the units sold, the conclusion is that the later Kinect v2 is not as popular as the Kinect v1. Even in late-2018, the Kinect v1 used by most researchers in their studies. It is not mentioned in the studies why the Kinect v1 is more popular than its later version. After assessing the two versions, the conclusion is that the pros of the first version of the Kinect outweigh the pros of the second version: its depth precision is higher, it has a weak correlation to the temperature and the depth estimation of the sensor is not affected by the environment color. In contrast, when using the second version of the Kinect, more attention must be paid to the test environment.

It can be concluded that the Kinects are two of the most accurate low-cost whole human body motion sensors that are available. Due to this, the two versions of the Kinect sensors are suitable in multiple fields of research such as rehabilitation, gesture recognition, education, entertainment, et cetera. Therefore, more expensive sensors in the industry or on the market can be replaced by the two Kinect sensors. Although, they are accurate sensors on the own, their accuracy can be improved by multiple software methods. Thus, the answer to the first goal of the author was found.

Even though the two versions of the Kinect can substitute more expensive sensors, two problems exist with them regarding gesture recognition: the first is that when the user stands at a different distance from the sensor than before, the (x, y, z) coordinates that are sensed by the sensor are also different. Therefore, when repeating the same gestures at another position, it may be recognized as another gesture. Another problem is the speed of the gesture: when the same gestures are done with a different speed, they may not be recognized as well.

Therefore, to solve this problem, the Asynchronous Prediction-Based Movement Recognition (APBMR) algorithm was developed [191] and was evaluated with four groups of people. The same four gestures had to be done by each of these groups at least ten times. Six different mean techniques (MTs) are used by the APBMR to decide whether the gestures could be accepted. Even though the APBMR is developed for the Kinects; in principle, it can be used with any sensor that returns movement descriptor data in real-time and evaluates gestures by using coordinates. The output of this research is to find the MT that gives the optimal gesture acceptance rate and to present an easy-to-use algorithm that can be used in the homes of the patient.

2.4 Summary of the state of the art

According to state of the art, a large number of spatial ability test types mostly exists on paper at the time of writing this PhD dissertation. However, VR-based versions of some test types are already created by a few researchers. Pre-tests and post-tests are included in their studies, while mostly using a DD. HMDs are only used in a few studies. According to the results, the use of VR can increase the spatial skills of the users and for this, the design of the VE is important as well. Thus, in this PhD dissertation, research is conducted in three different categories regarding spatial skills as mentioned in subsection 2.3.1.

Regarding both versions of the Kinect sensors, they are still used to this day. By assessing their accuracy, precision and by comparing them to other, more expensive sensors, the conclusion is that both can be used in telerehabilitation. Based on previous algorithms, a new algorithm was developed to help the users during physical rehabilitation and to retain motivation. Therefore, in this PhD dissertation, research is conducted in this regard as mentioned in subsection 2.3.2.

After conducting a literature review regarding the spatial skills of the users in VEs and the suitability of the Kinect, the following conclusion is made: both the spatial skills of the users in VEs and gesture classification are equally important in HCI. Therefore, by analyzing the design of a VR application, forming a recommendation and developing a user-centered algorithm, HCI can be enhanced.

Chapter 3

Materials and methods

In this chapter the materials and methods are presented regarding the spatial skills of the users in VEs and the APBMR to enhance HCI. The RQs and the Hs are shown in section 3.1 where four groups of RQs and Hs are presented. These are based on the planned outputs that are shown in section 2.3. The spatial skills of the users are in the focus of the first three groups, while the evaluation of the APBMR is in the focus of the fourth one. The methodology used is shown in section 3.2.

3.1 Research questions and hypotheses

During the research, four groups of RQs and Hs were formed. The first group (1RQ and 1H) is about finding the optimal user-centric VE design. The second group (2RQ and 2H) is about investigating the effects of the display device and the human skills on the completion time, while the correlation between the device used and the human skills is investigated in the third group (3RQ and 3H). In the last, fourth group (4RQ and 4H) the APBMR algorithm is assessed. These RQ and H groups are presented in subsections 3.1.1, 3.1.2, 3.1.3 and 3.1.4, respectively.

3.1.1 Finding the optimal user-centric virtual environment design

As it was mentioned in the previous sections, the first goal of the investigation was to examine if HCI could be positively or negatively affected by certain display parameters and display devices. To achieve this, the MRT, MCT and PSVT spatial ability test types were used as a starting point. In 1RQ and 1H, seven RQs and Hs were set up during the investigation, respectively. The RQs are as follows:

- 1RQ1: Is the probability of correct answers on the tests influenced by the different camera types?
- 1RQ2: Is the probability of correct answers on the tests influenced by the different camera FoVs?
- 1RQ3: Is the probability of correct answers on the tests influenced by the different camera rotations?

- 1RQ4: Is the probability of correct answers on the tests influenced by the different contrast ratios?
- 1RQ5: Is the probability of correct answers on the tests influenced by the presence of the shadows?
- 1RQ6: Is the probability of correct answers on the tests influenced by the device used?
- 1RQ7: What are the optimal preferences for these factors for achieving the largest probability of correct answers on the tests?

After the RQs were set up, the Hs were formulated based on them. Since the statistical H testing test the equality, and the alternative H is the nonequality, the following Hs are formulated:

- 1H1: The probability of correct answers is not affected by the camera type; opposite to: the probability of correct answers on the tests is positively influenced by the perspective camera type.
- 1H2: The probability of correct answers is not affected by the different FoVs; opposite to: the probability of correct answers on the tests is positively influenced by changing the camera FoV to a higher degree.
- 1H3: The probability of correct answers is not affected by the camera rotation; opposite to: the probability of correct answers on the tests is increased by changing the camera rotation.
- 1H4: The probability of correct answers is not affected by the contrast ratio; opposite to: the probability of correct answers on the tests is positively influenced by changing the contrast ratio from a higher ratio to a lower one.
- 1H5: The probability of correct answers on the tests is not affected by the presence of shadows; opposite to: the probability of correct answers on the tests is affected by the presence of shadows.
- 1H6: Using a DD or the GVR, the probabilities of correct answers are equal; opposite to: using the GVR the probability of correct answers is larger.
- 1H7: Based on the previous hypotheses, the optimal preferences are the perspective camera type, higher field of view, some rotation, lower contrast ratio while also using the GVR.

3.1.2 Investigating the effects of the display device and the human skills on the completion time

The second goal was to determine which skills of the users and the device used affect the completion times of the test types. For this goal, seven RQs were set up in 2RQ. These RQs are the following:

- 2RQ1: Are the completion times and the probabilities of correct answers independent of each other?
- 2RQ2: Are the completion times significantly affected by the gender of the user?
- 2RQ3: Are the completion times significantly affected by the primary hand of the user?
- 2RQ4: Are the completion times significantly affected by the type of the test?
- 2RQ5: Are the completion times significantly affected by the device used?
- 2RQ6: Which combination of the mentioned factors results in the largest test completion time?
- 2RQ7: Which combination of the mentioned factors results in the smallest test completion time?

Afterwards, the same number of Hs was formulated in 2H. The null-Hs and the alternatives are contained in these seven set-up Hs. These Hs are the following:

- 2H1: The completion times and the probabilities of the correct answers are independent of each other; opposite to: the completion times and the probabilities of the correct answers are dependent.
- 2H2: The test completion times are not affected significantly by the gender of the user; opposite to: the test completion times are significantly affected by the gender of the user.
- 2H3: The test completion times are not affected significantly by the primary hand of the user; opposite to: the test completion times are significantly affected by the primary hand of the user.
- 2H4: The test completion times are not affected significantly by the type of the test; opposite to: the test completion times are significantly affected by the type of the test.
- 2H5: The test completion times are not affected significantly by the device used; opposite to: the test completion times are significantly affected by the device used.
- 2H6: Based on the previous hypotheses, the same average completion time is required by all users and all test types; opposite to: significantly larger completion times are yielded by the male users, who are right-handed and do the MRT test type with the GVR.
- 2H7: Also based on the previous hypotheses; the same average completion time is required by all users and all test types; opposite to: significantly smaller completion times by the female users, who are left-handed and do the PSVT test type with the DD.

3.1.3 Looking for the correlation between the device used and the human skills

The third goal was to see whether correlation could be found between the device used and the human skills on all three test types. Therefore, 11 RQs and the same number of Hs were formulated in the 3RQ and 3H groups, respectively. The first five are about the tests on the DD, the next five are about the tests on the GVR HMD, and the last one is about comparing the display devices. The RQs regarding this goal are the following:

- 3RQ1: Which test mode is the easiest and the hardest when using a DD?
- 3RQ2: Is there any difference between the male and female performances using a DD?
- 3RQ3: Are the results on the tests influenced by the primary hand of the user when using a DD?
- 3RQ4: Are the results on the tests affected by the age of the user when using a DD?
- 3RQ5: Are the results on the tests affected by the major of the user when using a DD?
- 3RQ6: Which test mode is the easiest and the hardest with the GVR?
- 3RQ7: Is there any difference between the male and female performances on the tests with the GVR?
- 3RQ8: Are the results on the tests influenced by the primary hand of the user when using the GVR?
- 3RQ9: Are the results on the tests affected by the age of the user when using the GVR?
- 3RQ10: Are the results on the tests affected by the major of the user when using the GVR?
- 3RQ11: With which device are better results achieved on the tests?

After the RQs were set up, the same number of Hs was formulated. Similarly, to before, the null-Hs and the alternatives are contained in these Hs. The Hs are the following:

- 3H1: In the case of DD, the average rates of correct answers are the same in case of all types of tests; opposite to: they depend on the test type. In case of different values of average rates, are they the same with both display devices? Do the statistical evaluations reflect the subjective opinions of students: "The MCT mode is the hardest and the PSVT mode is the easiest when using a DD".

- 3H2: The performances of males and females are equal; opposite to: the performance of males is better on the tests when using a DD.
- 3H3: The performances of left-handed and right-handed people are equal; opposite to: the performance of left-handed people is better on the tests when using a DD.
- 3H4: The performance of older people is equal to younger people; opposite to: the performance of older people is better on the tests when using a DD.
- 3H5: The performances of the students with different major are equal; opposite to: they differ when using a DD.
- 3H6: The average rates of correct answers are the same in case of all types of test; opposite to: they depend on the test type when using the GVR. Do the statistical evaluations reflect the subjective opinions of students: "The MCT mode is the hardest and the PSVT mode is the easiest when using the GVR".
- 3H7: The performances of males and females are equal; opposite to: the performance of males is better on the tests when using the GVR.
- 3H8: The performances of left-handed and right-handed people are equal; opposite to: the performance of left-handed people is better on the tests with the GVR.
- 3H9: The performances of older people are equal to younger people; opposite to: the performance of older people is better on the tests with the GVR.
- 3H10: The performances of the students with different major are equal; opposite to: they differ when using the GVR.
- 3H11: The average rates of correct answers are equal if the DD and GVR are used; opposite to: better results are achieved by the users who use the GVR.

3.1.4 Assessing the APBMR algorithm

Afterward, the APBMR algorithm is evaluated using different perspectives. Therefore, eight RQs and eight Hs were formed in the fourth RQ and H group (4RQ and 4H). First, the RQs are defined:

- 4RQ1: Are the best average acceptance rates given by the same MTs in case of each gesture?
- 4RQ2: Are the worst average acceptance rates given by the same MTs in case of each gesture?
- 4RQ3: Is there a MT that gives an optimal average acceptance rate for gesture recognition using predictive motion analysis?
- 4RQ4: Is there a MT that should not be used for gesture recognition using predictive motion analysis?

- 4RQ5: Is there a difference between the average acceptance rates of different gestures?
- 4RQ6: Is there a difference between the average acceptance rates on different axes?
- 4RQ7: Is the file-based gesture prediction faster than real-time gesture prediction when using the APBMR algorithm?
- 4RQ8: Is a better average acceptance rate provided by the APBMR algorithm than the previous algorithm that it was based on?

The next to define are the Hs, which are the following:

- 4H1: The best average acceptance rate is given by the same MT in case of each gesture; opposite to: the best average acceptance rate is given by different MTs in case of each gesture.
- 4H2: The worst average acceptance rate is given by the same MT in case of each gesture; opposite to: the worst average acceptance rate is given by different MTs in case of each gesture.
- 4H3: There is a MT that gives an optimal average acceptance rate for gesture recognition using predictive motion analysis; opposite to: there is no MT that gives an optimal average acceptance rate for gesture recognition using predictive motion analysis.
- 4H4: The Geometric and the Cubic MTs should not be used for gesture recognition using predictive motion analysis; opposite to: the Geometric and the Cubic MTs should be used for gesture recognition using predictive motion analysis.
- 4H5: There are differences between the average acceptance rates of different gestures; opposite to: there are no difference between the average acceptance rates of different gestures.
- 4H6: There are differences between the average acceptance rates on different axes; opposite to: there is no difference between the average acceptance rates on different axes.
- 4H7: The file-based gesture prediction is faster than real-time gesture prediction; opposite to: real-time gesture prediction is faster than file-based gesture prediction.
- 4H8: A better average acceptance rate is provided by the APBMR algorithm than the previous algorithm that it was based on; opposite to: a better average acceptance rate is not provided by the APBMR algorithm than the previous algorithm that it was based on.

3.2 Methodology

After the RQs and the Hs were formulated, the methodology was created to provide answers to the Hs. The methodology is presented in the following subsections. Regarding the spatial skills of the users in VEs, the methodology is presented in detail in subsection 3.2.1, while regarding the APBMR algorithm it is presented in subsection 3.2.2.

3.2.1 Presenting the methodology for the spatial ability tests and measurements

The methodology regarding the spatial skills of the users in VEs consists of multiple parts and these are detailed in the following subsubsections. First, an application was developed for two platforms in which the MRT, MCT and PSVT test types are contained. Then, the colors had to be converted to the RGB color space as the sRGB color space is used by the Unity engine. After the conversion, the data could be collected and analyzed.

Summary of the developed application

Before the research commenced, the application for the measurements was designed and was developed in the first half of 2019 [192]. The application was programmed using the Unity game engine [193] with version 2018.3.14f and Visual Studio was used alongside Unity to write the codes in the C# programming language. As it is suggested by the "game engine" phrase, Unity is mainly used for game development. However, it is quite popular in the scientific community [194, 195, 196, 197, 198].

The spatial skills of the users are measured by the application as the aforementioned three types of spatial ability tests are in the focus of this research. Therefore, the three test types have been implemented inside the application, each type having ten different rounds of problems. The spatial ability problems and solutions that were implemented were based on existing paper-based ones. The following three examples of the tests are presented in the following figures: the MRT, the MCT and the PSVT test types are shown in Figures 3.1, 3.2 and 3.3, respectively. Their paper-based versions are taken from [199, 45, 51].

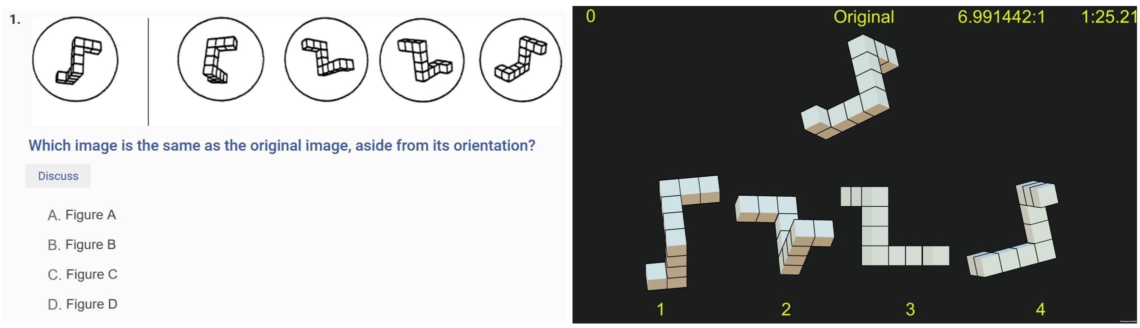

Figure 3.1: The paper-based MRT test (left) and its VR version with an orthographic camera, 7:1 contrast ratio, shadows turned on and no extra rotation (right).

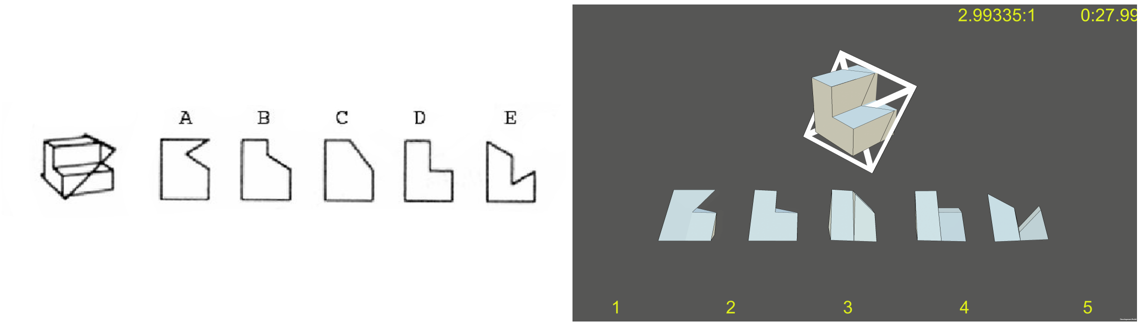

Figure 3.2: The paper-based MCT test (left) and its VR version with a perspective camera, a 60° FoV, 3:1 contrast ratio, shadows turned on and no extra rotation (right).

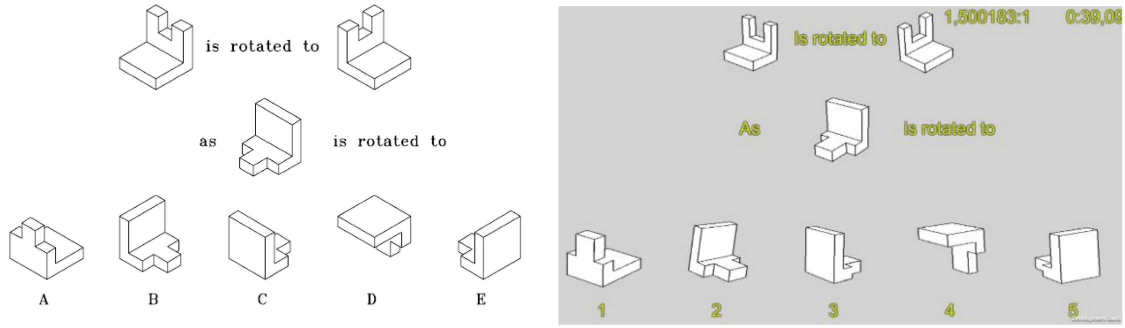

Figure 3.3: The paper-based PSVT test (left) and its VR version with a perspective camera, a 45° FoV, 1.5:1 contrast ratio, without shadows and extra rotation (right).

As can be seen in the previous figures (3.1 and 3.2), the tests have different display parameters: the virtual camera type, its FoV, its rotation and the contrast ratio can be changed, although the rotation was not changed in the figures. The existence of shadows can also be turned off or turned on. By turning off the shadows, the users get the white colored objects that are present in Figure 3.3. It should be noted that these white objects, a white background and an orthographic camera is the "standard" that is used in the case of the paper-based tests.

These display parameters could be changed by the users if they wanted to customize their tests. However, predefined options also exist in the application and during the testing, these predefined options were the ones that were used. The display parameters that can be selected in the main menu of the application are presented in Table 3.1. Naturally, these were also the display parameters that were measured by the application.

Since the goal was to investigate and compare the results on a DD and the GVR HMD, two versions of the application were developed. One is a desktop version which can be used with a Windows 7 or newer operating system and the other one is a GVR version which uses Android. Multiple versions of the GVR exist, but the SM-R322 version was used for the tests [200]. For Android, version 7.0 is required at least. Also, a smartphone had to be placed inside the GVR, and a Samsung Galaxy S6 Edge+ [201] was used for this purpose.

Table 3.1: The display parameters that can be selected.

Virtual camera type	Perspective, Orthographic
Virtual camera FoV	45°, 60°, 75°, 90°
Rotation of objects to the camera	-45°, -30°, -15°, 0°, 15°, 30°, 45°
Object to background contrast ratio	1.5:1, 3:1, 7:1, 14:1, 21:1
Shadows in the scene	Turned off, Turned on

Gathering the color data in Unity - Conversion to RGB color space

It is possible to obtain the color data of objects in Unity, but as mentioned, the sRGB color space is used by it. It is not a problem by default, but it has to be converted to the RGB color space to calculate the relative luminance values of the object and the background. The contrast ratio can be calculated with the relative luminance values. Before starting the conversion, the sRGB and RGB color spaces should be defined. These definitions can be seen in equations (3.1 and 3.2):

$$R_{sRGB} \in [0;1], G_{sRGB} \in [0;1], \mathrm{B}_{sRGB} \in [0;1] \tag{3.1}$$

$$R_{RGB} \in [0;1], G_{RGB} \in [0;1], \mathrm{B}_{RGB} \in [0;1] \tag{3.2}$$

The reader may notice that all color values are between 0 and 1, but usually they are between 0 and 255 as the hexadecimal notation is used. This is because float variable types are used by Unity to store the color values. These float variables can be converted to the standard hexadecimal notation, however this results in extra computational steps because the same results are yielded by using the hexadecimal notation as in the case of the float variables.

First, the albedo color of an object in Unity was determined using the *GameObject[i].GetComponent<Renderer>().material.GetColor("_ Color")* function, where i is the current round of the measurements. The albedo is a diffuse color without any lighting baked into it: in the case of the objects it is white. By looking at Figures 3.1 and 3.2 it can be observed that the color values are not exactly white. This is because there is ambient lighting in a scene in Unity and lighting is not baked into the objects in the case of the albedo color. Therefore, the returned albedo color values had to be corrected by using the following transformation as shown in equation (3.3):

$$w_{corr} = w_{obj} \times w_{al} \times Intensity_{al} \tag{3.3}$$

, where w_{obj} is the albedo color of the object, w_{corr} is its corrected value. The color value of the ambient light is marked with w_{al} and its intensity is $Intensity_{al}$. These two values can be returned easily with the *RenderSettings.ambientLight* and *RenderSettings.ambientIntensity* built-in functions.

After the color correction, the only step remaining was the conversion to the RGB color space. For this conversion, a new variable called q was defined. The R, G, B values are contained in this variable. This variable is similar to w as it contained the sR, sG and sB values. To convert from sRGB color space to the RGB color space, the following equation (3.4) that is defined in the IEC 61966-2-1:1999 standard [202] was used:

$$q = \begin{cases} \frac{w_{\text{corr}}}{12.92} & w_{\text{corr}} \leq 0.0405 \\ \left(\dfrac{w_{\text{corr}} + 0.055}{1.055}\right)^{2.4} & \text{otherwise} \end{cases} \tag{3.4}$$

After this conversion, the correct data of color values could be gathered. With this, the relative luminance values and the contrast ratio can be calculated. By using this conversion not only the color spaces are changed, but the color values are not gamma-compressed anymore as they are converted to gamma-expanded (or linear) color values. It should be noted that equation (3.4) was used because an older version of this equation exists as well [203], but it contains small rounding errors. This means that 0.03928 was used previously instead of the 0.0405 constant.

Calculating the correct values of the contrast ratio in Unity

Since the q has been calculated, it is possible to receive the contrast ratio between the foreground object and the background, However, the relative luminance values have to be calculated as it is essential to get the contrast ratio. This is possible by using a formula in which the luminosity function is reflected. This formula is defined in [203] and shown in the following equation (3.5):

$$L = 0.2126R + 0.7152G + 0.07272B \tag{3.5}$$

Afterward, the contrast ratio can be measured with the help of the relative luminance values as shown in [204] and in equation (3.6):

$$contrast = \frac{L_{\text{foreground}} + 0.05}{L_{\text{background}} + 0.05} \tag{3.6}$$

Using these formulas, the contrast ratio can be extracted correctly from the application, and thus, be measured.

Data collection

The testing and the data collection were conducted at two different universities. The first part of the testing was at the University of Pannonia in early-September 2019. In these tests, the GVR was used and the spatial ability of 61 students was measured. These testers were mainly consisted of IT students, but the spatial skills of other, non-IT students were measured as well. Non-IT students include chemical engineer, business and arts students; some were in their bachelor studies and some were in their master years. This means that the students who came to the tests were 23.5 years old on average with a dispersion of 3.1 years. The time of the tests was three weeks long as only one GVR was available at the University. This means that the testers had to come in a sequential order, one-by-one. Each tester required at least thirty minutes and an hour at most to complete the tests. Therefore, the skills of eight students were measured at most per day and the smallest number of testers per day was two. As they were students, their appointments were made according to their classes so they could come to the tests before or after their classes.

The second part of the tests was conducted at the University of Debrecen in late-September 2019. These tests were different than the ones at the University of Pannonia. First of all, they were greater in numbers as the spatial skills of 240 students were measured. Also, instead of the GVR, an LG 20M37A (19.5") DD device [205] was used by the students who either were architectural and social engineering students or mechanical engineering students. The skills of these students were measured mostly in their first years, thus, these students were 19.7 years old on average with a dispersion of 1.5 years. For the tests, a computer laboratory was used. Due to it being small, twelve groups of twenty students were made. Testing was done during the course of three weekdays and thus, were done within a week.

Regarding the students at the two universities, every person who was willing to do the tests could join. This means that there was no selection criteria applied. Also, since the spatial skills of the students were measured, no information was gathered of their height and body weight. To respect their anonymity, their names were not gathered.

During the measurements each test type had to be done three times: at the first sequence of testing, the MRT test type was the first to be done by the students, then the MCT test type was the second to be done and lastly the PSVT test type was done. As each test type consisted of ten rounds, this first test consisted of thirty questions in total. Then, the users could rest – if they wanted to – and after that they started the second sequence of testing. The second sequence consisted of the same test types, but the solutions to the spatial ability problems and the display parameters were changed using the randomization technique. This sequence also had thirty questions in total. After completing this second sequence, the students could rest again if they wanted to and afterward, the third – and final – sequence could be started. Similarly, to the second one, the solutions to the problems and the display parameters were changed by the randomization technique.

However, there was a reason to why the students had to do the tests in three different sequences. There were many parameters to test and testing three times were not enough to measure the spatial skills of the students when using these parameters. A huge amount of data was needed to investigate the influences and interactions of the display parameters and the devices. Therefore, the mentioned randomization technique was used. Using this technique, sufficient amount of data was gathered.

Technical information of the display parameters, the user-related information and information about the tests can be found in this data. These are saved into a .csv file by the application after a test type was completed. These are the following:

- Information about the display parameters, which are the virtual camera type, its FoV, its rotation and the contrast ratio in the scene. Also, it is logged whether the shadows are turned off.

- Information about the users which consists of their gender, age, and primary hand. The number of years spent at a university and their studies are gathered as well.

- Information about the type of the test, its completion time. The number of correct and incorrect answers of the user is also logged.

Due to this, a considerable amount of data is generated and saved either into the file of the DD or the GVR. To better visualize the data, its structure is shown in Table 3.2. In the mentioned table, the data of a user can be seen after completing a test type. As shown in Table 3.2, the data of one test is very descriptive and respects anonymity. However, the data does not contain the device used, instead it is stored in the name of the file: when using the GVR for the tests, the measurements are saved into a file called data_gearvr.csv and in case of the DD, it is saved into a file called data_desktop.csv. Thus, all information that is required for the three types of investigation can be seen in the data:

- The influence of the display parameters and display devices on the probabilities of correct answers to find the optimal user-centric preference.
- The effect of parameters and display devices on the test completion times.
- How the rates of correct answers are affected by the human factors and the display devices?

Table 3.2: Data structure of a user after completing a test type.

ID	3
Age	18
Gender	M
Hand	Right-handed
Majoring in	Architectural Engineering
University years	1st year
Camera type	Perspective
Camera field of view	45°
Camera rotation	30°
Contrast ratio	7:1
Shadows on or off	1
Test type	MRT
Test time	5:47.98
Correct answers - Round 1	2
Correct answers - Round 2	2
Correct answers - Round 3	2
Correct answers - Round 4	2
Correct answers - Round 5	0
Correct answers - Round 6	2
Correct answers - Round 7	1
Correct answers - Round 8	2
Correct answers - Round 9	2
Correct answers - Round 10	2
Total correct answers	17
Total incorrect answers	3

Data analysis of the spatial ability test results

After gathering sufficient amount of data, it was analyzed. Depending on the goal of the research, different analysis methods were used. When analyzing the influence of display parameters and display devices to find the optimal user-centric preference, logistic regression analysis was used [206]. When investigating the effects of parameters on the completion times, logistic regression and linear regression methods were used [207]. In the case of the correlation between the device used and the human skills, F-tests, t-tests or Welch-test was used [208]. These are presented in the following subsubsections.

Analyzing the influence of display parameters and display devices to find the optimal user-centric preference

The results of the measurements were grouped into groups of ten. In these groups of ten, nine results were of the same person. The number of display parameters and display devices were in another group, but their numbers are fixed. Due to this, the number of relative frequencies in the investigation was $240 \times 9 + 61 \times 9 = 2709$. In the knowledge of the values of the parameters, the probability of correct answers on the spatial ability tests can be estimated as the aim of this research was to clarify the influence of the display parameters and the display devices on the mentioned probabilities. To calculate the probabilities and verify the influence of the display parameters and the display devices, the logistic regression analysis was used.

Logistic regression analysis is a well-developed statistical method for detecting the effects of in themselves, additively or by taking their interactions into account. The probabilities are transformed into the interval by a monotone increase and invertible transformation $(-\infty, \infty)$, and linear regression models are fitted to the transformed values. The estimated coefficients of the variables are checked as to whether they can be interpreted as zero (no effect), or whether they vary significantly from zero (an effect exists). The direction of the effect can be revealed by the sign of the estimated value, for example it can tell whether there is an improvement or waste in the probabilities. The influences of the variables were analyzed one-by-one, in pairs, in triplets and also in a quartet. The numerical calculations were performed by statistical software called R [209].

The effects of parameters on the completion times

The goal was to identify the effects of certain factors on the test completion times. These factors are the gender of the user, the primary hand of the user, the test type and the used device. The connection between the probability of correct answers and the completion times of the test was also looked into. The logistic regression analysis method was used to determine the connection between the probabilities and the completion times. To examine the influence of the variables on the completion times, linear regression analysis methods were used. To help with the calculations for this goal, the statistical program package R was used in this investigation.

The correlation between the device used and the human skills

The ratios of the correct answers of the students were studied. The Hs of standard (Gauss) distribution was tested by Kolmogorov-Smirnov tests. Afterward, the equality of standard deviations (dispersions) and the expectations was checked. In the cases of the dispersions, F-tests and t-tests or Welch-test were applied to check the equality of expectations. Similarly, to the previous calculations, the statistical program package R was used in this investigation as well.

3.2.2 Presenting the APBMR algorithm and the methodology used during its evaluation

In this subsection the methodology regarding the evaluation of the APBMR algorithm is presented. In the first subsubsection the algorithm is presented in more detail than it was presented in [191] and its pros and cons are assessed as well. After defining the APBMR algorithm in the first subsubsection, the data collection and analysis are presented in the subsubsection afterwards.

How the algorithm works

Before the steps of the algorithm are defined, two things should be noted. The first is that only one axis is evaluated by the algorithm at a given time. After the evaluation of one axis is done, it is started on another. When all three are evaluated, then it is concluded whether the gesture is accepted. The second is that repeating gestures are looked for by the algorithm in the movement descriptors. The algorithm is built this way because repeating gestures are used in physical rehabilitation. Thus, from the starting coordinate the farthest and the closest coordinates are searched for on the axis that is currently evaluated. This is illustrated in Figure 3.4.

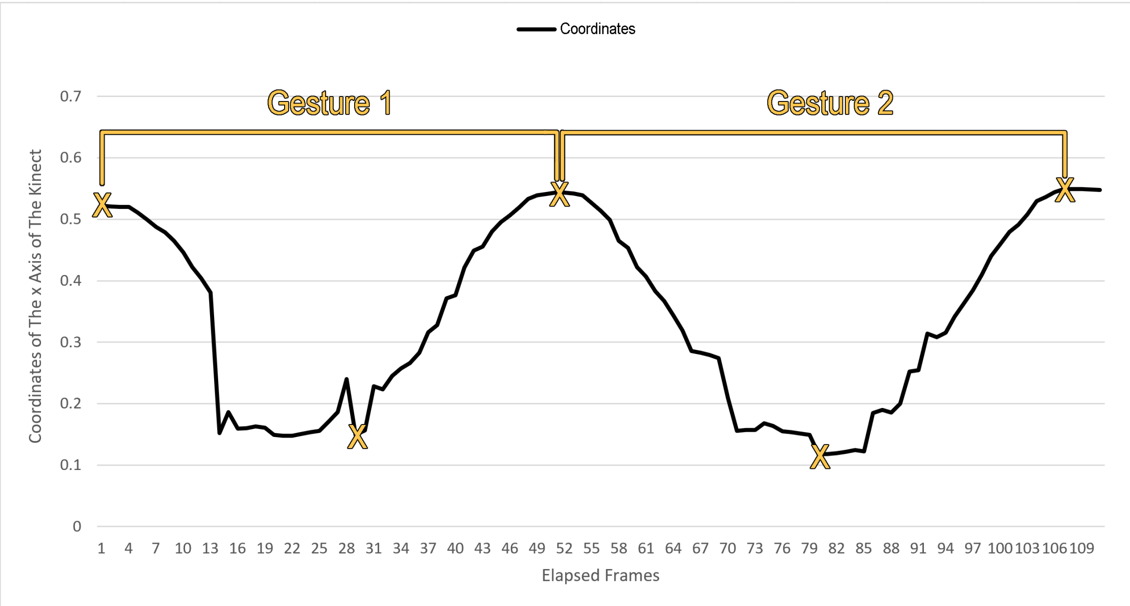

Figure 3.4: Calculating the number and the length of the gestures using the APBMR.

For example, the starting coordinate of the first gesture is denoted by the leftmost X in the figure. It is decided by the algorithm whether the starting coordinate is at the top or at the bottom of a "slope". Afterward, the farthest coordinate from the starting coordinate is searched for (shown with the second X): in Figure 3.4 the starting coordinate is at 0.5220696 and the farthest is at 0.1476125. In this example, the starting coordinate is at a top of a "slope", therefore the farthest coordinate is searched for at the bottom. The gesture is about halfway done when this coordinate is reached. Then, the coordinate which is the closest to the starting coordinate is investigated. This is shown with the third X in the figure. In this example, the coordinate is at 0.5437541. This coordinate is not the closest numerically, as finding this coordinate has a few criteria: when the starting coordinate is at the top of a "slope", then the closest coordinate also has to be at the top of another "slope". Naturally, when the starting coordinate is at the bottom of a "slope", then the closest coordinate has to be at the bottom of another "slope" as well. Another criterion is that the closest coordinate must be located after the previous farthest coordinate. When the closest coordinate is reached, then it can be concluded that the end of the first gesture is also reached. Then, this closest coordinate is renamed to the new starting coordinate of the next gesture. Afterward, these steps are repeated.

The steps of the algorithm are the following:

1. First, the number of gestures is calculated by searching for the closest and farthest coordinates from the starting coordinate. These points are referred to as $clofarpoint$.

2. Then, the average length of the found gestures is calculated.

3. Afterward, based on the last three gestures, the possible next movement of the user is predicted on the x axis and its acceptance domains (ADs) are calculated using MTs. This step consists of multiple substeps:

   (a) While $i < clofarpointnumber - 6$, the length of the previous three movements (x_1, x_2, x_3) are calculated using equations (3.7, 3.8 and 3.9):

   $$x_1 = clofarpoint_{i+2} - clofarpoint_i \tag{3.7}$$

   $$x_2 = clofarpoint_{i+4} - clofarpoint_{i+2} \tag{3.8}$$

   $$x_3 = clofarpoint_{i+6} - clofarpoint_{i+4} \tag{3.9}$$

   , where $clofarpointnumber$ equals to the number of all $clofarpoint$ in the movement descriptors. In each cycle, variable i is incremented by 2.

   (b) An average length is calculated from the length of the previous three gestures (x_1, x_2, x_3) by using one of the following operations (mt_k), which can be seen in equation (3.10). These operations are based on various MTs and can be selected by the user ($k \in [1, 6]$). mt_1 is based on the arithmetic mean, mt_2 is based on the geometric mean, mt_3 is based on the special case of harmonic mean in case of three numbers, mt_4 is based on the contraharmonic mean, mt_5 is based on the quadratic mean and mt_6 is based on the cubic mean. The use of the special case of harmonic mean was essential, because the use of the regular harmonic mean resulted in "Not a Number" (NaN) during prediction.

$$
\begin{aligned}
mt_1 &= \frac{1}{3}\sum_{l=1}^{3} x_l, & mt_2 &= \sqrt[3]{\prod_{l=1}^{3} x_l}, \\
mt_3 &= \frac{3x_1x_2x_3}{x_1x_2 + x_1x_3 + x_2x_3}, & mt_4 &= \frac{x_1^2 + x_2^2 + x_3^2}{x_1 + x_2 + x_3}, \\
mt_5 &= \sqrt{\frac{1}{3}\sum_{l=1}^{3} x_l^2}, & mt_6 &= \sqrt[3]{\frac{1}{3}\sum_{l=1}^{3} x_l^3}
\end{aligned}
\tag{3.10}
$$

(c) Afterward, a new coordinate named $predictedc_j$ is generated at every frame j, while $j < mt_k$ and $clofarpoint_{i+4} + j + 1 < x_i$. For this, one of the mentioned operations that are based on the MTs and the following rules are used as shown in equations (3.11 and 3.12):

$$
\begin{aligned}
y_1 &= \begin{cases} c_{clofarpoint_i + j}, & \dfrac{x_i - j}{mt_k} \geq \dfrac{mt_k - j}{mt_k} \\ \dfrac{c_{clofarpoint_i + j} + c_{clofarpoint_i + j + 1}}{2}, & \text{otherwise} \end{cases} \\
y_2 &= \begin{cases} c_{clofarpoint_{i+2} + j}, & \dfrac{x_i - j}{mt_k} \geq \dfrac{mt_k - j}{mt_k} \\ \dfrac{c_{clofarpoint_{i+2} + j} + c_{clofarpoint_{i+2} + j + 1}}{2}, & \text{otherwise} \end{cases} \\
y_3 &= \begin{cases} c_{clofarpoint_{i+4} + j}, & \dfrac{x_i - j}{mt_k} \geq \dfrac{mt_k - j}{mt_k} \\ \dfrac{c_{clofarpoint_{i+4} + j} + c_{clofarpoint_{i+4} + j + 1}}{2}, & \text{otherwise} \end{cases}
\end{aligned}
\tag{3.11}
$$

$$
\text{predictedc}_j = \begin{cases}
\dfrac{1}{3}\displaystyle\sum_{l=1}^{3} y_l, & k = 1 \\
\sqrt[3]{\displaystyle\prod_{l=1}^{3} y_l}, & k = 2 \\
\dfrac{3y_1y_2y_3}{y_1y_2 + y_1y_3 + y_2y_3}, & k = 3 \\
\dfrac{y_1^2 + y_2^2 + y_3^2}{y_1 + y_2 + y_3}, & k = 4 \\
\sqrt{\dfrac{1}{3}\displaystyle\sum_{l=1}^{3} y_l^2}, & k = 5 \\
\sqrt[3]{\dfrac{1}{3}\displaystyle\sum_{l=1}^{3} y_l^3}, & k = 6
\end{cases}
\tag{3.12}
$$

, where the coordinates of the previous gestures are contained by c.

(d) Then, for each $predictedc_j$ coordinate, three ADs are created. A very strict AD ($predictedc_j \pm 0.05$ m), a medium strict AD ($predictedc_j \pm 0.10$ m) and a least strict AD ($predictedc_j \pm 0.15$ m) are created.

(e) With the previous step, the evaluation on one axis is completed. Afterward, the evaluation is started on one of the remaining axes (y or z). The previous steps are repeated until all axes are evaluated.

(f) Afterward, the percentage of the coordinates is calculated inside all three ADs on all axes. It is also evaluated whether the gesture is accepted.

(g) Lastly, the APBMR waits until another gesture is done by the user. Then, the earliest gesture descriptor is pulled from the stack and the prediction is started over again with the remaining descriptors in the stack.

(h) The algorithm runs until it is turned off.

The sequence diagram of the APBMR algorithm is presented in Figure 3.5. In the figure, the interaction between the user and the algorithm can be seen.

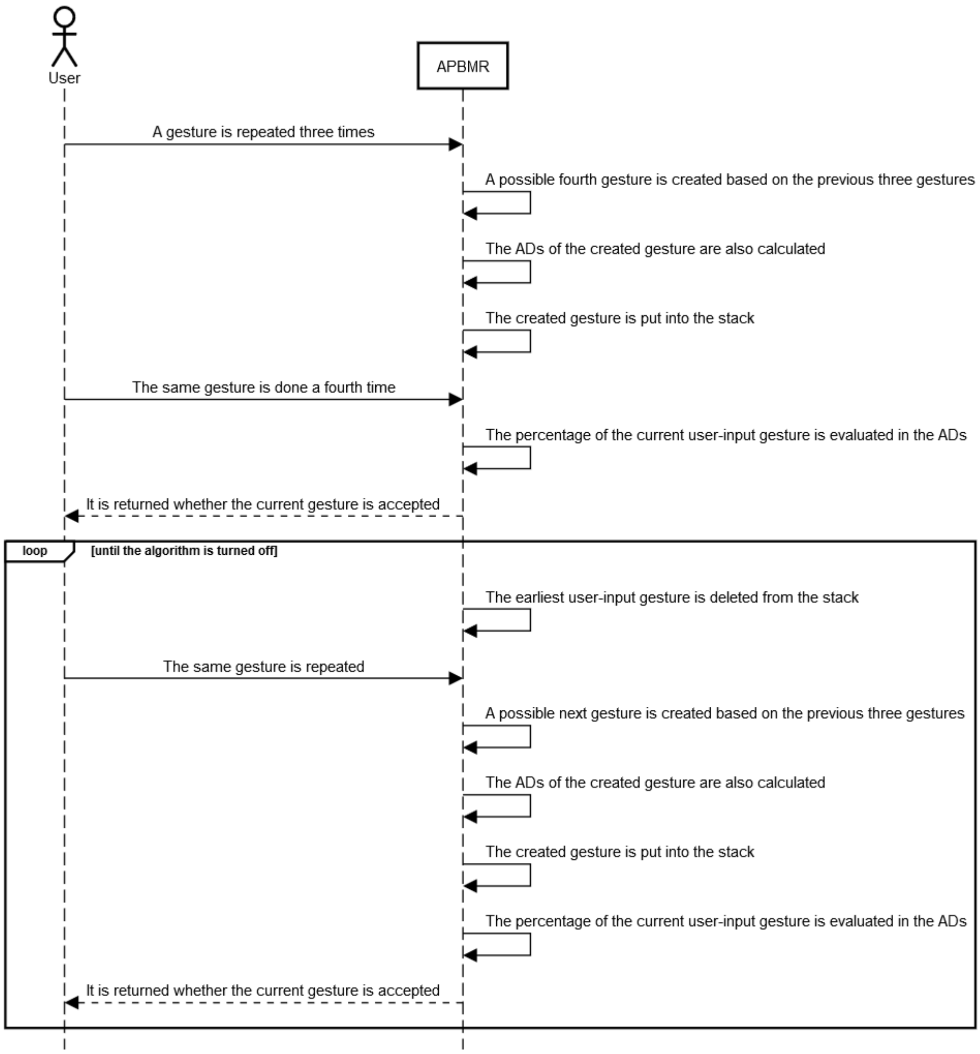

Figure 3.5: Sequence diagram of the APBMR algorithm.

According to substep 3/e, not only the next gesture is predicted, but its multiple ADs as well. In Figure 3.6 it is shown how the algorithm looks like with its ADs on the x axis using the arithmetic MT. In the figure, the ADs are represented by six thin blue lines, while the original gesture is shown with a black line and the predicted one with an orange line.

This algorithm has three strengths: the first is that to calculate the next movement and to create its ADs, six various MTs are used. The results of the predictions with each MT are presented in Figure 3.7, where the original movement descriptors are shown with a black line, while the predicted ones are shown with different colors. The ADs are not shown in Figure 3.7.

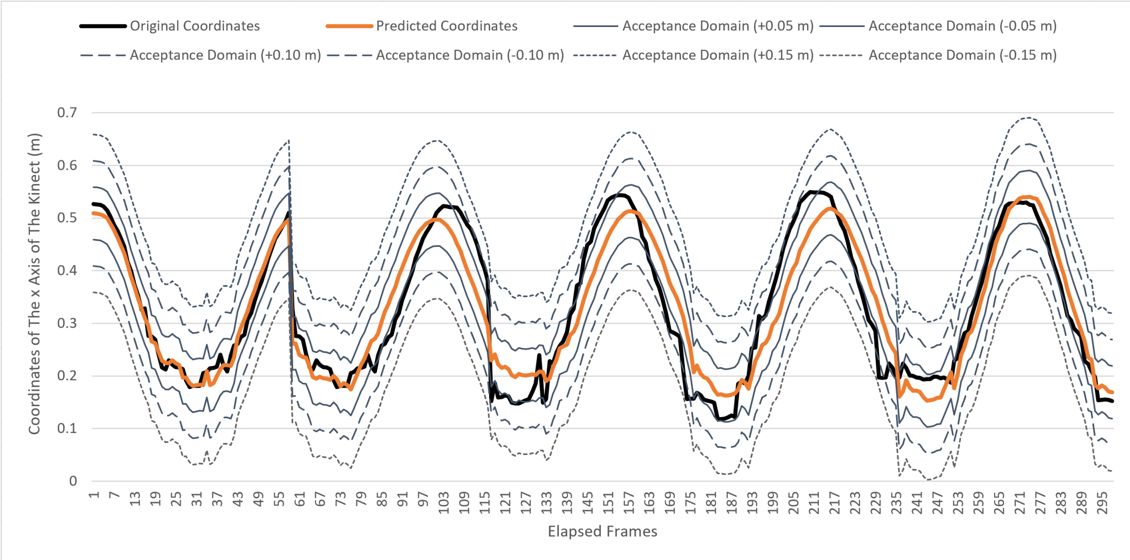

Figure 3.6: Comparing the user-input gesture to the predicted one, while also showing its ADs on the x axis.

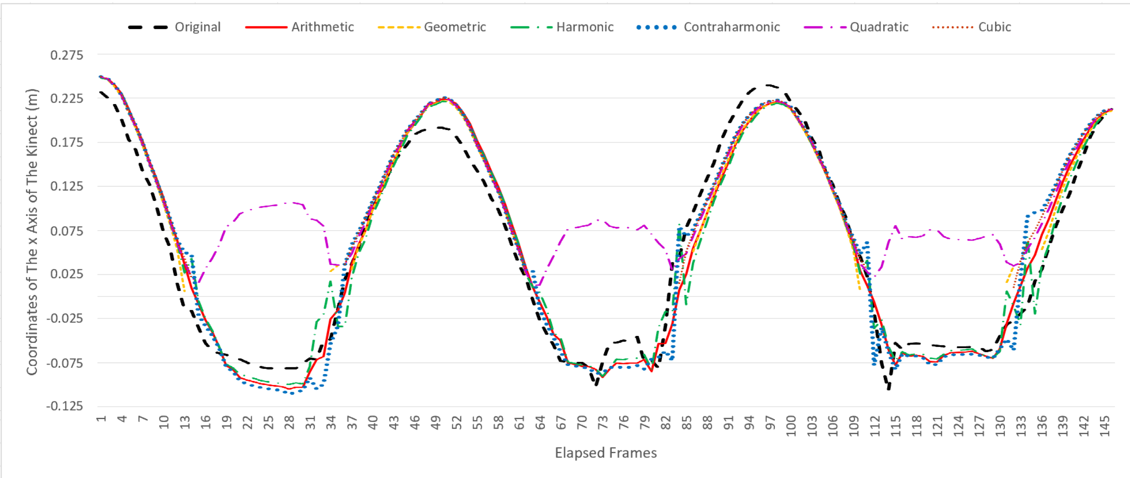

Figure 3.7: The results of the different MTs of the APBMR algorithm on the x axis.

According to Figure 3.7, the use of the geometric and the cubic MTs can present errors: they do not have gesture descriptors between frames 15 - 34, 63 - 85, 113 - 132. More errors are produced by the geometric MT, as gesture descriptors do not exist between frames 133 - 135 as well. These missing coordinates mean that the original movement descriptors are in the negative area.

The remaining two strengths of the algorithm are presented in Figures 3.8 and 3.9 showing only the strictest AD. The former of the strengths is that the same gestures can be accepted when done in a different position. This is a problem with the use of the Kinect, as only those gestures are accepted that are done in the same position as the previous gestures due to its built-in 3D coordinate system. In contrast, the position is followed by the APBMR algorithm. Therefore, the position of the possible following gesture can be predicted.

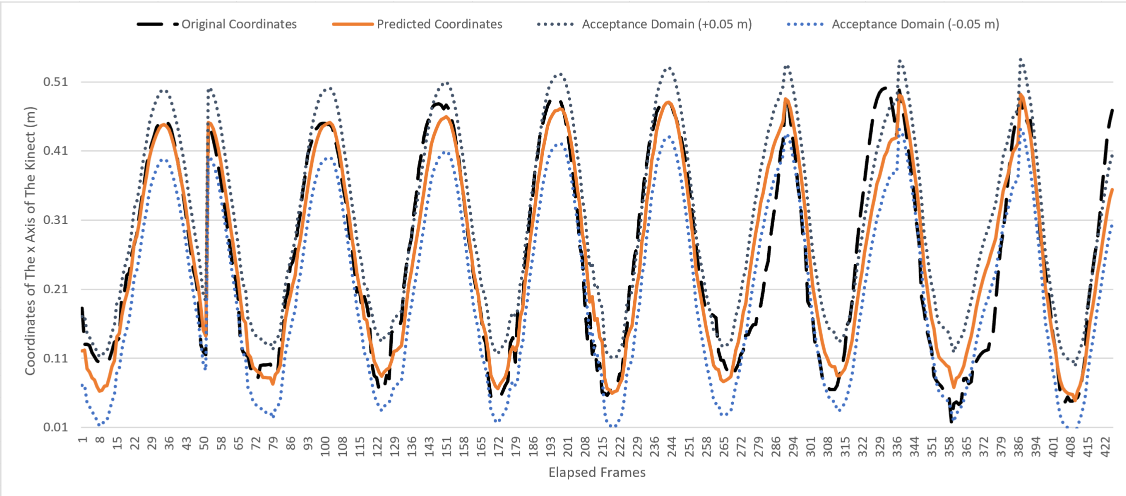

Figure 3.8: How the APBMR reacts to the changing position.

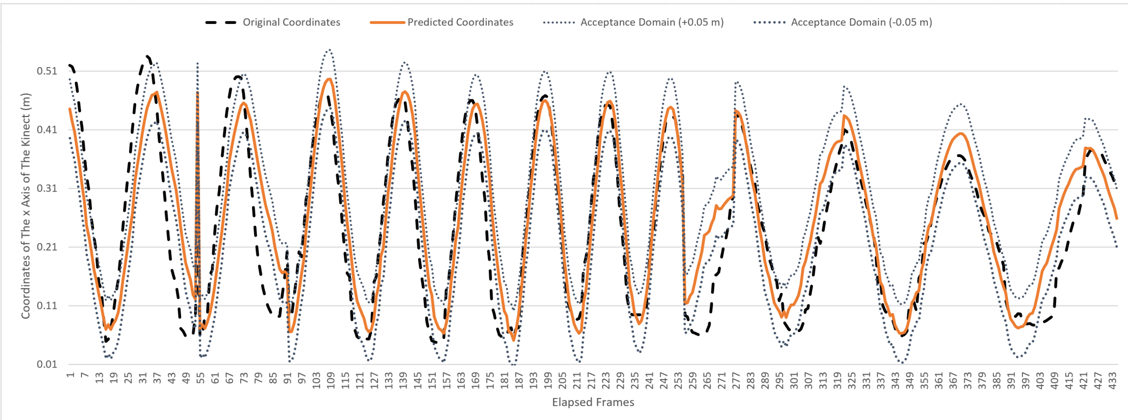

Figure 3.9: How the APBMR reacts to the changing speed.

The latter of the two strengths is that this algorithm is asynchronous. This means that when the movements have fewer frames due to various speeds, they are accepted as well. This is shown in Figure 3.9, where the first three gestures are done with "normal" speed, while the following six are done faster and the last few are slower. Only the strictest AD is shown in the figure.

Data collection

In the second half of 2019, data was collected at the University of Pannonia. The measurements were done with four groups of people: real-time data was collected with two groups and the data of the other two was logged into a file. The results of the measurements are presented in Table 3.3. The algorithm was used with two various computers. These will be referred to as general computer (GC) or advanced computer (AC). Their main specifications are the following, however, only their central processing unit (CPU), random access memory (RAM) and graphics processing unit (GPU) are listed:

- GC: CPU: Intel Core i7-720QM 1.60 GHz, RAM: 6 GB DDR3 1333 MHz, GPU: ATI Mobility Radeon HD 5850 1 GB
- AC: CPU: Intel Core i9-9900K 3.60 GHz, RAM: HyperX 32 GB Predator DDR4 3200 MHz, GPU: ASUS ROG Strix GeForce RTX 2080 8 GB GDDR6 SUPER

Table 3.3: Data collection using the Kinect.

Computer	**Number of people**	**Gestures**	**Repetition**	**Evaluation**	**Algorithm**
AC	16	4	10	File	APBMR
AC	32	4	10	Real-time	APBMR
GC	32	4	10	Real-time	APBMR
AC	32	4	10	File	APBMR and RDSMR/RDAMR

According to the data shown in Table 3.3, $16 \times 4 \times 10 + 32 \times 4 \times 10 + 32 \times 4 \times 10 + 32 \times 4 \times 10 = 4480$ cases were gathered and with these the accuracy and the speed of the APBMR algorithm can be evaluated. Each case was evaluated with six different MTs, meaning that the actual number of the cases is $4480 \times 6 = 26880$. Therefore, $32 \times 4 \times 10 \times 2 \times 6 = 15360$ cases were evaluated in real-time and the evaluation of $16 \times 4 \times 10 \times 6 + 32 \times 4 \times 10 \times 6 = 11520$ cases were file-based.

It should also be noted that in the second and the third row, the users who participated in the measurements were the same people. This means that 80 different people participated in the measurements. Out of the participants, 59 were male and 21 were female. They were 22.7 years old on average with a dispersion of 2.7 years as they were either in their bachelor studies or in their master studies. Their average height was 176.2 cm with a dispersion of 9.5 cm, while their average weight was 72.5 kg with a dispersion of 10.6 kg. There were no selection criteria applied to join the research, thus every person who was willing could help with the measurements.

During the measurements one Kinect v1 sensor was used, therefore the participants had to come sequentially. Similarly to the tests regarding the spatial skills of the users in VEs and because most of them were students, their appointments were made according to their classes. If only the gesture recognition is counted, the tests were 5 minutes long at most. However, instructions were given to the participants before each gesture, therefore one test could be as long as 15 minutes.

Regarding the measured gestures, they were also the same as well in each case: a circular gesture, a waving gesture, a diagonal gesture forwards and a diagonal gesture upwards. In the case of each gesture, the user starts the movement by standing in front of the sensor and facing it. The mentioned four gestures are shown in Figure 3.10 and are defined as the following:

- Circular gesture: the right hand of the participant is moved at shoulder height in a circular motion. When a circle is completed, a repetition ends.
- Waving gesture: the right hand of the participant is moved above shoulder height first from left to right and then from right to left with a waving motion. Afterward, a repetition ends.
- Forward-diagonal gesture: the right hand of the participant is at ease in the beginning. Then, the right hand is raised to shoulder height and extended forward, diagonally to the right. Lastly, the right hand has to be lowered. This is when one repetition ends.
- Upward-diagonal gesture: this gesture starts similarly to the forward-diagonal gesture. However, in this case the right hand is raised above shoulder height. Then, it is extended both forward and upward, diagonally to the right. Lastly, the right hand has to be lowered and then, one repetition is over.

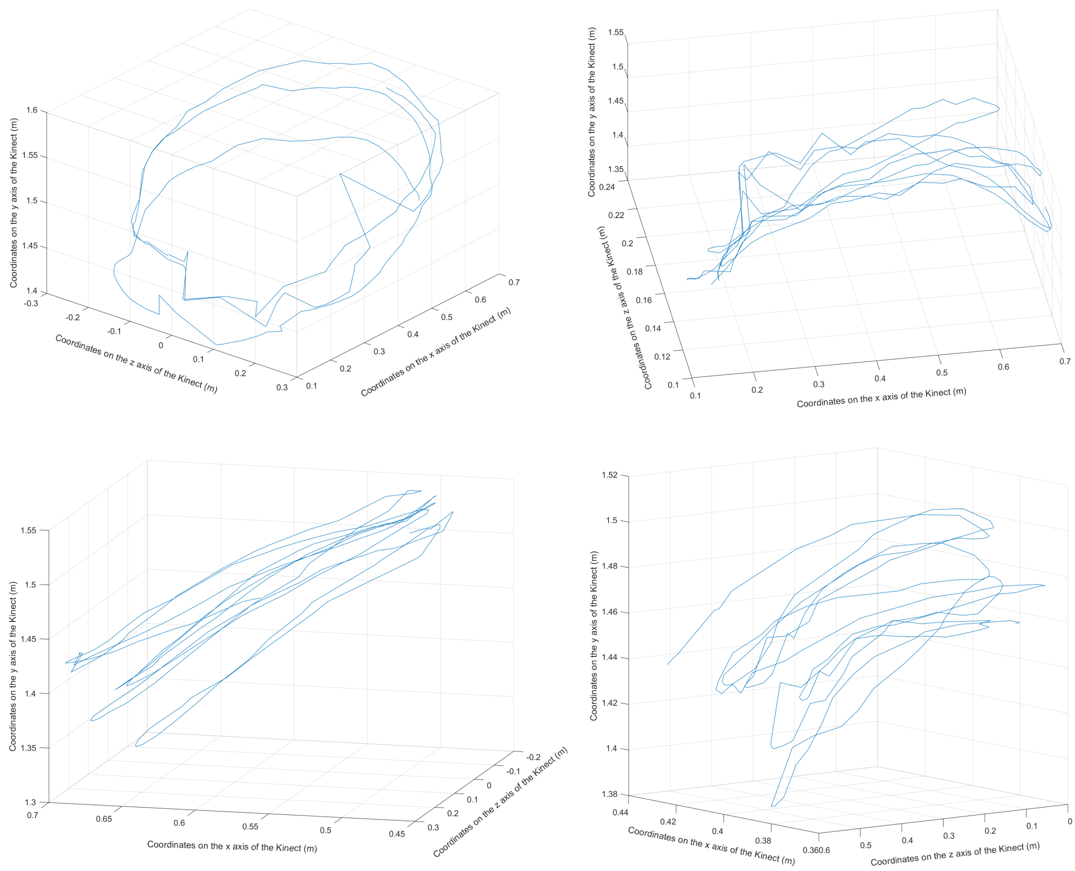

Figure 3.10: 200 frames of the circular (upper left), waving (upper right), forward-diagonal (lower left) and upward-diagonal (lower right) gestures.

In the fourth row of Table 3.3 it is shown that both algorithms were assessed in that case. During this comparison, the gesture descriptors were recorded and they were saved into a file, because it was critical to assess the same coordinates of the gestures. Therefore, to compare the two algorithms, the data was loaded from the mentioned file and was evaluated by both algorithms.

Data analysis

First, it was evaluated by the APBMR algorithm whether a gesture is accepted using each of the MTs. Then, the percentage of the acceptance (or rejection) inside all three ADs on all axes is logged into a .csv file. The time in milliseconds of each decision is logged as well. This means that 12 columns of data exist in the case of each MT. The format of the data is the following as presented in Table 3.4. Since six MTs are used by the algorithm, only one is shown in the table.

Table 3.4: Data structure of one MT, containing the acceptance percentage and the decision times (ms) of ten classified gestures in three ADs.

±0.05 m (x)	**±0.10 m** (x)	**±0.15 m** (x)	**Time** (x)	**±0.05 m** (y)	**±0.10 m** (y)	**±0.15 m** (y)	**Time** (y)	**±0.05 m** (z)	**±0.10 m** (z)	**±0.15 m** (z)	**Time** (z)
1.000	1.000	1.000	0.404	0.279	0.7441	0.953	0.173	0.777	0.866	0.866	8.765
1.000	1.000	1.000	0.195	0.342	0.789	1.000	0.161	0.755	1.000	1.000	0.454
1.000	1.000	1.000	0.522	0.097	0.219	0.365	0.173	0.526	1.000	1.000	0.147
1.000	1.000	1.000	0.269	0.431	0.727	0.954	0.221	0.976	1.000	1.000	0.180
1.000	1.000	1.000	0.311	0.125	0.175	0.250	0.200	0.775	1.000	1.000	0.223
1.000	1.000	1.000	0.181	0.076	0.128	0.179	0.261	0.960	1.000	1.000	0.445
0.381	0.981	1.000	0.302	0.196	0.470	0.901	0.235	0.843	1.000	1.000	0.204
0.666	1.000	1.000	0.206	0.735	1.000	1.000	0.230	0.895	1.000	1.000	0.184
0.596	1.000	1.000	0.249	0.173	0.288	0.442	0.282	1.000	1.000	1.000	0.212
1.000	1.000	1.000	0.261	0.893	1.000	1.000	0.205	0.900	1.000	1.000	0.201

As can be suspected from the previous table, the amount of gathered data is huge. Since no advanced statistical evaluations were required for the goal of this research, Microsoft Excel was used for the investigation instead of the R program package. This means that the data was analyzed inside of it using its built-in functions. Each gesture type was given its own worksheet for easier analysis.

The gestures in each worksheet were grouped by the ID of users (since no name was gathered to respect anonymity). As ten repetitions were done by a user, a row was left blank after ten rows for easier navigation through the data. The averages and the deviations were evaluated at the bottom of each worksheet. The percentages of acceptance were evaluated on one axis first, then on three pairs of two axes and lastly, on all three axes. After the evaluations on all worksheets were completed, a final worksheet was made. In this worksheet, all data were summarized and plotted.

Chapter 4

Results of the measurements

After defining the materials and methods in both parts of the research, the results of the measurements are presented in this chapter. There are two subsections in this chapter. The results on the spatial ability tests are presented and evaluated in section 4.1, while the APBMR algorithm is presented, tested and evaluated in section 4.2.

4.1 Analyzing the results on the spatial ability tests

In this section the results on the spatial ability tests are investigated. Since there were three goals in the research of the author regarding the spatial skills of the users in VEs, this section is split into three subsections. In subsection 4.1.1 the influence of display devices and display parameters on the spatial ability tests in VR is analyzed. The effects of display devices and human skills on the spatial ability test completion times are investigated in subsection 4.1.2. In subsection 4.1.3, the correlation between the used display devices and the human skills is assessed.

4.1.1 The influence of display devices and display parameters on the spatial ability tests in virtual reality

Results of the analyses of a single factor's effects

The virtual camera type was the first factor to be analyzed. Two types of virtual camera exist. The first type is called perspective camera which is similar to the human eye. The second type of virtual camera is orthographic, meaning orthographic projection is used by it: 3D objects are depicted in two dimensions [210]. Because of the randomization of the camera type variable, 1418 perspective camera tests were performed and 1291 orthographic camera tests were performed (as seen in Table B1 in the appendix). The logistic regression analysis results are shown in table A1 which can be found in the appendix as well. The estimated coefficients can be found in Figure 4.1 in the form of 95% confidence intervals (CIs) at the end of this subsection.

When using the logistic regression analysis method, a basis variable has to be selected. This variable is always indicated with the Intercept name. Here, when investigating the camera type, the orthographic camera type was used as the basis. The difference is significant, based on the *p-value* $= 2.57 \times 10^{-12}$. This value can be found in the $\Pr(>|z|)$ column as it defines whether the estimate of the coefficient point is significantly different from zero. Therefore, it can be concluded that the type of the camera has an effect on the probability of correct answers: better results are produced by the perspective camera type than the orthographic camera type.

The FoV of the virtual camera was the second factor to be analyzed. The default FoV in Unity is 60°, but this value can be easily modified in both the Unity editor and the developed spatial ability measuring application. Several FoVs, such as 45°, 60°, 75°, 90° were investigated. There were 1049 measurements performed using the 45° option, 120 using the 60° option, 134 using the 75° option, and 115 using the 90° option. The FoV is undefined in the case of an orthographic camera type due to its method of projection. With the orthographic camera type, 1291 measurements were made. This numerical data can be seen in Table B2 in the appendix. The results of the logistic regression analysis are presented in Table A2 in the appendix as well.

The basis value was -1 which means undefined in the measured data: the orthographic camera type is symbolized by this value of -1. All coefficients are estimated to be positive. Due to the *p-values*, every probability is significantly greater in the case of all FoVs of the perspective camera type. Moreover, due to the results shown in Table A1, it can be concluded that better probabilities are produced using the perspective camera type. Therefore, the data belonging to the orthographic camera type (value of -1) was eliminated. Then, it was investigated whether the results have different probabilities using various levels of FoVs in case of the perspective camera type. This means that the results were also investigated after the -1 was taken out. The results of the analysis based on this restricted data set are contained in Table A3 in the appendix, while the estimated coefficients can be seen in Figure 4.1.

For the analysis, the basic level was actually 45°. It is shown by the signs of the estimated coefficients that each further level is better. The difference is not significant, except for 90°. On the basis of *p-value* $= 0.0225$, a significantly better probability of correct answers was presented by the FoV of 90° than the others, at the level 0.05. However, the difference is not significant at the level 0.01. As the amount of data is quite large, it can be accepted that the effect of the variable named FoV is not significant in the case of the perspective camera type. Therefore, the variable named FoV is omitted from further analyses.

The next factor to be analyzed was the rotation of the camera. Thus, it was investigated whether the probabilities were affected by the virtual camera rotation. 106 measurements were carried out with a rotation of -45°, 294 with -30°, 294 with -15°, 1251 with no rotation, 312 with 15°, 313 with 30° and 139 with 45°. These numerical data can be seen in Table B3 in the appendix. The results of the logistic regression analysis are presented in Table A4 in the appendix.

The basic level was -15°. Significant increases in the probability of the correct answers are shown in the case of the -45°, 0°, 45° rotations to those of -15°, according to the *p-values*. The latter rotation is not significant at the 0.01 level. However, it is close to being significant. Two groups were created for the results, in order to further verify the findings. Those rotations which positively affected the probabilities are contained in the first group named "IMP_R". These are the rotations of -45°, 0°, 45°. In the other group which is called "NO_R", rotations which had no significant positive influence are included. 1496 tests are in the group IMP_R and 1213 tests are in the group NO_R (see Table B4 in the appendix). These two groups were investigated with the logistic regression analysis method. The results of the analysis are shown in Table A5 in the appendix, while the estimated coefficients can be seen in Figure 4.1. In this case the point of reference was NO_R. It is indicated by Table A5 that there were substantially different probabilities for the two groups previously described in this subsubsection. It is shown by these results that the two groups are distinguishable from each other. Therefore, from this point forward, these groups of camera rotations are used.

The next to analyze was the effect of the contrast ratio on the probabilities. Five contrast ratios were used for the calculations: 1.5:1, 3:1, 7:1, 14:1 and 21:1, The number of the results of students who tested with these contrast ratios are 1066, 167, 1121, 164, 191, respectively. This numerical data can be seen in Table B5 in the appendix. The regression coefficients were determined using logistic regression and the test statistics (testing their zero values). The respective *p-value*s of the results can be seen in Table A6 in the appendix.

As can be seen, the results were compared to the 1.5:1 contrast ratio. According to the comparison, significantly worse probabilities are produced by the 7:1, 14:1 contrast ratios, even the contrast ratio of 21:1 is on the 0.05 level. Therefore, two groups were created for the contrast ratios: IMP_C, containing the ratios of 1.5:1 and 3:1, and the other, NO_C, containing the ratios of 7:1, 14:1 and 21:1. There are 1233 results on the tests in IMP_C, while there are 1476 in NO_C. This numerical data is presented in Table B6 in the appendix. According to $p\text{-}value = 2.56 \times 10^{-6}$ in Table A7, a significantly better probability of correct answers is produced by the contrast ratio group of IMP_C than the group of NO_C. Therefore, better results are given by the VEs with a bright background and bright foreground objects. The estimated coefficients of these contrast ratio groups can also be seen in Figure 4.1.

The next factor analyzed was the presence of shadows in the VE. Only two levels exist of this variable: when their existence is turned on and when it is turned off. There were 1414 measurements of the former, and 1295 of the latter as can be seen in Table B7 in the appendix. The results of the logistic regression analysis are shown in Table A8 also in the appendix, while the estimated coefficients can be seen in Figure 4.1. The point of reference was "Turned off". According to $p\text{-}value = 0.204$, the probability of correct answers on the tests is not affected by the shadows.

The last factor to analyze was the display device used by the students during the tests. As mentioned in the "Data collection" subsubsection, two display devices were used during the spatial ability tests. One of the two devices was an LG 20M37A (19.5") DD, the other was the Samsung GVR HMD. 2160 measurements were done using the DD and 549 using the GVR as can be seen in Table B8 in the appendix. The results of the logistic regression analysis of the display devices used are presented in Table A9, where it can be seen that the point of reference was DD. According to the estimated coefficient of 0.07595 and the $p\text{-}value = 0.00677$, the probability of correct answers on the tests is significantly larger when the GVR is used in the VE.

To better visualize the data, the 95% CIs of the estimated coefficients of all mentioned variables are shown in Figure 4.1. It is possible to deduce which variables affect significantly the probabilities of correct answers: if the estimation of a variable reaches zero, then there is no significant influence. As could be seen from the results presented in the tables and from the following figure, there is no significant influence on the probabilities when the FoV is 60° or 75° or when the shadows are turned on.

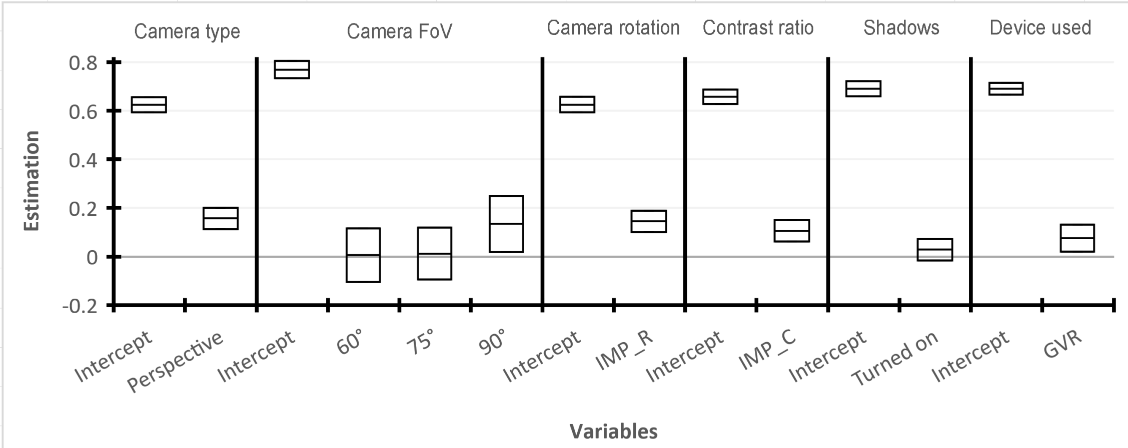

Figure 4.1: 95% CIs of the estimated coefficients in the case of one analyzed variable.

Results of the analyses of effects of two factors without interactions

In this subsubsection, the effects of the display parameters were analyzed in pairs without taking their interactions into account. Those variables were omitted that do not significantly affect the probabilities in themselves. Therefore, based on Tables A3 and A8, neither the influence of shadows nor the influence of camera FoV are examined further. The effects of the camera type, camera rotation, contrast ratio, and the device used variables were analyzed by pairing them in all possible combinations. This resulted in the following pairs: camera type and camera rotation; camera type and contrast ratio; camera type and the device used; camera rotation and contrast ratio; camera rotation and the device used; and contrast ratio and the device used. The numerical results of the possible pairs can be seen in Table B9, while the results of the logistic regression analysis are presented in Table A10. Both tables can be found in the appendix. 95% CIs of the estimated coefficients (Figure 4.2) can be found at the end of this subsubsection.

The pair of the camera type and camera rotation variables was the first to be analyzed. During the investigation, the previously defined camera rotation groups were used to make the calculations easier. Every level of camera type variable was paired with every level of camera rotation group. Therefore, the number of pairs equals to four; the first pair is called "Orthographic, NO_R". 611 measurements were made using this pair. The remaining pairs are "Orthographic, INC_R", with 680 measurements; "Perspective, NO_R", with 602 measurements; and "Perspective, INC_R", with 816 measurements. According to the *p-values*, significantly better results are produced by every other pair than the pair of "Orthographic, NO_R". No significant difference is detected between the pairs of "Orthographic, INC_R" and "Perspective, NO_R". This difference was calculated by means of a t-test and its *p-value* = 0.7126. The optimal result out of the pairs is provided by the pair of "Perspective, INC_R", and the improvement compared to "Perspective, NO_R" is significant (*p-value* = 0.0033).

The second pair to investigate was the type of camera and the contrast ratio. As in the case as the camera rotation groups, the groups of contrast ratios that were established previously were used. In "Orthographic, NO_R", 696 tests were performed; in "Orthographic, INC_C", 595 tests; in "Perspective, NO_C", 780; and in "Perspective, INC_C", 638 tests. According to results presented in Table A10, the pair of "Orthographic, NO_C", is significantly worse than the other pairs. However, no significant differences exist among the other three.

The pair of the camera type and the device used variables was the third to be examined. 1065 tests were performed in "Orthographic, DD", 226 tests in "Orthographic, GVR", 1095 tests in "Perspective, DD", and 323 tests in "Perspective, GVR". The worst results are produced when using a DD with an orthographic virtual camera type. When using the GVR with an Orthographic virtual camera type, there is no significant improvement either. However, the results of the measurements when using the DD or a GVR with a perspective virtual camera type are significantly better than when using the DD with an orthographic camera. The difference between a DD with an orthographic camera or a perspective camera and using a GVR with a perspective camera cannot be distinguished.

The pair of camera rotation and the contrast ratio was examined next. There are 1005 tests in the "NO_R, NO_C" group, 208 tests in the "NO_R, INC_C" group, 471 tests in the "INC_R, NO_C" group, and 1025 tests in the "INC_R, INC_C" group. Based on the results of the logistic regression analysis, better probabilities are yielded by every pair than by "NO_R, NO_C". Also, a difference was not indicated by the t-test between the pairs of "NO_R, NO_C" and "NO_R, INC_C", but differences can be seen among "NO_R, NO_C", "INC_R, NO_C" and "INC_R, INC_C". These last pairs are not distinguishable.

The next pair to be examined was the camera rotation and device used. The numbers of measurements were 1062 for the group named "NO_R, DD", 151 for the group "NO_R, GVR", 1098 for the group "INC_R, DD", and 398 for the group "INC_R, GVR". The group of "NO_R, DD" was the reference point. Since the sign of the estimation of the coefficient is negative, it can be observed that the probabilities of the correct answers are a bit smaller in the group called "NO_R, GVR", but the difference is not significant (*p-value* = 0.385). There are significantly greater probabilities in the remaining two groups and there is a significant difference between the groups of "NO_R, GVR" and "INC_R, GVR".

The pair of contrast ratio and the device used variables was the last one to be examined. After the creation of the pairs, the pairs of "NO_C, DD" consisted of 1183 tests, "NO_C, GVR" of 293 tests, "INC_C, DD" of 977 tests, and "INC_C, GVR" of 256 tests. In the case of the contrast ratio and the device used pair, the group "NO_C, DD" was the reference point. According to the results of the logistic regression analysis, a significant improvement is not produced by the use of the GVR in the case of the "NO_C" contrast group. However, in the case of the "INC_C" contrast group, significantly better results are yielded. Moreover, if the "INC_C, DD" and "INC_C, GVR" are compared, it can be seen that the differences between the average rates are large, however, these are not significant (*p-value* $= 0.06332$). This is due to the relatively small number of tests (256 tests). Moreover, a significant difference is provided by the difference between the average rates of "NO_C, DD" and "INC_C, DD" as the numbers of the samples are higher. It is suspected that if more data existed with the GVR, the difference would have become significant.

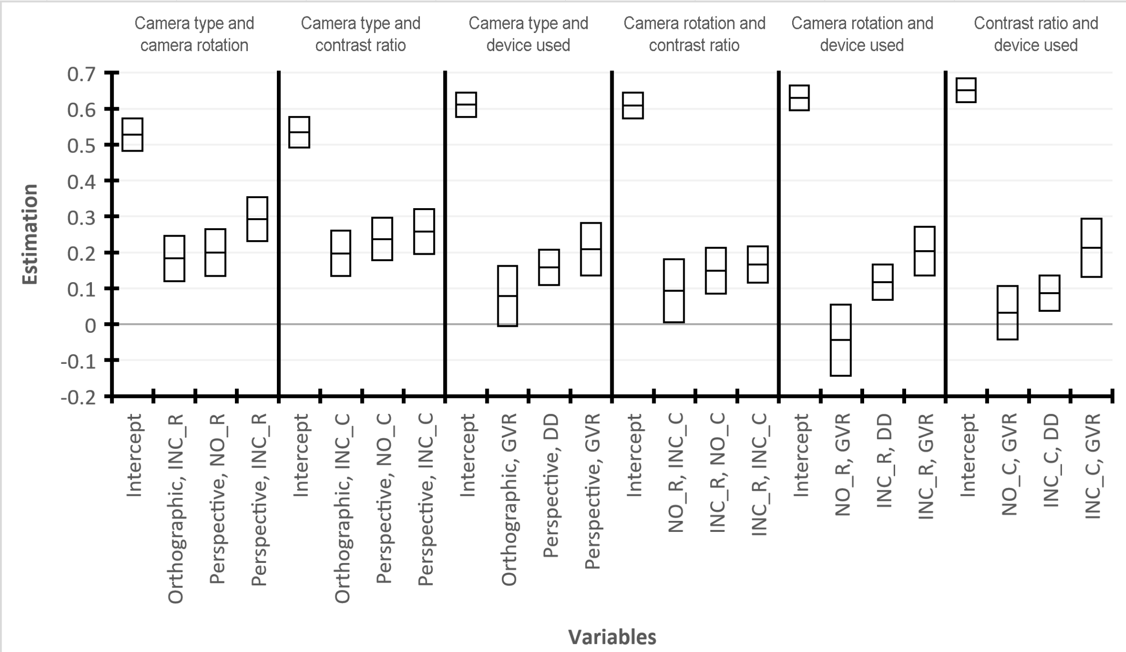

Figure 4.2: 95% CIs of the estimated coefficients in the case of the investigated variable pairs without interactions.

Results of the analyses of effects of two factors with interactions

The additive models were investigated on the basis of the pairs of the variables by allowing their interaction. Alike to the analyses without the interactions, the pairs are combined in every possible combination. The sequence of analyzing the pairs is the same as in the previous subsubsection: the first pair is the camera type and the camera rotation; the second is the camera type and the contrast ratio; the third is the camera type and the device used; the fourth is the camera rotation and contrast ratio; the fifth is the camera rotation and the device used; and the last one is the contrast ratio and the device used. The results of the logistic regression analyses of all possible pairs allowing interactions are shown in Table A11 in the appendix, while their estimated coefficients can be seen in Figure 4.3 on the next page.

Thus, the first pair to be analyzed is the camera type and the camera rotation variables. Due to the results of the logistic regression analysis shown in Table A11, an influence can be detected by both the camera type and the camera rotation. This is indicated by the estimated coefficients. Also, the *p-value* $= 0.0459$, meaning that a significant interaction exists as well. The negative sign of -0.08995 was a surprise, but this is due to the speed of improvement. The measures of the improvement cannot be summed, the result is a little bit lower than that caused additively.

The next to analyze was the pair of the camera type and the contrast ratio variables. As seen in the table, both the camera type and the contrast ratio variables have significant influences on the probability of correct answers. The significance of the analyzed contrast ratio group is stronger than the type of the camera. Also, there is a significant interaction ($p\text{-}value = 8.91 \times 10^{-5}$) between the two variables.

The camera type and the device used is the next pair to be investigated by the logistic regression analysis method. In addition to the camera type variable, no significant influence ($p\text{-}value = 0.0647$) has been detected in the results of the tests by investigating the device used. Significant interaction does not exist between the two variables ($p\text{-}value = 0.6164$).

The next pair to mention is the camera rotation and contrast ratio. The influence of INC_R is strong and is very significant with $p\text{-}value = 3.97 \times 10^{-6}$. In contrast to the INC_R, the effect of the INC_C is smaller: the estimation of the coefficient equals to 0.09323, but it is also significant with $p\text{-}value = 0.0376$. However, there is no significant interaction ($p\text{-}value = 0.1667$) between the two variables.

The next pair to analyze was the camera rotation and the device used. According to the logistic regression analysis results in the table, a significant influence ($p\text{-}value = 2.97 \times 10^{-6}$) is produced by the camera rotation in itself. When talking about only the influence of the device used variable, its influence is not significant ($p\text{-}value = 0.3847$). However, when the two variables are paired together, a significant interaction at the 0.05 level of significance appears ($p\text{-}value = 0.0326$).

The last pair to analyze is the contrast ratio and the device used variables. According to the last pair in Table A11, a significant effect ($p\text{-}value = 0.000583$) is produced by the contrast ratio. However, aside from the described influence, the influence of the device is not significant ($p\text{-}value = 0.396699$) and interaction between the two variables has not been detected ($p\text{-}value = 0.094199$).

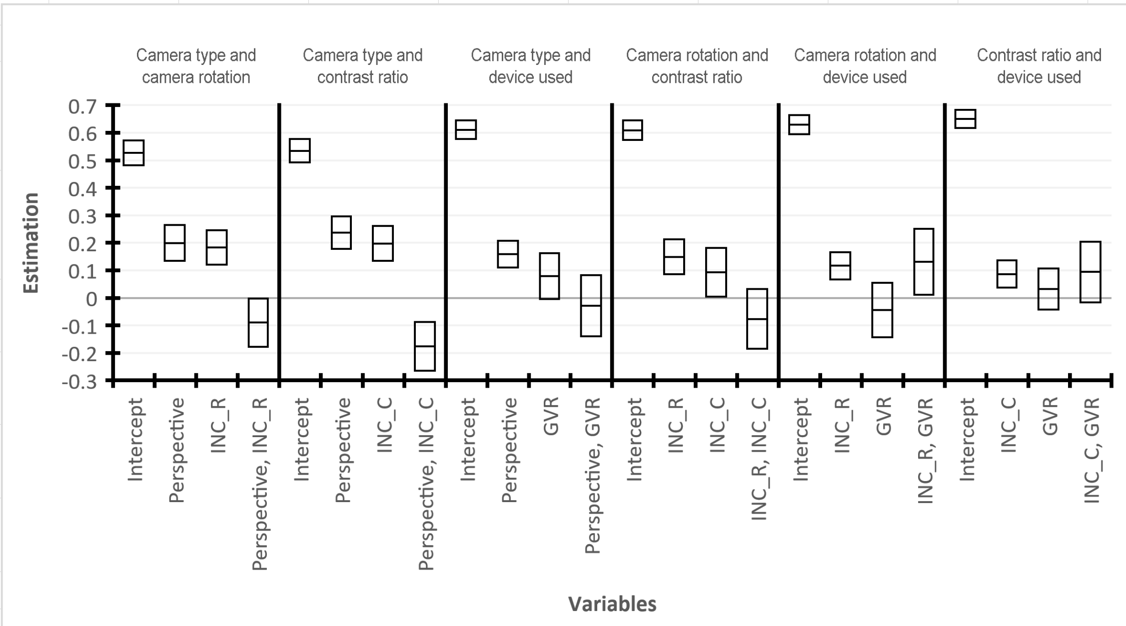

Figure 4.3: 95% CIs of the estimated coefficients in the case of the investigated variable pairs with interactions.

More investigations on the same pairs using the ANOVA dispersion analysis have also been done. With the ANOVA analysis, the interactions of the pairs of the variables can also be investigated. The logistic regression analysis and the ANOVA analysis were compared and similar results are produced. Using the results of the ANOVA analysis, it was also compared whether the interactions of the factors provide a stronger, but significant influence. The results of the ANOVA method can be observed in Table A12 which can be found in the appendix.

According to the results of Table A12, it can be concluded that a better probability (*p-value* $= 0.0459$) is provided by the model in which the interactions are taken into account in the case of the camera type and rotation variable pair. Significant probabilities are yielded by both the ANOVA analysis and the logistic regression analysis that are presented in Table A11.

In case of the next pair of the camera type and the contrast ratio variables, there is a little difference between the *p-value*s of the ANOVA analysis and the logistic regression analysis: in case of the former, the *p-value* $= 8.895 \times 10^{-5}$ and in the case of the latter, the *p-value* $= 8.91 \times 10^{-5}$. Both significances are very strong. According to these *p-value*s of the pair of the camera type and contrast ratio in case of both analyses, it can be concluded that significantly better probabilities are provided by the model that takes interactions into account than the additive model.

The third pair to analyze was the camera type and the device used variables. Due to *p-value* $= 0.6164$ in the column of Pr ($>$Chi) and to the same *p-value* from the results of the logistic regression in Table A11, the conclusion is the following: according to the results of both the ANOVA analysis and the logistic regression analysis in the case of the camera type and the device used variable pairs, the model in which the interactions are taken into account does not provide a better probability than the additive model.

In the case of the pair of the camera rotation and contrast ratio, the *p-value*s with the ANOVA analysis and the logistic regression analysis is different in the additive model. Using ANOVA analysis, the *p-value* $= 0.1661$ and in the additive model, the *p-value* $= 0.1667$. Therefore, based on the *p-value*s, in the case of the pairs of camera rotation and contrast ratio, it can be concluded that no significant probabilities are provided by both models. Since both probabilities are not significant, the model in which the interactions are taken into account does not provide a better probability than the additive model.

The next pair to analyze was the camera rotation and the device used variables. According to results of the ANOVA analysis, the *p-value* $= 0.0329$ and according to the results of the logistic regression, the *p-value* $= 0.0326$ in the additive model. Both *p-value*s are on the same level of significance. Therefore, it can be concluded that a significantly better probability is provided by the model in which the interactions are taken into account than it is provided by the additive model.

Lastly, in the case of the contrast ratio and the device used, the *p-value*s are not significant in each model. They are 0.094199 in the additive model and 0.09403 in the model where the interactions are taken into account. Since both *p-value*s are not significant, the conclusion is the following: in contrast to the additive model, a significantly better probability is not provided by the model in which the interactions are taken into account.

Results of the analyses of effects of three factors without interactions

When the investigation of all possible pairs were finished, the variables were grouped in triplets. Similarly, to the previous subsubsections, the analyses were performed in two different ways: firstly, without interactions, and afterward, with interactions. All possible triplets were created from the levels of variables and these triplets were the following:

- Camera type, camera rotation and contrast ratio
- Camera type, camera rotation and the device used
- Camera type, contrast ratio and the device used
- Camera rotation, contrast ratio and the device used

After the creation of the mentioned four triplets, the logistic regression analysis was performed on them. Both the results of the logistic regression analysis of the triples and the numerical results can be seen in the appendix, inside Tables A13 and B10, respectively. The 95% CIs of the estimated coefficients are shown in Figure 4.4 in the end of this subsubsection.

The camera type, camera rotation and the contrast ratio was the first variable triplet to be investigated. First, $2 \times 2 \times 2 = 8$ groups were formed according to the levels of the variables. The point of reference was "Orthographic, NO_R, NO_C" in the logistic regression analysis. According to the results presented in the table, the probability of "Orthographic, INC_R, INC_C" and the probability of all triplets containing the perspective camera type are significantly better than the "Orthographic, NO_R, NO_C". Therefore, the results were checked, and it was concluded that the "Orthographic, INC_R, INC_C" and the perspective rows are not distinguishable.

The investigation continued with the camera type, rotation, and device used variable triplet. The point of reference was "Orthographic, NO_R, DD". Due to the results presented in the table, the triplet of "Orthographic, NO_R, GVR" has some, but not significant decrease in the results compared to "Orthographic, NO_R, DD". The probabilities in the other combinations of the camera type, rotation and device used variable triplet are significantly stronger. The smallest improvement can be observed in the case of the "Perspective, NO_R, GVR" triplet and the greatest improvements can be noticed in "Perspective, INC_R, DD" and "Perspective, INC_R, GVR" triplets. It should be noted that the difference between the two best cases is not significant (*p-value* = 0.3012).

The camera type, contrast ratio, and the device used variable triplet was the following to be investigated. In this case, the point of reference was "Orthographic, NO_C, DD". According to the results of the logistic regression analysis, the "Orthographic, NO_C, GVR" tiplet was not significantly better than "Orthographic, NO_C, DD". The combinations that remain were significantly stronger than the point of reference and the previously mentioned "Orthographic, NO_C, GVR" triplet. However, the remaining combinations cannot be distinguished from each other.

Lastly, the camera rotation, contrast ratio and the device used variable triplet is the one remaining combination to be investigated. According to the results of the logistic regression analysis, no detectable significant difference is found between "NO_R, NO_C, DD" and "NO_R, NO_C, GVR" triplets as the $p\text{-}value = 0.82046$. Also, the cut is not significant in the row of "NO_R, INC_C, GVR". This fact requires clarification, since the average rates of correct answers are 0.602 and 0.561, respectively. However, in case of "NO_R, INC_C, GVR", the data that is available is very low. The number of the measurements was 23 and the dispersion of the correct answers is high. Thus, further measurements are needed to be able to reject the H of the equality of the average rates. In the remaining other cases, significant improvements exist. The largest improvement is in the case of the "INC_R, INC_C, GVR" triplet and it is the strongest significantly. This means that the results of this case are significantly better than the results of any other case.

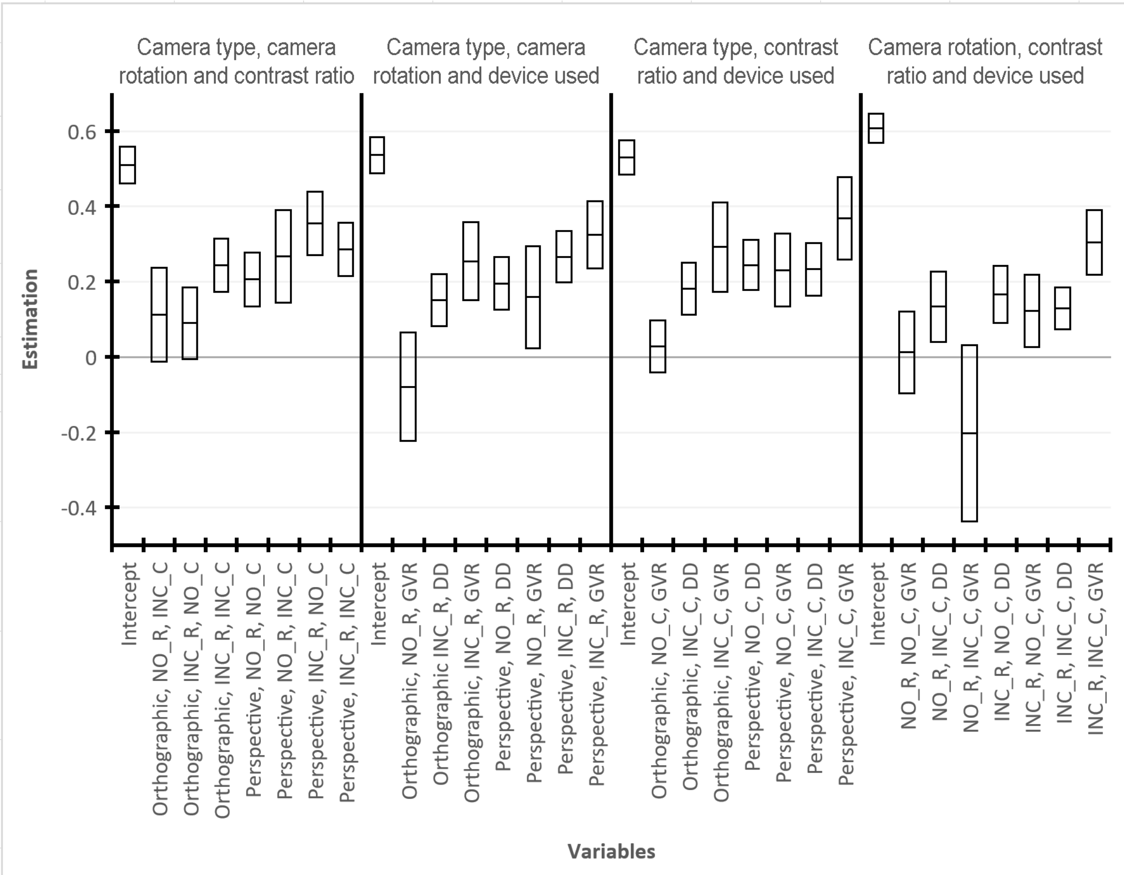

Figure 4.4: 95% CIs of the estimated coefficients in the case of the investigated variable triplets without interactions.

Results of the analyses of effects of three factors with interactions

The first variable triplet analyzed was the camera type, camera rotation and the contrast ratio. For this, the ANOVA dispersion analysis method was used. Similarly, to the analysis of pairs with interactions, comparisons were made between the models. However, in this case, the number of models was three. Therefore, a comparison was made between the logistic regression models with additive property without interactions (I), the models with interactions between two variables (II), and the models allowing interactions among all three variables (III).

According to the results of the comparison between models I and II, the model that took into account the interactions of the camera type and camera rotation, moreover the interaction of camera type and contrast ratio gave significantly better results than the logistic regression models without interactions (p-$value = 0.001258$). The interaction of camera rotation and contrast ratio was not significant, as was presented in the previous section; therefore, it was omitted in this section. Similar results (p-$value = 0.002387$) are yielded by the comparison of models I and III as the comparison between models I and II. Then, the comparison of models II and III commenced. However, after the comparison, it can be concluded that the model in which the interaction of all three variables is built in does not give better results than model II (p-$value = 0.2049$). The results of the ANOVA analysis method are presented Table A14 in the appendix.

Thus, the results of the logistic regression analysis are presented in the case of model II which is the most appropriate one. The results, and therefore the effects of the factors according to model II are presented in first block of Table A15. This table can be found in the appendix, while the 95% CIs of the estimated coefficients can be seen in Figure 4.5 on the next page. According to these results, every variable has an influence, and the interaction in the case of the camera type and the contrast ratio is significant. To double-check the calculations and the results as well, the model in which the interactions of the camera type and rotation, as well as the interactions of the camera type and contrast ratio were built in was compared to the model where the interaction of the camera type and contrast ratio were built in. According to the results of the double-check, worse results were not provided by the means of this reduction (p-$value = 0.99$).

The next variable triplet was the camera type, camera rotation and the device used. In the case of this triplet, similar models were compared as previously. When comparing the model in which interactions are not allowed (I), and the model in which the interactions of two variables are allowed (II) and the model in which the interaction of all variables are allowed (III), model II is significantly better than model I (p-$value = 0.01175$), and model III is not significantly better than model I (p-$value = 0.0502$). Between model II and model III, there is no significant difference (p-$value = 0.7445$). When analyzing model II by logistic regression, the results of the analysis are presented in Table A15 in the second block. According to the results presented in the table, it can be concluded that the influence of the device itself disappeared (p-$value = 0.2649$), but its interactions are still relevant (see p-$value$s 0.0322 and 0.0286). This means that the display device that was used during the tests should be taken into account when analyzing the data.

The third variable triplet was investigated in the same way. When comparing the model with the variables camera type, contrast ratio, and device used without interactions (I), the model in which interactions of pairs are allowed (II), and the model in which interactions of all three variables are allowed (III), the results were the following: model II is significantly better than model I (p-$value = 0.0001132$), model III is significantly better than model I (p-$value = 0.000546$), and no significant difference exists between models II and III (p-$value = 0.1792$). When using model II, the logistic regression analysis results are the following, as presented in the third block of Table A15. According to the results presented in the table, the influence of every parameter is significant, and the interaction of the camera type and contrast ratio is also detectable (p-$value = 0.000113$).

The last variable triplet to investigate was the camera rotation, contrast ratio and the device used. A similar way of analysis was used in the case of these variable triplets: when comparing the model investigating the effects of the variables without interactions (I), the model investigating the effects of variables allowing the interaction of rotation and used devices (II), and the model allowing the interactions of all three variables (III), the following conclusions can be drawn: II is significantly better than I (p-$value = 0.02568$), III is significantly better than I (p-$value = 0.0001853$) and is also significantly better than II (p-$value = 0.0006448$). When using model III, the logistic regression analysis yields the results that are presented in the last block of Table A15. According to these results, the unique influence of the device disappears, but it has an interaction with the contrast ratio, and even triple interactions can be detected.

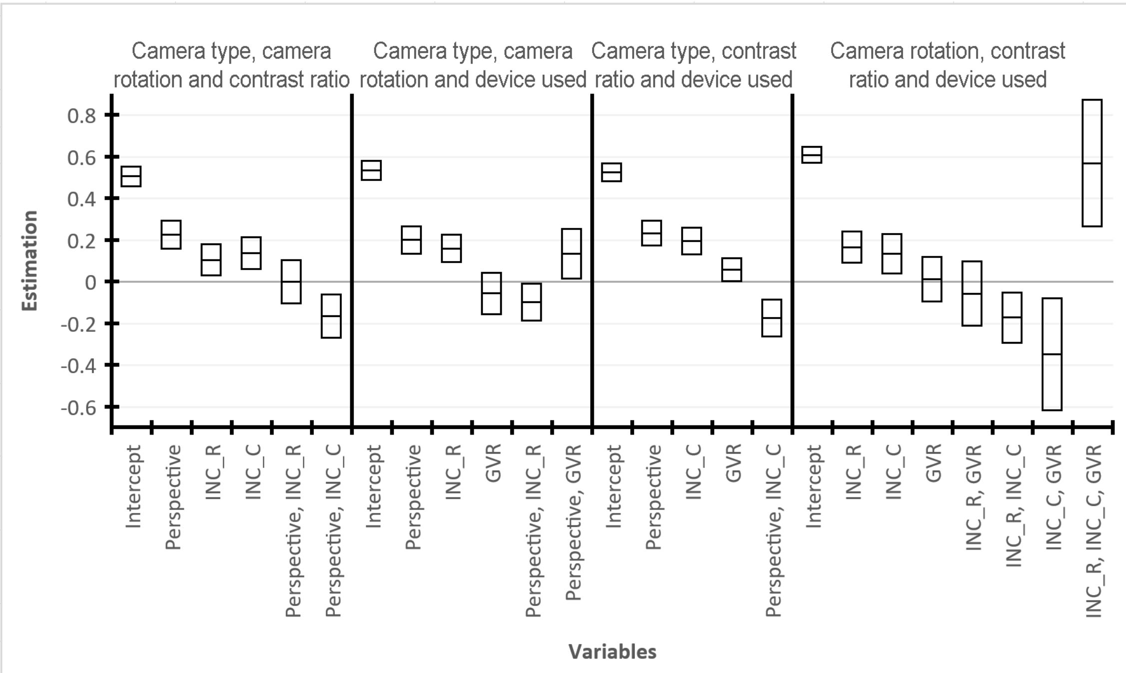

Figure 4.5: 95% CIs of the estimated coefficients in the case of the investigated variable triplets with interactions.

Concluding the analyses of the effects of triplets of variables, the following factors should be considered: camera type, rotation, contrast ratio, and the device used. According to the 95% CIs of the estimated coefficients in Figure 4.5 and to the results presented in Table A15, the most important interactions are between the following factors:

- Camera type – Camera rotation

- Camera type – Contrast ratio
- Camera rotation – Device used
- Camera rotation – Contrast ratio – Device used

Results of the analyses of effects of four factors without interactions

After concluding the analysis of triplets of factors, one more analysis remains: the analysis of all four significant factors. Therefore, the influence of the camera type, camera rotation, contrast ratio and the device used was assessed by combining these factor into a quartet. If the groups are constructed based on the possible quartets using the levels of the variables, $2^4 = 16$ groups are formed. Both the results of the logistic regression analysis and the numerical results are presented in the appendix in Tables A16 and B11, respectively. The 95% CIs of the estimated coefficients can be seen in Figure 4.6.

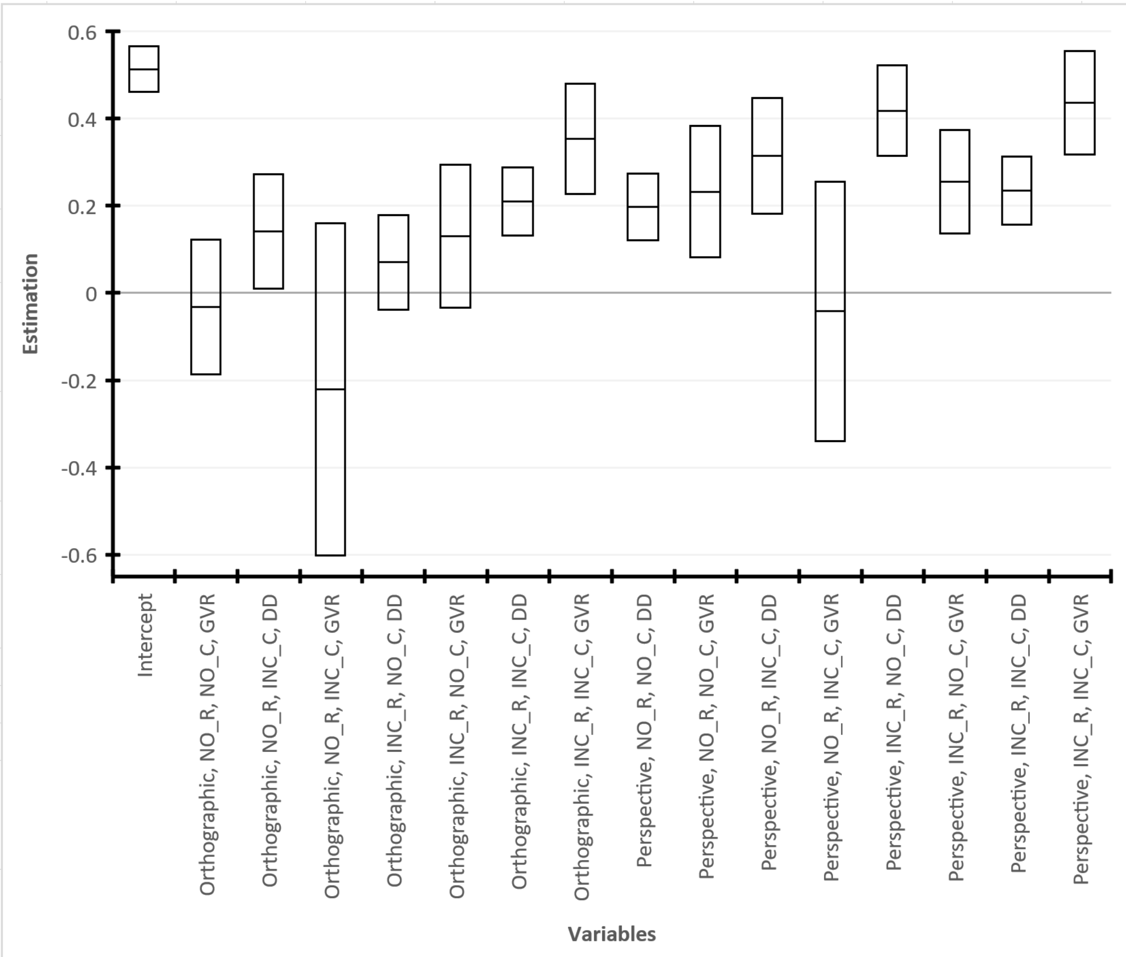

Figure 4.6: 95% CIs of the estimated coefficients in the case of the investigated variable quartet in which interactions are not allowed.

The point of reference was "Orthographic, NO_R, NO_C, DD". According to the results of logistic regression analysis, there is a significant improvement in every quartet that has a perspective value except "Perspective, NO_R, INC_C, GVR" compared to "Orthographic, NO_R, NO_C, DD". No significant influences are detected in the case of the variable quartets of "Orthographic, NO_R, NO_C, GVR", "Orthographic, NO_R, INC_C, GVR", "Orthographic, INC_R, NO_C, DD" and "Orthographic, INC_R, NO_C, GVR". The greatest improvements are in the case of "Perspective, INC_R, NO_C, DD" and "Perspective, INC_R, INC_C, GVR". The average rates belonging to these groups can be considered to be equal (*p-value* = 0.7627) in any case.

Results of the analyses of effects of four factors with interactions

To investigate the interaction of the variable quartet, a comparison was carried out between the different additive models. The model that uses 4 variables but in which interactions are not allowed (I), the model that uses 4 variables and in which the interactions of pairs are allowed (II), the model in which the interactions of three variables are allowed (III), and finally the model in which the interactions of all variables are allowed (IV). After comparison on the basis of the ANOVA analysis, model II proved to be significantly better than model I (*p-value* = 0.0004441), model III was significantly better than model I (*p-value* = 2.147×10^{-6}), and model III was also significantly better than model II (*p-value* = 0.0003342). Finally, IV was not significantly better than III (*p-value* = 0.1701). According to model III, the logistic regression analysis results are the following that are shown in Table A17 in the appendix. The 95% CIs of the estimated coefficients can be seen in Figure 4.7.

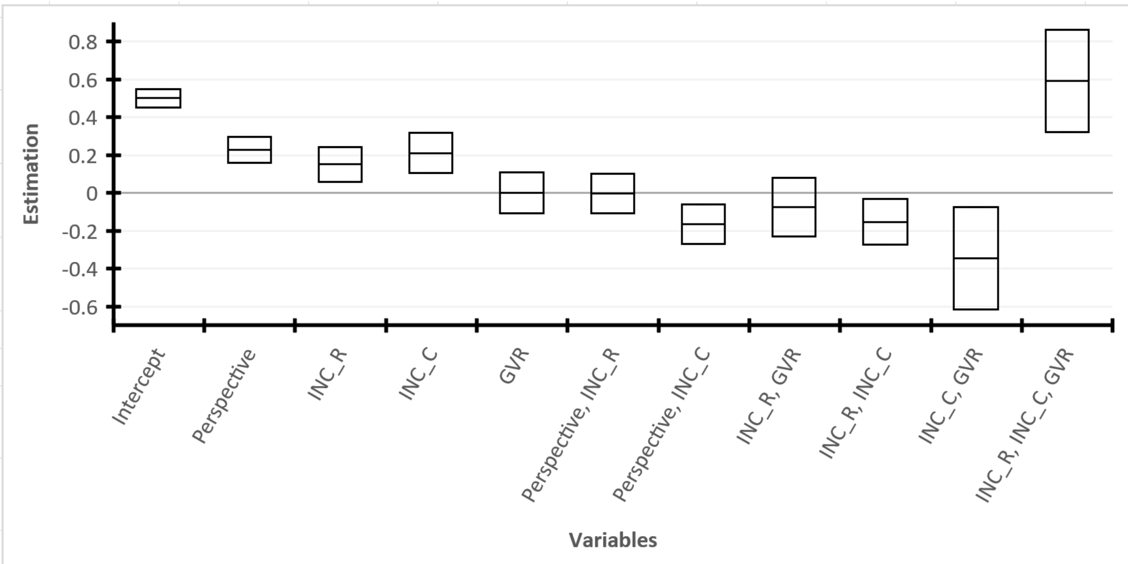

Figure 4.7: 95% CIs of the estimated coefficients in the case of the investigated variable quartet in which interactions are allowed.

Based on the logistic regression analysis results that are presented in Table A17, it can be concluded that the influence of the display devices in themselves cannot be detected (*p-value* = 0.987212). However, their interactions with other factors can be detected. In the end, after examining each factor on their own and in pairs, triplets and in a quartet, the final conclusion is the optimal results are provided by the quartet of "Perspective, INC_R, INC_C, GVR". This means that the spatial skills of those users who observe the VEs using the GVR HMD can be enhanced by the design of the VEs. Thus, the optimal spatial ability enhancement is achieved with the use of the GVR and with the following parameters in a VE: a virtual perspective camera with -45° or 0° or 45° rotations and 1.5:1 or 3:1 contrast ratios.

4.1.2 Investigating effects of display devices and human skills on the spatial ability test completion times

In this subsection, the effects of display devices and human skills on the spatial ability test completion times are analyzed. The test completion times are measured in seconds. A completion time of a test type is logged by the application after ten questions are completed by the user. This is due to one test type containing ten questions in the application. The smallest recorded completion time is 7.9 seconds and the largest recorded one is 1168.43 seconds which is approximately 20 minutes. The average of completion times on the tests is 200.388 seconds with a dispersion of 123.279 seconds.

The independence of time

It was analyzed whether the completion times are independent of the probabilities of correct answers on the tests. For this, the distribution of the completion times was investigated first. It is proven by the results of the investigation that the distribution of completion times is not normal as according to the results of the Kolmogorov-Smirnov test the *p-value* $< 2.2\times10^{-16}$. Due to this, the H of normal distribution is rejected. The histogram of the test completion times is presented in Figure 4.8.

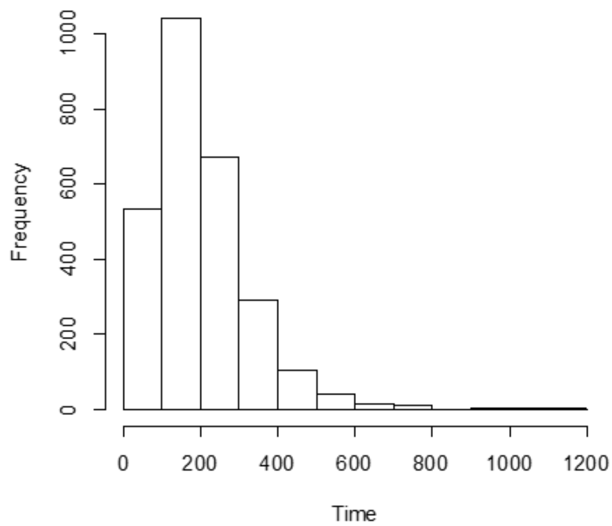

Figure 4.8: The histogram of the spatial ability test completion times.

Since both the probabilities of completion times and the correct answers are numerical values, the correlation coefficient can be used to evaluate whether there is a correlation between them or they are independent of each other. The numerical value of the correlation of completion times and probability of correct answers equals 0.223. Performing a test to check whether it can be considered zero or not, *p-value* $< 2.2\times10^{-16}$ was received as a result, therefore the H of the zero value of correlation is rejected. This rejection means that these variables are not independent of each other. The positive sign of the correlation coefficient means that when the completion time increases, the probability increases as well. It is shown by the value of the correlation coefficient that the linear relationship is not strong between the two variables. When the correlation coefficient of the logarithm of the time and the probabilities is looked at, a somewhat larger correlation is yielded, which is 0.299. The relations between the completion times and the factors are analyzed by regression as seen in Table 4.1.

Table 4.1: Results of the logistic regression analysis of the relation between the completion times and the probabilities of correct answers.

	Estimate	**Standard Error**	**z Value**	**Pr (>|z|)**
Intercept	0.232039	0.022724	10.21	$<2\times10^{-16}$
Time	0.002307	0.000099	3.19	$<2\times10^{-16}$

Due to the results presented in Table 4.1, the influence of the completion times is significant (*p-value* $<2 \times 10^{-16}$). Meanwhile, the positive sign of the estimated coefficient is 0.002307. It is shown by this estimated coefficient that when the completion time increases the probability increases as well, in tendency. Then, the logarithm of the test completion times was investigated with the logistic regression analysis and the coefficients are shown in Table 4.2. It is also presented in Tables 4.1 and 4.2 that the *p-value*s are equal. Both *p-value*s of the completion times and the logarithm of the completion times are the same (*p-value* $< 2\times10^{-16}$).

Table 4.2: Results of the logistic regression analysis of the logarithm of the completion times.

	Estimate	**Standard Error**	**z Value**	**Pr (>|z|)**
Intercept	-2.18981	0.08886	-24.64	$<2\times10^{-16}$
log(Time)	0.56389	0.01726	32.68	$<2\times10^{-16}$

Results of the analyses of a single factor's effects

In this subsubsection, the effect of a single factor on the completion times is evaluated. These factors are the following: the gender of the user, the primary hand of the user, the test type and the display device used. To calculate this, regression analysis was used in each case. Both the results of the regression analyses and the numerical results of all four variables can be seen in the appendix inside Tables A18 and B12-B15, respectively. The 95% CIs of the estimated coefficients can be seen in Figure 4.9 in the end of this subsubsection.

According to the results of the regression analysis, an effect on the completion times is produced by the gender of the user. This effect is also significant (*p-value* $= 2.97 \times 10^{-5}$). It is shown by the negative value of the coefficient that belongs to male students (-27.451) that the completion times of the male users are less than that of the female users.

The next to analyze was the primary hand of the user. The numerical averages of both the left-handed and right-handed users are almost the same: 199.535 seconds for the left-handed and 200.504 seconds for the right-handed students. Therefore, it is expected that there is no significant relation between the completion times and the primary hand of the users. After performing the regression analysis, these expectations are proved to be true (p-$value = 0.894$). Due to the results the completion times are not significantly affected by the primary hand of the users, thus this variable was omitted from further analyses. Therefore, it will not be paired or grouped in triplets with other factors.

When investigating the test types, it can be seen that the average is numerically larger in the case of the MRT test type. The numerical averages are 245.701 seconds, 176.156 seconds and 179.307 seconds, for the MRT, MCT and PSVT test types, respectively. It is shown by the results of the regression analysis that the average completion time of the MRT test is significantly larger (p-$value < 2 \times 10^{-16}$) than the average completion time of the others. When analyzing the average completion times of the MCT and the PSVT test types, no significant difference is detected between them (p-$value = 0.574$).

The last factor to be analyzed is the device used. It can be suspected from the numerical results as well, but according to the results of the regression analysis, a detectable difference in the test completion times exist in the case of the DD and the GVR. The spatial ability test completion times are significantly (p-$value < 2 \times 10^{-16}$) increased by the use of the GVR

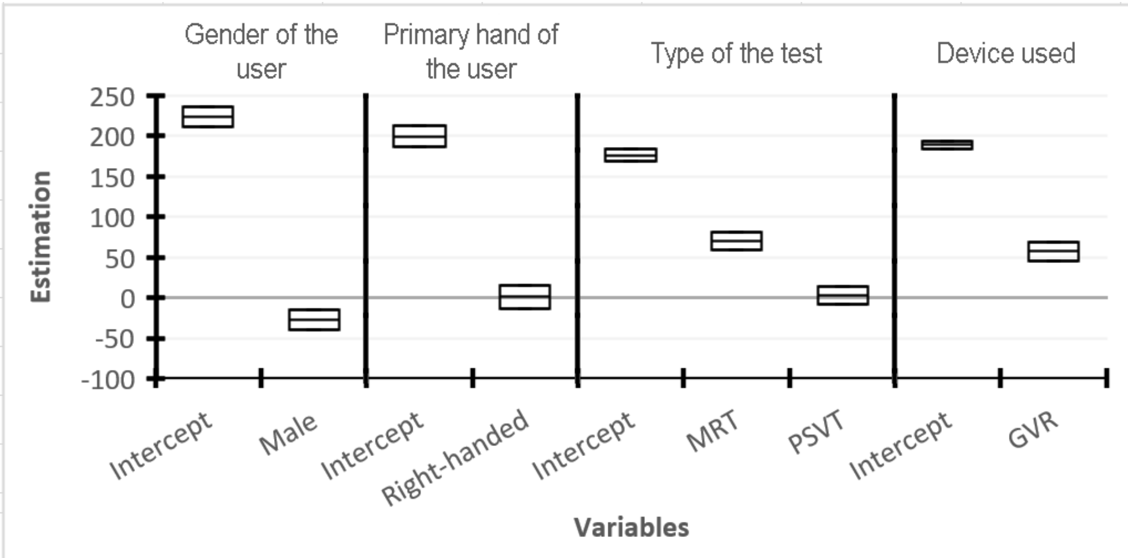

Figure 4.9: 95% CIs of the estimated coefficients in the case of one analyzed variable.

Results of the analyses of effects of two factors without interactions

After evaluating the effect of single variables, the investigation was continued in pairs. Since the primary hand of the user was omitted from further analyses due to having no significant effects on the completion times, only three pairs were made: the pair of gender and the test type; the pair of gender and the device used; and the pair of the test type and the device used. Similarly, to the previous subsubsection, both the results of the regression analyses and the numerical data are shown in the appendix in Tables A19 and B16, respectively. The 95% CIs of the estimated coefficients can be seen in Figure 4.10 in the end of this subsubsection.

In the case of the gender and the test type variable pairs, the results are the following: there are no significant differences in the completion times in the case of the MCT tests which are done by females and in the case of the PSVT tests which are done by females. Also, significant differences exist in the case of all the other pairs. Significantly smaller completion times are yielded by males who did the MCT tests and males who did the PSVT tests, while significantly larger completion times are yielded by females who did the MRT tests and males who did the MRT tests.

The next pair to investigate was the gender and the device used variable pair. According to the regression calculations, the females who used the GVR and the males who used the GVR have significantly ($p\text{-}value = 0.016046$ and $p\text{-}value = 0.000313$) increased completion times, when compared to the females who used the DD. However, significantly ($p\text{-}value = 0.000671$) smaller completion times are received by males who used the DD. Instead, by using the GVR, no significant difference could be found between the females who used the GVR and the males who used the GVR, as the $p\text{-}value = 0.3414$.

The last pair to investigate was the test type and the device used variable pair. In this case, the point of reference was the MCT and the DD pair. When comparing the other pairs to it, the following conclusion can be drawn: the completion times are significantly affected by every pair except the PSVT test type and use of the DD ("PSVT, DD").

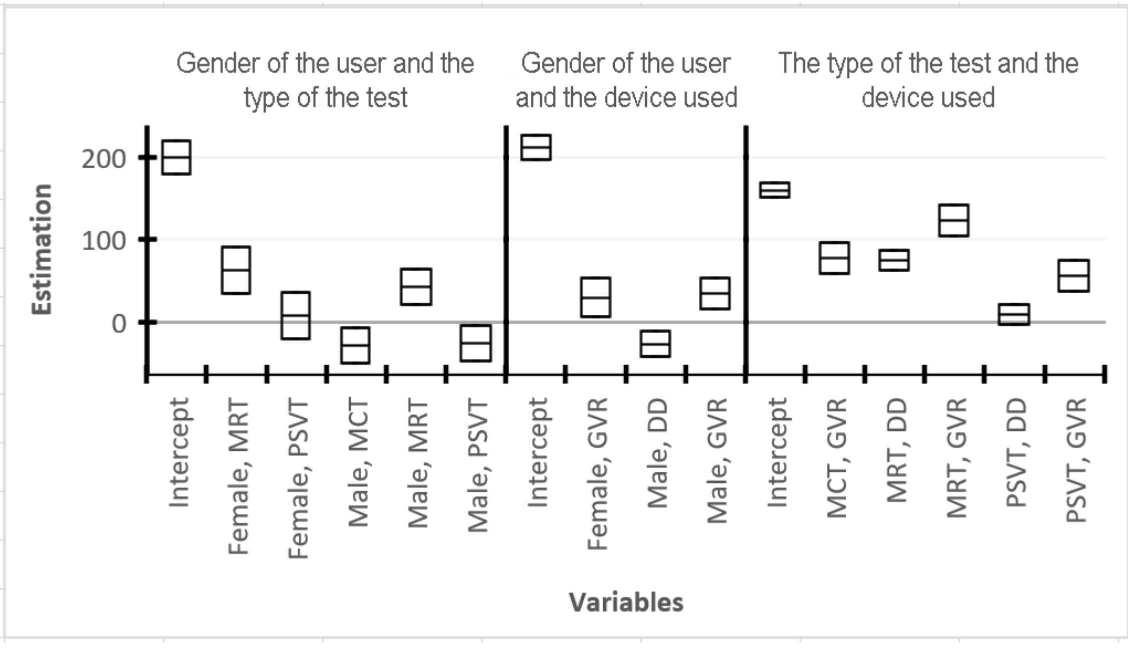

Figure 4.10: 95% CIs of the estimated coefficients in the case of the investigated variable pairs in which interactions are not allowed.

Results of the analyses of effects of two factors with interactions

The next to investigate was the interactions of the pairs. Similarly, to the previous subsubsection, the first pair to analyze was the gender of the user and the test type. After comparing the linear regression model which contains only the two factors to the model in which their interactions are taken into account, it can be concluded that the two models do not significantly differ from each other. The model in which the interactions are taken into account is analyzed by regression and the results are presented in Table A20 in the appendix, while the 95% CIs of the estimated coefficients can be seen in Figure 4.11.

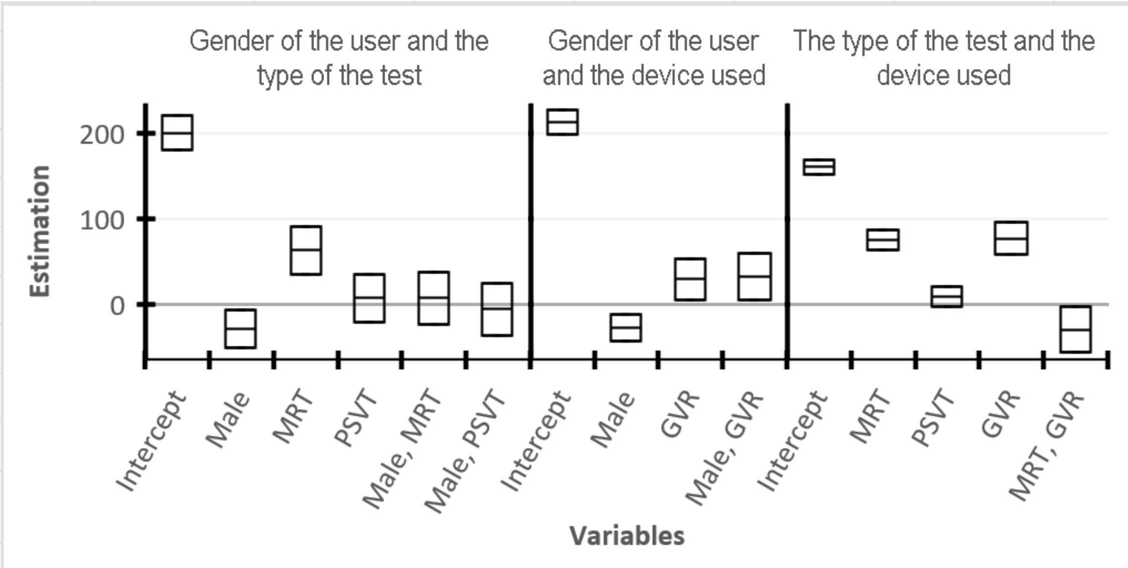

Figure 4.11: 95% CIs of the estimated coefficients in the case of the investigated variable pairs in which interactions are allowed.

The pair of the gender of the user and the test types was the first to be investigated by regression. According to the results, it can be concluded that the test completion times are significantly affected by the gender of the user and the MRT test type. However, no significant interactions exist between these two variables as can be seen from Table A20 and in Figure 4.11.

The pair of the gender of the user and the device used was investigated next. Thus, comparison was done between two models: in one model only the two factors are taken into account (I), while in the other their interactions are taken into account as well (II). Variance analysis resulted in $p\text{-}value = 0.02107$. This means that the model in which interactions are allowed is more appropriate. According to the regression analysis results presented in Table A20, the completion times are significantly affected by both the gender of the user and the device used. Besides the effects, interaction between them is also detectable. This coincides with the statement that the model allowing interactions describes better the phenomenon than the model in which interactions are not allowed. Moreover, due to the value of the estimated coefficient 32.309 the interaction of the gender and device used is very strong. Thus, it can be concluded that with use of the GVR, in case of male users, the completion times increase much more than in the case of the female users.

The last pair was the test type and the device used variable pair. Similarly, to the previous pair, the linear model in which only the factors are taken into account (I) was compared to the linear model in which their interactions are taken into account (II) as well. The latter model proved to be superior with *p-value* = 0.04093. The results of the regression analysis are presented in the third block of Table A20.

Therefore, according to the results presented in Table A20 it is shown that both factors have effects and interactions exist between them. The only one exception is when the PSVT test type and the DD are paired. It is shown by the negative sign of the estimated coefficient -29.243 that the increase in completion times is less in case of MRT test type using the GVR, than it would be expected by the sum of separate effects of test type and device used.

Results of the analyses of effects of all factors without interactions

After analyzing the pairs with and without interactions, the investigation of the effects of three variables is the next step. However, since only three variables remain after omitting the primary hand of the user, only one variable triplet can be created: the gender of the user, the test type and the device used. Both the results of the regression analysis and the numerical results are presented in the appendix in Tables A21 and B17, respectively. The 95% CIs of the estimated coefficients are shown in Figure 4.12.

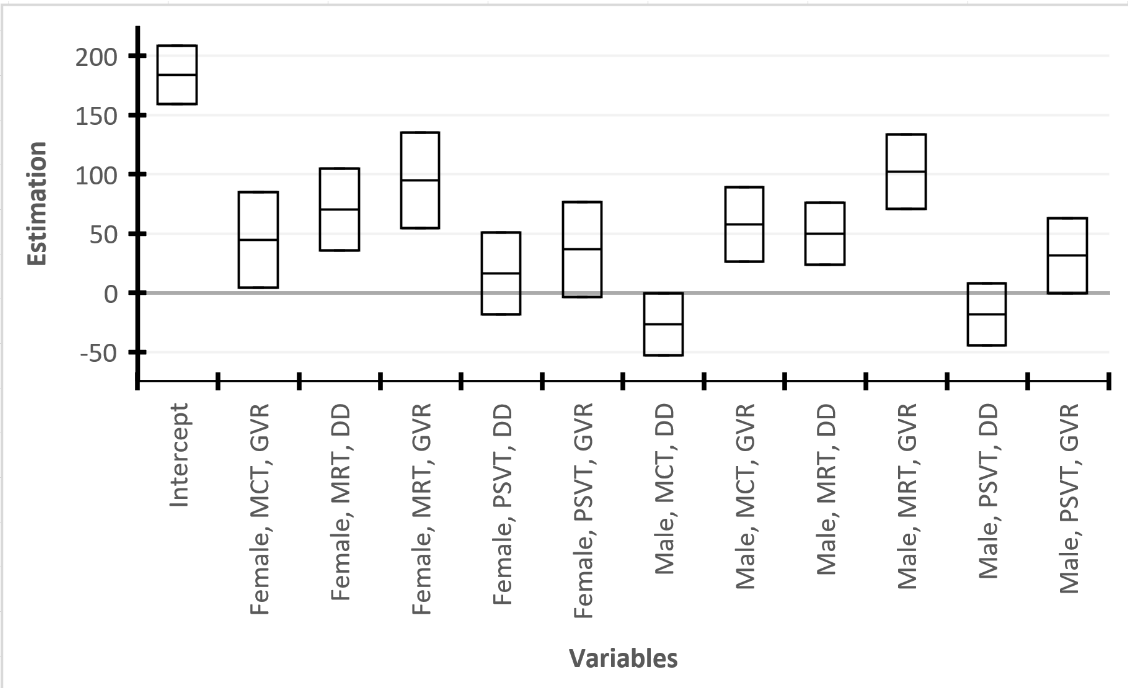

Figure 4.12: 95% CIs of the estimated coefficients in the case of the investigated variable triplets in which interactions are not allowed.

The point of reference was the "F, MCT, DD" triplet. Therefore, the remaining combinations were compared to it and the results are the following: the completion times are significantly smaller in the case of males who did the MCT test type with the DD (M, MCT, DD), while the completion times are significantly larger in the cases of females who did the MCT tests with the GVR (F, MCT, GVR), females who did the MRT tests with either the DD or the GVR (F, MRT, DD; F, MRT, GVR), males who did the MCT tests with the GVR (M, MCT, GVR) and males who did the MRT tests with either the DD and the GVR (M, MRT, DD; M, MRT, GVR). The largest increase of completion times is when the MRT test type is done by males using the GVR (M, MRT, GVR).

Results of the analyses of effects of all factors with interactions

What remains to be investigated is the only triplet with interactions. Therefore, comparisons were done between three models: the model in which the effects of the factors without interactions are investigated (I), the model in which the effects of variables are investigated, while interactions are also allowed (II), and the model in which the interactions of all factors are allowed (III). From the comparisons, it can be concluded that out of models I and II, model II is significantly better with $p\text{-}value = 0.0069$. However, there is no significant ($p\text{-}value = 0.9043$) difference between model II and III. Therefore, the results of the regression analysis of model II are presented in Table A22 in the appendix, while the 95% CIs of the estimated coefficients can be seen in Figure 4.13.

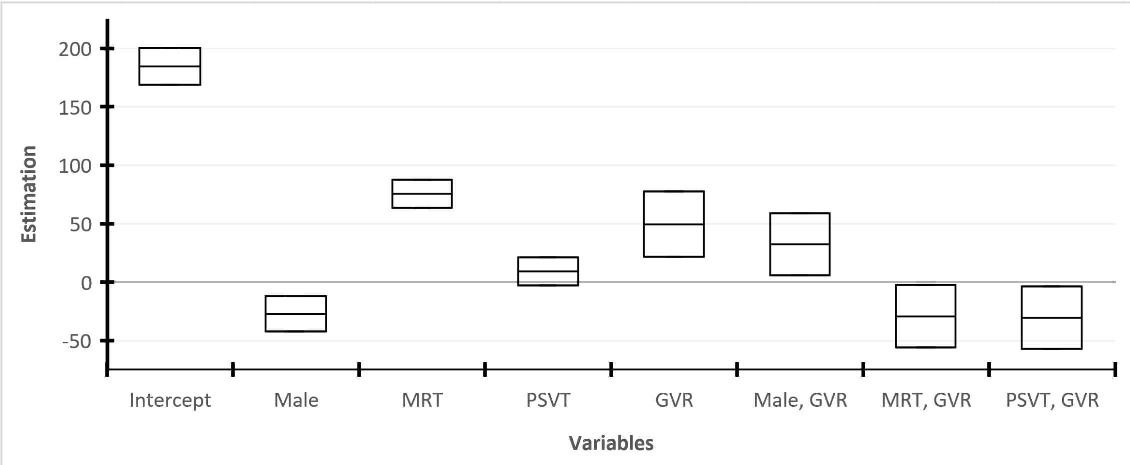

Figure 4.13: 95% CIs of the estimated coefficients in the case of the investigated variable triplets in which interactions are allowed.

It is shown by the results of the linear regression that out of the genders, the completion times are significantly decreased by the male users. Out of the test types, the completion times are significantly increased by the MRT test type. Similarly, out of the used display devices, the completion times are significantly increased by the use of the GVR. It can be concluded again that in the case of the male user, the completion times increases much more if the GVR is used, than in the case of a female user. Moreover, when using the GVR, the increase in completion times is smaller in the case of the MRT and the PSVT test types compared to the MCT test type.

4.1.3 Assessing the correlation between the used display devices and the human skills

In this subsection, the correlation between the used display devices and the human skills is analyzed. Five different aspects were investigated, grouped by display device: the difficulty of the tests; the rates of correct answers; the rates of correct answers regarding the gender; the rates of correct answers regarding the primary hand of the user; the rates of correct answers regarding age groups; and the rates of correct answers regarding the studies of the users. These are investigated in the next subsubsections.

The difficulty of the tests by display device

Based on the results on the tests, every test type is distinguishable from each other when using the DD or the GVR. There are significant differences between the test types. Although these differences can be seen in Tables C1-C4 in the appendix, it is suggested by the scores of the students as seen in Table 4.3 that the MCT is the hardest type of test. The MRT type is the easiest and the PSVT test stands between MRT and MCT in terms of difficulty. However, with the GVR, comparing only the dispersions, significant differences can be found between in some cases.

The equality of the standard deviations is accepted in the cases of "MRT, DD" and "MRT, GVR" (*p-value* $= 0.5476$), "MCT, DD" and "MCT, GVR" (*p-value* $= 0.2454$) and "PSVT, DD" and "PSVT, GVR" (*p-value* $= 0.8456$). However, the equality of average rates is only accepted in the first two cases: "MRT, DD" and "MRT, GVR" (*p-value* $= 0.8336$) and "MCT, DD" and "MCT, GVR" (*p-value* $= 0.1924$). In the case of "PSVT, DD" and "PSVT, GVR", the H of equality is rejected (*p-value* $= 0.0001766$), therefore the performance of the students who used the GVR is significantly better on the PSVT tests when compared to students who used the DD.

Table 4.3: The difficulty of the tests.

Variables	**Number of students**	**Min**	**Max**	**Average rates**	**Dispersion**
MRT, DD	240	0.4167	1.0000	0.8041	0.1334
MRT, GVR	61	0.4833	0.9833	0.8003	0.1248
MCT, DD	240	0.1333	0.9667	0.4389	0.1540
MCT, GVR	61	0.0000	0.8000	0.4071	0.1723
PSVT, DD	240	0.1333	0.9667	0.6168	0.1932
PSVT, GVR	61	0.0667	1.0000	0.7230	0.1886

The rates of correct answers by display device

The next to investigate was the rates of correct answers themselves. Normality analyses were done for both devices. The *p-value* $= 0.6335$ for DD and the *p-value* $= 0.2548$ for the GVR. This means that the Hs of Gauss distributions were accepted in the cases of both display devices. The Hs of the equality of the dispersions and the expectations are accepted (*p-value* $= 0.8786$, and *p-value* $= 0.3332$, respectively) thus there is no significant difference of the correct answers between the two devices. The rates of correct answers by display device are presented both in a numerical form in Table 4.4 and in a graphical form by plotting their empirical cumulative distribution functions (ECDFs) in Figure E1 in the appendix.

Table 4.4: The rates of correct answers by display device.

Variables	**Number of students**	**Min**	**Max**	**Average rates**	**Dispersion**
DD	240	0.3083	0.9667	0.6660	0.1185
GVR	61	0.3667	0.9000	0.6827	0.1199

The comparison of the rates of correct answers regarding the gender of the user by display device

The rates of correct answers regarding the gender of the user are also presented in a numerical form in Table 4.5, while the ECDFs can be seen in Figure E2 in the appendix. According to the results, in the case of the students who used the DD, the type of distribution was tested regarding their genders. The H of normal distribution was accepted with $p\text{-}value = 0.4846$ in the case of males and with $p\text{-}value = 0.9707$ in the case of females. Afterward, the equality of deviations is accepted with $p\text{-}value = 0.2213$, but the equality of average rates is rejected on the level of significance 0.00004. This means that the ratio of correct answers is significantly better for males in the case of the DD.

Next, the results on the GVR were investigated. A normality analysis was done with $p\text{-}value = 0.5377$ for males and with $p\text{-}value = 0.6657$ for females. The equality of the dispersions is accepted ($p\text{-}value = 0.5757$), and the equality of average rates is also accepted at a high-level of significance ($p\text{-}value = 0.6875$). Therefore, there is no significant difference between the results of the two genders.

Table 4.5: The comparison of the rates of correct answers regarding the gender of the user by display device

Variables	**Number of students**	**Min**	**Max**	**Average rates**	**Dispersion**
DD, Male	211	0.3083	0.9667	0.6769	0.1172
GVR, Male	44	0.3667	0.9000	0.6790	0.1247
DD, Female	29	0.4417	0.7833	0.5865	0.0965
GVR, Female	17	0.4417	0.8000	0.6922	0.1092

When comparing the results of males who used the DD to the results of males who used the GVR, the equality of the standard deviations ($p\text{-}value = 0.5609$) and the equality of the average rates ($p\text{-}value = 0.9193$) are both accepted. Moreover, the distribution of the two groups is tested, the equality of the distributions is accepted (*p-value* is 0.7931). Therefore, no significant difference exists between the two male groups. However, different results are yielded between the two female groups. When looking at the standard deviations, the equality is accepted with $p\text{-}value = 0.5513$, but when looking at the average rates, the equality is rejected with (*p-value* $= 0.0024$). According to the results, there is a significant improvement in the results when the GVR is used by female students. When the equality of the two distributions is tested, it is refused with *p-value* $= 0.02125$. Compared to the DD, the rates of correct answers of female students who used the GVR were significantly improved by 18.022% on average.

The comparison of the rates of correct answers regarding the primary hand of the user by display device

Next, the effect of the primary hand of the users was investigated and the results are presented in a numerical form in Table 4.6 in the end of this subsubsection and in a graphical form by plotting their ECDFs in Figure E3 in the appendix. In the table, the right-handed users are abbreviated as RH and the left-handed users are as LH. According to the results that are presented in the table, in the case of the DD, the H of normal distribution was accepted for both cases. For right-handed users, the H of normal distribution was accepted with *p-value* $= 0.5343$ and with *p-value* $= 0.9313$ for the left-handed users. The equality of dispersions was accepted with *p-value* $=$ 0.6567 and the equality of the expected values is also accepted with *p-value* $= 0.2796$. Due to the results, no significant difference was found between the right-handed and left-handed users when using a DD. In the case of the GVR, a normality analysis was also done: the H of Gauss distribution is accepted at high levels of significance (*p-value* $= 0.3623$ for the right-handed and *p-value* $= 0.9937$ for the left-handed users). The equality of the standard deviations is accepted (*p-value* $= 0.4826$), but the equality of average rates is rejected on the level of significance 0.05, due to *p-value* $= 0.02201$. Thus, with the use of the GVR, the performance of left-handed users was significantly better than their right-handed counterparts.

After analyzing the results separately by the device used, the next step was to compare them. Firstly, the two right-handed groups were compared to each other. The equality of the standard deviations and also of the average rates are accepted on very high levels (*p-value* $= 0.8633$ and *p-value* $= 0.9991$, respectively) and the equality of the distributions is also accepted (*p-value* $= 0.7086$). Therefore, no significant difference can be found between the two right-handed groups. However, different results are yielded between the two left-handed groups: the equality of the standard deviations is accepted (*p-value* $= 0.4164$), but the equality of average rates is rejected (*p-value* $= 0.006949$). This means that there is a significant improvement with the use of GVR in the case of the left-handed group.

Table 4.6: The comparison of the rates of correct answers regarding the primary hand of the user by display device.

Variables	**Number of students**	**Min**	**Max**	**Average rates**	**Dispersion**
DD, RH	213	0.3083	0.9667	0.6691	0.1177
GVR, RH	52	0.3667	0.8583	0.6691	0.1193
DD, LH	27	0.4417	0.8833	0.6414	0.1242
GVR, LH	9	0.6000	0.9000	0.7611	0.0938

The comparison of the rates of correct answers regarding the age groups by display device

The next to investigate was the age of the users. When analyzing the display devices separately, the following significant differences were found. In the case of the DD, significant differences appeared when comparing the results of the students who are less than or equal 18 years of age to students who are over 18 years of age. In the case of the GVR, the significant difference was between who are under or equal to 23 and who are over 23. Therefore, four age groups were made:

- DDU18: Students, who are aged 18 or are under and used the DD
- DDO18: Students, who are over 18 and used the DD
- GVRU18: Students, who are aged 23 or are under and used the GVR
- GVRO18: Students, who are over 23 and used the GVR

To investigate these age groups, the equality of standard deviations is accepted (*p-value* = 0.739) in category DDU18, GVRU23, and it is also accepted in category DDO18 and GVRO23 (*p-value* = 0.1794). The equality of expected rate values is accepted (*p-value* = 0.2784) in category DDU18, GVRU23. However, there is a significant difference between the DDO18 and GVRO23 categories (*p-value* = 0.02259). Therefore, the users who are over 23 and used the GVR during the tests are significantly better than the users who used the DD during the tests and are over 18. The results of the students in these age groups are presented in Table 4.7, while the age groups are more detailed in Tables C5-C8 in the appendix.

Table 4.7: The comparison of the rates of correct answers regarding the age groups by display device.

Variables	Number of students	Min	Max	Average rates	Dispersion
DDU18	34	0.3083	0.8583	0.6245	0.1224
GVRU23	37	0.3667	0.9000	0.6572	0.1297
DDO18	206	0.4083	0.9667	0.6728	0.1167
GVRO23	24	0.5417	0.8667	0.7219	0.0921

The comparison of the rates of correct answers regarding the studies of the user by display device

The last factor to compare was the rates of correct answers regarding the studies of the users. Since the students who tested using the DD were architectural/social engineering and mechanical engineering students, and the ones who tested with the GVR were IT and non-IT students, direct comparisons could not be made. However, the results on the different devices could be investigated separately. The rates of correct answers regarding the studies of the students can be seen in Table 4.8, where architectural/social engineering students are abbreviated as AE and the mechanical engineering students as ME.

Table 4.8: The comparison of the rates of correct answers regarding the studies of the user by display device.

Variables	Number of students	Min	Max	Average rates	Dispersion
AE	62	0.4083	0.8500	0.6460	0.1127
ME	178	0.3083	0.9667	0.6729	0.1200
IT	21	0.4417	0.8667	0.6845	0.1103
Non-IT	40	0.3667	0.9000	0.6817	0.1259

In the case of the use of the DD, normality analyses were executed. For architectural/social engineering students, the *p-value* = 0.8103 and for mechanical engineering students, the *p-value* = 0.8763. The results of the analyses are accepted. The equality of the standard deviations is accepted with *p-value* = 0.5774. The equality of average rates is also accepted with *p-value* = 0.1133. Therefore, as can be seen, no significant difference is detected between the results of architectural/social engineering and mechanical engineering students. In the case of the use of the GVR, the Hs of Gauss distribution were accepted (*p-value* = 0.9854 for IT students and *p-value* = 0.2599 for non-IT students). The equality of the dispersions is accepted (*p-value* = 0.5338), and so is the equality of the average rates (*p-value* = 0.9275). Therefore, there is no significant difference concerning the spatial skills measured by the tests in VEs between IT students and non-IT students.

4.2 Evaluating the APBMR algorithm

Since multiple types of evaluations were done, five subsections are contained in this section. In subsection 4.2.1, the real-time results of both computers are presented, while the results of the file-based evaluation is contained in subsection 4.2.2 and their execution times is compared in subsection 4.2.3. Afterward, the APBMR algorithm is compared to the previous one that it was based upon in subsection 4.2.4. Finally, all results are evaluated in subsection 4.2.5 using every piece of data.

Starting from this point, abbreviations will be used instead of the frequently occurring words or phrases. These abbreviations are the following:

- Average Gesture Acceptance Rate (AGAR)
- Arithmetic Mean Technique (AMT)
- Geometric Mean Technique (GMT)
- Harmonic Mean Technique (HMT)
- Contraharmonic Mean Technique (CHMT)
- Quadratic Mean Technique (QMT)
- Cubic Mean Technique (CMT)

4.2.1 Real-time results

As could be seen in Table 3.3 in the data collection subsubsection, 4 gestures were performed 10 times by each of the 32 people using the GC and by each of the 32 people using the AC. The results of both evaluations are shown in Tables D2 and D3, respectively, in the appendix. The results are also shown in Figures 4.14 and 4.15 in their respective subsubsections.

Real-time results with the general computer

In the case of the circular gestures as seen in the first block of D2, the best AGAR is provided by CHMT in the ±0.05 m AD (27.0%). In the ±0.10 m AD, the best AGAR is provided by the HMT (64.1%). The best AGAR in the ±0.15 m AD is provided by the AMT (87.1%). After the circular gesture, the waving gesture was investigated next and the results are shown in the second block of Table D2: in the ±0.05 m and ±0.10 m ADs the best AGARs are provided by the HMT with 76.2% and 95.7%, respectively. Contrary, the optimal AGAR of 97.3% is provided by the CHMT in the ±0.15 m AD. In the case of the forward-diagonal gestures, it is shown by the results that are presented in third block of Table D2 that the best AGARs are provided by the CHMT in both the ±0.05 m and the ±0.10 m ADs with 84.8% and 99.2%, respectively. In the case of the ±0.15 m AD, every forward-diagonal gesture is accepted with the HMT. The last gesture to be evaluated was the upward-diagonal gesture as seen in the last block of Table D2. According to the results, the best AGAR of 19.9% was provided by the CHMT in the ±0.05 m AD. In the remaining ADs of ±0.10 m and ±0.15 m, the best AGARs are returned by the AMT with 52.3% and 75.4%, respectively.

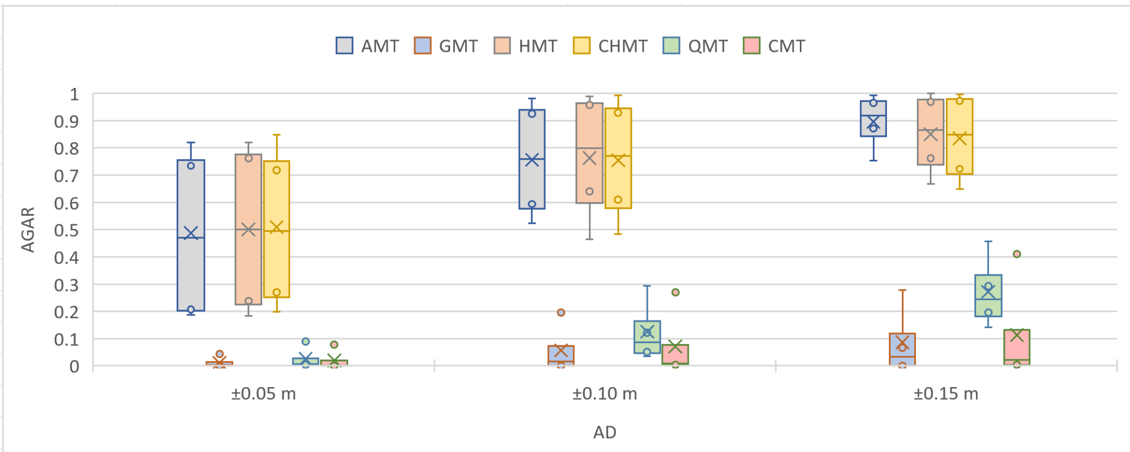

Figure 4.14: Real-time results using the GC.

Real-time results with the advanced computer

In the case of the circular gestures (first block of Table D3), the best AGAR is provided by the CHMT in the ±0.05 m AD with 23.4%. Contrary, the best AGARs of 73.0% and 98.4% are provided by the QMT in the ±0.10 m and the ±0.15 m

ADs, respectively. In the second block of Table D3 the case of the waving gesture is shown. According to the results, the best AGAR is provided by the CHMT in the ±0.05 m AD with 79.3%, while the best AGAR of 96.9% is provided by the AMT in ±0.10 m AD. In the case of the ±0.15 m AD, the best AGAR is provided by the HMT with 99.6%. In the case of the forward-diagonal gesture as seen in third block of Table D3, the best AGARs are provided by the CHMT in the ±0.05 m (65.6%) and in the ±0.10 m (92.2%) ADs, while the best AGAR is provided by the AMT in the case of the ±0.15 m AD with 96.1%. In the case of the upward-diagonal gestures, it is shown in the last block of Table D3 that the best AGARs are provided by the CHMT in all ADs: 48.8%, 89.8% and 94.9%, respectively. However, in the ±0.15 m AD, the same AGAR is provided by the HMT as the CHMT.

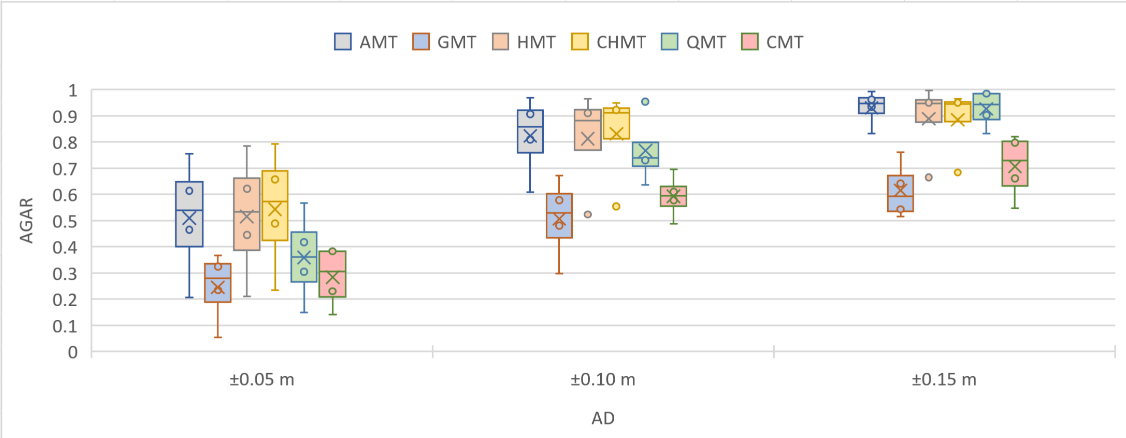

Figure 4.15: Real-time results using the AC.

4.2.2 File-based results

Contrary to the real-time measurements, during the file-based investigation 4 gestures were done 10 times by each of the 48 people as mentioned in the data collection subsubsection. In this case, the gesture descriptors were saved into a file and were evaluated by the APBMR algorithm. The results are shown in Figure 4.16 and are presented in a numerical form in Table D4 in the appendix.

In the case of the circular gestures, the best AGAR is provided by the CHMT in the ±0.05 m AD with 37.5%, while in the ±0.10 m AD, it is provided by the AMT with 64.3%. In the ±0.15 m AD it is provided by the QMT with 84.9%. In the case of the waving gestures, the best AGAR is provided by the HMT in the ±0.05 m AD with 62.9%, while in the remaining two ADs the best AGARs are provided by the CHMT with 91.5% and 97.1%, respectively. The latter is also equal to the AGARs in the case of the HMT and the CMT. According to the results of the forward-diagonal gestures, the AGARs are the best with the use of the CHMT in the ±0.05 m and ±0.10 m ADs with 72.1% and 88.3%, respectively. In the ±0.15 m AD, the best AGAR is provided by the AMT (94.8%). It is shown by the results of the upward-diagonal gestures that the best AGARs are provided by the CHMT in the ±0.05 m and ±0.10 m ADs with 45.5% and 76.6%, respectively. In contrast, the best AGAR is provided by the HMT in the ±0.15 m AD with 86.7%.

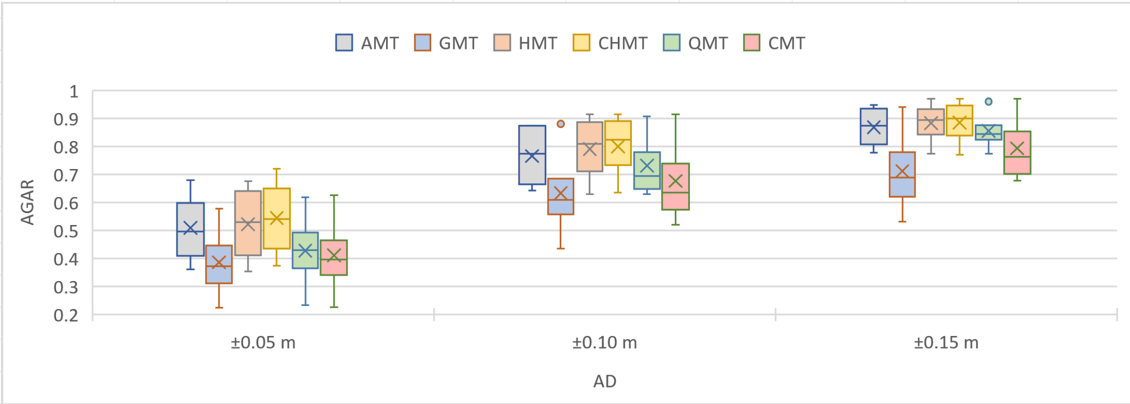

Figure 4.16: File-based results using the AC.

4.2.3 Comparing the real-time and the file-based execution time of the algorithm

The results of the execution time comparison can be seen in Figure 4.17. For the numerical form, see Table D5 in the appendix. According to Figure 4.17 and to the first block of Table D5, the real-time average execution times are faster than the file-based ones of the APBMR algorithm using the AC. This is an unexpected result. The largest execution time decrease is with the QMT (96.7%) and the least execution time decrease is 81.6% with the CMT. It is shown in the second block of Table D5 that even by using the GC, the algorithm is still superior in real-time, except in one case. The largest execution time decrease is with the QMT (87.5%), while the least execution time "decrease" is -23.8% with the CHMT which is the mentioned exception. Naturally, the minus sign means that the execution time increases. According to the results, there is a possibility that the file-based usage of the algorithm can be more time-consuming than when using it in real-time.

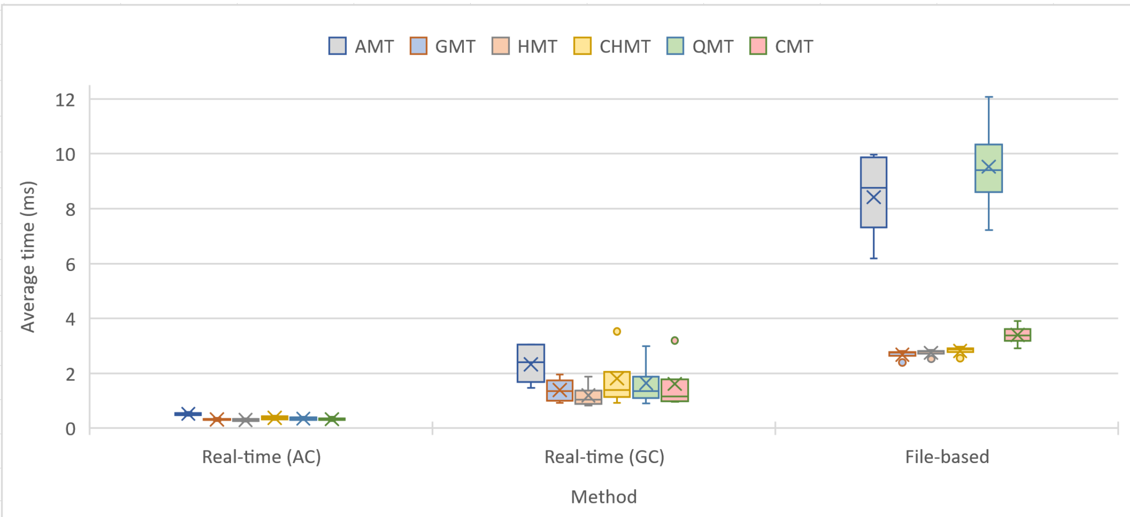

Figure 4.17: Comparing the execution time of the APBMR algorithm.

4.2.4 Comparing the APBMR to the RDAMR

The APBMR and the RDSMR/RDAMR algorithms are compared in this subsection. Although, the RDSMR is omitted from the comparison as its results can be influenced by the elapsed time between two gesture descriptors. Contrary, with the use of the RDAMR, the results cannot be influenced by the elapsed time. By this definition alone, a better rate of accepted gestures is provided by the RDAMR than the RDSMR.

As was mentioned earlier, the APBMR is built on the RDAMR. Therefore, the RDAMR works similarly: ADs are created as in the case of the APBMR, but only the first three gestures are used for their creation. Thus, during gesture recognition the ADs do not change. According to [76, 77] the RDAMR works and usable, however the not-changing ADs can cause problems, because the speed and the position of the user are not followed by the algorithm. In the case of the RDAMR, the same gesture is not accepted even if one of these factors changes. There is another difference between the two algorithms: with the APBMR, the gesture is evaluated when one is finished, while with the RDAMR, the gesture is evaluated during the movement.

To compare these two algorithms, the AGARs of the RDAMR algorithm and of each MTs of the APBMR algorithm were analyzed. It should be noted that only three ADs (±0.05 m, ±0.10 m and ±0.15 m) are generated and evaluated by the APBMR. However, it can be quickly observed that improved results are returned by the APBMR than by the RDAMR. Thus, the radius of the ADs with the RDAMR algorithm was increased until similar AGARs were received as with the APBMR. Since the gesture is evaluated differently, their execution times were not compared.

During the comparison, 4 gestures were done 10 times by each of the 32 people. To compare the algorithms, the results of the measurements were written into a file. Afterward, both algorithms evaluated the same movement descriptors from the file. The results of the comparisons can be seen in Figure 4.18 and they are also presented in a numerical form in Table D6 in the appendix.

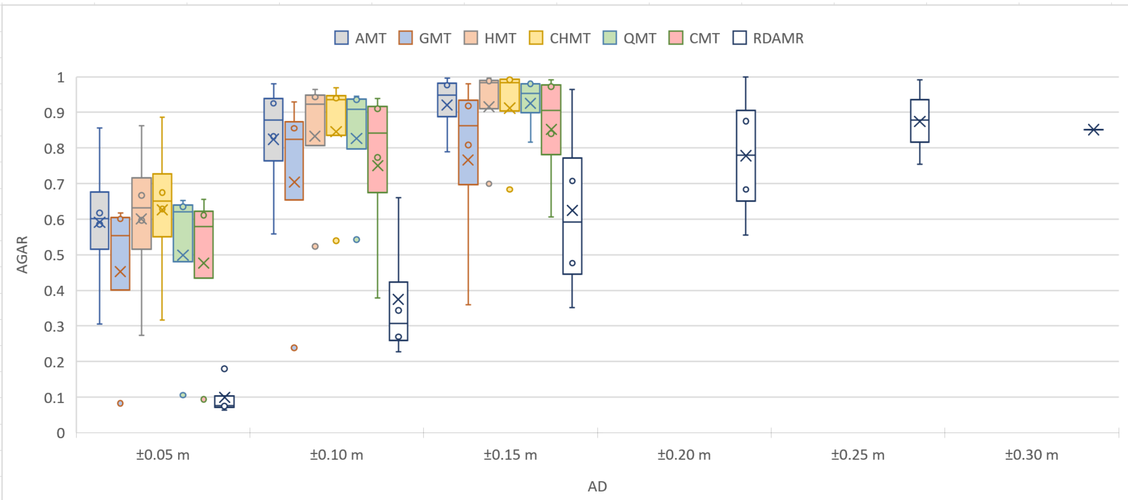

Figure 4.18: Comparing the MTs of the APBMR to the RDAMR algorithm.

Similarly, to before, the circular gesture was the first to be compared. According to the numerical data in the first block of Table D6, better AGARs are provided by all MTs of the APBMR than the by the use of the RDAMR algorithm. The difference between the AGARs of the two algorithms is very high in the cases of the AMT, HMT and CHMT. The AGARs when the GMT, QMT and the CMT are used are quite similar. Therefore, the APBMR is superior to the RDAMR in case of the circular gestures. Although, the results are more interesting in the case of the waving gesture. According to the numerical data in the second block of the table, the difference of the AGARs of the RDAMR between the ±0.05 m and ±0.10 m ADs is quite large. Also, the AGAR of the RDAMR in the ±0.10 m AD (66.0%) is similar to the AGARs of the APBMR algorithm in the ±0.05 m AD (61.7% - 67.5%, depending on the used MT). Contrary, in the case of the circular gesture, the ADs of the RDAMR are needed to be increased to ±0.15 m to have the same AGARs as the APBMR in the ±0.05 m AD. In the case of the forward-diagonal gesture as presented in the third block of Table D6, worse AGARs are returned by the RDAMR than in the case of the waving gesture. An AGAR of 88.7% is provided by the APBMR with the use of the CHMT in the ±0.05 m AD which is the optimal MT to be used in this case. Meanwhile, a similar AGAR is provided by the RDAMR with 87.5% in the ±0.20 m AD, which is quite a large AD. Similarly, to the forward-diagonal movement, the results of the upward-diagonal are alike to it. These results can be seen numerically in the fourth block of Table D6. In the ±0.20 m AD, an AGAR of 68.4% can be reached with the RDAMR algorithm which is slightly better than the ones in the ±0.05 m AD using the APBMR algorithm. The AGARs of the latter are between 58.6%-62.8% depending on the MT used.

According to Figure 4.18 and to the last block of Table D6, the results of all four gestures are summarized in case of both algorithms. As can be suspected, superior AGARs are provided by the APBMR. In the ±0.05 m AD, the increase of AGARs is between 358.2%-535.3% depending on the MT used, while in the ±0.10 m and ±0.15 m ADs it is 87.8%-125.4% and 22.7%-47.3%, respectively.

4.2.5 Evaluating all movement descriptors

All gesture descriptors that are evaluated with the APBMR algorithm are summarized in this subsection. This means that – according to Table 3.3 in the data collection subsubsection – 26880 lines of data are evaluated. The file-based measurements are included as well, meaning that the execution time of the algorithm can be a little higher. The results of the evaluation of all gesture descriptors are presented graphically in Figure 4.19, while the numerical results can be found in Table D7 in the appendix.

According to the results, it can be concluded that the best AGARs of 53.4% and 79.6% are provided by the CHMT in the ±0.05 m and ±0.10 m ADs. In the ±0.15 m AD, the optimal AGAR is of the AMT with 89.6%. In case of the AMT, HMT and CHMT, the dispersions are larger in the ±0.05 m AD, while they start to decrease when the AD is increased. In contrast, the dispersions of the other three MT increase when the ADs increase. Regarding their average execution times, the QMT and the AMT have the largest average with 4.655 ms and 4.424 ms, respectively. Although, their time dispersions are high as well.

The first AD to look at is the ±0.05 m AD. According to the numerical data, in case of the circular and upward-diagonal gestures, worse AGARs are returned than in the case of the waving and the forward-diagonal gestures. The best AGARs are provided by the CHMT in the ±0.05 m AD in the case of the circular, forward-diagonal and upward-diagonal gestures, which are 30.5%, 73.9% and 39.1%, respectively. In the case of the waving gesture, the best AGAR of 71.2% in the same AD is with the HMT.

The second AD to be evaluated is the ±0.10 m AD. In the case of the circular movements in the ±0.10 m AD, the best AGAR of 61.9% is provided by the AMT. In the same AD, the best AGARs are provided by the HMT in the case of the waving gesture and by the CHMT in the case of the forward-diagonal and upward-diagonal gestures with 94.1%, 92.5% and 72.3%, respectively. This means, that the optimal AGARs are provided by the CHMT in both the ±0.05 m AD and the ±0.10 m AD. It also can be seen that the optimal MT changes from HMT to AMT in the case of the waving gestures.

Lastly, in the ±0.15 m AD, the best AGARs of 83.7% and 96.4% are provided by the AMT in the case of the circular and forward-diagonal movements, respectively. For the waving and upward-diagonal gestures, the best AGARs are provided by the HMT with 97.8% and 83.4%, respectively. According to the results, the AGARs of the CHMT are better in the stricter ADs than in the ±0.15 m AD, while the AMT and the HMT prove to be superior in the ±0.15 m AD.

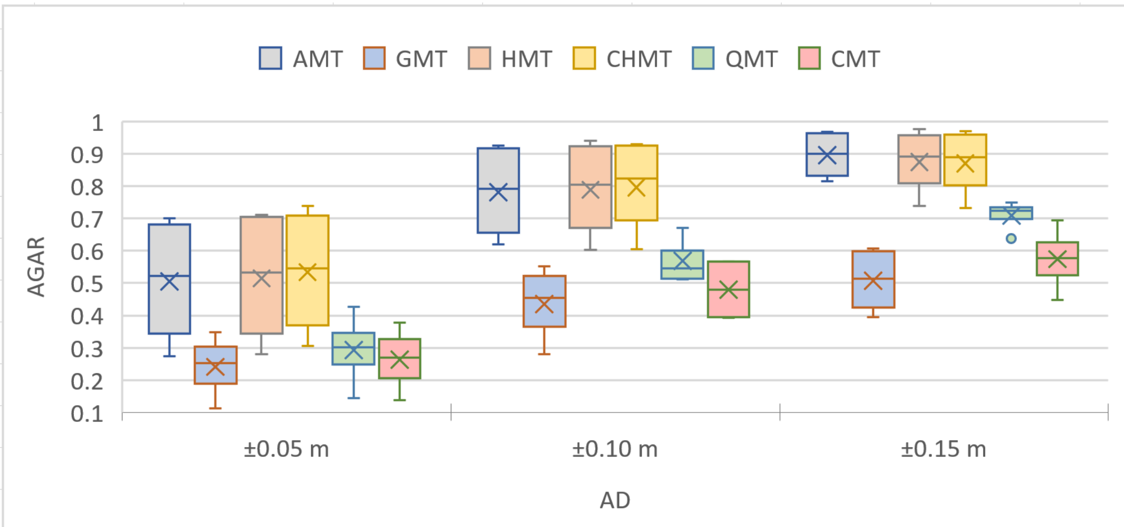

Figure 4.19: The results of the evaluation of all gesture descriptors with the APBMR.

Up to this point, all three axes were taken into account when evaluating the gesture descriptors. This was done because of the author's definition of accepted gestures. This definition is the following: the gestures are only considered accepted if more than 50% of their descriptors are inside the ADs on each axis. However, during this research only two axes at most are needed in the case of the evaluated gestures. Only the x, y axes are required by the circular, the waving and the upward-diagonal gestures, while only the x, z axes are required by the forward-diagonal gesture. Figure 4.20 presents the results of the evaluation on two axes, while the numerical data can be found in Table D8 in the appendix.

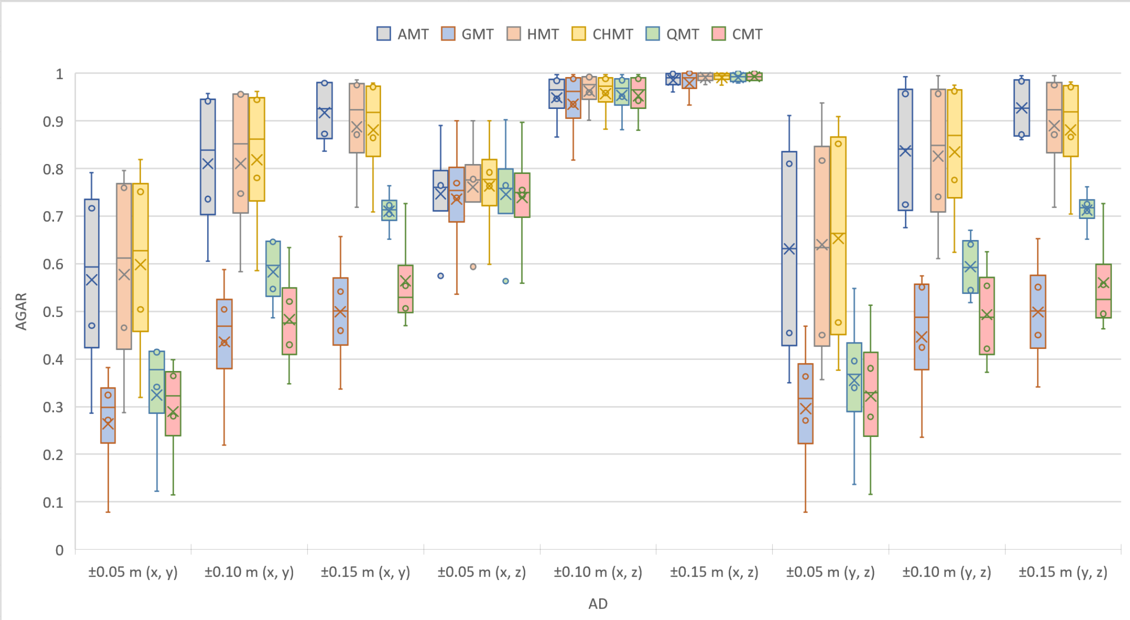

Figure 4.20: Evaluating the gesture descriptors with the APBMR on two axes.

What can be seen in the figure is that if only two axes are evaluated, the AGARs are higher than in the case of evaluating on three axes. The other thing that can be observed is that when the y axis is paired with another, the AGARs are lower than in the case of the x, z pair of axes. It is shown by the results of the evaluation of one axis (Figure 4.21) that the y axis has worse AGARs than the other axes. The numerical data can be seen in Table D9 in the appendix.

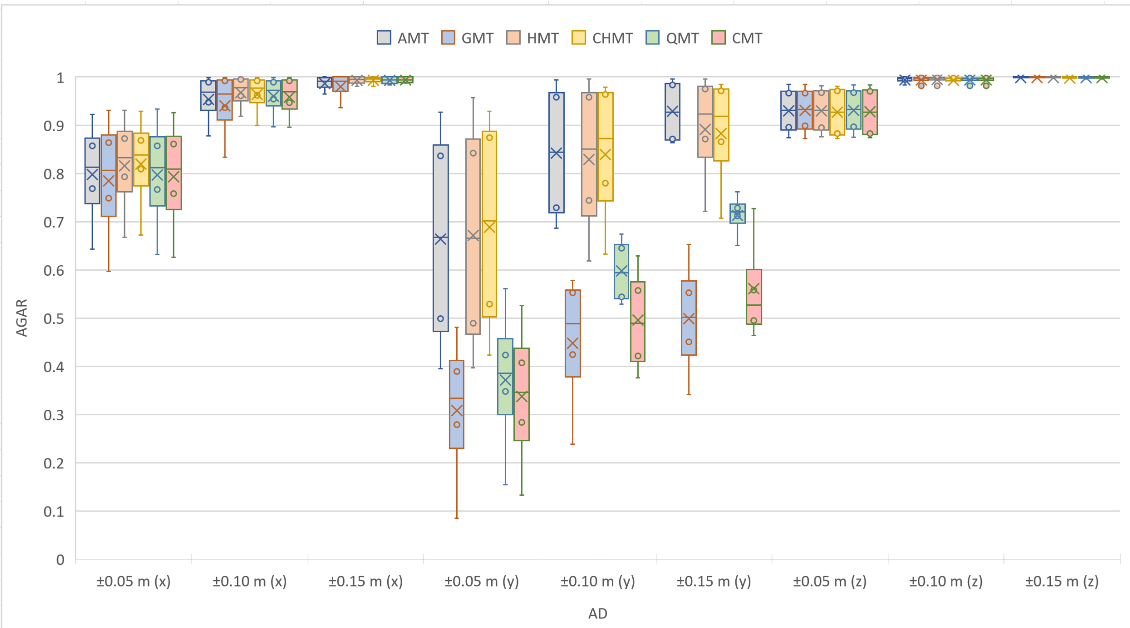

Figure 4.21: Evaluating the gesture descriptors with the APBMR on one axis.

Chapter 5

Discussion and conclusions

After presenting the results in the previous chapter, the discussion and the conclusions are done in this chapter. The results regarding the spatial skills of the users in VEs are discussed and concluded in section 5.1, while the APBMR is discussed and concluded in section 5.2. In each case, the RQs will be answered, and based on the Hs, theses (Ts) will be formulated.

5.1 Discussing the results on the spatial ability tests

In this section the results on the spatial ability tests in VEs are discussed. Since three groups of RQs and Hs were made regarding this part of HCI, this section is split into three subsections: in subsection 5.1.1 the factors that influence the probability of the results in VEs are discussed, in subsection 5.1.2 the factors that have an effect on the completion times are discussed and in subsection 5.1.3 the skills that can be affected by the display devices are discussed.

5.1.1 The factors that influence the probability of the results in virtual environments

Transitioning from paper to virtual is always difficult and this is not an exception in the case of the spatial ability tests. A different type of interaction is presented to the user by this new environment, namely by the VE. When designing VEs, it has to be kept in mind that interacting with virtual space and objects is not the same as interacting with real ones. These VEs are built differently and have graphics that are unlike reality. The goal of the author of this PhD dissertation was to make this HCI easier, and in order to achieve that, to present an optimal solution to make HCI in VR more effective.

According to the research data, an optimal solution was found, and the RQs were clearly answered in the first group. The effects of the parameters were demonstrated, and therefore, 1H5 was accepted, while 1H1, 1H2, 1H4, 1H6 were rejected and mixed cases are presented by 1H3, 1H7. These are elaborated in detail in the following subsubsections.

Rejected hypotheses - Detected influences

The first rejected hypothesis to discuss is 1H1. Originally, it was suspected that the probability of correct answers is influenced by the perspective camera type. The null H was that there is no effect, and the alternative H was that there exists some effect. According to Table A1, this latter proved to be the case. The probability of correct answers is positively influenced by the perspective camera type. However, as can be seen from Tables A10, A11, A12, A13, A15, A16 and A17, the results slightly changed when multiple factors were taken into account. This is due to VR being a complex, synthetic environment: in VR, no VE is made of only a single factor. Therefore, it is safe to assume that when the users are taken into the virtual space, multiple factors should be considered. That is why all factors are analyzed in pairs, triplets, and a quartet. After investigating every possible combination with the camera type, the perspective camera was demonstrated to be superior in all tests, and this was always an important factor. Therefore, *1T1* is formed*: when using a perspective camera, the performance of the users was significantly (p-value = 2.57×10^{-12}) influenced in terms of increasing their probability of answering correctly; and in pairs, it exhibited significant (p-value = 0.0459) interactions with -45°, 0°, 45° camera rotations or significant (p-value = 8.91×10^{-5}) interactions with the 1.5:1 and 3:1 contrast ratios; in triplets, it had no significant interactions; but in fours it exhibited significant (p-value = 0.000133) interactions with -45°, 0°, 45° camera rotations, 1.5:1, 3:1 contrast ratios, and the GVR.*

The existence of 1T1 is interesting, because when looking at the paper-based tests of all three types, the objects on the paper are drawn on the basis of orthographic projection. This fact can also lead to the question as to whether if the tests on the paper were changed to the projection of perspective type, would it also change the probability of the results of the testers?

The following rejected H is 1H2. Recall that the null H is that there is no effect; therefore, rejecting it means that its effect can be realized. This fact is interesting, because the camera has two types. For orthographic cameras, the FoV is undefined in VEs. For perspective cameras, the 45°, 60°, 75° and 90° FoVs are analyzed. Since it is stated by 1T1 that the perspective camera type is better than the orthographic type. It also is proven by Table A2 with respect to the FoVs, the main comparison was only carried out among the FoVs of the perspective camera as seen in Table A3. Due to these results, *1T2* is formed*: the performance of the users is influenced by the field of view of 90°, meaning that their probability of answering correctly is increased. This is a significant difference on the level 0.05, but not on the level 0.01 (p-value = 0.0225).*

The next rejected H to examine is 1H4. The contrast ratios are dealt by this H. According to logistic regression analysis results in Table A6, better probabilities of correct answers are produced by the smaller contrast ratios (1.5:1 and 3:1) than the larger ones (7:1, 14:1, 21:1). This was also confirmed by Table A7, in which the contrast ratios were grouped into two groups. In Tables A10, A11, A12, A13, A15, A16 and A17, the contrast ratios were examined in detail. Since this factor has a great influence on VEs, the interaction of the contrast ratio groups was also assessed. *1T4* is comprised by these facts*: the contrast ratios of 1.5:1 and 3:1 signifi-*

cantly (p-value = 2.56×10^{-6}) influence the performance of the users by increasing their probability of answering correctly, and in pairs they significantly (p-value = 8.91×10^{-5}) interact with the perspective camera type, and in triplets they significantly (p-value = 0.000237) interact with the -45°, 0°, 45° camera rotations and the Gear VR, while in fours they significantly (p-value = 0.000133) interact with the perspective camera type, the -45°, 0°, 45° camera rotations and the Gear VR.

After forming 1T4, let's think back to the paper-based tests. There are no contrast ratios on the paper-based tests. Everything is white, only the edges of the objects are black. A VE was made with brighter colors which is similar to the paper-based tests. However, the idea of using even brighter contrast ratios was discarded, as the eyes testers who used the GVR hurt after a few minutes with the use of the contrast ratio of 1.5:1. It is interesting to note that the best results are provided by using this contrast ratio in the VE, even numerically.

The last rejected H is 1H6. This was one of the most interesting Hs, as the aim was to compare the DD and the GVR HMD. According to Table A9, the GVR had a significant influence (and also, improvement) over the DD, as its HCI level was different. Since the display device used was one of the most important factors, and was influential, it was also examined in pairs, triplets, in fours, and its interactions were also assessed in Tables A10, A11, A12, A13, A15, A16 and A17. According to the results, when talking about unique influences of the display device used when investigated with more factors than one, it disappears in most cases, but its interactions remain. On the basis of these facts, *1T6* is formed*: in contrast to the desktop display, the use of the Gear VR significantly (p-value = 0.00677) increased the probability of correct answers on the tests and in pairs it significantly (p-value = 0.0326) interacted with the -45°, 0°, 45° camera rotations; in triplets it significantly (p-value = 0.000237) interacted with the -45°, 0°, 45° camera rotations and the 1.5:1 and 3:1 contrast ratios; and in fours it significantly (p-value = 0.000133) interacted with the perspective camera type, the -45°, 0°, 45° camera rotations, and the 1.5:1, 3:1 contrast ratios.*

Mixed cases

The first H that was mixed is 1H3. This is mixed because it was hypothesized that some rotation would help the users. This H was demonstrated to be true, but when no rotation occurred, it was also true. When the rotation was smaller than 45° in a given direction, the H is shown to be false by the results except in the case of 0°. It can be stated that the greatest influence on the results takes place when no, or a large rotation occurs. For this, see Table A4 and Table A5. Since the rotation is influential, it was also analyzed in pairs, triplets, and in fours, and its interactions were assessed in Tables A10, A11, A12, A13, A15, A16 and A17. Due to the results presented in the Tables, *1T3* was formed*: when rotating the camera -45°, 0°, 45°, the performance of the users will be significantly increased (p-value = 1.12×10^{-10}), increasing their probability of answering correctly and in pairs; it will significantly (p-value = 0.0459) interact with the perspective camera type; and in triplets it will significantly (p-value = 0.000237) interact with the 1:5.1, 3:1 contrast ratios and the Gear VR; and in fours it will have significant (p-value = 0.000133) interactions* *with the perspective camera type, the 1.5:1, 3:1 contrast ratios and the Gear VR.*

The second and final mixed-case H was 1H7, which is related to the optimal preferences in VEs for achieving the best HCI results. This H is only a mixed case, because originally the FoV of the camera was put into this H. However, for the optimal preferences due to similar results between the perspective camera type and the FoV, the latter was discarded and only the former was left in. Therefore, *1T7* is formed*: based on the previous theses, the optimal preference for the virtual environments to positively influence the correct answers on spatial ability tests by affecting the human–computer interaction is a perspective camera type, a camera rotation of -45° or 0° or 45°, a contrast ratio of 1.5:1 or 3:1, and the Gear VR display device.*

Accepted hypothesis - No differences detected

The first and only accepted H is 1H5, which deals with the presence of shadows in the VE. On the paper-based tests, there are no shadows, thus it was investigated whether the probability of correct answers were changed by their presence. According to Table A8, the probability of correct answers is not influenced significantly by the shadows. Therefore, the shadows were omitted from the multiple factor analyses. On this basis, *1T5* is formed*: the performance of the users is not significantly (p-value = 0.204) influenced by the shadows of the object in the virtual environment.*

5.1.2 The factors that have an effect on the completion times

According to the results in the previous chapter, only one H is accepted, four are rejected and two are mixed cases in the second group. The only H that is accepted is 2H3. The ones that are rejected are 2H1, 2H2, 2H4 and 2H5, while 2H6 and 2H7 are the Hs that are mixed cases. Therefore, this subsection is divided into three subsubsections: the rejected Hs, the mixed cases and the one accepted H.

Rejected hypotheses - Detected effects

The first H to be rejected is 2H1 which talks about the probabilities of the correct answers and the completion times being independent of each other. This was rejected since the correlation test resulted in 0.223 with *p-value* $< 2.2\times10^{-16}$, therefore it can be concluded that the two probabilities are not independent of each other. The same result (dependency) was concluded by logistic regression analyses in Tables 4.1 and 4.2. Therefore, *2T1* is formed*: the probabilities of the correct answers and the completion times are not independent of each other as the correlation coefficient is 0.223 and the p-value* $< 2.2 \times 10^{-16}$.

The second H to be rejected is 2H2 in which it is stated that a significant effect is produced by the gender of the user on the test completion times. This H was rejected due to the results presented in the first block of Table A18. It is proposed by the coefficient of -27.451 in the case of male users that the males have smaller test completion times than females. According to the *p-value* $= 2.97\times10^{-5}$, the effect of gender on test completion times is significant. Therefore, out of 2H2, *2T2* is formed*: not only the spatial ability average test completion times are smaller* *numerically by 12.274% in the case of male users, but the completion times are significantly (p-value* $= 2.97 \times 10^{-5}$*) affected by the gender of the users, making the completion times of male users smaller than that of female users.*

The third rejected H is 2H4. It is expressed by this H that the spatial ability test completion times of the users are affected by the type of the test. The largest average of time is provided by the MRT test type. There is a 39.479% increase in time between the MRT and the MCT test type, a 37.028% increase between the MRT and the PSVT test type and an 1.788% between the MCT and the PSVT test type. After the regression analysis, it can be proved by the results in the third block of Table A18 that the test completion time is significantly (*p-value* $< 2 \times 10^{-16}$) increased by the MRT test type. Therefore, *2T4* is formed*: the test completion times are significantly (p-value* $< 2 \times 10^{-16}$*) increased by the MRT spatial ability test type with an increment of 37.028% and 39.479%, when comparing it to the MCT and PSVT test types, respectively.*

The last H to reject is 2H5 in which the significant effect of different display devices is hypothesized. This is proved by the regression analysis with *p-value* $< 2 \times 10^{-16}$, which is seen in the fourth block of Table A18. Thus, *2T5* is formed*: compared to the desktop display, the spatial ability test completion times are significantly (p-value* $< 2 \times 10^{-16}$*) increased by 30.330% when using the Gear VR.*

Mixed cases

The first mixed case is 2H6, in which it is hypothesized that the largest completion times are achieved in the case of the MRT test type by male users who are right-handed and used the GVR. This is a mixed case, because from the first block of Table A18, it can be concluded that a significantly (*p-value* $= 2.97 \times 10^{-5}$) smaller average of completion times is achieved by the male users than the female users. Also, since it is proven by the results in the second block of Table A18 that the completion times are not significantly (*p-value* $= 0.894$) affected by the primary hand of the user, this can be omitted from this statement. However, according to the other part of the H in which it is suspected that the completion times are significantly increased by the MRT test type and by the use of the GVR. This suspicion is proved to be true in the third and fourth blocks of Table A18 with *p-value* $< 2 \times 10^{-16}$ in both cases.

Even though the completion time is decreased by the male gender in itself, when it is paired with the MRT test type (the first block of Table A19), the completion time of the tests is actually increased. When grouped into a triplet with the GVR and the MRT test type in Table A21, it has the largest increase in completion times as well. Therefore, due to these facts, *2T6* is formed: *the spatial ability test completion times are significantly (p-value* $= 2.51 \times 10^{-10}$*) increased by combining the male gender, the MRT test type and the use of the Gear VR. This triplet is also the largest increment.*

The second and last mixed case is 2H7, in which the inverse of 2H6 is suspected. As 2H6 is a mixed case, this is as well. As it is known from the previous results, the completion times are not significantly (*p-value* $= 0.894$) affected by the primary hand of the user. To form a thesis, the factors that decrease the completion times are needed to be known: in itself, the completion times are only decreased if the user is male (first block of Table A18). When paired with the MCT or the PSVT test type, the test completion times are significantly decreased with *p-value* = 0.009941 and *p-value* = 0.018158, respectively. If paired with the DD (Table A19), the completion times are significantly (*p-value* = 0.000671) decreased. To get the final result, it has to be grouped into a triplet. Therefore, from Table A21, it can be concluded that when the MCT and PSVT test types are done by using a DD, a decrease can be observed in the completion times. However, a significant (*p-value* = 0.047586) decrease in the completion times can only be found if the MCT test type is in the triplet as well. Also, this decrease is larger than in the case of the triplet with the PSVT test type. Therefore, from these facts, *2T7* is formed*: the spatial ability test completion times are significantly (p-value = 0.047586) decreased by the male gender, the MCT test type and the use of the DD which is also the largest decrease.*

Accepted hypothesis - No effect detected

In this group of Hs, 2H3 is the only one to be accepted. After the regression analysis, it is shown in the second block of Table A18 that the completion times are not significantly (*p-value* = 0.894) affected by the primary hand of the user. Due to this fact, *2T3* is formed*: the spatial ability test completion times are not significantly (p-value = 0.894) affected by the primary hand of the user.*

5.1.3 The skills that can be affected by the display devices

After analyzing the results on the correlation between the display devices and the human skills in the previous chapter, it can be concluded that a correlation was found. Regarding the Hs, the following can be concluded in the third H group: 3H3, 3H5, 3H7, 3H10, and 3H11 are accepted. Mixed cases are presented by 3H1 and 3H6, while 3H2, 3H4, 3H8, 3H9 are rejected.

Rejected hypotheses

The first H to reject is 3H2. During the investigation of 3H2, Table 4.5 is taken into account. Based on the statistical results, *3T2* is formed*: when using a desktop display, there is a significant difference between the results of males and females, meaning that the average performance of males on the tests is better than of females by 15.41%.* This fact is similar to older studies not featuring VR, such as [37], where it is proven that males have better spatial skills than females. When using the GVR, however, this significant difference disappears, meaning that the average performance of females on the tests is better than of males by 1.94%. This means that especially the women's achievements are improved by the GVR, which is an interesting result.

According to the results as can be seen in Table 4.7, the second H to be rejected is 3H4 in which the correlation between the age the DD is hypothesized. *3T4* is formulated according to the results*: the performance of the users who used a desktop display and are over 18 years of age is better by 7.73% on average than their younger counterparts.*

3H8 is also rejected. According to the results of the investigation of 3H8, *3T8* can be formed*: there is a significant difference in the performance of the left-handed and right-handed people who used the Gear VR. The difference is quite large, it is about 7%. This means that the performance of the left-handed people was increased significantly by the Gear VR: the increment is about 13%.* This result is quite different from older studies not featuring VR, such as [38] where it is concluded that the performance of right-handed people is better than of left-handed people. It seems like by using a DD their performance is made equal and by using the GVR, the performance of left-handed people is made better than of the right-handed ones.

Similarly, to 3H4, 3H9 is rejected as well due to the results presented in Table 4.5. *3T9* is formed*: the performance of the users who used the Gear VR and are over 23 years of age is significantly better than of their younger counterparts (The difference is 9.4% numerically). Comparing the displays, the performance of the users who used the Gear VR users and are over 23 years of age is significantly better than of the users who are over 18 years of age and used a desktop display (the difference is 7.29% on average).*

Mixed cases

Mixed cases are presented by both 3H1 and 3H6. What can easily be seen in Table 4.3 that the MCT test type is found the hardest and the MRT test type is found the easiest. Thus, the first half of both 3H1 and 3H6 is rejected, and the second half is accepted. Due to the results, *3T1* is formed*: the MCT test type is found the hardest, and the MRT is found the easiest in case of both display devices.*

When comparing the display devices, it can be concluded that a significantly better performance is achieved on the PSVT tests by the users who used the GVR than by their counterparts who used a DD. Therefore, *3T6* is the following*: while there are no significant changes in the ratio of correct answers in the MRT, MCT test types when comparing the desktop display to the Gear VR, the users who tested with the latter performed significantly better (17.217%) in the PSVT test type.*

Accepted hypotheses

Since the right side of the brain of the left-handed users is more often used, it is suspected that the performance of the left-handed people is better on the tests (alternative Hs of 3H3 and 3H8). However, according to Table 4.6 3H3 is accepted. Therefore, *3T3* can be formed*: there is no significant difference between the performances of left-handed and right-handed people who used a desktop display.* Concerning 3H5 and 3H10, the statistics can be seen in Table 4.8. Both 3H5 and 3H10 can be accepted. According to the results, *3T5* is formed*: there is no significant difference in the performance of architectural engineering and mechanical engineering students when using a desktop display.* Similarly, to 3H5, out of 3H10, *3T10* is formed*: there is no significant difference in the performance between IT and non-IT students when using the Gear VR.* Since the tests were conducted at two different universities, the majors of the students differ. However, it can be assumed after forming 3T5 and 3T10 that the performance of the users is not influenced by their majors.

Concerning 3H7, by using the GVR, no significant difference can be detected between the performances of male and female students. This is a different result in comparison to the results in the case of the DD. Thus, *3T7* is formed*: female users perform better numerically on the spatial ability tests than males by 1.94% on average with the Gear VR.*

Taking 3RQ11 and 3H11 into account when formulating 3T11, it can be seen from Table 4.4 that when using the GVR, better results are achieved on the tests. Therefore, 3H11 is accepted. Based on these facts, *3T11* is yielded*: there is no significant difference in the ratio of correct answers when comparing desktop displays with the Gear VR, but with the latter, the users produced numerically a better average of correct results by 2.5%.*

5.1.4 Comparing the results to the literature

As mentioned in chapter 2, the number of spatial ability tests in VEs is scarce and most studies only exist on paper. According to the paper-based results, the spatial skills of males are better than the spatial skills of females [37]. This is similar in the case of right-handed and left-handed people, meaning that the former group of people scored better on the tests than the latter group of people [38]. Another fact is that when using an HMD, significant improvements appear between two PSVT tests [99].

It is proven by the research that is presented in this PhD dissertation that the results which are similar to the paper-based tests can be gathered when using a DD. However, when using an HMD such as the GVR, the results change significantly. This means that when using a DD, there is a significant difference between the results of males and females. However, this significant difference disappears when using the GVR. A similar phenomenon can be observed between the right-handed and left-handed users. When using a DD, the performance of right-handed people on the tests is better than of their left-handed counterparts. However, with the use of the GVR, significantly better results are achieved on the tests by the left-handed users than the right-handed users. Also, according to the results presented in this PhD dissertation, the ratio of correct answers of the students who used the GVR is significantly better than their counterparts who used a DD.

According to the literature, the spatial skills of the users can be enhanced by a carefully designed VE: the camera should be carefully adjusted and binocular disparity should be provided to the users. Dynamic, blurry, faded, rich environments and motivation are also helpful in this regard. While different parts of VEs are analyzed in this PhD dissertation, it is proven by the results that the optimal preference for enhancing spatial skills is a combination of a perspective camera with a rotation of -45° or 0° or 45°, a contrast ratio of 1.5:1 or 3:1 and the use of the GVR.

Regarding the remaining results that are presented in this PhD dissertation, no study was found to compare them to. This is because the number of studies of spatial ability tests in VEs is small. However, according to the results, the spatial skills of the users can be enhanced by a carefully designed VE and with the use of the GVR. Therefore, HCI can be enhanced as well.

5.2 Discussing the results of the evaluation of the APBMR algorithm

The RQs in the fourth group were answered according to the results. In the previous section, the APBMR algorithm was evaluated with four gestures in real-time using the GC and the AC. File-based evaluation was also conducted. Therefore, based on the results, 4H2, 4H3, 4H5, 4H6 and 4H8 are accepted, while 4H1 is rejected. The mixed cases are 4H4 and 4H7. The one rejected H, the mixed cases and the accepted Hs are presented in subsections 5.2.1, 5.2.2 and 5.2.3, respectively.

5.2.1 Rejected hypothesis regarding the APBMR

The first and only rejected H is 4H1, because depending on the gestures, the best AGARs are produced by different MTs as can be seen in Table D7. In the case of the circular gesture in the ±0.05 m AD, the best AGAR of 30.5% is produced by the CHMT, while in the remaining two ADs, the best AGARs are provided by the AMT with 61.9% and 83.7%, respectively. In the case of the waving gesture, the best AGARs in every AD is provided by the HMT with 71.2%, 94.1% and 97.8%. With the forward-diagonal gesture, the best AGARs of 73.9% and 92.5% are provided by the CHMT in the ±0.05 m and ±0.10 m ADs, respectively. In contrast, the best AGAR is provided by the AMT in the ±0.15 m AD with 96.4%. In the case of the upward-diagonal gesture, the best AGARs in the ±0.05 m and ±0.10 m ADs are provided by the CHMT with 39.1% and 72.3%, respectively, while the best AGAR in the ±0.15 m AD is produced by the HMT with 83.4%. Due to these facts, *4T1* is formed*: the best AGARs are provided by different mean techniques in case of each gesture and each AD: in the ±0.05 m AD, the best AGARs in three out of four gestures are provided by the CHMT, while in the case of the remaining gesture, the best AGAR is provided by the HMT. In the ±0.10 m AD, the best AGARs in two out of four gestures are provided by the CHMT, while in the case of the remaining two gestures, they are provided by the AMT and the HMT, respectively. In the ±0.15 m AD, the best AGARs in two out of four gestures are provided by the HMT, while in the case of the remaining two gestures, they are provided by the AMT.*

5.2.2 Mixed cases regarding the APBMR

The first mixed case is 4H4. It was originally suspected that the GMT and CMT should not be used for prediction-based gesture recognition. However, similar results are produced by the QMT. It can be suspected that these bad results are due to the negative values of the movement descriptors. In the case of the Kinect, if the gestures are done left to the sensor or lower than it, the values of the gesture descriptors become negative values. One example is in the second block of Table D4, where the waving gesture was performed and similar AGARs were produced by the GMT as the other MTs. In the case of this measurement, the users stood in front of the Kinect, faced the sensor and the gestures were done on the right side of it, and in a higher position than it.

In contrast, the other waving gesture that is presented in the second block of Table D2 resulted in AGARs of 0% in all ADs. This is because the users were closer to the center the Kinect during these measurements, therefore mostly negative values were assigned to the gesture descriptors. Thus, *4T4* is formed*: the GMT, QMT and CMT should not be used for prediction-based gesture recognition as better AGARs are provided by the AMT, HMT and CHMT. From the ±0.05 m AD to the ±0.15 m AD, the AGARs of the former are 24.1%, 43.5% and 50.8%, while of the middle are 29.3%, 56.8% and 70.9%; and of the latter are 26.3%, 48.0% and 57.4%.*

The results regarding 4H7 are very interesting and came as a surprise. As can be seen in Table D5, this is only a mixed case because of the "GC 2 and File 2" columns. According to the results in the case of the waving gesture, the file-based gesture prediction and acceptation are faster by 23.8% using the CHMT. In the remaining cases, the execution times of the algorithm are decreased in real-time using either the GC and the AC. In the case of both computers, the decrease in the execution times is the smallest using the CMT, while the largest is in the case of the QMT. Therefore, *4T7* is formed*: except for using CHMT to evaluate the waving movement using the general computer, the use of the real-time feature of the APBMR algorithm can decrease the execution time compared to file-based usage by at least 6.9% to 96.7% at most.*

5.2.3 Accepted hypotheses regarding the APBMR

4H2 is the first H to be accepted. According to the first block of Table D7, the worst AGARs are produced by the GMT in every AD: for the ±0.05 m AD it is 24.1%, for the ±0.10 m AD it is 43.5% and for the ±0.15 m it is 50.8%. Worse AGARs are also returned by two other MTs than the remaining three. These two are the QMT (29.3%, 56.8%, 70.9%) and the CMT (26.3%, 48.0%, 57.4%). From the results presented in the blocks of Table D7, it can be concluded that the worst AGARs are provided by the GMT in case of every gesture. Due to the results, *4T2* is formed: *the worst AGARs are provided by the GMT in case of every gesture: for the circular movements, the AGARs of the GMT are 11.3%, 28.0% and 39.4% in the ±0.05 m, ±0.10 m and ±0.15 m ADs, respectively. In case of the waving gestures, the AGARs are 34.7%, 55.1% and 59.5% in the same respective ADs. For the forward-diagonal gestures, the AGARs are: 21.3%, 39.4% and 43.4% in the same respective ADs. In case of the upward-diagonal gesture, the AGARs are 29.0%, 51.3% and 60.8% in the same respective ADs.*

To assess 4H3, the first block of Table D7 should be taken into account. According to the table, the best AGARs are provided by the CHMT in both the ±0.05 m and ±0.10 m ADs with 53.4% and 79.6%, respectively. In the ±0.15 m AD, the best AGAR is provided by the use of the AMT with 89.6%. Due to these facts, *4T3* is formed*: when using the APBMR algorithm for prediction-based motion analysis, the optimal AGARs are provided by the CHMT in both the ±0.05 m and ±0.10 m ADs with 53.4% and 79.6%, respectively. In case of the ±0.15 m AD, the optimal AGAR of 89.6% is resulted by the use of the AMT.*

The next accepted H is 4H5, for which the answer can be calculated from each gesture that is presented in the blocks of Table D7. In the case of the circular gesture, the largest numerical difference between the ± 0.05 m and ± 0.15 m ADs is 0.584 when using QMT and the smallest is 0.281 when using the GMT. For the waving gesture the largest numerical difference between the same two respective ADs is 0.322 with the QMT and the smallest is 0.227 with the CMT. In case of the forward-diagonal gesture, the largest numerical difference is 0.357 with the QMT, while the smallest difference is 0.216 with the CHMT. For the upward-diagonal gesture, the largest numerical difference is 0.469 with the HMT and the smallest is 0.318 with the GMT. *4T5* is formed due to these results*: different numerical differences exist between the AGARs in case of each gesture: with the circular movements, the numerical differences are between 0.281 - 0.584. In the case of the waving gestures, the numerical differences are between 0.227 - 0.322. With the forward-diagonal gesture, the numerical differences are between 0.216 - 0.357, while the numerical differences are between 0.318 - 0.469 with the upward-diagonal gesture.*

4H6 is the following H to be accepted. All gestures had to be evaluated on each axis for it to be accepted and the data regarding it can be seen in Table D9. On the x axis, the largest numerical difference is 0.201 with the CMT, while the smallest numerical difference is in the case of the CHMT with 0.174. On the y axis, the largest numerical difference is in the case of the QMT with 0.341 and the smallest numerical difference is in the case of the GMT with 0.191. On the z axis, the largest numerical difference is with the CMT (0.072), however the CHMT is a close second (0.071). Here, the smallest numerical difference is in the case of the QMT with 0.068, however the numerical difference of 0.069 is yielded by the remaining three MTs. Thus, *4T6* is formed*: numerical differences exist between the AGARs on each axis: on the x axis, the numerical differences are between 0.174 - 0.201, while the numerical differences are between 0.191 - 0.341 and 0.068 - 0.072 on the y and z axes, respectively.*

The last H that is accepted is 4H8, due to the results presented in Table D6. According to the table, all MTs of the APBMR in case of all gestures have better AGARs in every AD than the AGARs of the RDAMR. Even with the worst AGAR in case of all gestures using the GMT, still better AGARs are provided by the APBMR than they are provided by the RDAMR in each AD. Due to the results, *4T8* is formed: *in the ± 0.05 m AD, the AGARs of the APBMR algorithm are between 45.2%-62.7% depending on the used MT, while the AGAR of the RDAMR algorithm is only 9.9%. In the ± 0.10 m AD, the AGARs of the APBMR algorithm are between 70.4%-84.5% depending on the used MT, while they are 37.5% in the case of the RDAMR algorithm. In* ± 0.15 *m AD, the AGARs of the APBMR are between 76.7%-92.6% depending on the used MT, while in the case of the RDAMR algorithm it is only 62.5%. However, in the ± 0.25 m AD, a similar AGAR to the AGAR of the APBMR in the* ± 0.15 *m AD is reached by the RDAMR algorithm.*

Therefore, worse AGARs are provided by the RDAMR than by the APBMR because the gestures are "taught" to the algorithm using only the first three gestures. In the case of each future gesture, only the first three gestures are referenced. Thus, the generated ADs will keep on repeating. Also, the same speed of the first three gestures has to be matched by the user. If the user can do this, a good AGAR could be reached with the RDAMR. However, if the speed is different, then the same gestures are not accepted by the RDAMR algorithm.

5.2.4 Comparing the APBMR algorithm to the literature

The APBMR and its predecessor (the RDAMR) were developed to maintain motivation because it is crucial in rehabilitation. A lack of motivation can appear as a generator and/or as a consequence of depression. A failure in rehabilitation can be caused by the effect of these two [211, 212]. Therefore, to increase or to maintain motivation it is important for the application to adapt to the skills of the users. According to [213], if the application is calibrated beforehand or correction values are set, the application could not adapt well to the skills of the users. However, since all users are different and all can be in different phases of rehabilitation, not to mention that telerehabilitation is in the focus of this part of this research, an algorithm was needed to adapt to the skills of the users "on the fly". This means that an algorithm which can run on a "general computer" and which does not require a prior data set (or only a small amount of data should be used) to not demotivate the users.

Therefore, before the APBMR the RDAMR was developed [76, 77] and according to the referenced studies, it is a viable algorithm to maintain motivation. However, there were problems regarding the position of the user in front of the Kinect sensors and the speed of the gestures. To solve these problems, the APBMR was developed. Not only the problems were solved by the APBMR, but its AGARs are far superior to the RDAMR algorithm.

5.3 Conclusions

It is important to develop spatial skills in the modern day and age. In the last century, mostly paper-based methods were available, although their number is many. Since the world is transitioning into a digital age, virtual versions of these tests can be created or they can even be improved upon. Therefore, a VR application was designed and developed. Three old methods (the MRT, MCT and the PSVT test types) are presented in a new context by this application. The measurements could be gathered in real-time from the spatial ability tests and the application is also available on two platforms: PC and Android. This application was tested by 61 students on the Android platform with the GVR and by 240 students on the PC platform with an LG DD. Data were gathered from the users and were analyzed. The data regarding the users consisted of the age of the user, gender of the user, primary hand of the user, the number of years spent at the university and what the student is majoring in. Technical data was also gathered such as the virtual camera type, its FoV, its rotation, the contrast ratio and the existence of shadows. The test type, its completion time and the number of correct answers were logged as well.

After presenting and investigating the results of these measurements, three things can be concluded:

1. The optimal user-centric preference in VEs, presented in subsection 5.3.1.
2. The independence and the effect on completion times in VEs, shown in subsection 5.3.2.
3. The correlation between the display devices and the human skills, as can be seen in subsection 5.3.3.

Not only the spatial skills of the users are important in VEs, but the HCI – mainly in rehabilitation – as well. Therefore, in this PhD dissertation a new method is proposed. This is called the Asynchronous Prediction-Based Movement Recognition algorithm with the aim to help the physical rehabilitation of people with movement disabilities using sensors. After the implementation, it was tested with the Kinect v1 at the laboratory. While it was only tested with the Kinect v1 sensor, in principle, it can be used with any motion sensor that returns coordinates in real-time. For the research, four gestures had to be done by four groups of users and each gesture had to be repeated ten times. Afterward, the data was evaluated with six MTs. After presenting and evaluating the results, it can be concluded in subsection 5.3.4 whether the APBMR is viable and usable for telerehabilitation.

5.3.1 The optimal user-centric preference in virtual environments

Designing VEs is not an easy task, even if the VE is a virtual version of something in reality. The aim of this research was to find the factors which positively influence users in VR. The virtual camera types, FoVs, and rotations, the contrast ratios were analyzed between the foreground object and the background, the existence of shadows, and the display device used. This was investigated with the logistic regression analysis. It is shown by the results that HCI can be affected by the display factors and devices. While these factors and devices all have a unique influence, it has to be kept in mind that no VE exists comprising only one of these factors. Therefore, these factors will always be in effect with multiple others. On this basis, these factors were analyzed in pairs, in triplets and in fours. Some lost their unique influences, but interactions emerged between them. These interactions change depending on the number of examined factors. Many retain their interactions, but most of their significances are lost. Thus, VEs should be carefully designed.

In conclusion, the users can be positively influenced in their tasks by a carefully designed VE: according to the results, a perspective camera type, a camera rotation of -45° or 0° or 45°, a contrast ratio of 1.5:1 or 3:1 and the Gear VR HMD proved to be the optimal factors in VEs. When the user is in the VE with these factors and display devices, their probability of correct interaction, and even the results, can be increased.

5.3.2 The independence and the effect of completion times in virtual environments

The completion time of the spatial ability is not affected by one investigated factor only. Therefore, the factors were evaluated one-by-one, in pairs and in a triplet. From the one-by-one analyses, it was found out that the test completion times of males is significantly less than of females. Also, the completion times of the users are significantly increased by the MRT test type or the use of the GVR.

While the skills of the users are improved with the use of the GVR, the completion time of the tests is also increased by it. This means, that the interaction is less time-consuming with the traditional keyboard and mouse than with the touchpad on the right side of the GVR. However, it is possible that this interaction time can still be decreased in the future with new input devices. Even though the interaction with the GVR is more time consuming than with a DD, it should be noted that the increase in completion times between the test types is less in the case of the GVR.

However, no VE exists with only one factor and even the human factors should be considered: in this case, this human factor is the gender of the user. The other factors that should be considered are the test type and the used device. When adding these factors together, the following can be concluded: the largest increase in interaction time in VR during the spatial ability tests is when the user is male and the MRT tests are done with the use of the GVR, while the largest decrease in interaction time is when the user is male and the MCT tests are done with the use of DD. These are the factors that have significant effects on the spatial ability test completion times. It also has to be kept in mind that the probabilities of correct answers and the completion times are not independent of each other. A larger completion time results in a higher probability of correct answers. Due to the results, it can be concluded that the interaction times of males in VEs are increased by the GVR in a larger degree than in the case of females. Moreover, with the GVR, the interaction time of males and females became similar to each other.

5.3.3 The correlation between the display devices and the human skills

According to the results on the spatial ability tests, a correlation can be found between the used display devices and the human skills. It can be concluded that the spatial performances of female users or left-handed users or older users are significantly improved by the use of GVR. The difficulty of the PSVT test type is also made significantly easier when using the GVR. Also with the GVR, the number of correct answers increased in the case of the female, left-handed and older users. The results of female students significantly improved and reached the level of the results of the male students. The spatial skills of left-handed students are increased the by use of the GVR so much that right-handed students are outperformed by them. Lastly, in the case of students who are over 23 and used the GVR, they outperformed the ones who are over 18 and used the DD.

According to the results, the spatial skills of the users are positively influenced by the GVR HMD. This is good, as most education for engineers at universities contains subjects such as technical representation or descriptive geometry and a well-developed spatial ability is necessary for successful studies. In conclusion, it can be safely stated that the use of VEs and VR can help with enhancing the spatial skills of students. Also, these results strengthen the fact that VR has a future in education.

5.3.4 The usability of the Asynchronous Prediction-Based Movement Recognition algorithm

The APBMR algorithm works by predicting the next gesture of the user. The prediction is based on the previous three gestures of the user and the next gesture is predicted by using six different MTs. Not only the next gesture is predicted, but three different ADs are generated for it was well. Afterward, it is decided by the algorithm whether the next user-input movement is accepted based on its percentage inside the ADs. The speed and the position of the user can also be followed. Due to this feature, a decision could be made whether to accept the next gesture easier.

According to the results, to get the optimal AGARs, the MT that should be used differs from gesture to gesture as well as from AD to AD, while the MTs not to use are the same in each case. It can be concluded that the AMT, HMT and CHMT should be used for prediction-based gesture recognition, but depending on the gesture and the AD, these MTs should be changed. It is summarized by Table 5.1 that which MT should be used in the case of different gestures and ADs when evaluating on all three axes.

Table 5.1: Which MT to use in case of different gestures and ADs when evaluating on all three axes?

AD	**Circular**	**Waving**	**Forward-diagonal**	**Upward-diagonal**
± 0.05 m	CHMT (0.305)	HMT (0.712)	CHMT (0.739)	CHMT (0.392)
± 0.10 m	AMT (0.619)	HMT (0.941)	CHMT (0.925)	CHMT (0.723)
± 0.15 m	AMT (0.837)	HMT (0.978)	AMT (0.964)	HMT (0.834)

After the evaluation of all three axes, it was concluded that in the case of simpler gestures such as the waving movement where only one axis is necessary most of the time, that the best AGARs are provided by the HMT. In the remaining cases, where more axes are required by the gestures to be correctly evaluated, the best AGARs are provided by the CHMT and the AMT in most cases. If the whole database of gestures that was gathered during this research are considered, then it can be concluded that the optimal AGARs out of the six MTs in the ± 0.05 m and ± 0.10 m ADs are provided by the CHMT with 53.4% and 79.6%, respectively. Contrary, in the ± 0.15 m AD, the optimal AGAR is provided by the AMT with 89.6%.

It should be noted that the Kinect was used for the file-based and real-time measurements and the Kinect has its own coordinate system. Positive and negative values exist in this coordinate system. Due to the possible negative values, the results of the GMT, QMT and CMT were worse than the results of the other three MTs. There is a possibility that better results may be provided by using other sensors – which do not return negative coordinate values – or with the creation of some methods that shift the returned coordinates of the Kinect.

It can be concluded that the prediction-based gesture recognition method is superior to the older (RDAMR) algorithm and therefore, it is more accurate as well as it can adapt to the current capabilities of the user, which is a criterion for maintaining motivation in the patients and for successful physical rehabilitation. Since the RDAMR algorithm could be used at home, the APBMR algorithm can be as well, making the rehabilitation process easier for both the therapist and the patient. By developing this algorithm, an alternative method was presented. Thus, the workers in the field of healthcare could have one more, easy-to-use gesture recognition method to choose from. Based on the results, the APBMR is both viable and suitable for gesture recognition in telerehabilitation with the Kinect v1.

Chapter 6

Application of the new scientific results

In the PhD dissertation of the author, two parts of HCI were investigated: the spatial skills of the users in VEs and the use of the APBMR in gesture classification and in physical rehabilitation. According to the reviewed literature and to the results of this research, both are equally important parts of HCI. Therefore, based on the results of the author, four thesis groups were formulated.

It is shown by the 1st thesis group that HCI can be affected by the display factors and devices. While these factors and devices all have a unique influence, it has to be kept in mind that no VE exists comprising only one of these factors. Therefore, these factors will always be in effect with multiple others, meaning that VEs should be carefully designed. Since spatial ability is important in VR as the user is placed into a four-dimensional space, the designers of VEs can take these results into account. With the use of these parameters, VEs can be designed in a way that the spatial skills of the users can be enhanced. With these results, the education of engineers and the rehabilitation of people with movement difficulties can be more effective.

Several factors that increase the test completion times are presented by the 2nd thesis group. One of the most important factors that increase the times is the use of the GVR. Also, the completion times and the probabilities of correct answers are not independent of each other. This means that if the users are given too much time for test completion, their probability of answering correctly becomes higher. However, since the completion times are increased by the use of the GVR as it needs to be interacted with differently than a traditional keyboard and mouse, more time on the tests have to be given to the users. This means that the test examination committee has to give enough, but not too much time for the testers. This also presents the following future research possibility: which is the optimal test deadline? Therefore, the results presented in this thesis group can also be used in the education of engineers and other jobs that require spatial skills.

It is shown by the 3rd thesis group that there is a correlation between the used display devices and the human skills. It can be concluded that significant improvements are made by using the GVR. This means that the spatial performances of female users, left-handed users, and older users are improved with its use. Also, the PSVT test type is made significantly easier with the use of the GVR. This means that the spatial skills of the user are positively influenced by the GVR HMD. This is good, as most education for engineers at universities contain subjects such as technical representation or descriptive geometry and a well-developed spatial ability is necessary for successful studies. The fact that VR has a future in education is strengthened by these results. Therefore, according to the results presented in the form of theses in this thesis group, the GVR should be used in education.

Based on the results that the 4$^{\text{th}}$ thesis group is formulated on, it can be concluded that the APBMR algorithm is easy-to-use and viable in gesture recognition with the Kinect v1. This algorithm can adapt to the current capabilities of the user without the help of any external factors. Therefore, motivation is not lost by the user and telerehabilitation can be conducted using this algorithm. This means that the presence of a therapist is not required during the actual physical rehabilitation, only when consultation is needed. Thus, the physical rehabilitation of people with movement disabilities are made more convenient and safer due to being in a home environment.

Summarizing the application of the results that are presented in the form of four thesis groups, it can be said that all results of the research of the author can help the users in some way. The results of the research regarding the spatial skills of the users in VEs can be used in the design of VEs, in education and in cognitive rehabilitation. Regarding the research of gesture classification, it can be concluded that the APBMR algorithm can used in physical telerehabilitation. With it, by time, the burden could be taken off of the hospitals.

Chapter 7

Theses summary

In this chapter, new scientific results are presented in the form of thesis groups since the theses were discussed in the previous discussion sections. Afterward, the possible future work is also presented.

7.1 New scientific results

1$^{\text{st}}$ **thesis group:** *The optimal preference for the virtual environments to positively influence the correct answers on spatial ability tests by affecting the human–computer interaction is a perspective camera type, a camera rotation of -45° or 0° or 45°, a contrast ratio of 1.5:1 or 3:1, and the Gear VR display device.*

(Own publications regarding this thesis group: [P1], [P2], [P3])

The results on the spatial ability tests were investigated with logistic regression analysis to determine how the probabilities of correct answers on the tests were influenced by the display parameters and display devices. The factors were investigated by themselves, in pairs, in triplets and in a quartet. According to the results, 7 theses were formed which can be found in subsection 5.1.1. The conclusion is that the probabilities of correct answers are not significantly influenced by the field of view of the virtual camera and the existence of shadows. However, they are significantly ($p\text{-}value = 4.62 \times 10^{-13}$) influenced by the perspective camera type, a camera rotation of -45° or 0° or 45°, a contrast ratio of 1.5:1 or 3:1, and the Gear VR. Significant ($p\text{-}value = 0.00013$) interaction can also occur between them.

2$^{\text{nd}}$ **thesis group:** *The probabilities of the correct answers and test completion times are not independent and the latter is significantly affected by the used display device, the test type and the gender of the user.*

(Own publication regarding this thesis group: [P5])

The completion times on the spatial ability tests were analyzed and the effects of multiple factors on them were analyzed as well. The effects of factors were investigated one by one, in pairs and in a triplet. According to the results, 7 theses were formed which can be found in subsection 5.1.2. Due to the correlation coefficient of 0.223 and *p-value* $< 2.2 \times 10^{-16}$, the probabilities of the correct answers and the completion times are not independent of each other. Although the latter is not significantly affected by the primary hand of the user, it is significantly influenced by the gender of the user, the test type and the display device which even interact with each other. The completion times are significantly (*p-value* $= 2.51 \times 10^{-10}$) increased by the combining factors of the male gender, the MRT test type and the use of the Gear VR, while it is significantly (*p-value* $= 0.047586$) decreased by the combining factors of the male gender, the MCT test type and the use of the desktop display.

3rd thesis group: *The ratio of correct answers of female, left-handed, older students are significantly improved and the PSVT test type is made significantly easier by using the Gear VR.*

(Own publications regarding this thesis group: [P2], [P3], [P6])

The results of 240 and 61 students were investigated who used the desktop display and the Gear VR, respectively. According to the results, 11 theses were formed which can be seen in subsection 5.1.3. The conclusion of this research is that compared to the desktop display, the results of female, left-handed, older students are significantly improved with the use of the Gear VR by 18.022%, 13%, 7.29%, respectively. With it, the PSVT test type is also made significantly easier by 17.217%.

4th thesis group: *The APBMR algorithm is viable and usable in a home environment for telerehabilitation with the Kinect v1 sensor.*

(Own publications regarding this thesis group: [P4], [P7])

A new gesture recognition method was designed and developed by building on a previous algorithm (RDAMR) and it was tested with four groups of people using the Kinect v1. It was also compared to the RDAMR algorithm that it was based upon. Since mean techniques are used by the new, prediction-based method to predict the next gesture of the user, the average gesture acceptance rate was evaluated with each mean technique. According to the results, 8 theses were formed which can be found in section 5.2. The optimal mean techniques are found in the case of each evaluated four gestures. Compared to the previous algorithm, the prediction-based algorithm has an increased average gesture acceptance rate by 358.2%-535.3% in the ± 0.05 m acceptance domain depending on the used mean technique. The increase in the ± 0.10 m and the ± 0.15 m acceptance domains are 87.8%-125.4% and 22.7%-47.3%, respectively. Therefore, the prediction-based gesture recognition method is viable and usable in a home environment for telerehabilitation with the Kinect v1.

7.2 Future plans

Naturally, the results that were presented and the theses that were formed are not the end of this scientific work. Thus, it can be continued in the following ways:

1. The effects of the display parameters on the test completion times:
   - While the display parameters can improve the probabilities of correct answers by enhancing the spatial skills of the users, they can also affect the completion times. Further examination of this fact is possible.
2. Post-test with the spatial ability application:
   - By performing a post-test in the end of the same semester with the same architectural and mechanical engineering students, it can be investigated whether their studies at the university improved their spatial skills.
   - This research is in progress as the post-test has been performed. Only the data have to be analyzed.
3. Introducing filtering into the APBMR algorithm:
   - The movement descriptors that are returned by the Kinect are noisy by default, therefore filtering should be used. Based on the literature, the Kalman filter has the possibility to improve the average gesture acceptance rate of an algorithm. Therefore, it is planned to integrate the Kalman filtering into the APBMR.
   - While the Kalman filter has the possibility to improve the average gesture acceptance rate, it can increase the execution time of the algorithm. It can also be investigated whether the use of the filter is worth it based on the required computational power.

7.3 Publications of the author

The main results of this PhD dissertation are published in multiple international journals and some are presented at national and international conferences. In the following group those publications are shown in which the results of this research are presented. The number of the respective thesis group is shown in parentheses. The publications are sorted by year in a decreasing order.

[P1] **Tibor Guzsvinecz**, Cecilia Sik-Lanyi, Eva Orban-Mihalyko, Erika Perge: The Influence of Display Parameters and Display Devices over Spatial Ability Test Answers in Virtual Reality Environments, *Applied Sciences*, 10(2): 526, 2020. DOI: 10.3390/app10020526 **(IF$_{\mathbf{2019}}$: 2.474)** (**1st thesis group**)

[P2] **Tibor Guzsvinecz**, Cecilia Sik-Lányi, Éva Orbán-Mihálykó, Erika Perge: Improvements on Spatial Ability Tests Using Virtual Reality. *7th Winter School for PhD Students in Informatics and Mathematics*. Association of Hungarian PhD and DLA Students, pp. 26, 2020. (**1st and 3rd thesis groups**)

[P3] **Guzsvinecz Tibor**: A megjelenítési paraméterek és eszközök által befolyásolt eredmények a virtuális valóság alapú térérzékelési teszteken, *XXIII. Tavaszi Szél Konferencia*, pp. 321, 2020. (**1st and 3rd thesis groups**)

[P4] **Tibor Guzsvinecz**, Veronika Szucs, Attila Magyar: Preliminary results of evaluating a prediction-based algorithm for movement pattern recognition and classification. *11th IEEE International Conference on Cognitive Infocommunications*, pp. 39-44, 2020. (**4th thesis group**)

[P5] **Tibor Guzsvinecz**, Éva Orbán-Mihálykó, Cecília Sik-Lányi, Erika Perge: Investigation of Spatial Ability Test Completion Times in Virtual Reality using a Desktop Display and the Gear VR, *Virtual Reality* **(Under review, $\mathbf{IF_{2019}}$: 3.634)** (**2nd thesis group**)

[P6] **Tibor Guzsvinecz**, Éva Orbán-Mihálykó, Erika Perge, Cecília Sik-Lányi: Analyzing the Spatial Skills of University Students with a Virtual Reality Application using a Desktop Display and the Gear VR, *Acta Polytechnica Hungarica*, 17(2), pp. 35-56, 2020. DOI: 10.12700/APH.17.2.2020.2.3 **($\mathbf{IF_{2019}}$: 1.219)** (**3rd thesis group**)

[P7] **Tibor Guzsvinecz**, Veronika Szucs, Attila Magyar: Evaluating the APBMR Algorithm with the Kinect for Gesture Recognition in Physical Rehabilitation, *Applied Sciences* **(Under review, $\mathbf{IF_{2019}}$: 2.474)** (**4th thesis group**)

The following group of publications is not directly related to the theses of the author:

[N1] Veronika Szucs, **Tibor Guzsvinecz**, Attila Magyar: Movement Pattern Recognition in Physical Rehabilitation - Cognitive Motivation-based IT Method and Algorithms, *Acta Polytechnica Hungarica*, 17(2), pp. 211-235, 2020. DOI: 10.12700/APH.17.2.2020.2.12 **($\mathbf{IF_{2019}}$: 1.219)**

[N2] Veronika Szucs, Cecilia Sik-Lanyi, **Tibor Guzsvinecz**: Presenting the User's Focus in Needs & Development (UFIND) method and its comparison to other design methods. *11th IEEE International Conference on Cognitive Infocommunications*, pp. 89-95, 2020.

[N3] **Tibor Guzsvinecz**, Veronika Szucs, Cecilia Sik-Lanyi: Suitability of the Kinect Sensor and Leap Motion Controller - A Literature Review, *Sensors*, 19(5), 1072, 2019. DOI: 10.3390/s19051072 **($\mathbf{IF_{2019}}$: 3.275)**

[N4] Cecilia Sik-Lanyi, Veronika Szucs, Shervin Shirmohammadi, Petya Grudeva, Boris Abersek, **Tibor Guzsvinecz**, Karel Van Isacker: How to Develop Serious Games for Social and Cognitive Competence of Children with Learning Difficulties, *Acta Polytechnica Hungarica*, 16(6), pp. 149-169, 2019. DOI: 10.12700/APH.16.6.2019.6.10 **($\mathbf{IF_{2019}}$: 1.219)**

[N5] Veronika Szucs, **Tibor Guzsvinecz**, Attila Magyar: Improved algorithms for movement pattern recognition and classification in physical rehabilitation, *10th IEEE International Conference on Cognitive Infocommunications*, pp. 417-424, 2019. DOI: 10.1109/CogInfoCom47531.2019.9089987

[N6] **Tibor Guzsvinecz**, Monika Szeles, Erika Perge, Cecilia Sik-Lanyi: Preparing spatial ability tests in a virtual reality application, *10th IEEE International Conference on Cognitive Infocommunications*, pp. 363-368, 2019. DOI: 10.1109/CogInfoCom47531.2019.9089919

[N7] **Tibor Guzsvinecz**, Veronika Szucs, Cecilia Sik-Lanyi: Designing gamified virtual reality applications with sensors - A gamification study, *Proceedings of the Pannonian Conference on Advances in Information Technology (PCIT'2019)*, pp. 105-112, 2019.

[N8] **Tibor Guzsvinecz**, Bence Jandas, Veronika Szucs, Cecilia Sik-Lanyi: Development of a Wingsuit-style gamified application, *Orvosi Informatika 2018. A XXXI. Neumann Kollokvium konferencia-kiadványa*, pp. 122-127, 2018.

[N9] **Tibor Guzsvinecz**, Csaba Kovacs, Dominik Reich, Veronika Szucs, Cecilia Sik-Lanyi: Developing a virtual reality application for the improvement of depth perception, *9th IEEE International Conference on Cognitive Infocommunications*, pp. 417-424, 2018. DOI: 10.1109/CogInfoCom.2018.8639935

[N10] Metka Abersek, Boris Abersek, Kosta Dolenc, Cecilia Sik-Lányi, Shervin Shirmohammadi, Karel Van Isacker, Petya Grudeva, Veronika Szűcs, **Tibor Guzsvinecz**: Intelligent Serious Games for Learning in Informal Learning Environments, *2nd International Scientific Conference on Philosophy of Mind and Cognitive Modelling in Education*, 2018.

[N11] Sikné Lányi Cecília, Szücs Veronika, **Guzsvinecz Tibor**: A VR/AR jelenlegi, illetve prognosztizált felhasználási területei az egészségügyben, *XV. Jubileumi Országos Infokommunikációs Konferencia*, 2017.

[N12] Cecilia Sik-Lanyi, Shervin Shirmohammadi, **Tibor Guzsvinecz**, Boris Abersek, Veronika Szucs, Karel Van Isacker, Petya Grudeva, Andrean Lazarov: How to Develop Serious Games for Social and Cognitive Competence of Children with Learning Difficulties, *8th IEEE International Conference on Cognitive Infocommunications*, pp. 321-326, 2017. DOI: 10.1109/CogInfoCom.2017.8268264

[N13] Eva A. Barta, **Tibor Guzsvinecz**, Cecilia Sik Lanyi, Veronika Szucs: Android-Based Daily Routine Organizing Application for Elementary School Students Living with ASD, *Harnessing the Power of Technology to Improve Lives*, pp. 283-290, 2017. DOI: 10.3233/978-1-61499-798-6-283

[N14] Gyula Hajdics, **Tibor Guzsvinecz**, Veronika Szucs, Cecilia Sik-Lanyi: Development of Mathematical Skills Developing Game Software, *Harnessing the Power of Technology to Improve Lives*, pp. 1005-1008, 2017. DOI: 10.3233/978-1-

61499-798-6-1005

[N15] **Tibor Guzsvinecz**, David Koszegi-Vigh, Veronika Szucs, Sik-Lanyi Cecilia: "Sliders" Android Game - Improving Logical Skills of People with Disabilities., *Harnessing the Power of Technology to Improve Lives*, pp. 279-282, 2017. DOI: 10.3233/978-1-61499-798-6-279

The last group of the publications is not related to the theses of the author:

[O1] Cecilia Sik-Lanyi, **Tibor Guzsvinecz**, Norbert Doszkocs, Imre Mark Meszaros, Gabor Lajos Somogyi: Test software development of size and contrast effect research. *11th IEEE International Conference on Cognitive Infocommunications*, pp. 45-50, 2020.

[O2] Sikné Dr. Lányi Cecília, **Guzsvinecz Tibor**, Halmosi Bence: Programozási nyelvek oktatási tapasztalatai a COVID-19 járvány idején, *26th Multimedia in Education Online Conference, XXVI. Multimédia az oktatásban online nemzetközi konferencia*, pp. 34-35, 2020.

[O3] **Tibor Guzsvinecz**, Tibor Medvegy: Comparing the Reflection and the Refraction of Light Phenomena in Reality to their Counterparts in a Virtual Environment Rendered by Cycles, *Proceedings of the Pannonian Conference in Advances of Information Technology (PCIT'2020)*, pp. 62-69, 2020.

[O4] Mostafa Elgendy, **Tibor Guzsvinecz**, Cecilia Sik-Lanyi: Identification of Markers in Challenging Conditions for People with Visual Impairment Using Convolutional Neural Network, *Applied Sciences*, 9(23): 5110, 2019. DOI: 10.3390/app9235110 **(IF$_{2019}$: 2.474)**

[O5] Mostafa Elgendy, Miklós Herperger, **Tibor Guzsvinecz**, Cecilia Sik-Lanyi: Indoor Navigation for People with Visual Impairment using Augmented Reality Markers, *10th IEEE International Conference on Cognitive Infocommunications*, pp. 425-430, 2019. DOI: 10.1109/CogInfoCom47531.2019.9089960

[O6] **Tibor Guzsvinecz**, Tibor Medvegy, Veronika Szucs: A brief review on the challenges of Internet of Things and their solutions, *Proceedings of the Pannonian Conference on Advances in Information Technology (PCIT'2019)*, pp. 181-186, 2019.

[O7] Cecilia Sik-Lanyi, **Tibor Guzsvinecz**, Laszlo Czuni: Teaching renewed Multimedia subjects at the University of Pannonia, *Journal of Applied Multimedia*, 14, pp. 1-6, 2019.

[O8] Veronika Szucs, **Tibor Guzsvinecz**, Daniel Bor, Cecilia Sik-Lanyi: Development of colour vision test game for android devices, *Proc. 12th ICDVRAT with ITAG*, pp. 290-293, 2018.

[O9] **Tibor Guzsvinecz**, Balázs Ruzsonyi, Veronika Szücs, Cecilia Sik-Lanyi: Open world memory game, *Proc. 12th ICDVRAT with ITAG*, pp. 213-216, 2018.

[O10] Sikné Lányi Cecília, **Guzsvinecz Tibor**, Czúni László: Megújult multimédia tárgyak oktatása a Pannon Egyetemen, *XXIV. Multimedia in Education Conferences: XXIV. Multimédia az oktatásban konferencia*, pp. 204-209, 2018.

[O11] Szilvia Paxian, Veronika Szücs, Shervin Shirmohammadi, Boris Abersek, Petya Grudeva, Karel Van Isacker, **Tibor Guzsvinecz**, Cecilia Sik-Lanyi: Designing Trainer's Manual for the ISG for Competence Project, *Computers Helping People with Special Needs: 16th International Conference, ICCHP*, pp. 284-288, 2018. DOI: 10.1007/978-3-319-94277-3_45

[O12] Cecilia Sik-Lányi, Szűcs Veronika, **Guzsvinecz Tibor**: Usability and colour-check of a healthcare WEB-site, *2017 IEEE 30th Neumann Colloquium (NC)*, pp. 111-116, 2017. DOI: 10.1109/NC.2017.8263263

Appendix

Tables regarding the research of spatial skills in virtual environments

This section of the appendix is split into three subsections. In the first the results of the logistic regression, linear regression analysis methods can be found and the results of the ANOVA dispersion analyses are also contained. In the second subsection the numerical results can be found which were a basis for the analyses that are done in the first subsection. In the third subsection, supplementary data regarding the rates of correct answers are presented.

Results of various analysis methods

It should be noted that in the tables where the results of linear regression analysis are presented, the estimated coefficients, the standard error, the test statistics (t value) and the *p-value* (the probability of the type I. error (Pr(>t))) are shown. In the those tables where the results of logistic regression analysis are presented, the z values are shown instead of t values.

Table A1: Logistic regression results by investigating the effect of camera type.

Camera Type	**Estimate**	**Standard Error**	z **Value**	**Pr** $(>\|z\|)$
Intercept	0.62452	0.01597	39.113	$<2 \times 10^{-16}$
Perspective	0.15670	0.02239	6.999	2.57×10^{-12}

Table A2: Logistic regression results of the effect of the FoV of the virtual camera.

FoV	**Estimate**	**Standard Error**	z **Value**	**Pr** $(>\|z\|)$
Intercept	0.62452	0.01597	39.113	$<2 \times 10^{-16}$
45°	0.14423	0.02423	5.952	2.64×10^{-9}
60°	0.14988	0.05595	2.679	0.00739
75°	0.15676	0.05351	2.93	0.00339
90°	0.27871	0.05830	4.781	1.75×10^{-6}

Table A3: Logistic regression analysis results of the camera FoV without the orthographic FoV.

FoV	**Estimate**	**Standard Error**	**z Value**	**Pr ($>\|z\|$)**
Intercept	0.768746	0.018226	42.179	$<2 \times 10^{-16}$
60°	0.005655	0.056640	0.100	0.9205
75°	0.012530	0.054223	0.231	0.8172
90°	0.134485	0.058957	2.281	0.0225

Table A4: Logistic regression analysis results of the camera rotation.

Camera Rotation	**Estimate**	**Standard Error**	**z Value**	**Pr ($>\|z\|$)**
Intercept	0.620184	0.33839	18.327	$<2 \times 10^{-16}$
-45°	0.176147	0.06557	2.686	0.00722
-30°	-0.009832	0.047978	-0.205	0.83763
0°	0.147985	0.03769	3.926	8.62×10^{-5}
15°	0.015393	0.047061	0.327	0.7436
30°	0.012121	0.046995	0.258	0.79646
45°	0.145807	0.059162	2.465	0.01372

Table A5: Logistic regression analysis results of the camera rotation groups.

Groups	**Estimate**	**Standard Error**	**z Value**	**Pr ($>\|z\|$)**
Intercept	0.62498	0.01664	37.57	$<2 \times 10^{-16}$
IMP_R	0.14503	0.02248	6.45	1.12×10^{-10}

Table A6: Logistic regression analysis results of the contrast ratio.

Contrast Ratio	**Estimate**	**Standard Error**	**z Value**	**Pr ($>\|z\|$)**
Intercept	0.77059	0.01801	42.779	$<2 \times 10^{-16}$
3:1	-0.05359	0.04876	-1.099	0.2717
7:1	-0.11437	0.02493	-4.588	4.47×10^{-6}
14:1	-0.12867	0.04912	-2.620	0.0088
21:1	-0.09264	0.04554	-2.034	0.0419

Table A7: Logistic regression analysis results of the contrast ratio groups.

Groups	**Estimate**	**Standard Error**	**z Value**	**Pr ($>\|z\|$)**
Intercept	0.65750	0.01504	43.712	$<2 \times 10^{-16}$
IMP_C	0.10584	0.02250	4.703	2.56×10^{-6}

Table A8: Logistic regression analysis results of the existence of shadows.

Shadows	**Estimate**	**Standard Error**	**z Value**	**Pr ($>\|z\|$)**
Intercept	0.69046	0.01612	42.83	$<2 \times 10^{-16}$
Turned on	0.02864	0.02239	1.271	0.204

Table A9: Logistic regression results of the device used.

Device Used	**Estimate**	**Standard Error**	**z Value**	**Pr ($>\|z\|$)**
Intercept	0.69002	0.01249	55.231	$<2 \times 10^{-16}$
GVR	0.07595	0.02805	2.708	0.00677

Table A10: Logistic regression analysis results of the pairs without interactions.

Camera type and its rotation				
Variables	**Estimate**	**Standard error**	z **Value**	**Pr** $(>\|z\|)$
Intercept	0.52784	0.02309	22.864	$<2 \times 10^{-16}$
Orthographic, INC_R	0.18323	0.03199	5.728	1.02×10^{-8}
Perspective, NO_R	0.19943	0.03334	5.982	2.21×10^{-9}
Perspective, INC_R	0.29271	0.03102	9.437	$<2 \times 10^{-16}$

Camera type and contrast ratio				
Variables	**Estimate**	**Standard error**	z **Value**	**Pr** $(>\|z\|)$
Intercept	0.53467	0.02150	24.873	$<2 \times 10^{-16}$
Orthographic, INC_C	0.19765	0.03215	6.148	7.85×10^{-10}
Perspective, NO_C	0.23707	0.03014	7.867	3.64×10^{-15}
Perspective, INC_C	0.25819	0.03181	8.117	4.80×10^{-16}

Camera type and the device used				
Variables	**Estimate**	**Standard error**	z **Value**	**Pr** $(>\|z\|)$
Intercept	0.61105	0.01752	34.87	$<2 \times 10^{-16}$
Orthographic, GVR	0.07859	0.04255	1.847	0.0647
Perspective, DD	0.15872	0.02501	6.347	2.20×10^{-10}
Perspective, GVR	0.20891	0.03735	5.593	2.23×10^{-8}

Camera rotation and contrast ratio				
Variables	**Estimate**	**Standard error**	z **Value**	**Pr** $(>\|z\|)$
Intercept	0.60941	0.01821	33.469	$<2 \times 10^{-16}$
NO_R, INC_C	0.09323	0.04483	2.08	0.0376
INC_R, NO_C	0.14922	0.03235	4.613	3.97×10^{-6}
INC_R, INC_C	0.16594	0.02584	6.421	1.36×10^{-10}

Camera rotation and the device used				
Variables	**Estimate**	**Standard error**	z **Value**	**Pr** $(>\|z\|)$
Intercept	0.63039	0.01778	35.465	$<2 \times 10^{-16}$
NO_R, GVR	-0.04388	0.05049	-0.869	0.385
INC_R, DD	0.11682	0.0250	4.673	2.97×10^{-6}
INC_R, GVR	0.20364	0.03461	5.883	4.02×10^{-9}

Contrast ratio and the device used				
Variables	**Estimate**	**Standard error**	z **Value**	**Pr** $(>\|z\|)$
Intercept	0.65115	0.01679	38.786	$<2 \times 10^{-16}$
NO_C, GVR	0.03204	0.0378	0.848	0.396699
INC_C, DD	0.08647	0.02514	3.44	0.000583
INC_C, GVR	0.21296	0.04108	5.184	2.18×10^{-7}

Table A11: Logistic regression analysis results of the pairs with interactions (additive model).

Camera type and rotation				
Variables	**Estimate**	**Standard error**	z **Value**	**Pr** $(>\|z\|)$
Intercept	0.52784	0.02309	22.864	$<2 \times 10^{-16}$
Perspective	0.19943	0.03334	5.982	2.21×10^{-9}
INC_R	0.18323	0.03199	5.728	1.02×10^{-8}
Perspective and INC_R	-0.08995	0.04507	-1.996	0.0459

Camera type and contrast ratio				
Variables	**Estimate**	**Standard error**	z **Value**	**Pr** $(>\|z\|)$
Intercept	0.53467	0.0215	24.873	$<2 \times 10^{-16}$
Perspective	0.23707	0.03014	7.867	3.64×10^{-15}
INC_C	0.19765	0.03215	6.148	7.85×10^{-10}
Perspective and INC_C	-0.17653	0.04505	-3.919	8.91×10^{-5}

Camera type and the device used				
Variables	**Estimate**	**Standard error**	z **Value**	**Pr** $(>\|z\|)$
Intercept	0.61105	0.01752	34.87	$<2 \times 10^{-16}$
Perspective	0.15872	0.02501	6.347	2.2×10^{-10}
GVR	0.07859	0.04255	1.847	0.0647
Perspective and GVR	-0.02841	0.05672	-0.501	0.6164

Camera rotation and contrast ratio				
Variables	**Estimate**	**Standard error**	z **Value**	**Pr** $(>\|z\|)$
Intercept	0.60941	0.01821	33.469	$<2 \times 10^{-16}$
INC_R	0.14922	0.03235	4.613	3.97×10^{-6}
INC_C	0.09323	0.04483	2.08	0.0376
INC_R and INC_C	-0.07651	0.05533	-1.383	0.1667

Camera rotation and the device used				
Variables	**Estimate**	**Standard error**	z **Value**	**Pr** $(>\|z\|)$
Intercept	0.63039	0.01778	35.465	$<2 \times 10^{-16}$
INC_R	0.11682	0.025	4.673	2.97×10^{-6}
GVR	-0.04388	0.05049	-0.869	0.3847
INC_R and GVR	0.1307	0.06115	2.137	0.0326

Contrast ratio and the device used				
Variables	**Estimate**	**Standard error**	z **Value**	**Pr** $(>\|z\|)$
Intercept	0.65115	0.01679	38.786	$<2 \times 10^{-16}$
INC_C	0.08647	0.02514	3.44	0.000583
GVR	0.03204	0.0378	0.848	0.396699
INC_C and GVR	0.09445	0.05644	1.674	0.094199

Table A12: Comparison of the variable pairs by ANOVA.

Camera type and rotation					
Resid.	**Df Resid.**	**Dev**	**Df**	**Deviance**	**Pr (>Chi)**
1	2706	10.070	-	-	-
2	2705	10.066	1	3.9852	0.0459

Camera type and contrast ratio					
Resid.	**Df Resid.**	**Dev**	**Df**	**Deviance**	**Pr (>Chi)**
1	2706	10.085	-	-	-
2	2705	10.069	1	15.358	8.895×10^{-5}

Camera type and the device used					
Resid.	**Df Resid.**	**Dev**	**Df**	**Deviance**	**Pr (>Chi)**
1	2706	10.103	-	-	-
2	2705	10.102	1	0.25097	0.6164

Camera rotation and contrast ratio					
Resid.	**Df Resid.**	**Dev**	**Df**	**Deviance**	**Pr (>Chi)**
1	2706	10.112	-	-	-
2	2705	10.111	1	1.9179	0.1661

Camera rotation and the device used					
Resid.	**Df Resid.**	**Dev**	**Df**	**Deviance**	**Pr (>Chi)**
1	2706	10.113	-	-	-
2	2705	10.108	1	4.5511	0.0329

Contrast ratio and the device used					
Resid.	**Df Resid.**	**Dev**	**Df**	**Deviance**	**Pr (>Chi)**
1	2706	10.128	-	-	-
2	2705	10.125	1	2.804	0.09403

Table A13: Logistic regression analysis results of the triplets without interactions.

Camera type, camera rotation and contrast ratio				
Variables	**Estimate**	**Standard error**	**z Value**	**Pr ($>\|z\|$)**
Intercept	0.51019	0.02512	20.309	$<2 \times 10^{-16}$
Ortho., NO_R, INC_C	0.11228	0.06379	1.760	0.0784
Ortho., INC_R, NO_C	0.09065	0.04857	1.866	0.0620
Ortho., INC_R, INC_C	0.24358	0.03629	6.712	1.92×10^{-11}
Persp., NO_R, NO_C	0.20626	0.03651	5.649	1.61×10^{-8}
Persp., NO_R, INC_C	0.26714	0.06260	4.268	1.98×10^{-5}
Persp., INC_R, NO_C	0.35544	0.04310	8.246	$<2 \times 10^{-16}$
Persp., INC_R, INC_C	0.28576	0.03593	7.952	1.83×10^{-15}

Camera type, camera rotation and the device used				
Variables	**Estimate**	**Standard error**	**z Value**	**Pr ($>\|z\|$)**
Intercept	0.53653	0.02449	21.907	$<2 \times 10^{-16}$
Ortho., NO_R, GVR	-0.07870	0.07340	-1.072	0.2836
Ortho. INC_R, DD	0.15114	0.03508	4.308	1.65×10^{-5}
Ortho., INC_R, GVR	0.25496	0.05299	4.811	1.50×10^{-6}
Persp., NO_R, DD	0.19570	0.03564	5.490	4.01×10^{-8}
Persp., NO_R, GVR	0.15942	0.06935	2.299	0.0215
Persp., INC_R, DD	0.26674	0.03473	7.680	1.59×10^{-14}
Persp., INC_R, GVR	0.32541	0.04549	7.154	8.44×10^{-13}

Camera type, contrast ratio and the device used				
Variables	**Estimate**	**Standard error**	**z Value**	**Pr ($>\|z\|$)**
Intercept	0.53024	0.02342	22.644	$<2 \times 10^{-16}$
Ortho., NO_C, GVR	0.02804	0.03534	0.475	0.635
Ortho., INC_C, DD	0.18131	0.03534	5.130	2.90×10^{-7}
Ortho., INC_C, GVR	0.29212	0.06042	4.835	1.33×10^{-7}
Persp., NO_C, DD	0.24482	0.03365	7.275	3.45×10^{-13}
Persp., NO_C, GVR	0.23070	0.04936	4.674	2.96×10^{-6}
Persp., INC_C, DD	0.23317	0.03533	6.600	4.11×10^{-11}
Persp., INC_C, GVR	0.36797	0.05587	6.587	4.50×10^{-11}

Camera rotation, contrast ratio and the device used				
Variables	**Estimate**	**Standard error**	**z Value**	**Pr ($>\|z\|$)**
Intercept	0.60785	0.01946	31.238	$<2 \times 10^{-16}$
NO_R, NO_C, GVR	0.01252	0.05517	0.227	0.82046
NO_R, INC_C, DD	0.13403	0.04789	2.799	0.00513
NO_R, INC_C, GVR	-0.20239	0.11945	-1.694	0.0902
INC_R, NO_C, DD	0.16645	0.03855	4.317	1.58×10^{-5}
INC_R, NO_C, GVR	0.12203	0.04894	2.494	0.01265
INC_R, INC_C, DD	0.12882	0.02841	4.534	5.78×10^{-6}
INC_R, INC_C, GVR	0.30462	0.04418	6.895	5.37×10^{-12}

Table A14: Comparison of model II and III by ANOVA.

Resid.	**Df Resid.**	**Dev**	**Df**	**Deviance**	**Pr (>Chi)**
1	2703	10.053	-	-	-
2	2701	10.050	2	3.1703	0.2049

Table A15: Logistic regression analysis results of the triplets with interactions.

Camera type, camera rotation and contrast ratio (model II)				
Variables	**Estimate**	**Standard error**	z **Value**	**Pr (>|z|)**
Intercept	0.5062478	0.0238618	21.216	$<2 \times 10^{-16}$
Persp.	0.2259740	0.0344853	6.553	5.65×10^{-11}
INC_R	0.1054279	0.0388003	2.717	0.006584
INC_C	0.1378076	0.0389859	3.535	0.000408
Persp. and INC_R	0.0006636	0.0528045	0.013	0.989973
Persp. and INC_C	-0.1653400	0.0528044	-3.131	0.001741

Camera type, camera rotation and the used device (model II)				
Variables	**Estimate**	**Standard error**	z **Value**	**Pr (>|z|)**
Intercept	0.53406	0.02376	22.479	$<2 \times 10^{-16}$
Persp.	0.20095	0.03337	6.022	1.72×10^{-9}
INC_R	0.15942	0.03339	4.774	1.80×10^{-6}
GVR	-0.05639	0.05059	-1.115	0.2649
Persp. and INC_R	-0.09672	0.04515	-2.142	0.0322
Persp. and GVR	0.13418	0.06130	2.189	0.0286

Camera type, contrast ratio and the used device (model II)				
Variables	**Estimate**	**Standard error**	z **Value**	**Pr (>|z|)**
Intercept	0.53024	0.02342	22.644	$<2 \times 10^{-16}$
Ortho., NO_C, GVR	0.02804	0.03534	0.475	0.635
Ortho., INC_C, DD	0.18131	0.03534	5.13	2.90×10^{-7}
Ortho., INC_C, GVR	0.29212	0.06042	4.835	1.33×10^{-7}
Persp., NO_C, DD	0.24482	0.03365	7.275	3.45×10^{-13}
Persp., NO_C, GVR	0.23070	0.04936	4.674	2.96×10^{-6}
Persp., INC_C, DD	0.23317	0.03533	6.600	4.11×10^{-11}
Persp., INC_C, GVR	0.36797	0.05587	6.587	4.50×10^{-11}

Camera rotation, contrast ratio and the used device (model III)				
Variables	**Estimate**	**Standard error**	z **Value**	**Pr (>|z|)**
Intercept	0.60785	0.01946	31.238	$<2 \times 10^{-16}$
NO_R, NO_C, GVR	0.01252	0.05517	0.227	0.82046
NO_R, INC_C, DD	0.13403	0.04789	2.799	0.00513
NO_R, INC_C, GVR	-0.20239	0.11945	-1.694	0.0902
INC_R, NO_C, DD	0.16645	0.03855	4.317	1.58×10^{-5}
INC_R, NO_C, GVR	0.12203	0.04894	2.494	0.01265
INC_R, INC_C, DD	0.12882	0.02841	4.534	5.78×10^{-6}
INC_R, INC_C, GVR	0.30462	0.04418	6.895	5.37×10^{-12}

Table A16: Logistic regression analysis results investigating the effects of the camera type, rotation, contrast ratio, and device used without interactions.

Variables	**Estimate**	**Standard error**	z **Value**	**Pr** ($>\|z\|$)
Intercept	0.51385	0.02670	19.246	$<2 \times 10^{-16}$
Ortho., NO_R, NO_C, GVR	-0.03202	0.07884	-0.406	0.6847
Ortho., NO_R, INC_C, DD	0.14107	0.06717	2.100	0.0357
Ortho., NO_R, INC_C, GVR	-0.22087	0.19458	-1.135	0.2563
Ortho., INC_R, NO_C, DD	0.07053	0.05559	1.269	0.2045
Ortho., INC_R, NO_C, GVR	0.13040	0.08391	1.554	0.1202
Ortho., INC_R, INC_C, DD	0.21036	0.03966	5.305	1.13×10^{-7}
Ortho., INC_R, INC_C, GVR	0.35379	0.06417	5.513	3.52×10^{-8}
Persp., NO_R, NO_C, DD	0.19790	0.03904	5.069	4.00×10^{-7}
Persp., NO_R, NO_C, GVR	0.23262	0.07695	3.023	0.0025
Persp., NO_R, INC_C, DD	0.31400	0.06772	4.637	3.54×10^{-6}
Persp., NO_R, INC_C, GVR	-0.04214	0.15152	-0.278	0.7809
Persp., INC_R, NO_C, DD	0.41798	0.05303	7.882	3.22×10^{-15}
Persp., INC_R, NO_C, GVR	0.25528	0.06062	4.211	2.54×10^{-5}
Persp., INC_R, INC_C, DD	0.23513	0.03959	5.94	2.86×10^{-9}
Persp., INC_R, INC_C, GVR	0.43645	0.06032	7.236	4.62×10^{-13}

Table A17: Logistic regression analysis results investigating the effects of the camera type, rotation, contrast ratio, and device used with interactions.

Variables	**Estimate**	**Standard error**	z **Value**	**Pr** ($>\|z\|$)
Intercept	0.4997428	0.0252612	19.783	$<2 \times 10^{-16}$
Persp.	0.2281487	0.0345391	6.606	3.96×10^{-11}
INC_R	0.1501627	0.0475111	3.161	0.001575
INC_C	0.2106181	0.0547860	3.844	0.000121
GVR	0.0008861	0.0552796	0.016	0.987212
Persp., INC_R	-0.0028185	0.0532272	-0.053	0.95777
Persp., INC_C	-0.1660764	0.0532649	-3.118	0.001821
INC_R, GVR	-0.0746026	0.0788879	-0.946	0.344313
INC_R, INC_C	-0.1535519	0.0619693	-2.478	0.013217
INC_C, GVR	-0.3450059	0.1374857	-2.509	0.012094
INC_R, INC_C, GVR	0.5920143	0.1374857	3.821	0.000133

Table A18: Regression analysis results of the influence of one factor on the completion times.

Gender of the user				
Variable	**Estimate**	**Standard Error**	t **Value**	**Pr(>$\|t\|$)**
Intercept	223.644	6.040	37.024	$<2\times10^{-16}$
Male	-27.451	6.563	-4.183	2.97×10^{-5}

Primary hand of the user				
Variable	**Estimate**	**Standard Error**	t **Value**	**Pr(>$\|t\|$)**
Intercept	199.5352	6.8501	29.129	$<2\times10^{-16}$
RH	0.9685	7.3006	0.133	0.894

Test type				
Variable	**Estimate**	**Standard Error**	t **Value**	**Pr(>$\|t\|$)**
Intercept	176.156	3.963	44.454	$<2\times10^{-16}$
MRT	69.545	5.604	12.410	$<2\times10^{-16}$
PSVT	3.151	5.604	0.562	0.574

The device used				
Variable	**Estimate**	**Standard Error**	t **Value**	**Pr(>$\|t\|$)**
Intercept	188.784	2.606	72.43	$<2\times10^{-16}$
GVR	57.259	5.790	9.89	$<2\times10^{-16}$

Table A19: Results of the regression analysis of the influence of the pairs without interactions on the completion times.

Gender of the user and the test type				
Variables	**Estimate**	**Standard Error**	t **Value**	**Pr(>|t|)**
Intercept	200.152	10.106	19.805	$<2\times10^{-16}$
F, MRT	63.000	14.292	4.408	1.08×10^{-5}
F, PSVT	7.476	14.292	0.523	0.600945
M, MCT	-28.324	10.980	-2.580	0.009941
M, MRT	42.401	10.980	3.862	0.000115
M, PSVT	-25.954	10.980	-2.364	0.018158

Gender of the user and the device used				
Variables	**Estimate**	**Standard Error**	t **Value**	**Pr(>|t|)**
Intercept	212.681	7.484	28.417	$<2\times10^{-16}$
F, GVR	29.663	12.312	2.409	0.016046
M, DD	-27.182	7.982	-3.405	0.000671
M, GVR	34.790	9.640	3.609	0.000313

Test type and the device used				
Variables	**Estimate**	**Standard Error**	t **Value**	**Pr(>|t|)**
Intercept	160.512	4.351	36.889	$<2\times10^{-16}$
MCT, GVR	77.197	9.666	7.987	2.03×10^{-15}
MRT, DD	75.471	6.154	12.265	$<2\times10^{-16}$
MRT, GVR	123.425	9.666	12.769	$<2\times10^{-16}$
PSVT, DD	9.346	6.154	1.519	0.129
PSVT, GVR	55.971	9.666	5.791	7.82×10^{-9}

Table A20: Results of the regression analysis of the influence of pairs with interactions on the completion times.

Gender of the user and the test type				
Variables	**Estimate**	**Standard Error**	**t Value**	**Pr(>|t|)**
Intercept	200.152	10.106	19.805	$<2\times10^{-16}$
M	-28.324	10.980	-2.580	0.00994
MRT	63.000	14.292	4.408	1.08×10^{-5}
PSVT	7.476	14.292	0.523	0.60095
M, MRT	7.725	15.528	0.498	0.61885
M, PSVT	-5.106	15.528	-0.329	0.74231

Gender of the user and the device used				
Variables	**Estimate**	**Standard Error**	**t Value**	**Pr(>|t|)**
Intercept	212.681	7.484	28.417	$<2\times10^{-16}$
M	-27.182	7.982	-3.405	0.000671
GVR	29.663	12.312	2.409	0.016046
M, GVR	32.309	14.007	2.307	0.021150

Test type and the device used				
Variables	**Estimate**	**Standard Error**	**t Value**	**Pr(>|t|)**
Intercept	160.512	4.351	36.889	$<2\times10^{-16}$
MRT	75.471	6.154	12.265	$<2\times10^{-16}$
PSVT	9.346	6.154	1.519	0.1289
GVR	77.197	9.666	7.897	2.03×10^{-15}
MRT, GVR	-29.243	13.669	-2.139	0.0325

Table A21: Results of the regression analysis of the influence of all triplets without interactions on the completion times.

Variables	**Estimate**	**Standard Error**	**t Value**	**Pr(>|t|)**
Intercept	183.74	12.50	14.700	$<2\times10^{-16}$
F, MCT, GVR	44.41	20.56	2.160	0.030886
F, MRT, DD	70.33	17.68	3.979	7.11×10^{-5}
F, MRT, GVR	94.90	20.56	4.615	4.11×10^{-6}
F, PSVT, DD	16.49	17.68	0.933	0.351047
F, PSVT, GVR	36.51	20.56	1.776	0.075911
M, MCT, DD	-26.42	13.33	-1.982	0.047586
M, MCT, GVR	57.66	16.10	3.581	0.000348
M, MRT, DD	49.76	13.33	3.732	0.000194
M, MRT, GVR	102.24	16.10	6.350	2.51×10^{-10}
M, PSVT, DD	-18.06	13.33	-1.355	0.175688
M, PSVT, GVR	31.29	16.10	1.943	0.052091

Table A22: Results of the regression analysis of the influence of all triplets with interactions on the completion times.

Variables	Estimate	Standard Error	t Value	Pr(>|t|)
Intercept	184.409	8.037	22.944	$<2\times10^{-16}$
M	-27.182	7.692	-3.534	0.000417
MRT	75.471	6.141	12.289	$<2\times10^{-16}$
PSVT	9.346	6.141	1.522	0.128172
GVR	49.601	14.241	3.483	0.000504
M, GVR	32.309	13.499	2.394	0.016756
MRT, GVR	-29.243	13.642	-2.144	0.032161
PSVT, GVR	-30.572	13.642	-2.241	0.025112

Numerical results of the users

Table B1: Numerical results of the users regarding the camera type.

Camera type	Number of tests	Average rate	Dispersion
Orthographic	1291	0.606	0.240
Perspective	1418	0.642	0.247

Table B2: Numerical results of the users regarding the camera FoV.

FoV	Number of tests	Average rate	Dispersion
-1[1]	1291	0.606	0.240
45°	1049	0.639	0.245
60°	120	0.637	0.280
75°	134	0.640	0.256
90°	115	0.673	0.218

[1]-1 is undefined in the case of the orthographic camera.

Table B3: Numerical results of the users regarding the camera rotation.

Camera rotation	Number of tests	Average rate	Dispersion
-45°	106	0.651	0.235
-30°	294	0.603	0.241
-15°	294	0.606	0.232
0°	1251	0.639	0.247
15°	312	0.611	0.247
30°	313	0.607	0.240
45°	139	0.632	0.259

Table B4: Numerical results of the users regarding the camera rotation groups.

Camera rotation groups	Number of tests	Average rate	Dispersion
IMP_R	1496	0.639	0.247
NO_R	1213	0.607	0.240

Table B5: Numerical results of the users regarding the contrast ratio.

Contrast ratio	**Number of tests**	**Average rate**	**Dispersion**
1.5:1	1066	0.639	0.247
3:1	167	0.628	0.249
7:1	1121	0.615	0.239
14:1	164	0.609	0.247
21:1	191	0.616	0.248

Table B6: Numerical results of the users regarding the contrast ratio groups.

Contrast ratio groups	**Number of tests**	**Average rate**	**Dispersion**
IMP_C	1233	0.637	0.241
NO_C	1476	0.614	0.247

Table B7: Numerical results of the users regarding the shadows in the scene.

Shadows	**Number of tests**	**Average rate**	**Dispersion**
Turned on	1414	0.628	0.242
Turned off	1295	0.621	0.247

Table B8: Numerical results of the users regarding the device used.

Device used	**Number of tests**	**Average rate**	**Dispersion**
DD	2160	0.620	0.242
GVR	549	0.643	0.252

Table B9: Numerical results of the users regarding the pairs of variables.

Variables	**Number of tests**	**Average rate**	**Dispersion**
Orthographic, NO_R	611	0.583	0.235
Orthographic, INC_R	680	0.626	0.243
Perspective, NO_R	602	0.631	0.243
Perspective, INC_R	816	0.650	0.249
Orthographic, NO_C	696	0.585	0.235
Orthographic, INC_C	595	0.630	0.244
Perspective, NO_C	780	0.640	0.244
Perspective, INC_C	638	0.644	0.250
Orthographic, DD	1065	0.601	0.237
Orthographic, GVR	226	0.626	0.252
Perspective, DD	1095	0.638	0.246
NO_R, NO_C	1005	0.603	0.237
NO_R, INC_C	208	0.625	0.253
INC_R, NO_C	471	0.638	0.248
INC_R, INC_C	1025	0.640	0.246
NO_R, DD	1062	0.607	0.238
NO_R, GVR	151	0.603	0.251
INC_R, DD	1098	0.632	0.245
INC_R, GVR	398	0.659	0.251
NO_C, DD	1183	0.611	0.239
NO_C, GVR	293	0.626	0.249
INC_C, DD	977	0.630	0.245
INC_C, GVR	256	0.663	0.254

Table B10: Numerical results of the users regarding the variable triplets.

Variables	**Number of tests**	**Average rate**	**Dispersion**
Orthographic, NO_R, NO_C	511	0.579	0.234
Orthographic, NO_R, INC_C	100	0.606	0.235
Orthographic, INC_R, NO_C	185	0.603	0.235
Orthographic, INC_R, INC_C	495	0.635	0.246
Perspective, NO_R, NO_C	494	0.628	0.238
Perspective, NO_R, INC_C	108	0.643	0.268
Perspective, INC_R, NO_C	286	0.661	0.254
Perspective, INC_R, INC_C	530	0.664	0.247
Orthographic, NO_R, DD	541	0.584	0.232
Orthographic, NO_R, GVR	70	0.574	0.255
Orthographic INC_R, DD	524	0.619	0.242
Orthographic, INC_R, GVR	156	0.650	0.248
Perspective, NO_R, DD	521	0.631	0.243
Perspective, NO_R, GVR	81	0.629	0.246
Perspective, INC_R, DD	574	0.644	0.248
Perspective, INC_R, GVR	242	0.664	0.252
Orthographic, NO_C, DD	585	0.583	0.232
Orthographic, NO_C, GVR	111	0.595	0.251
Orthographic, INC_C, DD	480	0.623	0.242
Orthographic, INC_C, GVR	115	0.657	0.251
Perspective, NO_C, DD	598	0.639	0.243
Perspective, NO_C, GVR	182	0.645	0.247
Perspective, INC_C, DD	497	0.637	0.248
Perspective, INC_C, GVR	141	0.668	0.257
NO_R, NO_C, DD	877	0.602	0.236
NO_R, NO_C, GVR	128	0.611	0.245
NO_R, INC_C, DD	185	0.633	0.248
NO_R, INC_C, GVR	23	0.561	0.287
INC_R, NO_C, DD	306	0.638	0.246
INC_R, NO_C, GVR	165	0.638	0.253
INC_R, INC_C, DD	792	0.630	0.245
INC_R, INC_C, GVR	233	0.673	0.248

Table B11: Numerical results of the users regarding the camera type, rotation, contrast ratio, device used.

Variables	Number of tests	Average rate	Dispersion
Orthographic, NO_R, NO_C, DD	450	0.579	0.232
Orthographic, NO_R, NO_C, GVR	61	0.578	0.251
Orthographic, NO_R, INC_C, DD	91	0.612	0.230
Orthographic, NO_R, INC_C, GVR	9	0.544	0.296
Orthographic, INC_R, NO_C, DD	135	0.598	0.229
Orthographic, INC_R, NO_C, GVR	50	0.615	0.253
Orthographic, INC_R, INC_C, DD	389	0.626	0.246
Orthographic, INC_R, INC_C, GVR	106	0.667	0.246
Perspective, NO_R, NO_C, DD	427	0.626	0.238
Perspective, NO_R, NO_C, GVR	67	0.641	0.236
Perspective, NO_R, INC_C, DD	94	0.654	0.264
Perspective, NO_R, INC_C, GVR	14	0.571	0.291
Perspective, INC_R, NO_C, DD	171	0.670	0.255
Perspective, INC_R, NO_C, GVR	115	0.648	0.253
Perspective, INC_R, INC_C, DD	403	0.633	0.244
Perspective, INC_R, INC_C, GVR	127	0.679	0.252

Table B12: Numerical results of the users regarding their gender.

Gender of the user	Number of tests	Average time (sec)	Dispersion (sec)
Female	414	223.644	123.214
Male	2295	196.193	122.850

Table B13: Numerical results of the users regarding their primary hand.

Primary hand	Number of tests	Average time (sec)	Dispersion (sec)
Left-handed	324	199.535	118.589
Right-handed	2385	200.504	123.926

Table B14: Numerical results of the users regarding the type of the tEstimate

Test type	Number of tests	Average time (sec)	Dispersion (sec)
MCT	903	176.156	129.959
MRT	903	245.701	127.911
PSVT	903	179.307	96.374

Table B15: Numerical results of the users regarding the device used.

Device used	Number of tests	Average time (sec)	Dispersion (sec)
DD	2160	188.784	121.389
GVR	549	246.043	120.117

Table B16: Numerical results of the users regarding the pairs of variables.

Variables	**Number of tests**	**Average time (sec)**	**Dispersion (sec)**
F, MCT	138	200.152	135.476
F, MRT	138	263.152	123.841
F, PSVT	138	207.628	98.435
M, MCT	765	171.827	128.554
M, MRT	765	242.553	128.458
M, PSVT	765	174.198	95.169
F, DD	261	212.681	125.440
F, GVR	153	242.344	117.373
M, DD	1899	185.499	120.486
M, GVR	396	247.472	121.276
MCT, DD	720	160.512	122.476
MCT, GVR	183	237.708	140.238
MRT, DD	720	235.983	130.113
MRT, GVR	183	283.937	111.186
PSVT, DD	720	169.858	94.485
PSVT, GVR	183	216.483	94.980

Table B17: Numerical results of the users regarding all factors.

Variables	**Number of tests**	**Average time (sec)**	**Dispersion (sec)**
F, MCT, DD	87	183.741	135.609
F, MCT, GVR	51	228.147	131.875
F, MRT, DD	87	254.074	128.061
F, MRT, GVR	51	278.636	115.875
F, PSVT, DD	87	200.229	100.221
F, PSVT, GVR	51	220.250	94.948
M, MCT, DD	633	157.319	120.327
M, MCT, GVR	132	241.403	143.653
M, MRT, DD	633	233.496	130.296
M, MRT, GVR	132	285.985	109.705
M, PSVT, DD	633	165.683	92.979
M, PSVT, GVR	132	215.028	95.313

Supplementary data regarding the rates of correct answers

Table C1: Comparisons of standard deviations of the ratios of correct answers with a DD.

	MRT			MCT			PSVT		
	Trial stat.	**Significance**	**Q?**	**Trial stat.**	**Significance**	**Q?**	**Trial stat.**	**Significance**	**Q?**
MRT	0	1	Yes	0.7503	0.0268	No	0.4764	0	No
MCT	0.7503	0.0268	No	0	1	Yes	0.6349	0.0005	No
PSVT	0.4764	0	No	0.6349	0.0005	No	0	1	Yes

In the header of the table, "Q?" stands for "Distinguishable?".

Table C2: Comparisons of the average rates of the ratios of correct answers with a DD.

	MRT			MCT			PSVT		
	Trial stat.	**Significance**	**Q?**	**Trial stat.**	**Significance**	**Q?**	**Trial stat.**	**Significance**	**Q?**
MRT	0	1	Yes	27.775	0	No	12.358	0	No
MCT	27.775	0	No	0	1	Yes	-11.155	0	No
PSVT	12.358	0	No	-11.155	0	No	0	1	Yes

In the header of the table, "Q?" stands for "Distinguishable?".

Table C3: Standard deviations of rates of correct answers with the GVR.

	MRT			MCT			PSVT		
	Trial stat.	**Significance**	**Q?**	**Trial stat.**	**Significance**	**Q?**	**Trial stat.**	**Significance**	**Q?**
MRT	1	1	Yes	0.5247	0.0136	No	0.4376	0.0016	No
MCT	0.5247	0.0136	No	1	1	Yes	0.8340	0.4842	Yes
PSVT	0.4376	0.0016	No	0.8340	0.4842	Yes	1	1	Yes

In the header of the table, "Q?" stands for "Distinguishable?".

Table C4: Comparison of average rates of correct answers with the GVR.

	MRT			MCT			PSVT		
	Trial stat.	**Significance**	**Q?**	**Trial stat.**	**Significance**	**Q?**	**Trial stat.**	**Significance**	**Q?**
MRT	0	1	Yes	14.437	0	No	2.6704	0.0087	No
MCT	14.437	0	No	0	1	Yes	-9.657	0	No
PSVT	2.6704	0.0087	No	-9.657	0	No	0	1	Yes

In the header of the table, "Q?" stands for "Distinguishable?".

Table C5: Statistical data of the rates of correct answers by age groups using a DD.

Age	**Students in group**	**Group average**	**Group dispersion**
17	1	0.666	0.000
18	33	0.623	0.124
19	89	0.667	0.116
20	75	0.680	0.111
21	29	0.673	0.125
22	6	0.573	0.087
23	2	0.750	0.087
24	2	0.612	0.205
25	1	0.808	0.000
27	1	0.758	0.000
32	1	0.866	0.000

Table C6: Comparing different age groups who used a DD.

Age	**G1**	**Average rate of G1**	**G2**	**Average rate of G2**	**p-value**	**Significant difference?**
<= **17** & >**17**	1	0.6667	239	0.6660	0.9277	No
<= **18** & >**18**	34	0.6245	206	0.6728	0.0375	Yes
<= **19** & >**19**	123	0.6556	117	0.6769	0.1656	No
<= **20** & >**20**	198	0.6652	42	0.6696	0.8357	No
<= **21** & >**21**	227	0.6662	13	0.6615	0.9047	No
<= **22** & >**22**	233	0.6638	7	0.7369	0.1801	No
<= **23** & >**23**	235	0.6646	5	0.7317	0.3885	No
<= **24** & >**24**	237	0.6641	3	0.8111	0.0359	Yes
<= **25** & >**25**	238	0.6647	2	0.8125	0.2179	No
<= **27** & >**27**	239	0.6651	1	0.8667	0.0000	Yes

In the header of the table, "G1" refers to Group 1 which is on the left side of the "&" symbol, while "G2" refers to Group 2 which is on the right side of the "&" symbol.

Table C7: Statistical data of the rates of correct answers by age groups using the GVR.

Age	Students in group	Group average	Group dispersion
19	3	0.7306	0.1008
20	4	0.6000	0.1763
21	9	0.6296	0.1176
22	10	0.6667	0.1153
23	11	0.6720	0.1492
24	8	0.6938	0.1006
25	4	0.7250	0.1369
26	3	0.6944	0.0376
27	3	0.7500	0.1228
28	2	0.7542	0.0412
30	1	0.8083	0
31	1	0.6583	0
32	1	0.8083	0
34	1	0.7583	0

Table C8: Comparing different age groups who used the GVR.

Age	G1	Average rate of G1	G2	Average rate of G2	p-value	Significant difference?
<= **19** & >**19**	3	0.7306	58	0.6802	0.4812	No
<= **20** & >**20**	7	0.6560	54	0.6861	0.6330	No
<= **21** & >**21**	16	0.6411	45	0.6974	0.1396	No
<= **22** & >**22**	26	0.6510	35	0.7062	0.0792	No
<= **23** & >**23**	37	0.6572	24	0.7219	0.0266	Yes
<= **24** & >**24**	45	0.6637	16	0.6637	0.0163	Yes
<= **25** & >**25**	49	0.6687	12	0.6687	0.0155	Yes
<= **26** & >**26**	52	0.6702	9	0.6702	0.0140	Yes
<= **27** & >**27**	55	0.6745	6	0.6745	0.0157	Yes
<= **28** & >**28**	57	0.6773	4	0.6773	0.0996	No
<= **30** & >**30**	58	0.6796	3	0.6796	0.2915	No
<= **31** & >**31**	59	0.6792	2	0.6792	0.0753	No
<= **32** & >**32**	60	0.6814	1	0.6814	0.0000	Yes

In the header of the table, "G1" refers to Group 1 which is on the left side of the "&" symbol, while "G2" refers to Group 2 which is on the right side of the "&" symbol.

Tables regarding the research of gesture classification

Table D1: Comparing the technical specifications of both Kinect sensors.

	Kinect v1	**Kinect v2**
Dimensions	27.94 cm × 6.35 cm × 3.81 cm	24.9 cm × 6.6 cm × 6.7 cm
Color camera resolution and fps	640 × 380 at 30 fps, 1280 × 720 at 12 fps	1920 × 1080 at 30 fps
IR resolution and fps	640 × 480 at 30 fps	512 × 424 at 30 fps
Depth resolution and fps	320 × 240 at 30 fps	512 × 424 at 30 fps
Field of view	57° horizontal, 43° vertical	70° horizontal, 60° vertical
Specified min. distance	0.4 m or 0.8 m	0.5 m
Recommended min. distance	1.8 m	1.4 m
Tested min. distance	1 m	0.7 m
Specified max. distance	4 m	4.5 m
Tested max. distance	6 m	4 m
Active infrared	Not available	Available
Measurement method	Infrared structured light	Time of Flight
Minimum latency	102 ms	20 ms
Microphone array	4 microphones, 16 kHz	4 microphones, 48 kHz
Tilt-motor	Available, ±27°	Not available
Temperature	Weak correlation	Strong correlation
More distance	Less accuracy	Same accuracy
Striped depth image	Increases with depth	No stripes on image
Depth precision	Higher	Less
Flying pixels	Not present	Present if surface is not flat
Environment color	Depth estimation is unaffected	Affects depth estimation
Multipath interference	Not present	Present
Angles affect precision	No	No
Precision decreasing	Second order polynomial	No mathematical behavior

Table D2: Real-time results using the general computer.

Circular movements (GC)				
MT	**AGAR ±0.05 m, ±0.10 m, ±0.15 m**	**Dispersion ±0.05 m, ±0.10 m, ±0.15 m**	**Average Time (ms)**	**Dispersion of Time (ms)**
AMT	0.207, 0.594, 0.871	0.202, 0.227, 0.163	3.047	14.476
GMT	0.004, 0.031, 0.066	0.022, 0.071, 0.095	1.673	3.811
HMT	0.238, 0.641, 0.762	0.232, 0.375, 0.297	1.194	0.377
CHMT	0.270, 0.609, 0.723	0.244, 0.362, 0.322	1.575	0.533
QMT	0.008, 0.121, 0.293	0.031, 0.166, 0.341	1.525	0.470
CMT	0.000, 0.012, 0.039	0.000, 0.037, 0.081	0.985	0.511

Waving movements (GC)				
MT	**AGAR ±0.05 m, ±0.10 m, ±0.15 m**	**Dispersion ±0.05 m, ±0.10 m, ±0.15 m**	**Average Time (ms)**	**Dispersion of Time (ms)**
AMT	0.734, 0.926, 0.965	0.224, 0.126, 0.057	3.048	6.061
GMT	0.000, 0.000, 0.000	0.000, 0.000, 0.000	1.965	1.443
HMT	0.762, 0.957, 0.969	0.225, 0.098, 0.078	1.878	1.754
CHMT	0.719, 0.930, 0.973	0.262, 0.100, 0.061	3.529	3.872
QMT	0.000, 0.035, 0.195	0.000, 0.177, 0.324	2.983	2.915
CMT	0.000, 0.000, 0.000	0.000, 0.000, 0.000	3.199	2.895

Forward-diagonal movements (GC)				
MT	**AGAR ±0.05 m, ±0.10 m, ±0.15 m**	**Dispersion ±0.05 m, ±0.10 m, ±0.15 m**	**Average Time (ms)**	**Dispersion of Time (ms)**
AMT	0.820, 0.980, 0.992	0.177, 0.056, 0.031	1.466	5.720
GMT	0.000, 0.000, 0.000	0.000, 0.000, 0.000	0.919	0.647
HMT	0.820, 0.988, 1.000	0.177, 0.037, 0.000	0.825	0.366
CHMT	0.848, 0.992, 0.996	0.158, 0.031, 0.022	0.925	0.337
QMT	0.004, 0.051, 0.141	0.022, 0.179, 0.297	0.901	0.365
CMT	0.000, 0.004, 0.004	0.000, 0.022, 0.022	0.955	0.288

Upward-diagonal movements (GC)				
MT	**AGAR ±0.05 m, ±0.10 m, ±0.15 m**	**Dispersion ±0.05 m, ±0.10 m, ±0.15 m**	**Average Time (ms)**	**Dispersion of Time (ms)**
AMT	0.188, 0.523, 0.754	0.165, 0.157, 0.154	1.771	7.031
GMT	0.043, 0.195, 0.277	0.093, 0.215, 0.296	1.023	0.372
HMT	0.184, 0.465, 0.668	0.155, 0.188, 0.151	0.896	0.381
CHMT	0.199, 0.484, 0.648	0.168, 0.212, 0.202	1.221	0.560
QMT	0.090, 0.293, 0.457	0.143, 0.299, 0.392	1.164	0.502
CMT	0.078, 0.270, 0.410	0.122, 0.303, 0.431	1.324	0.813

Table D3: Real-time results using the advanced computer.

Circular movements (AC)				
MT	**AGAR** **±0.05 m, ±0.10 m, ±0.15 m**	**Dispersion** **±0.05 m, ±0.10 m, ±0.15 m**	**Average Time (ms)**	**Dispersion of Time (ms)**
AMT	0.207, 0.609, 0.832	0.200, 0.224, 0.189	0.603	2.255
GMT	0.055, 0.300, 0.516	0.100, 0.237, 0.273	0.384	0.127
HMT	0.211, 0.523, 0.664	0.186, 0.255, 0.255	0.355	0.313
CHMT	0.234, 0.555, 0.684	0.228, 0.242, 0.244	0.454	0.166
QMT	0.148, 0.730, 0.984	0.221, 0.227, 0.042	0.410	0.137
CMT	0.140, 0.578, 0.820	0.210, 0.299, 0.206	0.272	0.135

Waving movements (AC)				
MT	**AGAR** **±0.05 m, ±0.10 m, ±0.15 m**	**Dispersion** **±0.05 m, ±0.10 m, ±0.15 m**	**Average Time (ms)**	**Dispersion of Time (ms)**
AMT	0.754, 0.969, 0.992	0.276, 0.064, 0.031	0.549	2.011
GMT	0.324, 0.578, 0.641	0.306, 0.386, 0.419	0.340	0.125
HMT	0.785, 0.965, 0.996	0.238, 0.085, 0.022	0.323	0.268
CHMT	0.793, 0.949, 0.965	0.237, 0.089, 0.073	0.433	0.172
QMT	0.566, 0.953, 0.984	0.331, 0.082, 0.042	0.393	0.151
CMT	0.383, 0.609, 0.660	0.328, 0.403, 0.423	0.414	0.163

Forward-diagonal movements (AC)				
MT	**AGAR** **±0.05 m, ±0.10 m, ±0.15 m**	**Dispersion** **±0.05 m, ±0.10 m, ±0.15 m**	**Average Time (ms)**	**Dispersion of Time (ms)**
AMT	0.613, 0.906, 0.961	0.186, 0.156, 0.103	0.486	1.987
GMT	0.234, 0.480, 0.543	0.257, 0.410, 0.423	0.304	0.265
HMT	0.621, 0.910, 0.945	0.214, 0.146, 0.110	0.266	0.074
CHMT	0.656, 0.922, 0.941	0.203, 0.141, 0.110	0.326	0.081
QMT	0.305, 0.637, 0.832	0.267, 0.387, 0.235	0.309	0.095
CMT	0.230, 0.488, 0.547	0.244, 0.406, 0.430	0.333	0.171

Upward-diagonal movements (AC)				
MT	**AGAR** **±0.05 m, ±0.10 m, ±0.15 m**	**Dispersion** **±0.05 m, ±0.10 m, ±0.15 m**	**Average Time (ms)**	**Dispersion of Time (ms)**
AMT	0.465, 0.809, 0.934	0.248, 0.203, 0.152	0.452	1.888
GMT	0.367, 0.672, 0.762	0.286, 0.312, 0.261	0.271	0.187
HMT	0.445, 0.852, 0.949	0.218, 0.189, 0.134	0.245	0.075
CHMT	0.488, 0.898, 0.949	0.251, 0.172, 0.130	0.306	0.193
QMT	0.418, 0.746, 0.902	0.278, 0.251, 0.145	0.297	0.187
CMT	0.383, 0.695, 0.797	0.280, 0.313, 0.276	0.315	0.176

Table D4: File-based results using the advanced computer.

Circular movements (AC)				
MT	**AGAR** **±0.05 m, ±0.10 m, ±0.15 m**	**Dispersion** **±0.05 m, ±0.10 m, ±0.15 m**	**Average Time (ms)**	**Dispersion of Time (ms)**
AMT	0.362, 0.643, 0.818	0.238, 0.247, 0.161	9.826	7.096
GMT	0.224, 0.435, 0.531	0.287, 0.361, 0.331	2.393	4.068
HMT	0.354, 0.630, 0.773	0.250, 0.263, 0.214	2.523	4.273
CHMT	0.375, 0.635, 0.771	0.247, 0.244, 0.224	2.556	3.980
QMT	0.234, 0.630, 0.849	0.275, 0.249, 0.179	9.763	7.108
CMT	0.227, 0.521, 0.711	0.288, 0.327, 0.328	3.274	4.632

Waving movements (AC)				
MT	**AGAR** **±0.05 m, ±0.10 m, ±0.15 m**	**Dispersion** **±0.05 m, ±0.10 m, ±0.15 m**	**Average Time (ms)**	**Dispersion of Time (ms)**
AMT	0.570, 0.875, 0.932	0.241, 0.144, 0.108	9.979	7.600
GMT	0.578, 0.880, 0.940	0.244, 0.141, 0.104	2.812	4.270
HMT	0.629, 0.915, 0.971	0.251, 0.141, 0.083	2.786	4.055
CHMT	0.627, 0.915, 0.971	0.235, 0.141, 0.087	2.850	4.193
QMT	0.619, 0.907, 0.961	0.239, 0.144, 0.105	12.072	9.624
CMT	0.627, 0.915, 0.971	0.248, 0.141, 0.087	3.503	4.705

Forward-diagonal movements (AC)				
MT	**AGAR** **±0.05 m, ±0.10 m, ±0.15 m**	**Dispersion** **±0.05 m, ±0.10 m, ±0.15 m**	**Average Time (ms)**	**Dispersion of Time (ms)**
AMT	0.680, 0.875, 0.948	0.319, 0.208, 0.131	6.185	4.938
GMT	0.341, 0.599, 0.651	0.342, 0.393, 0.373	2.722	4.107
HMT	0.677, 0.878, 0.922	0.327, 0.212, 0.176	2.863	3.899
CHMT	0.721, 0.883, 0.938	0.304, 0.215, 0.161	2.977	3.996
QMT	0.451, 0.737, 0.841	0.356, 0.359, 0.275	7.230	5.344
CMT	0.381, 0.592, 0.679	0.332, 0.390, 0.375	2.917	4.273

Upward-diagonal movements (AC)				
MT	**AGAR** **±0.05 m, ±0.10 m, ±0.15 m**	**Dispersion** **±0.05 m, ±0.10 m, ±0.15 m**	**Average Time (ms)**	**Dispersion of Time (ms)**
AMT	0.424, 0.672, 0.779	0.324, 0.332, 0.289	7.687	6.528
GMT	0.403, 0.620, 0.727	0.351, 0.381, 0.324	2.775	4.002
HMT	0.432, 0.740, 0.867	0.317, 0.312, 0.260	2.806	4.330
CHMT	0.455, 0.766, 0.862	0.336, 0.318, 0.287	2.903	4.079
QMT	0.408, 0.654, 0.773	0.354, 0.367, 0.290	9.058	7.208
CMT	0.412, 0.680, 0.815	0.355, 0.361, 0.279	3.915	5.503

Table D5: Comparing the execution time of the APBMR algorithm (ms).

Comparing the averages of time.								
MT	**AC 1**	**AC 2**	**AC 3**	**AC 4**	**File 1**	**File 2**	**File 3**	**File 4**
AMT	0.603	0.549	0.486	0.452	9.826	9.979	6.185	7.687
GMT	0.384	0.340	0.304	0.271	2.393	2.812	2.722	2.775
HMT	0.355	0.323	0.266	0.245	2.523	2.786	2.863	2.806
CHMT	0.454	0.433	0.326	0.306	2.556	2.850	2.977	2.903
QMT	0.410	0.393	0.309	0.297	9.763	12.072	7.230	9.058
CMT	0.272	0.414	0.333	0.315	3.274	3.503	2.917	3.915

Comparing the averages of time.								
MT	**GC 1**	**GC 2**	**GC 3**	**GC 4**	**File 1**	**File 2**	**File 3**	**File 4**
AMT	3.047	3.048	1.466	1.771	9.826	9.979	6.185	7.687
GMT	1.673	1.965	0.919	1.023	2.393	2.812	2.722	2.775
HMT	1.194	1.878	0.825	0.896	2.523	2.786	2.863	2.806
CHMT	1.575	3.529	0.925	1.221	2.556	2.850	2.977	2.903
QMT	1.525	2.983	0.901	1.164	9.763	12.072	7.230	9.058
CMT	0.985	3.199	0.955	1.324	3.274	3.503	2.917	3.915

In the header of the table, “1” refers to the circular, “2” to the waving, “3” to the forward-diagonal and “4” to the upward-diagonal gestures.

Table D6: Comparing the APBMR to the RDAMR algorithm.

Comparing the AGARs of the circular movements.							
AD	**AMT**	**GMT**	**HMT**	**CHMT**	**QMT**	**CMT**	**RDAMR**
±0.05 m	0.305	0.082	0.273	0.316	0.105	0.094	0.078
±0.10 m	0.559	0.238	0.523	0.539	0.543	0.379	0.227
±0.15 m	0.789	0.359	0.699	0.684	0.816	0.605	0.352
±0.20 m	-	-	-	-	-	-	0.555
±0.25 m	-	-	-	-	-	-	0.754
±0.30 m	-	-	-	-	-	-	0.852

Comparing the AGARs of the waving movements.							
AD	**AMT**	**GMT**	**HMT**	**CHMT**	**QMT**	**CMT**	**RDAMR**
±0.05 m	0.617	0.617	0.667	0.675	0.635	0.655	0.180
±0.10 m	0.926	0.930	0.943	0.940	0.935	0.940	0.660
±0.15 m	0.977	0.980	0.988	0.992	0.980	0.992	0.965
±0.20 m	-	-	-	-	-	-	1.000

Comparing the AGARs of the forward-diagonal movements.							
AD	**AMT**	**GMT**	**HMT**	**CHMT**	**QMT**	**CMT**	**RDAMR**
±0.05 m	0.855	0.508	0.863	0.887	0.652	0.547	0.062
±0.10 m	0.980	0.793	0.965	0.969	0.945	0.773	0.344
±0.15 m	0.996	0.809	0.980	0.977	0.980	0.840	0.707
±0.20 m	-	-	-	-	-	-	0.875
±0.25 m	-	-	-	-	-	-	0.992

Comparing the AGARs of the upward-diagonal movements.							
AD	**AMT**	**GMT**	**HMT**	**CHMT**	**QMT**	**CMT**	**RDAMR**
±0.05 m	0.586	0.600	0.597	0.628	0.605	0.611	0.074
±0.10 m	0.832	0.855	0.902	0.934	0.883	0.910	0.270
±0.15 m	0.922	0.918	0.996	0.996	0.926	0.973	0.477
±0.20 m	-	-	-	-	-	-	0.684
±0.25 m	-	-	-	-	-	-	0.879

Comparing the AGARs of all movements.							
AD	**AMT**	**GMT**	**HMT**	**CHMT**	**QMT**	**CMT**	**RDAMR**
±0.05 m	0.591	0.452	0.600	0.627	0.499	0.477	0.099
±0.10 m	0.824	0.704	0.833	0.845	0.827	0.750	0.375
±0.15 m	0.921	0.767	0.916	0.912	0.926	0.852	0.625
±0.20 m	-	-	-	-	-	-	0.778
±0.25 m	-	-	-	-	-	-	0.875

Table D7: The results of all movement descriptors.

All movements				
MT	**AGAR** **±0.05 m, ±0.10 m, ±0.15 m**	**Dispersion** **±0.05 m, ±0.10 m, ±0.15 m**	**Average Time (ms)**	**Dispersion of Time (ms)**
AMT	0.505, 0.782, 0.896	0.319, 0.250, 0.173	4.424	7.546
GMT	0.241, 0.435, 0.508	0.308, 0.404, 0.412	1.638	3.083
HMT	0.515, 0.789, 0.876	0.323, 0.275, 0.213	1.604	2.958
CHMT	0.534, 0.796, 0.871	0.324, 0.270, 0.222	1.835	3.084
QMT	0.293, 0.568, 0.709	0.328, 0.399, 0.380	4.655	6.642
CMT	0.263, 0.480, 0.574	0.316, 0.410, 0.432	2.076	3.521

Circular movements				
MT	**AGAR** **±0.05 m, ±0.10 m, ±0.15 m**	**Dispersion** **±0.05 m, ±0.10 m, ±0.15 m**	**Average Time (ms)**	**Dispersion of Time (ms)**
AMT	0.273, 0.619, 0.837	0.235, 0.237, 0.172	5.254	9.962
GMT	0.113, 0.280, 0.394	0.245, 0.323, 0.307	1.613	3.452
HMT	0.280, 0.603, 0.739	0.236, 0.263, 0.236	1.524	2.955
CHMT	0.305, 0.605, 0.732	0.248, 0.245, 0.234	1.678	2.762
QMT	0.145, 0.513, 0.729	0.257, 0.248, 0.155	4.737	6.390
CMT	0.137, 0.392, 0.550	0.262, 0.315, 0.289	2.005	3.270

Waving movements				
MT	**AGAR** **±0.05 m, ±0.10 m, ±0.15 m**	**Dispersion** **±0.05 m, ±0.10 m, ±0.15 m**	**Average Time (ms)**	**Dispersion of Time (ms)**
AMT	0.676, 0.925, 0.968	0.267, 0.124, 0.088	5.305	7.324
GMT	0.347, 0.551, 0.595	0.300, 0.310, 0.316	1.864	3.075
HMT	0.712, 0.941, 0.978	0.256, 0.123, 0.067	1.823	2.996
CHMT	0.701, 0.923, 0.970	0.248, 0.123, 0.081	2.353	3.655
QMT	0.427, 0.671, 0.749	0.279, 0.124, 0.085	6.138	8.334
CMT	0.378, 0.566, 0.605	0.306, 0.314, 0.318	2.534	3.699

Forward-diagonal movements				
MT	**AGAR** **±0.05 m, ±0.10 m, ±0.15 m**	**Dispersion** **±0.05 m, ±0.10 m, ±0.15 m**	**Average Time (ms)**	**Dispersion of Time (ms)**
AMT	0.701, 0.914, 0.964	0.274, 0.188, 0.120	3.208	5.260
GMT	0.213, 0.394, 0.434	0.314, 0.402, 0.395	1.516	2.916
HMT	0.702, 0.919, 0.951	0.287, 0.188, 0.153	1.539	2.812
CHMT	0.739, 0.925, 0.955	0.269, 0.189, 0.142	1.633	2.877
QMT	0.281, 0.512, 0.638	0.323, 0.371, 0.258	3.444	4.804
CMT	0.229, 0.394, 0.448	0.307, 0.397, 0.400	1.618	3.028

Upward-diagonal movements				
MT	**AGAR** **±0.05 m, ±0.10 m, ±0.15 m**	**Dispersion** **±0.05 m, ±0.10 m, ±0.15 m**	**Average Time (ms)**	**Dispersion of Time (ms)**
AMT	0.368, 0.669, 0.816	0.295, 0.294, 0.254	3.929	6.648
GMT	0.290, 0.513, 0.608	0.325, 0.354, 0.299	1.559	2.845
HMT	0.365, 0.693, 0.834	0.280, 0.274, 0.221	1.529	3.058
CHMT	0.391, 0.723, 0.826	0.304, 0.276, 0.240	1.681	2.909
QMT	0.320, 0.577, 0.720	0.324, 0.327, 0.250	4.300	6.279
CMT	0.308, 0.567, 0.694	0.325, 0.340, 0.276	2.146	3.956

Table D8: Evaluating the gesture descriptors with the APBMR on two axes.

MT	**AGAR** (x, y) **±0.05 m, ±0.10 m, ±0.15 m**	**AGAR** (x, z) **±0.05 m, ±0.10 m, ±0.15 m**	**AGAR** (y, z) **±0.05 m, ±0.10 m, ±0.15 m**
AMT	0.566, 0.810, 0.918	0.746, 0.949, 0.985	0.632, 0.837, 0.927
GMT	0.264, 0.436, 0.499	0.736, 0.935, 0.979	0.295, 0.446, 0.499
HMT	0.577, 0.811, 0.888	0.761, 0.962, 0.991	0.640, 0.826, 0.890
CHMT	0.598, 0.818, 0.881	0.763, 0.957, 0.991	0.653, 0.834, 0.881
QMT	0.324, 0.582, 0.711	0.746, 0.954, 0.991	0.355, 0.593, 0.712
CMT	0.290, 0.483, 0.564	0.739, 0.952, 0.993	0.322, 0.493, 0.560

Table D9: Evaluating the gesture descriptors with the APBMR on one axis.

MT	**AGAR** (x) **±0.05 m, ±0.10 m, ±0.15 m**	**AGAR** (y) **±0.05 m, ±0.10 m, ±0.15 m**	**AGAR** (z) **±0.05 m, ±0.10 m, ±0.15 m**
AMT	0.798, 0.953, 0.987	0.664, 0.842, 0.928	0.930, 0.994, 0.999
GMT	0.785, 0.940, 0.979	0.308, 0.448, 0.499	0.930, 0.994, 0.999
HMT	0.816, 0.968, 0.993	0.672, 0.829, 0.891	0.930, 0.994, 0.999
CHMT	0.819, 0.963, 0.993	0.688, 0.839, 0.882	0.927, 0.993, 0.998
QMT	0.797, 0.960, 0.993	0.372, 0.598, 0.713	0.931, 0.993, 0.999
CMT	0.793, 0.958, 0.994	0.338, 0.496, 0.561	0.927, 0.993, 0.999

Supplementary figures

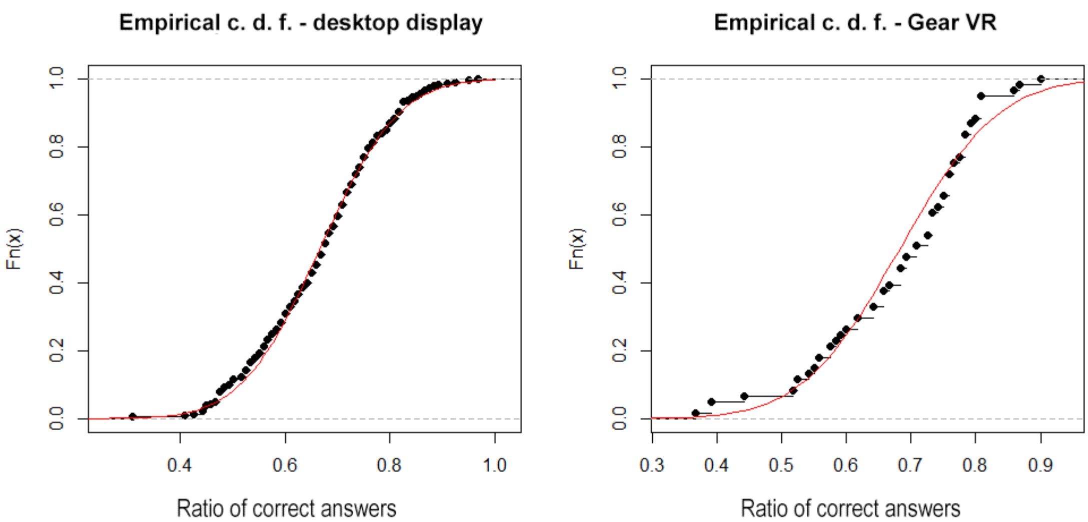

Figure E1: ECDFs of the rates of correct answers in the case of the DD (left) and the GVR (right).

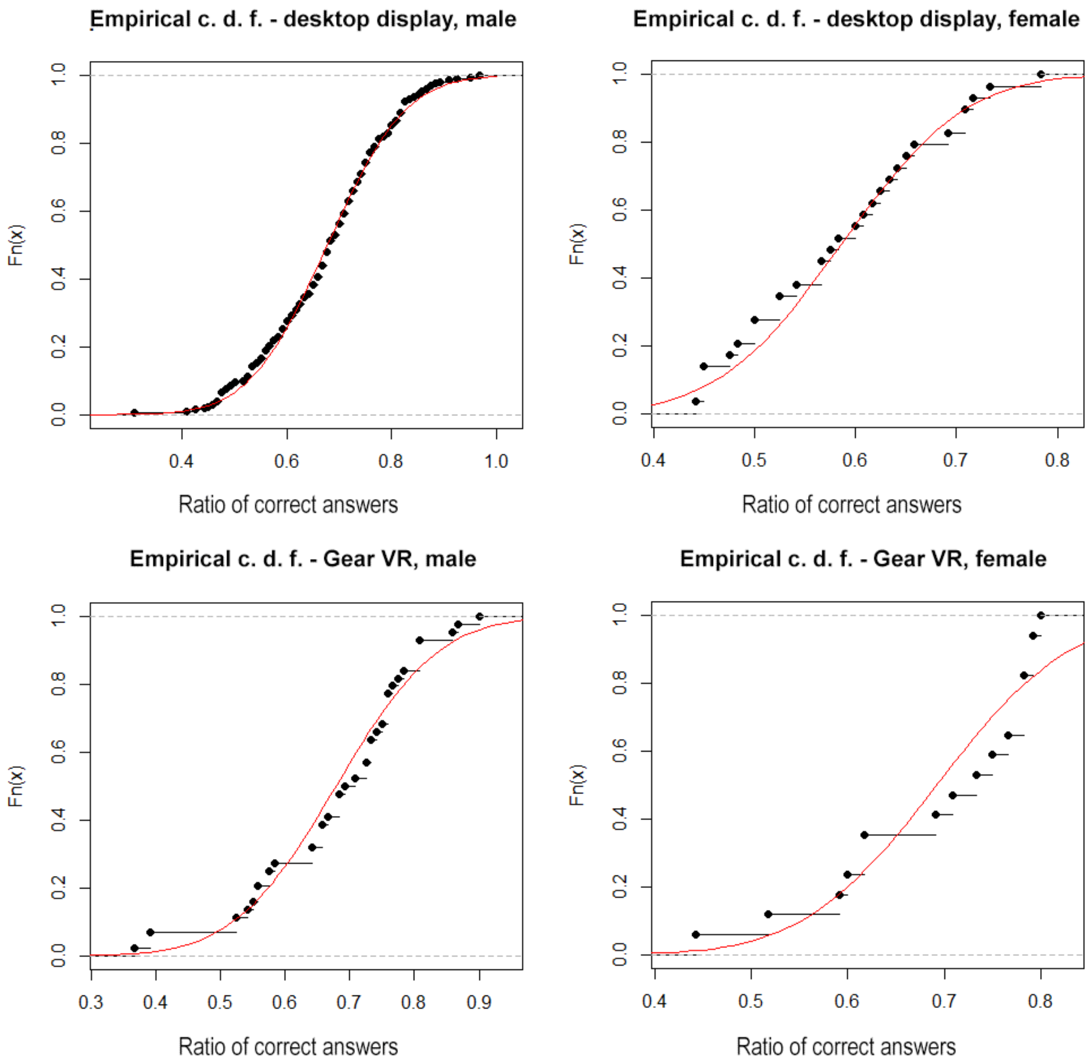

Figure E2: ECDFs of the rates of correct answers in the case of the genders.

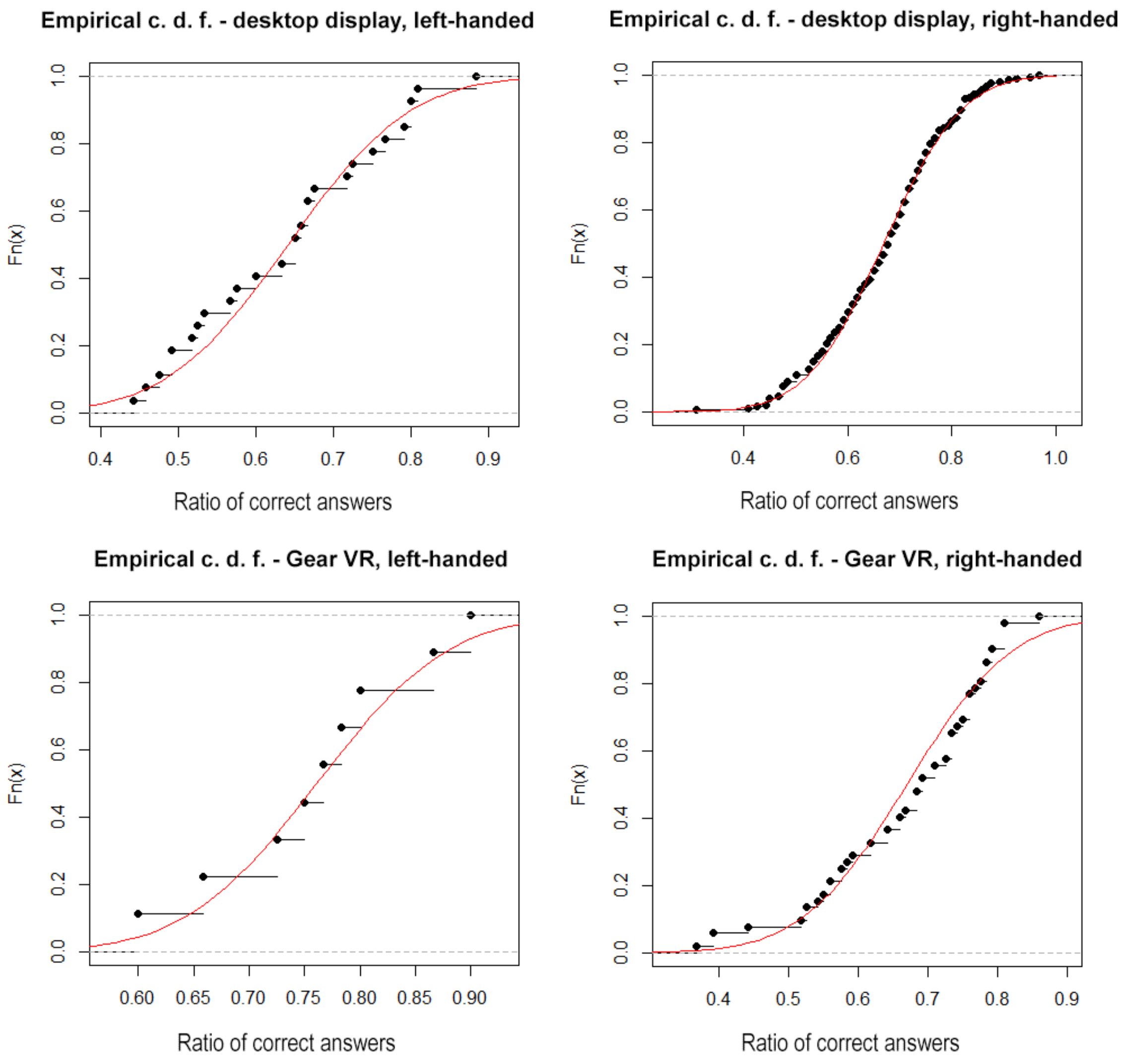

Figure E3: ECDFs of the rates of correct answers in the case of the primary hand of the users.

Bibliography

[1] William R. Sherman and Alan B. Craig. *Understanding virtual reality: Interface, application, and design.* Morgan Kaufmann Publishers: San Francisco, CA, USA, 2002.

[2] Greg Kipper and Joseph Rampolla. *Augmented Reality: an emerging technologies guide to AR.* Syngress: Waltham, MA, USA, 2012.

[3] Hideyuki Tamura, Hiroyuki Yamamoto, and Akihiro Katayama. Mixed reality: Future dreams seen at the border between real and virtual worlds. *IEEE Computer Graphics and Applications*, 21(6):64–70, 2001.

[4] Virtual and Augmented Reality Users 2019 - eMarketer Trends, Forecast & Statistics. `https://www.emarketer.com/content/virtual-and-augmented-reality-users-2019`.

[5] Global Augmented Reality & Mixed Reality Market: Growth, Trends and Forecasts to 2025 - ResearchAndMarkets.com | Business Wire. `https://www.businesswire.com/news/home/20200515005267/en/Global-Augmented-Reality-Mixed-Reality-Market-Growth`.

[6] Joseph Psotka. Immersive training systems: Virtual reality and education and training. *Instructional science*, 23(5-6):405–431, 1995.

[7] Jannat Falah, Soheeb Khan, Tasneem Alfalah, Salsabeel F.M. Alfalah, Warren Chan, David K. Harrison, and Vassilis Charissis. Virtual reality medical training system for anatomy education. In *2014 Science and Information Conference*, pages 752–758. IEEE, 2014.

[8] Laura Freina and Michela Ott. A literature review on immersive virtual reality in education: state of the art and perspectives. In *The International Scientific Conference eLearning and Software for Education*, volume 1, 2015.

[9] Richard M. Satava. Medical applications of virtual reality. *Journal of Medical Systems*, 19(3):275–280, 1995.

[10] Willem I.M. Willaert, Rajesh Aggarwal, Isabelle Van Herzeele, Nicholas J. Cheshire, and Frank E. Vermassen. Recent advancements in medical simulation: patient-specific virtual reality simulation. *World journal of surgery*, 36(7):1703–1712, 2012.

[11] Neal E. Seymour, Anthony G. Gallagher, Sanziana A. Roman, Michael K. O'brien, Vipin K. Bansal, Dana K. Andersen, and Richard M. Satava. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Annals of surgery*, 236(4):458, 2002.

[12] Gunnar Ahlberg, Lars Enochsson, Anthony G. Gallagher, Leif Hedman, Christian Hogman, David A. McClusky III, Stig Ramel, C. Daniel Smith, and Dag Arvidsson. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *The American journal of surgery*, 193(6):797–804, 2007.

[13] Julie Dugdale, Bernard Pavard, Nico Pallamin, Mehdi El Jed, and Commander Laurent Maugan. Emergency fire incident training in a virtual world. *Proceedings ISCRAM*, 167, 2004.

[14] Ajey Lele. Virtual reality and its military utility. *Journal of Ambient Intelligence and Humanized Computing*, 4(1):17–26, 2013.

[15] Joseph Bates. Virtual reality, art, and entertainment. *Presence: Teleoperators & Virtual Environments*, 1(1):133–138, 1992.

[16] Michael Zyda. From visual simulation to virtual reality to games. *Computer*, 38(9):25–32, 2005.

[17] Bimo Sunarfri Hantono, Lukito Edi Nugroho, and P. Insap Santosa. Review of augmented reality agent in education. In *2016 6th International Annual Engineering Seminar (InAES)*, pages 150–153. IEEE, 2016.

[18] Philip Kortum. *HCI beyond the GUI: Design for haptic, speech, olfactory, and other nontraditional interfaces.* Elsevier: Burlington, MA, USA, 2008.

[19] Domenica Mirauda, Nicola Capece, and Ugo Erra. Streamflowvl: A virtual fieldwork laboratory that supports traditional hydraulics engineering learning. *Applied Sciences*, 9(22):4972, 2019.

[20] Zahra Al Mahdi, Vikas Rao Naidu, and Preethy Kurian. Analyzing the role of human computer interaction principles for e-learning solution design. In *Smart Technologies and Innovation for a Sustainable Future*, pages 41–44. Springer, 2019.

[21] Pu Liu, Sidney Fels, Nicholas West, and Matthias Görges. Human Computer Interaction Design for Mobile Devices Based on a Smart Healthcare Architecture. *arXiv e-prints*, page arXiv:1902.03541, February 2019.

[22] Zhijuan Zhu, Wenzhen Pan, Xin Ai, and Renjun Zhen. Research on human-computer interaction design of bed rehabilitation equipment for the elderly. In *International Conference on Applied Human Factors and Ergonomics*, pages 275–286. Springer, 2019.

[23] Cecilia Sik-Lanyi, Veronika Szucs, Shervin Shirmohammadi, Petya Grudeva, Boris Abersek, Tibor Guzsvinecz, and Karel Van Isacker. How to develop serious games for social and cognitive competence of children with learning difficulties. *Acta Polytechnica Hungarica*, 16(9):149–169, 2019.

[24] Tibor Guzsvinecz, Veronika Szucs, and Cecilia Sik-Lanyi. Designing gamified virtual reality applications with sensors - a gamification study. In *Proceedings of the Pannonian Conference on Advances in Information Technology (PCIT'2019)*, pages 105–112, 2019.

[25] Tao Ding and Duanzhen Zhu. Applications of the human-computer interaction interface to moba mobile games. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 11071, 2019.

[26] Hind Kharoub, Mohammed Lataifeh, and Naveed Ahmed. 3d user interface design and usability for immersive vr. *Applied Sciences*, 9(22):4861, 2019.

[27] Alistair G. Sutcliffe, Charalambos Poullis, Andreas Gregoriades, Irene Katsouri, Aimilia Tzanavari, and Kyriakos Herakleous. Reflecting on the design process for virtual reality applications. *International Journal of Human–Computer Interaction*, 35(2):168–179, 2019.

[28] George Drettakis, Maria Roussou, Alex Reche, and Nicolas Tsingos. Design and evaluation of a real-world virtual environment for architecture and urban planning. *Presence: Teleoperators and Virtual Environments*, 16(3):318–332, 2007.

[29] Kunwar Aditya, Praise Chacko, Deeksha Kumari, Divya Kumari, and Saurabh Bilgaiyan. Recent trends in hci: A survey on data glove, leap motion and microsoft kinect. In *2018 IEEE International Conference on System, Computation, Automation and Networking (ICSCA)*, pages 1–5. IEEE, 2018.

[30] Lucas Medeiros Souza do Nascimento, Lucas Vacilotto Bonfati, Melissa La Banca Freitas, José Jair Alves Mendes Junior, Hugo Valadares Siqueira, and Sergio Luiz Stevan. Sensors and systems for physical rehabilitation and health monitoring—a review. *Sensors*, 20(15):4063, 2020.

[31] Marcia C. Linn and Anne C. Petersen. Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child development*, pages 1479–1498, 1985.

[32] Jeffrey M. Zacks. Neuroimaging studies of mental rotation: a meta-analysis and review. *Journal of cognitive neuroscience*, 20(1):1–19, 2008.

[33] Marc Jeannerod and Pierre Jacob. Visual cognition: a new look at the two-visual systems model. *Neuropsychologia*, 43(2):301–312, 2005.

[34] Michael A. Motes, Rafael Malach, and Maria Kozhevnikov. Object-processing neural efficiency differentiates object from spatial visualizers. *Neuroreport*, 19(17):1727–1731, 2008.

[35] Peter Herbert Maier. Spatial geometry and spatial ability–how to make solid geometry solid. In *Selected papers from the Annual Conference of Didactics of Mathematics*, pages 63–75, 1996.

[36] Michael Peters, Peter Chisholm, and Bruno Laeng. Spatial ability, student gender, and academic performance. *Journal of Engineering Education*, 84(1):69–73, 1995.

[37] Sarah A. Burnett, David M. Lane, and Lewis M. Dratt. Spatial ability and handedness. *Intelligence*, 6(1):57–68, 1982.

[38] Walter F. McKeever. The influences of handedness, sex, familial sinistrality and androgyny on language laterality, verbal ability, and spatial ability. *Cortex*, 22(4):521–537, 1986.

[39] S.B. O'Sullivan, T.J. Schmitz, and G.G. Fulk. Physical rehabilitation 6th ed (p. 661). *Philadelphia, PA: FA Davis*, 2014.

[40] Hye-Sun Lee, Jae-Heon Lim, Byeong-Hyeon Jeon, and Chiang-Soon Song. Non-immersive virtual reality rehabilitation applied to a task-oriented approach for stroke patients: A randomized controlled trial. *Restorative Neurology and Neuroscience*, (Preprint):1–8, 2020.

[41] Steven G. Vandenberg and Allan R. Kuse. Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and motor skills*, 47(2):599–604, 1978.

[42] Holly K. Ault and Samuel John. Assessing and enhancing visualization skills of engineering students in africa: A comparative study. *The Engineering Design Graphics Journal*, 74(2), 2010.

[43] Rita Nagy-Kondor and Csilla Sörös. Engineering students' spatial abilities in budapest and debrecen. In *Annales Mathematicae et Informaticae*, volume 40, pages 187–201, 2012.

[44] Kumiko Shiina, Dennis R. Short, Craig L. Miller, and Kenjiro Suzuki. Development of software to record solving process of a mental rotations test. *Journal for Geometry and Graphics*, 5(2):193–202, 2001.

[45] College Entrance Examination Board (CEEB). Special aptitude test in spatial relations, developed by the college entrance examination board, 1939.

[46] Agnes Bosnyak and Rita Nagy-Kondor. The spatial ability and spatial geometrical knowledge of university students majored in mathematics. *Acta Didactica Universitatis Comenianae*, 8:1–25, 2008.

[47] Folkert Hendrik Haanstra. Effects of art education on visual-spatial ability and aesthetic perception: two meta-analyses. 1994.

[48] Brigitta Németh, Csilla Sörös, and Miklós Hoffmann. Typical mistakes in mental cutting test and their consequences in gender differences. *Teaching Mathematics and Computer Science*, 5(2):385–392, 2007.

[49] Emiko Tsutsumi, Kanakao Shiina, Ayako Suzaki, Kyoko Yamanouchi, Saito Takaaki, and Kenjiro Suzuki. A mental cutting test on female students using a stereographic system. *Journal for Geometry and Graphics*, 3(1):111–119, 1999.

[50] Melih Turgut and Rita Nagy-Kondor. Comparison of hungarian and turkish prospective mathematics teachers' mental cutting performances. *Acta Didactica Universitatis Comenianae*, 13:47–58, 2013.

[51] Roland B. Guay. Purdue spatial visualization test - visualization of rotations. *W. Lafayette, IN. Purdue Research Foundation*, 1977.

[52] Theodore J. Branoff and Patrick E. Connolly. The addition of coordinate axes to the purdue spatial visualization test-visualization of rotations: A study at two universities. In *Proc. of ASEE Annual Conference & Exposition*, pages 1–9, 1999.

[53] Vivian Lee Seiver Heinrich. *The development and validation of a spatial perception test for selection purposes.* PhD thesis, The Ohio State University, 1989.

[54] Ruth B. Ekstrom, Diran Dermen, and Harry Horace Harman. *Manual for kit of factor-referenced cognitive tests*, volume 102. Educational testing service Princeton, NJ, 1976.

[55] Christopher A. Sanchez. Enhancing visuospatial performance through video game training to increase learning in visuospatial science domains. *Psychonomic Bulletin & Review*, 19(1):58–65, 2012.

[56] Sijing Wu and Ian Spence. Playing shooter and driving videogames improves top-down guidance in visual search. *Attention, Perception, & Psychophysics*, 75(4):673–686, 2013.

[57] Veronika Szücs and Cecilia Sik-Lanyi. Abilities and limitations of assistive technologies in post-stroke therapy based on virtual/augmented reality. In *Assistive Technology: From Research to practice, 12th European AAATE conference, IOS Press*, pages 1087–1091, 2013.

[58] Cecilia Sik-Lanyi and Veronika Szücs. Stroke care systems games applied for therapy in stroke tele-rehabilitation: Wsc-1477. *International Journal of Stroke*, 9, 2014.

[59] Cecilia Sik-Lanyi and Veronika Szucs. Motivating rehabilitation through competitive gaming. In *Modern Stroke Rehabilitation Through e-Health-Based Entertainment*, pages 137–167. Springer, 2016.

[60] Veronika Szucs, Silvia Paxian, and Cecilia Sik-Lanyi. Augmented reality–where it started from and where it's going. *The Thousand Faces of Virtual Reality. London: Intech Open*, pages 37–63, 2014.

[61] Michael Yates, Arpad Kelemen, and Cecilia Sik-Lanyi. Virtual reality gaming in the rehabilitation of the upper extremities post-stroke. *Brain injury*, 30(7):855–863, 2016.

[62] Chuan-Jun Su, Chang-Yu Chiang, and Jing-Yan Huang. Kinect-enabled home-based rehabilitation system using dynamic time warping and fuzzy logic. *Applied Soft Computing*, 22:652–666, 2014.

[63] Kathryn LaBelle. Evaluation of kinect joint tracking for clinical and in-home stroke rehabilitation tools. *Undergraduate Thesis, University of Notre Dame*, 2011.

[64] Wenbing Zhao, Hai Feng, Roanna Lun, Deborah D. Espy, and M. Ann Reinthal. A kinect-based rehabilitation exercise monitoring and guidance system. In *2014 IEEE 5th International Conference on Software Engineering and Service Science*, pages 762–765. IEEE, 2014.

[65] Tibor Guzsvinecz, Veronika Szucs, and Cecilia Sik-Lanyi. Suitability of the kinect sensor and leap motion controller—a literature review. *Sensors*, 19(5):1072, 2019.

[66] Veronika Szucs, Cecilia Sik-Lanyi, and Tibor Guzsvinecz. Presenting the user's focus in needs & development (UFIND) method and its comparison to other design methods. In *11th IEEE International Conference on Cognitive Infocommunications*, pages 89–95, 2020.

[67] Barbara Olasov Rothbaum, Larry Hodges, Samantha Smith, Jeong Hwan Lee, and Larry Price. A controlled study of virtual reality exposure therapy for the fear of flying. *Year Book of Psychiatry and Applied Mental Health*, 2002(1):109–111, 2002.

[68] Mónica S Cameirão, Sergi Bermúdez i Badia, Esther Duarte Oller, and Paul FMJ Verschure. Neurorehabilitation using the virtual reality based rehabilitation gaming system: methodology, design, psychometrics, usability and validation. *Journal of neuroengineering and rehabilitation*, 7(1):48, 2010.

[69] Amy Henderson, Nicol Korner-Bitensky, and Mindy Levin. Virtual reality in stroke rehabilitation: a systematic review of its effectiveness for upper limb motor recovery. *Topics in stroke rehabilitation*, 14(2):52–61, 2007.

[70] Joong Hwi Kim, Sung Ho Jang, Chung Sun Kim, Ji Hee Jung, and Joshua H You. Use of virtual reality to enhance balance and ambulation in chronic stroke: a double-blind, randomized controlled study. *American Journal of physical medicine & rehabilitation*, 88(9):693–701, 2009.

[71] Kate E Laver, Belinda Lange, Stacey George, Judith E Deutsch, Gustavo Saposnik, and Maria Crotty. Virtual reality for stroke rehabilitation. *Stroke*, 49(4):e160–e161, 2018.

[72] Alma S Merians, Eugene Tunik, and Sergei V Adamovich. Virtual reality to maximize function for hand and arm rehabilitation: exploration of neural mechanisms. *Studies in health technology and informatics*, 145:109, 2009.

[73] Yu-Hyung Park, Chi-ho Lee, and Byoung-Hee Lee. Clinical usefulness of the virtual reality-based postural control training on the gait ability in patients with stroke. *Journal of exercise rehabilitation*, 9(5):489, 2013.

[74] Albert A Rizzo, Maria Schultheis, Kimberly A Kerns, and Catherine Mateer. Analysis of assets for virtual reality applications in neuropsychology. *Neuropsychological rehabilitation*, 14(1-2):207–239, 2004.

[75] Ksenia I Ustinova, Wesley A Leonard, Nicholas D Cassavaugh, and Christopher D Ingersoll. Development of a 3d immersive videogame to improve arm-postural coordination in patients with tbi. *Journal of NeuroEngineering and Rehabilitation*, 8(1):61, 2011.

[76] Veronika Szucs, Tibor Guzsvinecz, and Attila Magyar. Improved algorithms for movement pattern recognition and classification in physical rehabilitation. In *10th IEEE International Conference on Cognitive Infocommunications*, pages 417–424, 2019.

[77] Veronika Szücs, Tibor Guzsvinecz, and Attila Magyar. Movement pattern recognition in physical rehabilitation-cognitive motivation-based it method and algorithms. *Acta Polytechnica Hungarica*, 17(2), 2020.

[78] Howard Gardner. *Frames of mind: The theory of multiple intelligences*. Basic Books, New York, 1983.

[79] Craig L. Miller and Gary R. Bertoline. Spatial visualization research and theories: Their importance in the development of an engineering and technical design graphics curriculum model. *Engineering Design Graphics Journal*, 55(3):5–14, 1991.

[80] Craig L. Miller. Enhancing visual literacy of engineering students through the use of real and computer generated models. *Engineering Design Graphics Journal*, 56(1):27–38, 1992.

[81] Edwin E. Ghiselli. The validity of aptitude tests in personnel selection. *Personnel Psychology*, 26(4):461–477, 1973.

[82] Sheryl Sorby, Thomas Drummer, Kedmon Hungwe, and Paul Charlesworth. Developing 3 d spatial visualization skills for non engineering students. In *2005 Annual Conference*, pages 10–428, 2005.

[83] Herman A. Witkin. Individual differences in ease of perception of embedded figures. *Journal of personality*, 19(1):1–15, 1950.

[84] Jean Piaget and Bärbet Inhelder. *Child's Conception of Space: Selected Works vol 4*, volume 4. Routledge: Abingdon, Oxon, United Kingdom & New York, NY, USA, 2013.

[85] Richard E. Stafford. Sex differences in spatial visualization as evidence of sex-linked inheritance. *Perceptual and motor skills*, 13(3):428–428, 1961.

[86] Herman A. Witkin and Solomon E. Asch. Studies in space orientation. iv. further experiments on perception of the upright with displaced visual fields. *Journal of experimental psychology*, 38(6):762, 1948.

[87] William H Quasha and Rensis Likert. The revised minnesota paper form board test. *Journal of Educational Psychology*, 28(3):197, 1937.

[88] Mona Mohamed Kamal Hijazi. Attention, visual perception and their relationship to sport performance in fencing. *Journal of human kinetics*, 39(1):195–201, 2013.

[89] Thomas Romeas and Jocelyn Faubert. Assessment of sport specific and non-specific biological motion perception in soccer athletes shows a fundamental perceptual ability advantage over non-athletes for reorganizing body kinematics. *Journal of vision*, 15(12):504, 2015.

[90] Maryam Delavar, Iryna V. Kolesnikova, Beheshte Rahimzade, Mahdi Ghahhari, Alimorad Mosapuor, and Ali Moradi. Development of mental rotation ability at primary school level. *World Scientific News*, 101:77–88, 2018.

[91] Andrew J. Latham, Lucy L.M. Patston, and Lynette J. Tippett. The virtual brain: 30 years of video-game play and cognitive abilities. *Frontiers in psychology*, 4:629, 2013.

[92] Best Jobs with Good Visual and Spatial Skills|LoveToKnow. `https://jobs.lovetoknow.com/Best_Jobs_with_Good_Visual_and_Spatial_\Skills`.

[93] Károly Hercegfi, Anita Komlódi, Bálint Szabó, Máté Köles, Emma Lógó, Balázs P. Hámornik, and Gyöngyi Rózsa. Experiences of virtual desktop collaboration experiments. In *2015 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pages 375–379. IEEE, 2015.

[94] Ildikó Horváth. Innovative engineering education in the cooperative vr environment. In *2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pages 359–364. IEEE, 2016.

[95] Attila Kovari. Coginfocom supported education: A review of coginfocom based conference papers. In *2018 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pages 233–236. IEEE, 2018.

[96] Adam Wilson. Analysis of current virtual reality methods to enhance learning in education. *Sel. Comput. Res. Pap*, 8:61–66, 2019.

[97] Jordi Torner Ribé, Francesc Alpiste Penalba, and Miguel Ángel Brigos Hermida. Virtual reality application to improve spatial ability of engineering students. In *WSCG'2016-24th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2016 proceedings are published in Computer Science Research Notes [CSRN] ISSN 2464-4617*, pages 69–77, 2016.

[98] Rafael Molina-Carmona, María Luisa Pertegal-Felices, Antonio Jimeno-Morenilla, and Higinio Mora-Mora. Assessing the impact of virtual reality on engineering students' spatial ability. *The Future of Innovation and Technology in Education: Policies and Practices for Teaching and Learning Excellence*, pages 171–185, 2018.

[99] Rafael Molina-Carmona, María Luisa Pertegal-Felices, Antonio Jimeno-Morenilla, and Higinio Mora-Mora. Virtual reality learning activities for multimedia students to enhance spatial ability. *Sustainability*, 10(4):1074, 2018.

[100] Thomas D. Parsons, Peter Larson, Kris Kratz, Marcus Thiebaux, Brendon Bluestein, J. Galen Buckwalter, and Albert A. Rizzo. Sex differences in mental rotation and spatial rotation in a virtual environment. *Neuropsychologia*, 42(4):555–562, 2004.

[101] Miroslav Macik. Cognitive aspects of spatial orientation. *Acta Polytechnica Hungarica*, 15(5):149–167, 2018.

[102] Andreas Dünser, Karin Steinbügl, Hannes Kaufmann, and Judith Glück. Virtual and augmented reality as spatial ability training tools. In *Proceedings of the 7th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction: design centered HCI*, pages 125–132, 2006.

[103] Hilary Mclellan. Cognitive issues in virtual reality. *Journal of Visual literacy*, 18(2):175–199, 1998.

[104] Peter Baranyi, Adam Csapo, and Peter Varlaki. An overview of research trends in coginfocom. In *IEEE 18th International Conference on Intelligent Engineering Systems INES 2014*, pages 181–186. IEEE, 2014.

[105] Péter Baranyi, Adam Csapo, and Gyula Sallai. *Cognitive Infocommunications (CogInfoCom)*. Springer, 2015.

[106] Ádám B. Csapó, Ildikó Horváth, Péter Galambos, and Péter Baranyi. Vr as a medium of communication: from memory palaces to comprehensive memory management. In *2018 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pages 389–394. IEEE, 2018.

[107] Gergely Sziladi, Tibor Ujbanyi, Jozsef Katona, and Attila Kovari. The analysis of hand gesture based cursor position control during solve an it related task. In *2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pages 413–418. IEEE, 2017.

[108] Cristina Costescu, Adrian Rosan, Brigitta Nagy, Ilona Heldal, Carsten Helgesen, Attila Kővári, József Katona, Serge Thill, Róbert Demeter, and Igor Efrem. Assessing visual attention in children using gp3 eye tracker. In *Proceedings of the 10th IEEE International Conference on Cognitive Infocommunications*, pages 343–348, 2019.

[109] Jozsef Katona, Tibor Ujbanyi, Gergely Sziladi, and Attila Kovari. Electroencephalogram-based brain-computer interface for internet of robotic things. In *Cognitive Infocommunications, Theory and Applications*, pages 253–275. Springer, 2019.

[110] Attila Kovari, Jozsef Katona, and Cristina Costescu. Evaluation of eye-movement metrics ina software debbugingtask using gp3 eye tracker. *Acta Polytechnica Hungarica*, 17(2), 2020.

[111] Attila Kovari, Jozsef Katona, and Cristina Costescu. Quantitative analysis of relationship between visual attention and eye-hand coordination. *Acta Polytechnica Hungarica*, 17(2), 2020.

[112] Tamás Budai and Miklós Kuczmann. Towards a modern, integrated virtual laboratory system. *Acta Polytechnica Hungarica*, 15(3):191–204, 2018.

[113] László Bognár, Éva Hamar Fáncsikné, Péter Horváth, Antal Joós, Bálint Nagy, and Györgyi Strauber. Improved learning environment for calculus courses. *Journal of Applied Technical and Educational Sciences*, 8(4):35–43, 2018.

[114] Tuukka M. Takala. Ruis: A toolkit for developing virtual reality applications with spatial interaction. In *Proceedings of the 2nd ACM symposium on Spatial user interaction*, pages 94–103, 2014.

[115] Ahmad Rafi, Khairul Anuar, Abdul Samad, Maizatul Hayati, and Mazlan Mahadzir. Improving spatial ability using a web-based virtual environment (wbve). *Automation in construction*, 14(6):707–715, 2005.

[116] Jack Shen-Kuen Chang, Georgina Yeboah, Alison Doucette, Paul Clifton, Michael Nitsche, Timothy Welsh, and Ali Mazalek. Evaluating the effect of tangible virtual reality on spatial perspective taking ability. In *Proceedings of the 5th Symposium on Spatial User Interaction*, pages 68–77, 2017.

[117] Elaine Jiang and David H. Laidlaw. Practicing in virtual reality improves mental rotation ability: Lower scorers benefit more.

[118] Charles M. Oman, Wayne L. Shebilske, Jason T. Richards, Travis C. Tubré, Andrew C. Beall, and Alan Natapoff. Three dimensional spatial memory and learning in real and virtual environments. *Spatial Cognition and computation*, 2(4):355–372, 2000.

[119] Marc A. Schnabel and Thomas Kvan. Spatial understanding in immersive virtual environments. *International Journal of Architectural Computing*, 1(4):435–448, 2003.

[120] David Passig and Sigal Eden. Virtual reality as a tool for improving spatial rotation among deaf and hard-of-hearing children. *CyberPsychology & Behavior*, 4(6):681–686, 2001.

[121] A. Michael Johnson. Speed of mental rotation as a function of problem-solving strategies. *Perceptual and motor skills*, 71(3):803–806, 1990.

[122] Nathan W. Hartman, Patrick E. Connolly, Jeffrey W. Gilger, Gary R. Bertoline, and Justin Heisler. Virtual reality-based spatial skills assessment and its role in computer graphics education. In *ACM SIGGRAPH 2006 Educators program*, pages 46–53. 2006.

[123] Albert A. Rizzo, J. Galen Buckwalter, Ulrich Neumann, Carl Kesselman, Marcus Thiébaux, Peter Larson, and Andre van Rooyen. The virtual reality mental rotation spatial skills project. *CyberPsychology & Behavior*, 1(2):113–119, 1998.

[124] Albert "Skip" Rizzo, J. Galen Buckwalter, Peter Larson, Andre van Rooyen, K. Kratz, Ulrich Neumann, Carl Kesselman, and Marcus Thiébaux. Preliminary findings on a virtual environment targeting human mental rotation / spatial abilities. 2002.

[125] I.F. Capanema, F.L. Santos Garcia, and G. Tissiani. Implications of virtual reality in education. *Virtual Reality in Education: Online Survey*, 2001.

[126] Bruce W. Field. A course in spatial visualisation. *Journal for Geometry and Graphics*, 3(2):201–209, 1999.

[127] Tim R.H. Cutmore, Trevor J. Hine, Kerry J. Maberly, Nicole M. Langford, and Grant Hawgood. Cognitive and gender factors influencing navigation in a virtual environment. *International Journal of Human-Computer Studies*, 53(2):223–249, 2000.

[128] Abdeldjallil Naceri, Ryad Chellali, and Thierry Hoinville. Depth perception within peripersonal space using head-mounted display. *Presence: Teleoperators and Virtual Environments*, 20(3):254–272, 2011.

[129] Marina A. Cidota, Rory M.S. Clifford, Stephan G. Lukosch, and Mark Billinghurst. Using visual effects to facilitate depth perception for spatial tasks in virtual and augmented reality. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pages 172–177. IEEE, 2016.

[130] Rebekka S. Renner, Boris M. Velichkovsky, and Jens R. Helmert. The perception of egocentric distances in virtual environments-a review. *ACM Computing Surveys (CSUR)*, 46(2):1–40, 2013.

[131] Claudia Armbrüster, Marc Wolter, Torsten Kuhlen, Will Spijkers, and Bruno Fimm. Depth perception in virtual reality: distance estimations in peri-and extrapersonal space. *Cyberpsychology & Behavior*, 11(1):9–15, 2008.

[132] Marc Rébillat, Xavier Boutillon, Étienne Corteel, and Brian F.G. Katz. Audio, visual, and audio-visual egocentric distance perception by moving subjects in virtual environments. *ACM Transactions on Applied Perception (TAP)*, 9(4):1–17, 2012.

[133] Che Abdullah, Glyn Lawson, and Tessa Roper. A virtual environment with haptic feedback for better distance estimation. In *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, pages 87–92, 2017.

[134] Adrian K.T. Ng, Leith K.Y. Chan, and Henry Y.K. Lau. Depth perception in virtual environment: The effects of immersive system and freedom of movement. In *International Conference on Virtual, Augmented and Mixed Reality*, pages 173–183. Springer, 2016.

[135] Nicolas Gerig, Johnathan Mayo, Kilian Baur, Frieder Wittmann, Robert Riener, and Peter Wolf. Missing depth cues in virtual reality limit performance and quality of three dimensional reaching movements. *PLoS one*, 13(1), 2018.

[136] Huiyu Zhou and Huosheng Hu. Human motion tracking for rehabilitation—a survey. *Biomedical Signal Processing and Control*, 3(1):1–18, 2008.

[137] Hossein Mousavi Hondori and Maryam Khademi. A review on technical and clinical impact of microsoft kinect on physical therapy and rehabilitation. *Journal of medical engineering*, 2014, 2014.

[138] Hui-mei Justina Hsu. The potential of kinect in education. *International Journal of Information and Education Technology*, 1(5):365, 2011.

[139] Jorge Bacca, Silvia Baldiris, Ramon Fabregat, Sabine Graf, et al. Augmented reality trends in education: a systematic review of research and applications. 2014.

[140] Alana Da Gama, Pascal Fallavollita, Veronica Teichrieb, and Nassir Navab. Motor rehabilitation using kinect: a systematic review. *Games for health journal*, 4(2):123–135, 2015.

[141] Helena Reis, Seiji Isotani, and Isabela Gasparini. Rehabilitation using kinect and an outlook on its educational applications: A review of the state of the art. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 26, page 802, 2015.

[142] Mingshao Zhang, Zhou Zhang, Yizhe Chang, El-Sayed Aziz, Sven Esche, and Constantin Chassapis. Recent developments in game-based virtual reality educational laboratories using the microsoft kinect. *International Journal of Emerging Technologies in Learning (iJET)*, 13(1):138–159, 2018.

[143] Maria Kourakli, Ioannis Altanis, Symeon Retalis, Michail Boloudakis, Dimitrios Zbainos, and Katerina Antonopoulou. Towards the improvement of the cognitive, motoric and academic skills of students with special educational

needs using kinect learning games. *International Journal of Child-Computer Interaction*, 11:28–39, 2017.

[144] Imran Amjad, Hamza Toor, Imran Khan Niazi, Sanna Pervaiz, Mads Jochumsen, Muhammad Shafique, Heidi Haavik, and Touqeer Ahmed. Xbox 360 kinect cognitive games improve slowness, complexity of eeg, and cognitive functions in subjects with mild cognitive impairment: A randomized control trial. *Games for health journal*, 8(2):144–152, 2019.

[145] Amir Matallaoui, Jonna Koivisto, Juho Hamari, and Ruediger Zarnekow. How effective is “exergamification”? a systematic review on the effectiveness of gamification features in exergames. In *Proceedings of the 50th Hawaii International Conference on System Sciences*, 2017.

[146] HyeonHui Sin and GyuChang Lee. Additional virtual reality training using xbox kinect in stroke survivors with hemiplegia. *American journal of physical medicine & rehabilitation*, 92(10):871–880, 2013.

[147] GyuChang Lee. Effects of training using video games on the muscle strength, muscle tone, and activities of daily living of chronic stroke patients. *Journal of physical therapy science*, 25(5):595–597, 2013.

[148] Fernando Mateo, Emilio Soria-Olivas, Juan J. Carrasco, Santiago Bonanad, Felipe Querol, and Sofía Pérez-Alenda. Hemokinect: a microsoft kinect v2 based exergaming software to supervise physical exercise of patients with hemophilia. *Sensors*, 18(8):2439, 2018.

[149] Oskar M. Szczepaniak and Dariusz J. Sawicki. Gesture controlled human–computer interface for the disabled. *Med. Pr*, 68(1):11–21, 2017.

[150] Laurie A. Malone, Jennifer L. Rowland, Rebecca Rogers, Tapan Mehta, Sangeetha Padalabalanarayanan, Mohanraj Thirumalai, and James H. Rimmer. Active videogaming in youth with physical disability: Gameplay and enjoyment. *Games for health journal*, 5(5):333–341, 2016.

[151] Péter Müller, Anett Nagyváradi, Levente Szabó, Miklós Gerzson, and Ádám Schiffer. Gait cycle recording using kinect one sensor. In *Pannonian Conference on Advances in Information Technology (PCIT 2020)*, pages 56–61, 2020.

[152] Wei Song, Liying Liu, Yifei Tian, Guodong Sun, Simon Fong, and Kyungeun Cho. A 3d localisation method in indoor environments for virtual reality applications. *Human-centric Computing and Information Sciences*, 7(1):39, 2017.

[153] How Does the Kinect Work? – kinect.pdf. `ftp://labattmot.ele.ita.br/ele/jricardo/Leitura/Kinect/kinect.pdf`.

[154] Slide 1 – Lecture 22 – How the Kinect works – CP Fall 2017.pdf. `https://courses.engr.illinois.edu/cs445/fa2017/lectures/Lecture%2022%20-%20How%20the%20Kinect%20Works%20-%20CP%20Fall%202017.pdf`.

[155] Kinect Sensor for Xbox Gaming – download. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.476.2368&rep=rep1&type=pdf.

[156] Xbox One Kinect Teardown—iFixit. https://www.ifixit.com/Teardown/Xbox+One+Kinect+Teardown/19725.

[157] How It Works: Xbox Kinect. https://www.jameco.com/jameco/workshop/howitworks/xboxkinect.html.

[158] Gamasutra: Daniel Lau's Blog—The Science Behind Kinects or Kinect 1.0 versus Kinect 2.0. http://www.gamasutra.com/blogs/DanielLau/20131127/205820/The_Science_Behind_Kinects_or_Kinect_10_versus_20.php.

[159] Kinect Sales Reach 24 Million—GameSpot. https://www.gamespot.com/articles/kinect-sales-reach-24-million/1100-6403766/.

[160] Why Xbox Kinect didn't Take Off—Business Insider. https://www.businessinsider.com/why-microsoft-xbox-kinect-didnt-take-off-2015-9.

[161] Higinio Gonzalez-Jorge, Belén Riveiro, Esteban Vazquez-Fernandez, Joaquín Martínez-Sánchez, and Pedro Arias. Metrological evaluation of microsoft kinect and asus xtion sensors. *Measurement*, 46(6):1800–1806, 2013.

[162] Xtion. 3D Sensor. ASUS Global. https://www.asus.com/3D-Sensor/Xtion/specifications/.

[163] Philip Breedon, Bill Byrom, Luke Siena, and Willie Muehlhausen. Enhancing the measurement of clinical outcomes using microsoft kinect. In *2016 International Conference on Interactive Technologies and Games (ITAG)*, pages 61–69. IEEE, 2016.

[164] Monica Carfagni, Rocco Furferi, Lapo Governi, Chiara Santarelli, Michaela Servi, Francesca Uccheddu, and Yary Volpe. Metrological and critical characterization of the intel d415 stereo depth camera. *Sensors*, 19(3):489, 2019.

[165] Veronica Romero, Joseph Amaral, Paula Fitzpatrick, RC Schmidt, Amie W Duncan, and Michael J Richardson. Can low-cost motion-tracking systems substitute a polhemus system when researching social motor coordination in children? *Behavior research methods*, 49(2):588–601, 2017.

[166] Liberty Latus Brochure. https://polhemus.com/_assets/img/LIBERTY_LATUS_brochure_1.pdf.

[167] Ying Sun, Cuiqiao Li, Gongfa Li, Guozhang Jiang, Du Jiang, Honghai Liu, Zhigao Zheng, and Wanneng Shu. Gesture recognition based on kinect and semg signal fusion. *Mobile Networks and Applications*, 23(4):797–805, 2018.

[168] Dimitar Bogatinov, Petre Lameski, Vladimir Trajkovik, and Katerina Mitkovska Trendova. Firearms training simulator based on low cost motion tracking sensor. *Multimedia tools and applications*, 76(1):1403–1418, 2017.

[169] Evan A. Suma, Belinda Lange, Albert Skip Rizzo, David M. Krum, and Mark Bolas. Faast: The flexible action and articulated skeleton toolkit. In *2011 IEEE Virtual Reality Conference*, pages 247–248. IEEE, 2011.

[170] Helene Fournier, Jean-Francois. Lapointe, Irina Kondratova, Bruno Emond, and Cosmin Munteanu. Crossing the barrier: a scalable simulator for course of fire training. In *Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*, number 1, 2012.

[171] Joan R. Rosell-Polo, Eduard Gregorio, Jordi Gené, Jordi Llorens, Xavier Torrent, Jaume Arnó, and Alexandre Escola. Kinect v2 sensor-based mobile terrestrial laser scanner for agricultural outdoor applications. *IEEE/ASME Transactions on Mechatronics*, 22(6):2420–2427, 2017.

[172] Keir E. Keightley and Gerald W. Bawden. 3d volumetric modeling of grapevine biomass using tripod lidar. *Computers and Electronics in Agriculture*, 74(2):305–312, 2010.

[173] Oliver Wasenmüller and Didier Stricker. Comparison of kinect v1 and v2 depth images in terms of accuracy and precision. In *Asian Conference on Computer Vision*, pages 34–45. Springer, 2016.

[174] Higinio Gonzalez-Jorge, Pablo Rodríguez-Gonzálvez, Joaquín Martínez-Sánchez, Diego González-Aguilera, Pedro Arias, Manuel Gesto, and Lucía Díaz-Vilariño. Metrological comparison between kinect i and kinect ii sensors. *Measurement*, 70:21–26, 2015.

[175] Kourosh Khoshelham and Sander Oude Elberink. Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*, 12(2):1437–1454, 2012.

[176] Lin Yang, Longyu Zhang, Haiwei Dong, Abdulhameed Alelaiwi, and Abdulmotaleb El Saddik. Evaluating and improving the depth accuracy of kinect for windows v2. *IEEE Sensors Journal*, 15(8):4275–4285, 2015.

[177] Sara Bragança, Pedro Arezes, Miguel Carvalho, Susan P Ashdown, Ignacio Castellucci, and Celina Leão. A comparison of manual anthropometric measurements with kinect-based scanned measurements in terms of precision and reliability. *Work*, 59(3):325–339, 2018.

[178] Kenneth David Mankoff and Tess Alethea Russo. The kinect: a low-cost, high-resolution, short-range 3d camera. *Earth Surface Processes and Landforms*, 38(9):926–936, 2013.

[179] Ting On Chan, Derek D Lichti, Adam Jahraus, Hooman Esfandiari, Herve Lahamy, Jeremy Steward, and Matthew Glanzer. An egg volume measurement system based on the microsoft kinect. *Sensors*, 18(8):2454, 2018.

[180] Achuta Kadambi, Ayush Bhandari, and Ramesh Raskar. 3d depth cameras in vision: Benefits and limitations of the hardware. In *Computer Vision and Machine Learning with RGB-D Sensors*, pages 3–26. Springer, 2014.

[181] Mark A. Livingston, Jay Sebastian, Zhuming Ai, and Jonathan W. Decker. Performance measurements for the microsoft kinect skeleton. In *2012 IEEE Virtual Reality Workshops (VRW)*, pages 119–120. IEEE, 2012.

[182] Karen Otte, Bastian Kayser, Sebastian Mansow-Model, Julius Verrel, Friedemann Paul, Alexander U Brandt, and Tanja Schmitz-Hübsch. Accuracy and reliability of the kinect version 2 for clinical measurement of motor function. *PloS one*, 11(11), 2016.

[183] Lydia R. Reither, Matthew H. Foreman, Nicole Migotsky, Chase Haddix, and Jack R. Engsberg. Upper extremity movement reliability and validity of the kinect version 2. *Disability and Rehabilitation: Assistive Technology*, 13(1):54–59, 2018.

[184] Meghan E. Huber, Amee L. Seitz, Miriam Leeser, and Dagmar Sternad. Validity and reliability of kinect skeleton for measuring shoulder joint angles: a feasibility study. *Physiotherapy*, 101(4):389–393, 2015.

[185] Mohamed Elgendi, Flavien Picon, and Nadia Magenant-Thalmann. Real-time speed detection of hand gesture using, kinect. In *Proc. Workshop on Autonomous Social Robots and Virtual Humans, The 25th Annual Conference on Computer Animation and Social Agents (CASA 2012)*, 2012.

[186] Carlos Gutiérrez López de la Franca, Ramón Hervás, Esperanza Johnson, Tania Mondéjar, and José Bravo. Extended body-angles algorithm to recognize activities within intelligent environments. *Journal of Ambient Intelligence and Humanized Computing*, 8(4):531–549, 2017.

[187] Carlos Gutiérrez-López-Franca, Ramón Hervás, and Esperanza Johnson. Strategies to improve activity recognition based on skeletal tracking: Applying restrictions regarding body parts and similarity boundaries. *Sensors*, 18(5):1665, 2018.

[188] Tibor Guzsvinecz, Cecilia Sik-Lanyi, Eva Orban-Mihalyko, and Erika Perge. The influence of display parameters and display devices over spatial ability test answers in virtual reality environments. *Applied Sciences*, 10(2):526, 2020.

[189] Tibor Guzsvinecz, Éva Orbán-Mihálykó, Cecília Sik-Lányi, and Erika Perge. Investigation of spatial ability test completion times in virtual reality using a desktop display and the gear vr. *Virtual Reality*. Under review.

[190] Tibor Guzsvinecz, Éva Orbán-Mihálykó, Erika Perge, and Cecilia Sik-Lányi. Analyzing the spatial skills of university students with a virtual reality application using a desktop display and the gear vr. *Acta Polytechnica Hungarica*, 17(2), 2020.

[191] Tibor Guzsvinecz, Veronika Szucs, and Attila Magyar. Preliminary results of evaluating a prediction-based algorithm for movement pattern recognition and classification. In *11th IEEE International Conference on Cognitive Infocommunications*, pages 39–44, 2020.

[192] Tibor Guzsvinecz, Monika Szeles, Erika Perge, and Cecilia Sik-Lanyi. Preparing spatial ability tests in a virtual reality application. In *10th IEEE International Conference on Cognitive Infocommunications*, pages 363–368, 2019.

[193] Unity Real-Time Development Platform|3D, 2D VR & AR Visualizations. `https://unity.com`.

[194] Sa Wang, Zhengli Mao, Changhai Zeng, Huili Gong, Shanshan Li, and Beibei Chen. A new method of virtual reality based on unity3d. In *2010 18th international conference on Geoinformatics*, pages 1–5. IEEE, 2010.

[195] Jonathan Linowes. *Unity virtual reality projects*. Packt Publishing Ltd: Birmingham, United Kingdom, 2015.

[196] Vinh T Nguyen and Tommy Dang. Setting up virtual reality and augmented reality learning environment in unity. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pages 315–320. IEEE, 2017.

[197] Christopher G. Coogan and Bin He. Brain-computer interface control in a virtual reality environment and applications for the internet of things. *IEEE Access*, 6:10840–10849, 2018.

[198] Moch Fachri, Ali Khumaidi, Nur Hikmah, and Nuke L Chusna. Performance analysis of navigation ai on commercial game engine: Autodesk stingray and unity3d. *Jurnal Mantik*, 4(1, May):61–68, 2020.

[199] Mental Rotation Quiz Questions - ProProfs Quiz. `https://www.proprofs.com/quiz-school/story.php?title=mental-rotation-task`.

[200] Gear VR SM-R322 Support & Manual|Samsung Business. `https://www.samsung.com/us/business/support/owners/product/gear-vr-sm-r322/`.

[201] Samsung Galaxy S6 Edge Plus—The Official Samsung Galaxy Site. `https://www.samsung.com/global/galaxy/galaxy-s6-edge-plus/`.

[202] How to interpret the sRGB color space (specified in IEC 61966-2-1) for ICC profiles. `http://color.org/chardata/rgb/sRGB.pdf`.

[203] Matthew Anderson, Ricardo Motta, Srinivasan Chandrasekar, and Michael Stokes. Proposal for a standard default color space for the internet—srgb. In *Color and imaging conference*, volume 1996, pages 238–245. Society for Imaging Science and Technology, 1996.

[204] Cecilia Sik-Lanyi. Choosing effective colours for websites. In *Colour Design*, pages 619–640. Elsevier, 2017.

[205] LG LED Monitor 20M37A|19.5 LG LED Monitor—LG Electronics UK. `https://www.lg.com/uk/monitors/lg-20M37A`.

[206] David W. Hosmer Jr, Stanley Lemeshow, and Rodney X. Sturdivant. *Applied logistic regression*, volume 398. John Wiley & Sons, 2013.

[207] Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying Ye. *Probability & statistics for engineers & scientists*. Pearson Prentice Hall, 2011.

[208] András Prékopa. *Valószínűségelmélet műszaki alkalmazásokkal*. Műszaki Kiadó, 1974.

[209] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.

[210] Patrick Maynard. *Drawing distinctions: the varieties of graphic expression*. Cornell University Press, 2018.

[211] David Webster and Ozkan Celik. Systematic review of kinect applications in elderly care and stroke rehabilitation. *Journal of neuroengineering and rehabilitation*, 11(1):108, 2014.

[212] Kelly J Bower, Julie Louie, Yoseph Landesrocha, Paul Seedy, Alexandra Gorelik, and Julie Bernhardt. Clinical feasibility of interactive motion-controlled games for stroke rehabilitation. *Journal of neuroengineering and rehabilitation*, 12(1):63, 2015.

[213] Arin Ghazarian and S Majid Noorhosseini. Automatic detection of users' skill levels using high-frequency user interface events. *User Modeling and User-Adapted Interaction*, 20(2):109–146, 2010.