UNIVERSITY OF TARTU
Institute of Computer Science
Computer Science Curriculum

Kristjan-Julius Laak

From the brain to intelligent systems: The attenuation of sensation of self-generated movement

Master's Thesis (30 ECTS)

Supervisor: Jaan Aru, PhD
Co-supervisor: Raul Vicente, PhD

Tartu 2016

From the brain to intelligent systems: The attenuation of sensation of self-generated movement

Summary: Despite the recent achievements of the artificial intelligence systems, humans are still remarkably more elegant in performing a variety of sensorimotor tasks in complex and dynamically changing environments. To build machines that could learn and think like people, one needs to understand the algorithms the human brain implements to interact with the world. For an intelligent machine to independently and flexibly cope with the highly dynamical environment, discriminating self-generated changes in the environment from those generated by external agents is of critical importance. In this study, we investigated a putative mechanism of how the sensory consequences of self-generated movements are processed in the human brain. The general idea with some experimental support is that the brain actively dampens the sensory consequences of movement produced by the brain itself. To test the generality of this mechanism we conducted virtual reality (VR) experiments with human subjects where - with the help of a hand tracking device - moving targets were presented behind their own moving (but for them invisible) hand. The data from two experiments indicate attenuation of movement signals when the targets were presented behind the hand. These insights about how to cope with the sensory consequences of self-generated movement are important for building intelligent autonomous systems.

Keywords: sensory attenuation, virtual reality, Leap Motion Controller, Oculus Rift, self-generated movements, intelligent systems

CERCS: P176, Artificial Intelligence

Ajust intelligentsete süsteemideni: enese tekitatud liikumis-aistingute pidurdus

Summary: Hoolimata viimaste aastate kiiretest arengutest tehisintellekti valdkonnas on inimesed endiselt märkimisväärselt osavamad ülesannetes, mis puudutavad hakkamasaamist keerulises ja dünaamiliselt muutuvas keskkonnas. Inimsarnase õppimis- ja mõtlemisvõimega masinate ehitamiseks on vajalik kõigepealt mõista, kuidas inimaju maailmaga vastastikmõjus on. Selleks, et intelligentne masin suudaks pidevas muutumises olevas maailmas iseseisvalt ja paindlikult toimida, on masina jaoks oluline eristada iseenda poolt põhjustatud muutusi välise keskkonna mõjurite poolt tekitatud sisendist. Antud töös uurime mehhanismi, mida inimaju oletatavalt kasutab enda liigutustest põhjustatud tajukogemuse töötlemisel. Varasematest töödest on teada, et aju pidurdab aktiivselt sensoorseid signaale, mis on põhjustatud aju enda poolt kontrollitud jäsemete liikumisest. Antud töös testisime selle teooria üldkehtivust, viies katseisikutega läbi virtuaalreaalsuseksperimendid, kus katseisikud pidid tuvastama liikuvaid stiimuleid iseenda liikuva (kuid neile nähtamatu) käe tagant. Kahe eksperimendi andmed viitavad pidurdatud liikumistajule, kui eesmärkstiimulid kuvati liikuva käe taha. Teadmised selle kohta, kuidas inimaju töötleb iseenda tekitatud liikumisest tingitud sensoorseid tagajärgi on olulised autonoomsete masinate ehitamiseks.

Keywords: pidurdatud liikumistaju, virtuaalreaalsus, Leap Motion Controller, Oculus Rift, intelligentsed süsteemid

CERCS: P176, Tehisintellekt

Contents

Introduction

Scientists, engineers, and philosophers over centuries have been fascinated with intelligent humanoid machines. From the early science fiction films of androids (such as Karel Capek's "R.U.R.") to the non-fiction books about the future of artificial intelligence (e.g. Ray Kurzweil's "The Singularity Is Near", 2005), intelligent systems have been predicted to outstrip human intelligence in the near future. Recent developments in artificial intelligence (AI) and machine learning have indeed indicated the validity of at least some of these predictions. In 2016, Google's AI program AlphaGo won the world champion in the Go, a game that has long been viewed as one of the most challenging games for AI (Silver et al., 2016). However, these seemingly intelligent programs have limited capacity even in their own fixed environments. A major challenge in AI research and computer science is building a general artificial intelligence (GAI) system that would be capable of successfully performing any task that humans can (Kurzweil, 2005).

Humans are remarkably elegant in predicting, controlling, and learning complex behaviour. Hence, one way of building a GAI would be to transfer the principles of human cognitive systems to machines. For example, the field of cognitive developmental robotics has emerged that aims to understand the processes that intelligent machines require to be able to interact with complex and dynamic environments (Asada et al., 2001, 2009). One possibility to reach the goal of building such machines is to acquire comprehensive understanding of the algorithms of the higher cognitive functions of the human brain (Lake et al., 2016).

In the current study, we are interested in an important computation of the brain - how the brain processes its own movement. Evolutionarily, detecting motion in the environment is of critical importance. For example, it makes sense to pull off your hand when something next to it moves fast. If you see movement in the bush you better direct attention to it and most likely you need to run. However, in many cases, the "movement" that we see is actually produced by our own body. Examples of this self-generated movement would be the saccadic movement of eyes, moving the head, or raising a glass of water to drink. In the latter situation, a relatively big object - your own arm - moves through your visual field. It would be a problem if every time we grab for a beer in a pub we would be afraid that there is a large moving object that might hit us. In fact, if an agent (e.g. an animal or a robot) is moving, most of the movement in the environment is caused by the movement of the agent. For an intelligent machine, discriminating self-generated from external stimuli is crucial as otherwise the machine might mistake its own movement for movement in the environment.

What is known about the computations the brain does to predict the sensory consequences of its own movement? In recent decades, it has been postulated that the transformation from motor commands to their sensory consequences is represented in the brain by an internal "forward model" (Wolpert et al., 1995). When a motor command is sent to muscles performing a task, a copy of the motor command ("efferent copy", Holst & Mittelstaedt, 1950) is sent in parallel to the "forward model". The "forward model" then uses the copy to predict the sensory outcome of the movement. By computing the difference between the predicted signal and the sensory feedback, the brain can distinguish the sensory activation of self-generated movements from those of the external agents.

A classical effect of this prediction has been studied in reference to the saccadic movement of our eyes. At first it looks like a paradox that we have a stable vision at all although the eyes move multiple times per second. The brain can solve this problem by learning to predict the input change in retina caused by brain's own motor commands about eye movements, resulting in our visual field staying essentially stable (Von Holst, 1954; von Helmholtz, 1867). However, the ability to recognize and provide perceptual stability of all self-generated actions suggest a more general mechanism of how the brain works (Frith, 1992; Friston, 2010; Clark, 2016): The brain constantly predicts the sensory consequences of its own movement. This predictive coding or active inference account explains why self-generated movement results in attenuation of the sensory signal (Friston, 2010; Brown et al., 2013; Clark, 2016).

Previous studies have shown that such attenuation of the sensory signal is the reason why we cannot tickle ourselves (Blakemore et al., 1998). In a clever experiment where the subjects had to move a robotic arm to tickle themselves, the experimenters demonstrated the existence of a central mechanism that cancels out the sensation of the self-generated movement, resulting in the inability to tickle oneself (Blakemore et al., 1998). Such suppression of self-generated stimuli is widely observed from insects to human. The central nervous system of male crickets for example uses this mechanism to maintain auditory sensitivity during own singing (J. Poulet & Hedwig, 2003). To do so, the massive auditory influx of singing frequencies is dampened by a neuron that gets input from the singing pattern generator responsible for moving the wings that make the sound (J. F. Poulet & Hedwig, 2006). Other examples involve some fish sensing disturbance in their electric fields to detect prey (Bell et al., 2008), and the phenomenon of "force escalation" in which each person feels the other hit them harder (Shergill et al., 2003). Importantly, it has been shown that the reduction of sensitivity to self-generated stimuli is not affected by sensorimotor contingencies altering the response bias of the subjects, but that the anticipation of the sensory consequences itself attenuates the perception of the stimuli (Cardoso-Leite et al., 2010).

Following the examples above, many aspects of the computational algorithms of the human sensorimotor system have been already studied. However, the more low-level fundamental modalities of this predictive processing in the brain are yet to be experimentally tested. In this study, we test whether the well-predicted visual sensation of the motion of the self-generated hand movement is attenuated. Based on the previous work summarized above the clear hypothesis is that the brain suppresses visual motion signals in the area of the visual field where the hand currently is moving. The basic idea to test this hypothesis is simple: have objects move behind the hand and show that the cognitive processing of these moving objects is impaired. However, it is complicated to test this hypothesis with conventional tools of experimental psychology, as this test would require the experimental subjects to see the moving objects behind the moving hand but not to see the moving hand itself.

To test our hypothesis, we designed and conducted two virtual reality (VR) experiments where the subjects were shown a field of moving objects from which they had to find a salient moving target object. At the same time, subjects had to move their hand in front of their eyes. With methods for tracking the hand position we captured the coordinates of hand movement without showing the hand to the subject in the VR environment. Critically, in the experimental condition, targets appeared behind the (invisible) hand during the hand movement and moved in the same direction with the hand movement. We measured the reaction time (RT) of noticing the targets with a clear prediction: targets that are currently directly behind the (invisible) hand are processed more slowly.

The author of the thesis developed the VR environment for the experiments, conducted the experiments with human subjects, analysed the data, and partially designed the experiments.

In the first chapter of the thesis, we give overview of the technical background of the software and hardware devices used, and describe the general design and implementation of the experiments. The technical solutions are described in such details that would allow the assessment of the amount of work done. However, we leave out the direct description of the classes and functions of software implementation itself and instead describe the complex behaviour of the experimental environment the code produces. For a full review of the code written, please see the Github project in Appendix IV. In the second chapter, the full methods and results of the first experiment are given. The methods and results of the second experiment are presented in the third chapter. Finally, the results of the two experiments are discussed in greater detail in the fourth chapter. In the final paragraph we also discuss whether the VR technology is ready for conducting such experiments, and propose future work. In the Appendix I and II, additional figures and tables, respectively, are given that aid the interpretation of the results. The methods and results of the pilot study conducted before the two experiments is given in the Appendix III.

1 General methods

1.1 Unity

We use Unity Personal Edition 5.3.4p3 (Unity Technologies) to develop the virtual reality environment of the experiment. Unity is a game engine meant for creating 2D and 3D games and experiences, optimization, and deployment. The engine has a feature-rich and highly flexible Editor with multiple interfaces for interactive game development. Besides the Editor, Unity supports C# and Javascript scripting to interact with the player and arrange events in the gameplay (*Scripting in Unity*, 2016). The implementation of this study was done in C. Using a game engine such as Unity can reduce time it takes to build VR experiences for the Oculus Rift (*Oculus Rift Documentation Overview*, 2016).

The experiment is run in an virtual environment similar to 3D VR game, and we call the environment a "scene". Although Oculus SDK has to be installed to the computer for the Oculus Rift headset (see below) to work properly, the version of Unity used contains a built-in support for the Oculus Rift. That is, Unity provides base API and feature set and adds the ability to directly target Oculus from the Editor and Unity scripts (*Unity Manual VR Overview*, 2016). When VR is enabled in the Editor, Unity automatically renders any camera in stereo to the Oculus displays, and applies head tracking and appropriate field of view to the camera.

Unity works with 3D models both created with modelling software and directly within Unity (*Unity Manual Primitive and Placeholder Objects*, 2016). The types of primitive objects used in the VR environment of this study are spheres and planes. A sphere has a diameter and a natural texture wrapping around the surface. A plane can be positioned, rotated and expanded in two dimensions as long as needed. Both of these primitives are accompanied by a Collider component that allows the detection of collisions. Unity uses units that are equivalent to meters in real life (Unity meters, or shortly Um). Thus, if a plane or sphere is 2 units afar in the VR environment, it seems like it is 2 meters far in real life.

1.2 Oculus Rift

We use a new-generation VR device, the Oculus Rift DK2 (Oculus VR, LLC) virtual reality headset, a developer-release of the Oculus Rift. The DK2 has a resolution of 960 x 1080 pixels per eye, a refresh rate of 75 Hz, and a 100 degrees field of view (*Oculus Rift Specs - DK1 vs DK2 comparison*, 2016; see Figure 1 below). The headset has multiple near infrared sensors that are tracked by the Oculus infrared camera and allow, together with the gyroscope and accelerometer, precise, low-latency positional tracking (*Oculus Rift Development Kit 2 homepage*, 2016). The software for this study was developed in accordance with the best practices of developing VR experiences (*Oculus Rift Documentation Intoduction to Best Practices*, 2016).

The Oculus uses two OLED displays to create the illusion of depth of the scene by the use of stereopsis. The stereoscopic method works by presenting two offset images to the left and right eye. The two views of a scene are produced by Unity via built-in VR support which automatically renders any camera in the scene in stereo. This results in the effect of voluntary diplopia, the simultaneous perception of two images of a single object. This had to be considered in finding objects from behind of the hand, as for the subject, there are actually two images of a hand in the scene.

1.3 Leap Motion Controller

We use a novel hand recognition system Leap Motion Controller (Leap Motion, Inc; shortly Leap) Orion Beta SDK version 3.1.2 and the app Unity Core Assets version 4.0.2. The Leap tracks the position, velocity, and orientation of hands and fingers with low latency and an average of 1.2 mm position accuracy (Weichert et al., 2013). The Leap system consists of a hardware device and a software component which runs as a service on the host computer (*Leap Motion Unity Plugin Overview*, 2016). The hardware device uses two infrared optical sensors and three infrared lights to detect the hand, and has a view of 150 degrees long side and 120 degrees short side. The average effective distance of the Leap is about 0.03 to 0.6 meters (Figure 1).

The hardware device produces infrared images of the environment and sends tracking information to software applications such as Unity. The latter receives the tracking data via the Leap Motion Unity plugin that connects to the hardware device. The software component translates the tracking data from the right-handed coordinate system of Leap Motion API to left-handed convention of the Unity. The hands are rendered in Unity by the Leap software component that combines the sensory data with an internal model of the hands. The internal model is used by the software to maintain the most plausible position of the hand even in

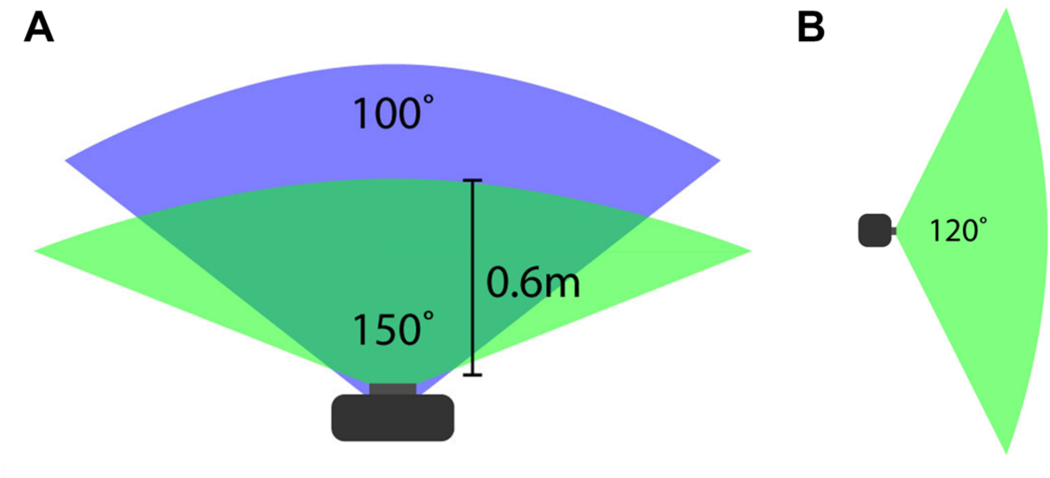

Figure 1: The field of view of the Oculus Rift (blue) and the Leap Motion Controller (green). A. Top view of the long side of the FOVs of the Leap controller (grey) and the Oculus Rift (dark grey). B. Side view of the short side of the FOV of Leap Motion Controller.

challenging tracking conditions (e.g. cramped hand).

In order to hide the visual models of the hands and arms while preserving the physical tracking of the hands, the rendering of the forearm and hand had to be disabled in the VR environment. The Unity assets provided by Leap had a built-in Unity Editor GUI checkbox to toggle the visibility of the forearm, but not of the hand. These functionalities were both missing from the API of the Leap Motion Controller. To control the visibility of the hands, we implemented a custom extension for the Leap API that allowed toggling the rendering of the hand and forearm in real time from the script.

1.4 General design of experiments

We tested our hypothesis using Oculus Rift virtual reality (VR) headset together with the Leap Motion Controller for detecting hand gestures. Using the novel VR headset allowed us to have absolute control over the visual environment the subject perceived. Throughout the experiments, the hands of the subject were not rendered in the VR environment. However, the mathematical-physical parameters (position, velocity, orientation) of the hand were constantly monitored by the experiment software through the Leap Motion API. This allowed us to study

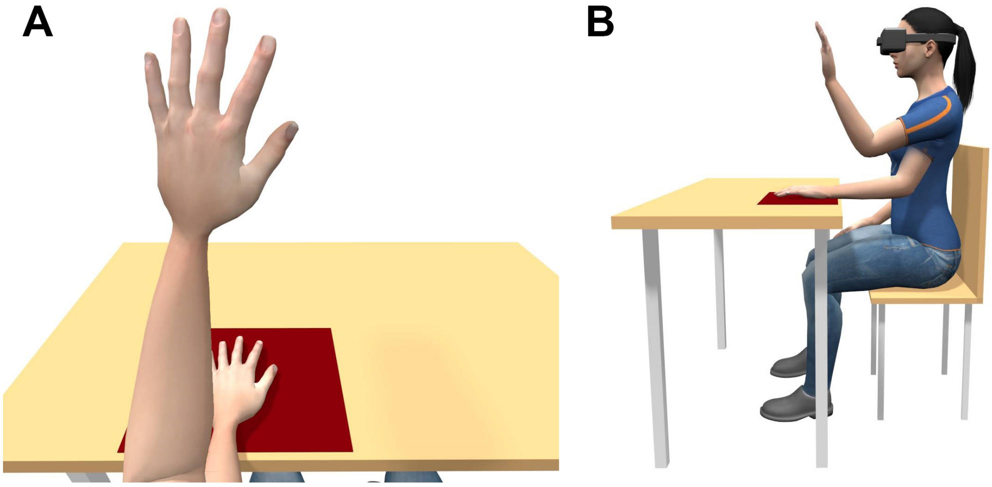

Figure 2: 3D model of the left hand position in the beginning of the hand movement and in the transition point from moving upwards to moving downwards. A. First person view of the two hand positions. Note that the camera angle is distorted to allow better overview of the hand. B. Side view of the subject sitting behind the table, right hand on the lap holding computer mouse and left hand making the movement. For detailed description of the movement, see Instructions to the subject below.

whether the brain uses the knowledge about the hand position to suppress movement information from that particular area of the visual field that would have been currently covered by the hand (that was not rendered in the scene). Hence note that whenever we throughout the manuscript talk about "target being behind the hand" the hand is actually invisible for the subject.

The subjects were shown a grid of horizontally moving objects and their main task was to react as fast as possible if they noticed a tachistoscopically presented target object that, for a short period of time, moved vertically. The subjects were instructed to always look at the fixation point (a cross in the middle of the field of view), partly to avoid the failure of detection of the target during saccadic eye movement, an effect called saccadic suppression (Bridgeman et al., 1975) The vertical motion of the target object was initiated, or triggered, by the hand movement of the subject. Specifically, the left hand was raised from the table upwards and back on the table (Figure 2). The response to the target ("reaction time", shortly "RT") was registered by the press of the left mouse button of the wired mouse held in the right hand.

In all the experiments, we compared the RTs from the condition where the moving target was behind the hand with a condition where the hand was not in front of the target. The temporal and spatial distribution of the targets in different conditions was kept constant in all the experiments. Prior studies suggest a clear prediction that the area of the visual field where the hand currently is during movement is dampened for motion, thus, the mean RT in the condition where the hand is in front of the target should be slower than in the control condition.

1.5 Physical setup

The experiments were conducted in a room with a table, a 24-inch screen, and four-legged chair without armrests and a static backrest. The room was windowless to avoid daylight that could affect the performance of the Leap hand tracking. Throughout the experiment, subjects were seated behind the table with their left hand on the table and right hand relaxed on the lap, holding a wired USB mouse (see Figure 2 above). The hand rested on a hard gaming mousepad (Logitech G440) that was found to increase the detection rate of the hand by the Leap. The gaming pad was located in front of the subject with the right edge approximately in the center of the subject.

1.6 Virtual scene setup

The virtual scene of the experiment consisted of a 20x19 grid of spheres with a cross-shaped fixation point in the middle (Figure 3). The grid and the fixation point were both initialised on a two-dimensional (2D) plane with a fixed distance of 2 Um from the cameras in the scene. The spheres of the grid had all random offset in both the horizontal and vertical axis, and the grid was re-generated with new random offsets in the beginning of each trail. The fixation point was represented as a small cross in the absolute middle of the grid.

1.7 The animations

Throughout the experiment, the spheres were oscillating horizontally on a sinusoidal trajectory (Figure 4A) in the approximate radius equal to the half of the distance between the spheres in the grid. The horizontal animation started from an offset equal to the horizontal random offset of the given sphere. Thus, the spheres moved around their initiation points and never overlapped with each other. The

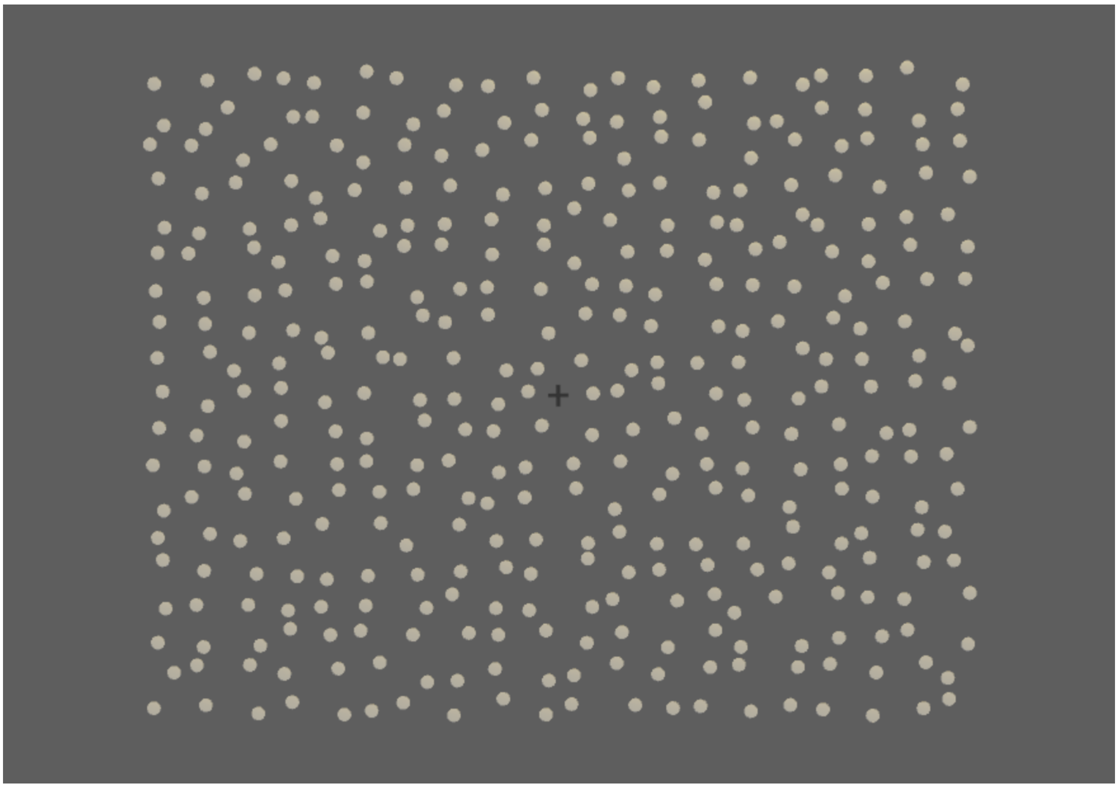

Figure 3: The grid of spheres oscillating horizontally and the fixation point in the middle that the subjects had to focus on throughout the experiments.

animation of the target objects followed a linear vertical trajectory in the same direction with the hand movement, e.g. if the hand moved up, then the target also moved only up (Figure 4B).

The animations of the spheres were controlled in script through an Animator component of each sphere. The Animator is part of the Unity's Mecanim animation system and used to assign animations to the objects in the scene (*Unity Manual Animator Component*, 2016). The Animator manages animations through an Animator Controller which manages the animations and transitions inbetween using a flowchart-like state machine (*Unity Manual Animator Controller*, 2016). The state machine is used to visualise the sequence of the animations according to the parameters modified through the script. The layers of the Animator component allow multiple animations to be run simultaneously. The spheres in the experiment have two layers for the horizontal and vertical animation, respectively, and float parameters to control the velocity of the animations.

The target animation was played by triggering the condition of the state machine

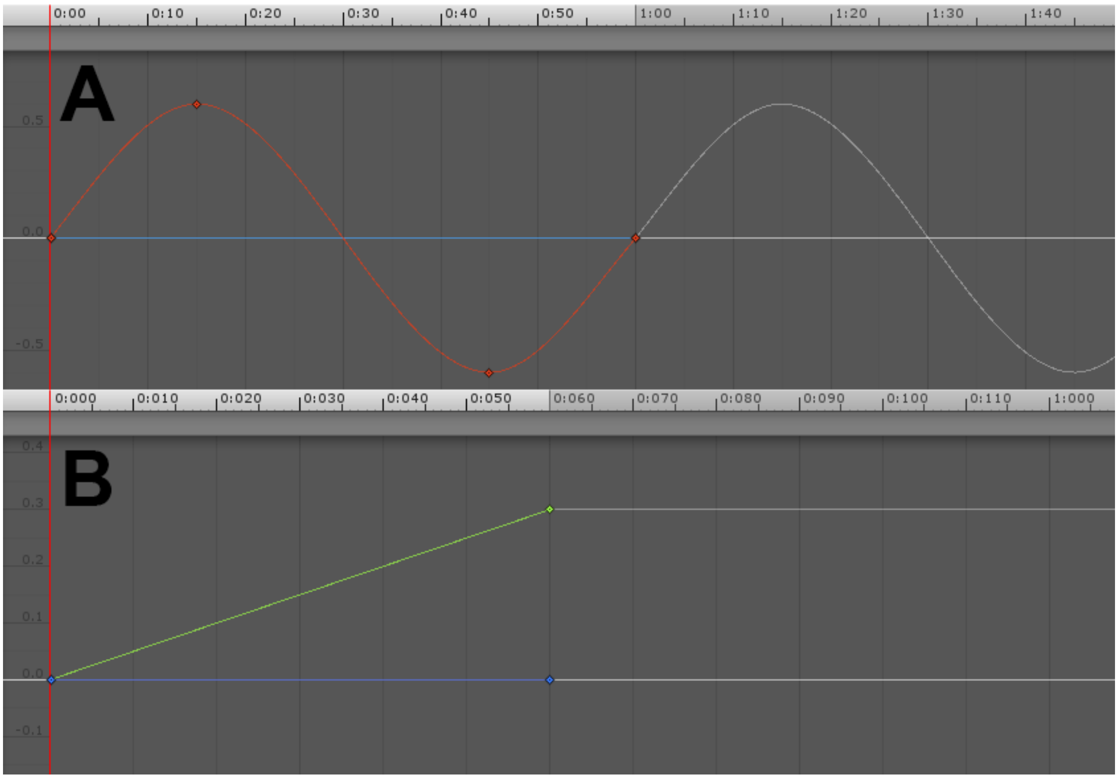

Figure 4: The animation curves of the animations of the spheres in the GUI of the Animation interface in the Unity Editor. A. The horizontal animation followed a sinusoidal trajectory (red line) around the initialization point of the sphere. B. The linear vertical animation of the target object. The target moved in the same direction as the hand for a short distance in approximately the same velocity as the hand.

to enter the "vertical" state. As the animation of the vertical state did not loop, the vertical state switched back to the idle state after the animation finished. To lower the threshold of perceiving the vertical movement of the target, the horizontal movement was paused when the vertical animation started and continued from the same position when the vertical animation finished.

1.8 Locating target objects

The Oculus Rift provides the user's visual system with two offset images to create the illusion of depth. This binocular parallax is also known as stereoscopy and, as already noticed by Leonardo Da Vinci, allows to see objects behind obstacles.

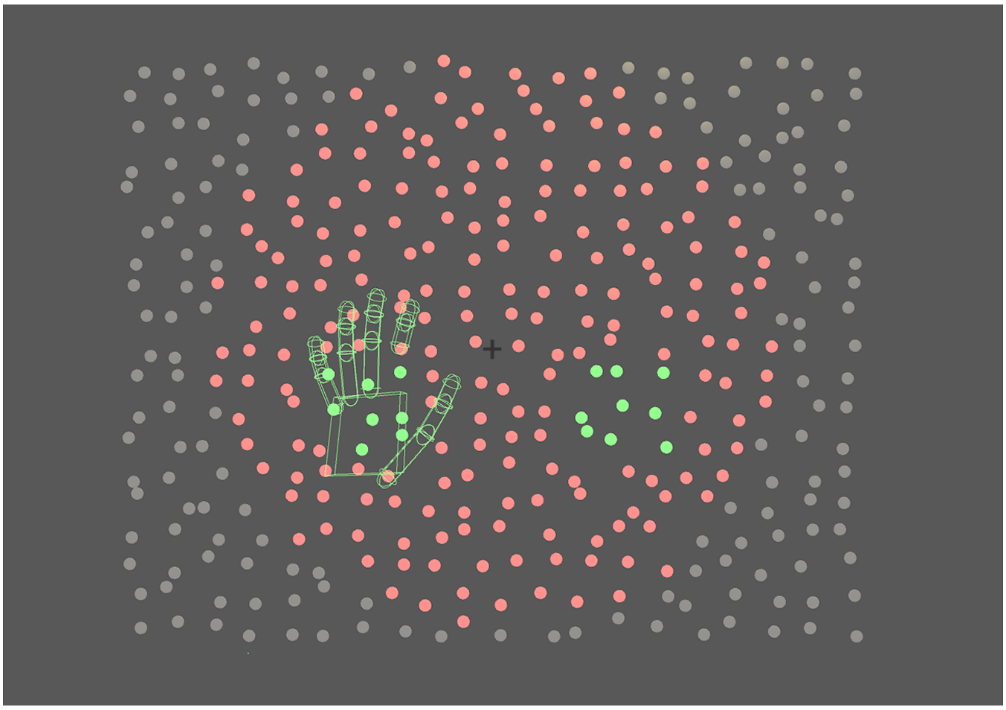

Figure 5: The test scene for Experiment 1 with target groups colored and the colliders of the hand visible. The targets in the experimental condition were chosen from the potential targets behind the hand (green spheres behind the hand). The choice among the potential targets was done randomly to allow for natural variation in the spatial distribution of the targets. In the control condition, the targets were chosen from the vertical reflection of the exact same spot where the hand targets located (green spheres on the right). Random targets were chosen from the narrow virtual field of view (nVFOV) that was visible to the subject (red spheres).

Therefore, although an object might be hidden in the view of one eye, the object might be visible from the other. Hence, to find potential targets from behind the hand, the objects that are in the area that is behind the hand from both of the eyes had to be found. To do that, a small spherical area was found that was behind the palm of the hand when looking through both of the eyes. Technically, a sphere sweep was casted from the mean position of the left and right camera view (or the "center eye") in the scene in direction of the center of the left hand. The radius of the sweep was chosen visually by colorizing the objects that were hit by the spherecast in a test scene where the hand was visible to the developer (Figure 5). Finally, the spheres intersecting the sweep were selected as "potential targets".

The targets for any condition were chosen from within the narrow virtual field of view (VFOV) to ensure that the targets were visible to the subject. (By "virtual field of view" we mean the part of the scene that is visible to the subject at the given frame through the VR headset.) To do that, a spherecast was produced from the center eye in the forward direction (Figure 5 above). The radius of the latter spherecast was chosen by visualizing the objects that were hit by the spherecast with the purpose of limiting the targets to the VFOV of the subject. The radius was chosen smaller than the actual VFOV because of the narrow focus radius of the Oculus Rift, resulting in blurry edges of the VFOV. Thus, we call this area the "narrow" VFOV (nVFOV). The random targets were chosen from the objects in the nVFOV minus the regions where the targets for the other conditions were (behind the hand and reflected).

1.9 Notifications displayed to the subject

In order the hand movement to be similar between subjects, in the first experiment, notifications were displayed to the subject in situations where the hand movement deviated from the desired trajectory or velocity.

- If, after the hand movement beginning was detected, the minimum velocity was not exceeded within 800ms, the subject was displayed a message: "Move your hand a bit faster".
- If the vertical coordinate of the palm at the top turn was below 0.2 Um from the fixation point, the subject was displayed a message: "Raise your hand a bit higher".
- If the hand stopped before the target animation was played, the subject was displayed a message: "Move your hand a bit slower".
- If the subject responded in the condition where no targets were shown, the following message was displayed: "There was not target!".
- If the hand movement was not detected at all within 3 seconds after the red "go" signal, the message "Did you forget to move your hand?" was displayed (see Figure 6).

In the second experiment, most of the notifications were omitted for the subject and instead displayed on a separate screen directly to the experimenter. During

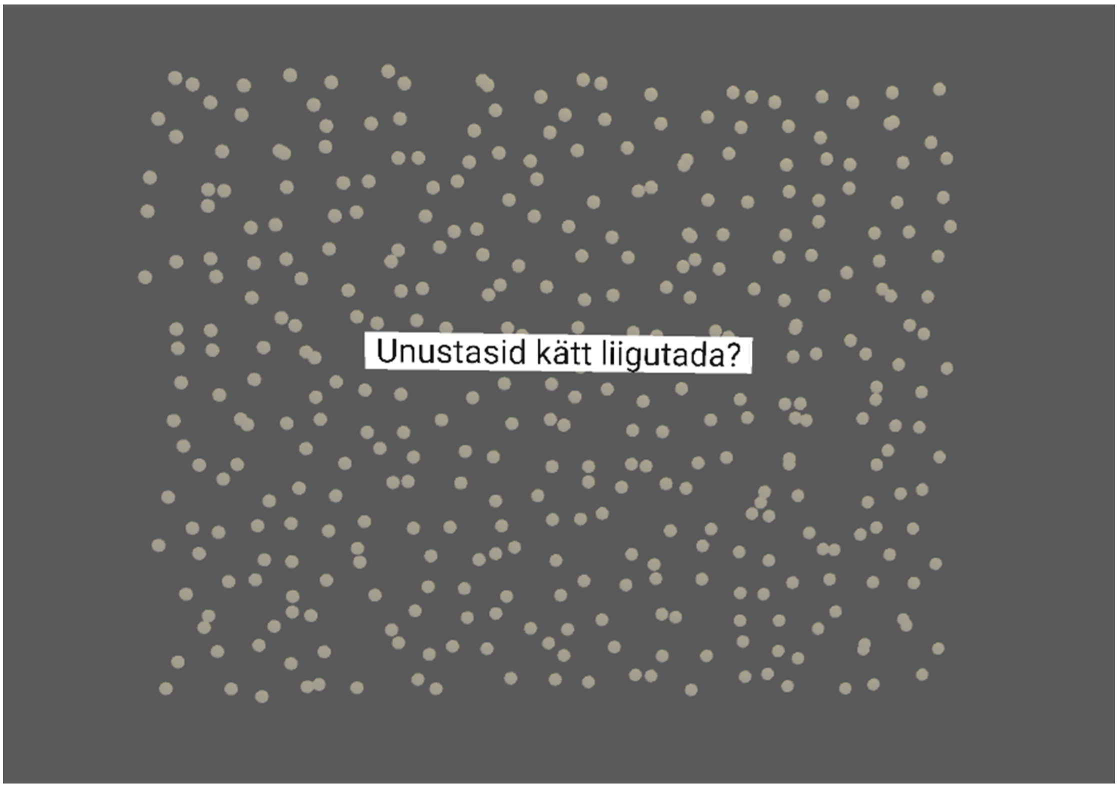

Figure 6: The notification to the subject in case no hand movement was detected. The text translates from Estonian as: "Did you forget to move your hand?".

the experiment, the subjects were instructed verbally when e.g. the hand movement was too low for several consecutive trials. Yet, notification for the cases of no movement and false positive reaction were kept. Most of the messages were excluded because in the pilot study (see Appendix III) and the Experiment 1 (see below) the subjects verbally reported paying subjectively considerable amount of attention to the hand movement after getting a notification.

1.10 Instructions to the subjects

To achieve best hand tracking performance, the subjects were assisted to remove all the accessories (jewelry, bracelets, rings, etc.) and expose their hands up to the elbows. First, the virtual reality headset and hand tracking device were generally introduced to the subject. To minimize disclosure of the objectives of the experiment, the hand tracking device was described as: "A device that detects the direction of the motion of the hands." The subjects were then guided through a training which consisted of (i) practice of the hand movement, (ii) practice of the RT task without the hand movement, (iii) and practice of the combination of the hand movement and the RT task, i.e. the experiment itself. The latter phase differed from the experiment only in length (the subjects were exposed to only 30-70 trials). The order of targets in the final training phase was randomized.

The hand training was conducted without the headset on. The subjects were instructed to practice the hand movement that was used in the experiment by the following instructions: "Move the hand from the start position back to the start position with homogeneous speed. The motion has to be smooth and ought not to stop at the top. The hand should rise to about the height where the palm is in line with your eyes. The motion has to be straight up from the start position to avoid getting too close to the tracking device. In addition, the fingers should be pointed upwards and a bit separate from each other." The reason of the training of the hand was explained by the experimenter (a) as a scientific rigour to standardise the hand movement between the subjects, and (b) to make sure the hand tracking device could recognize the hand as accurately as possible. This reasoning was important to minimize the attention of the hand training on the real objectives of the movement. In the Experiment 2, the subjects were also told a cover story: "We previously measured the effect of one movement on the RT, now we want to see what happens when we use two movements." The movement was visually presented by the experimented as long as the experimenter decided the subject was ready to move on. This method of training was found superior to other methods e.g. trying to follow a ball movement with a visible hand inside the VR.

After the hand training, the VR headset was put on and the scene was recentered by the subject via pressing the left mouse button. The subjects were instructed to lean on the backrest of the chair, relax and look straight. Before the reaction time training, the subjects were one-by-one visually introduced to the elements of the experiment scene: the fixation point that the subject had to pay attention to throughout the experiment, the briefly appearing red "go" signal (red fixation point), horizontally moving objects, and the task of finding a single vertically moving object. The motion of the vertically moving objects was explicitly shown by rendering the target blue for a brief moment before the vertical animation, and the subjects were instructed to turn their attention on the blue target.

During the RT training, the targets were not coloured, the subjects had to focus on the fixation point and react as fast as possible to the target object by pressing the left mouse button. In the training of the experiment, the subjects were instructed to move the hand as practiced within 3 seconds after the red "go" signal was visible. The subjects were informed that the target will be triggered by their hand movement, without specifying exactly, at which exact point of the movement. Again, the subjects had to focus on the fixation point and react as fast as possible to noticing the target.

After the trainings, the main experiment was started. The subjects were motivated to give their best effort to concentrate and be as accurate as possible; to strongly blink during the pauses if they experienced eyestrain; and to easily bring the attention back to the task if they discovered their thoughts wandering around.

1.11 Data collection

For each of the successful trials, the following data was recorded:

- Trial ID
- The condition the subject was exposed to
- Time point the start of the hand motion was detected
- Time point the target appeared
- Time point the target was noticed by the subject, specifically, when the subject pressed the left mouse button
- The reaction time as the difference of the target appear and target notice time
- A boolean of whether the subject moved the hand (true or false)
- A boolean of whether the target was shown (true or false), false in case the trial was aborted
- A boolean of whether the subject reacted to the target (true or false)
- The 3D coordinates of the hit of the raycast in direction of the palm on the grid plane
- The 3D coordinates of the target position on the grid plane
- The distance of the target from the fixation point (in Um)
- The distance of the target from the hit of the raycast in direction of the palm on the grid plane, i.e.

- distance from the most hidden point of the hand from both the left and right VFOV of the subject
- The distance of the target from the hit of the raycast in direction of the tip of the middle finger on grid plane
- Time point the top turn was registered
- Time point the hand stopped

All the measurements of time were made relative to the time in seconds since the start of the experiment. In the cases where the trial was aborted (e.g. the top turn of the hand was detected before the target animation played), only the time the hand motion begun and target condition were recorded.

1.12 Data preprocessing

The data preprocessing and analysis were performed in GNU R (The R Foundation) version 3.1.2. We preprocessed the data to control for the non-genuine processes entangled in the RT of the subjects. First, trials with RTs less than 100ms - the minimum time needed for physiological processes such stimulus perception (Luce, 1986) - were eliminated. To control for the spurious slow RTs in the data, we used a cutoff of 3 standard deviations (SDs) from the mean of the subject response times.

The control conditions of the experiments were specifically chosen to have maximum control over the latent variables that could affect the reaction time. Nevertheless, to make sure that any two conditions are comparable, we balanced the groups for the variability of the distances from the mean of the target positions in the group (or cluster). More specifically, all the targets positioned more than 3 SDs of the smallest variance of the groups from the mean were excluded from the further analysis.

1.13 Data analysis

Although many variables were recorded, in the current study, we were interested in only the response time to the target (RT). We checked for the statistical equality of the mean distances of the targets from the (a) fixation point, and (b) mean of the condition using t-test. Data of subjects with unequal means of the these distances were removed from the further analysis.

For statistical testing we used a paired t-test to test against the null hypothesis that the mean difference between the RTs from the conditions is zero. However, RT distributions of empirical data are found to fit not normal distribution, but rather a mixture of a Gaussian and an exponential distribution, resulting in an ex-Gaussian distribution (Balota & Spieler, 1999). This ex-Gaussian has a long positive tail, or skew, to the right. Using a t-test on data that are skewed can result in failure to discriminate between e.g. slow and fast responses across the conditions. To check for these potential drawbacks, we also analysed the whole ex-Gaussian distribution using the extra parameter tau (t) of the RT distribution, a single value explaining the mean and standard deviation of the exponential part of the distribution (Hervey et al., 2006). The exponential distribution explains the commonly seen positive skewness in empirical RT data (Luce, 1986). The ex-Gaussian measures of mean, standard deviation, and tau were evaluated using two-sided paired t-tests.

1.14 Debriefing

All the subjects were debriefed after the end of the experiment. The subjects were asked about their theory of objectives of the experiment, and whether they focused more on the task of moving the hand or on the RT task. The experimenter also tried to understand whether the subjects recognized any patterns in the spatial distribution of the targets (i.e. whether they became aware that some targets were behind the hand). The experiments were explained in detail to all the subjects after all the experiments with all the subjects were finished.

1.15 Pilot study

Before the experiment, a pilot study was conducted to test out an early version of the experimental design, and find warnings where the implementation or instructions may fail or the protocol not be followed (see Appendix III for the thorough description of the study). A total number of 10 subjects participated in the pilot study. In addition, 8 subject participated in different pilot experiments that were not part of the pilot study but were used to refine the experiments (including pilot study).

1.16 Ethics

All the subject read and signed written informed consent and participated in the experiments voluntarily. The VR experiments were approved by the Ethics Committee of the University of Tartu.

2 Experiment 1

2.1 Methods

2.1.1 Experimental design

The experiment consisted of 150 trials of which 30% were random targets, 30% targets behind hand, and 30% reflected from the vertical line going (imaginably) through the fixation point (see Figure 5 above and further explanation below). The conditions were balanced and randomized for each subject. Each subject was exposed to all of the conditions, hence, we used repeated measures design. In 10% of the trials, no target was presented. After every 50 trials (that is, two times during the whole experiment), a pause of 10s was made to provide time for rest. Each trial had the following algorithmic structure (For a detailed flowchart, see Figure 1 in Appendix I):

1. Display the spheres and the fixation point;
2. Show red "go" signal
3. If hand is moved, start short delay, otherwise end the trial
4. If delay end, choose a target and play the vertical animation;
5. Wait for subject reaction
6. Save trial and end

2.1.2 The delay period

The random delay period of 220 to 500 ms after the beginning of the hand movement was chosen to allow for a spread out spatial distribution of the targets. The time window was chosen by the experimenter as a balance between the maximum spatial distribution of the target objects and the average time the top turn was registered. This resulted in the targets being presented at the moment of the hand movement when the hand velocity was the highest. The delay period was the same for both experimental and control condition.

2.1.3 Locating reflected targets

To find the targets that would represent the position behind the hand but reflected from the vertical line of the fixation point (center of the VFOV), the center hit of the spherecast on grid plane was found and the horizontal x-coordinate of the hit point was multiplied by -1. The reason of reflecting from the vertical line of the fixation point (and not e.g. just the vertical line of the center of the eyes) was the main task of the subject: to look all the time at the fixation point. The random targets were chosen arbitrarily from the nVFOV described above. To add, the potential targets behind the hand and the area of reflected targets were excluded from the set of potential random targets.

2.1.4 Detection of the hand movement

In order to control the flow of the experiment, the specific swiping motion of the hand had to be detected. Specifically, the reliable beginning of the hand movement, and the point in the motion when the hand changes the direction from going up to coming back down (the "top turn") had to be registred. For the API of the stable release (version 2.3) of the Leap Motion Controller the SwipeGesture class was included, representing the swiping motion of each finger, allowing detection of continuous swipes of minimum length and velocity. For the Orion Beta API used in this study, these functionalities had to be implemented by ourselves.

The beginning of the up-motion was registered when the vertical component of the palm velocity exceeded velocity of 100mm/s. To avoid false positive events, the top turn of the hand motion was registered only after the minimum length of 150mm and velocity of 600mm/s were exceeded. The top turn was chosen in the event of the vertical component of the palm velocity changing from positive to a negative value. The stop event of the hand movement was registered when the velocity was again below a certain threshold. All of the constants were found prior the experiments by visual testing of the palm velocity during hand movement similar to the movement used in the experiment.

2.2 Results

2.2.1 Descriptive analysis

Eight healthy subjects with normal or corrected to normal vision took part in the first VR experiment (3 female, 5 male, age 22-24 [mean = 23.4]). A total number of 839 trials with non-random targets were collected where the RT was measured. The RT of the subject was not measured in cases where (i) the subject did not respond via clicking the left mouse button, (ii) the target was shown but the subject did not notice the target, or (iii) the subject raised her hand too low or too fast/slow, resulting rejection of the trial (see methods above). Preprocessing excluded another 24 trials from further analysis (see Methods; for overall distribution of trials per conditions see Table 1 in Appendix II).

The average reaction time to the target was 361 ms (SD = 67 ms; for RTs per subject see Table 2 in Appendix II). In the experimental and control condition targets were visually equally distributed within groups and from equal distance from the fixation point for all the subjects except one (see Figure 2 in Appendix I).

2.2.2 Difference of mean distances

We first checked whether the targets behind the hand and reflected from the center have equal distance from the fixation point. The test for the equality of the mean distances resulted insignificant except for one subject (t(42) = 9.429, p ¡ .0001; Figure 2 in Appendix I). As the equal distance of the target groups from the fixation point was important for a valid comparison, the data of this subject was excluded from the further analysis. The tests for the equal mean distances from the center of the clusters also indicated inequality for only the latter subject (t(61) = 5.321, p ¡ .0001).

2.2.3 Reaction time analysis

With the remaining 7 subjects, we tested for the reaction time (RT) as a dependent variable in the condition of targets behind the hand as compared to targets reflected from the vertical line of the fixation point. Note that we did not test for the difference with random targets as the random targets had uncomparable

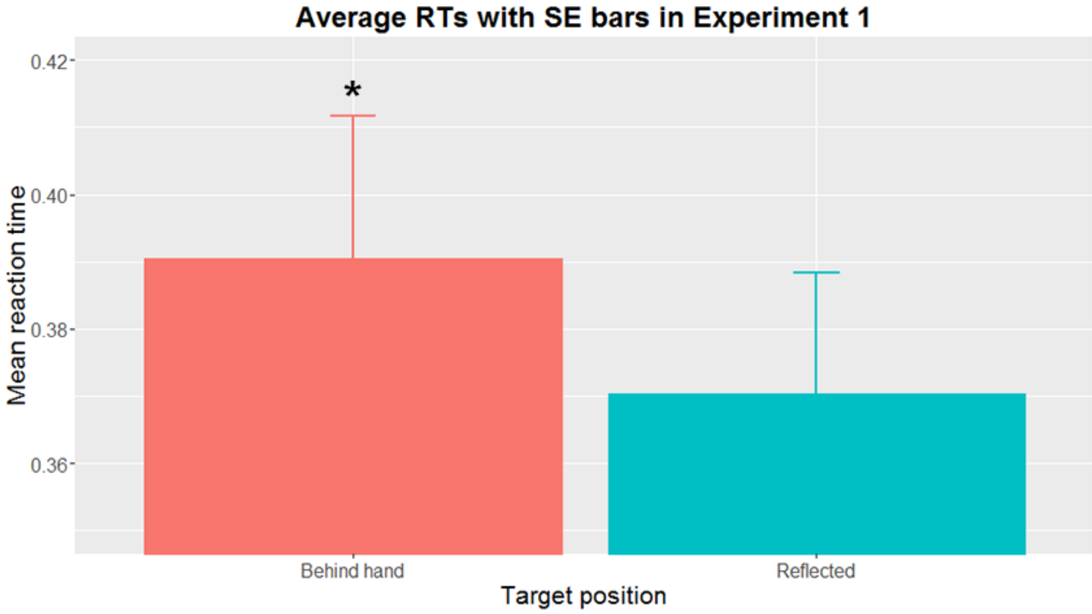

Figure 7: The mean RTs in the Experiment 1 for both conditions with standard error (SE) bars, all subjects pooled. The RTs in the experimental condition (behind the hand, red) were significantly slower than in the control condition (reflected, blue).

distribution compared to the other conditions. We opted for one-way paired t-test as we had a clear prediction: RTs from targets behind the hand are slower. We observed significantly lower RTs in the condition where targets were behind the hand (mean difference 20ms; $t(6) = 2.532$, $p = .022$, $d = .200$; Figure 7).

Using the ex-Gaussian measures described above, a similar pattern of results emerged. RTs from the targets behind the hand were lower on the mean of the normal component of the ex-Gaussian RT curve ($p = .024$). The variance of the normal part, and the skewness measure (tau) of the exponential part of the RTs curve did not reveal significant differences ($p = .132$, $p = .717$, respectively).

2.2.4 Debriefing

No subject guessed the objectives of the experiment. In one case, in a longer discussion and with the assistance of the experimenter, a subject suggested that we are measuring the differences of the hemispheres for the hand movement and reaction time for the visual stimulus. This subject, contrary to the theory, hypothesised that the reaction time in the left side of the VFOV should be faster. The other subjects prevailingly suggested the distance from fixation point to be the objective of the study.

2.3 Discussion

These results suggest that the movement of hand in front of the experimental targets is related to slower RTs in the experimental condition compared to control condition. This result nicely fits with the idea that self-generated hand movement attenuates the visual processing of their respective parts of the visual field. However, it cannot be discounted that these findings could have resulted from the bilaterality of the locations of the targets in the VFOV. The vertical motion of the targets in the two conditions, in other words, was largely processed by the different brain hemispheres, while the hand movement was controlled only by the right hemisphere. Hence, one could argue that while the left hemisphere was occupied with the targets in the right field of view, the right hemisphere was overloaded with both moving the hand and processing the movement of the targets appearing on the left side of the field of view, behind the hand. It could be that only because of this unspecific interference we observed slower RTs to targets in the experimental condition. To address this issue, we conducted Experiment 2 similar to Experiment 1, using additional control conditions.

3 Experiment 2

3.1 Methods

3.1.1 Experimental design

Similarly to the Experiment 1, the subjects had to move their hand as practiced and the target was shown after a short random delay when the hand was moving upwards. However, the subject now had to move the hand in two different positions: in the same place as in the previous experiment, and in a place shifted to the left from the initial position (the "right" and "left" condition, respectively). Importantly, the position of the target was kept the same for both of these conditions (see about finding targets below). Hence, the difference between those conditions was only the position of the hand: In the critical case ("right"), the hand was in front of the moving target, and in the control condition ("left") the hand was shifted away from the spot where the target was presented, while, importantly, still being in the same side of the visual field (Figure 8). The width of the shift was approximately the width of the palm of the subject, allowing the target not to be occluded by the hand while keeping it in the view of Leap Motion Controller. The hand movement start position alternated between left and right in turns between the trials.

The reflected targets were also included as a control condition. Note that comparison between the trials where the target appeared behind the hand ("right") and the reflected condition was similar to the first experiment, hence we expected similar findings. Furthermore, comparing trials where the target appeared in the same side of the visual field, but not behind the hand ("left") and the reflected condition provides a control condition for the alternative explanation we suggested to the results of the first experiment. Namely, if it is true that the results of the first experiment were caused by the fact that "the right hemisphere was overloaded with both moving the hand and processing the movement of the targets appearing on the left side of the field of view" then the condition "left" should have slower reaction times than the reflected condition.

All in all, 154 trials per subject were measured of which there were 10 trials with no targets, with the rest distributed between conditions of "behind hand" ("right"), "hand left", and reflected control. We had twice as many trials in the reflected control condition than in the other two conditions as otherwise the subjects would have been biased to search for targets on the left side of the visual field. After

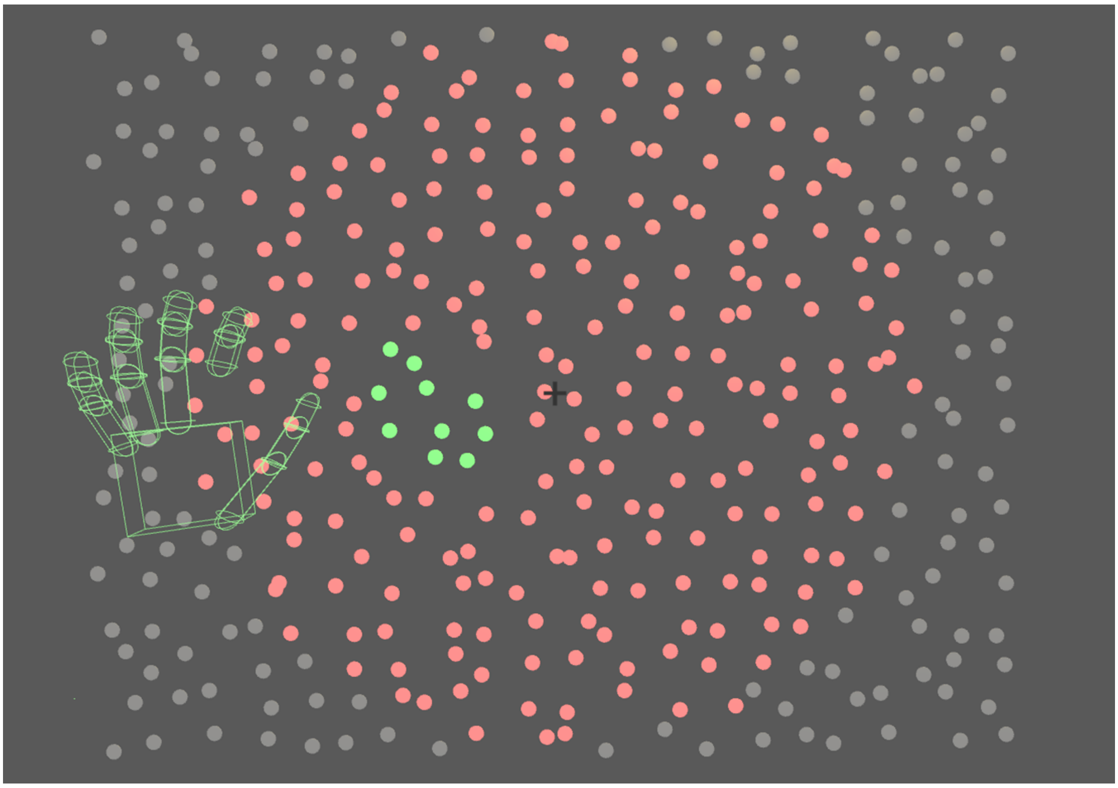

Figure 8: In the control condition of the Experiment 3, the hand movement was shifted to the left while the targets were shown from a similar place as in the experimental condition (green spheres). The potential targets were at the same vertical height as the hand but the horizontal position was chosen randomly from a set of positions where the hand had previously been during the same movement from the right side. Note that in the experimental settings, the targets were not coloured and the hand was not visible to the subject.

every 40 trials, a pause of 10s was made to provide time for rest.

The experiment employed the same algorithmic procedure within a trial as in Experiment 1, however, with several differences. First, the delay period was started not when the hand motion was detected, but when the first horizontal objects were found to be behind the palm. This was to normalize the spatial distribution of the targets between subjects. The delay period was set to 140-340 ms to accommodate the latter modification. Second, notifications were displayed to the subject only for the cases the hand was not detected or a false positive response was made.

3.1.2 Finding the control targets

The targets behind the hand were found the same way as in Experiment 1. However, the positions for the targets were recorded (saved in a list) in all the cases where the hand was on the right and the subject reacted to the target. In the "hand left" condition, the palm position was mimicked in the same area where it would appear when the hand was on the right, but with the vertical position taken from the real height of the hand. More specifically, in the "left" condition, the horizontal coordinate of the palm was chosen arbitrarily from the list of previously recorded palm positions, and the vertical coordinate from the "real" position of the palm. Then, the same processes as in Experiment 1 were followed for finding the potential targets in that region on the grid. To minimize the event of two targets being in a very similar position in sequential trials, we recorded the target positions already in the training phase.

3.2 Results

Descriptive analysis Eight healthy subjects with normal or corrected vision took part in the experiment (4 female, 4 male, age 18-28 [mean = 24.1]). A total number of 677 non-random trials were collected where the RT was measured. Further processing excluded another 73 trials from the analysis (see Table 3 in Appendix for overall distribution of the trials per condition). The mean RT across the subjects was 380 ms (SD = 73 ms; see Table 4 in Appendix for RTs per subject and condition).

3.2.1 Difference of mean distances

We tested for the pairwise equality of mean distances of the targets from the center of respective groups and from the fixation point using t-tests. The tests resulted in equal mean RTs for both measures for all the subjects except two. The data of these subjects were excluded from the further analysis (see Figure 3 in Appendix I).

3.2.2 Reaction time analysis

To control for the effect of the side of the visual field on the reaction times (See discussion of experiment 1), we first tested for the difference between the RTs to the targets in the condition where the targets were presented to the left-half of the visual field, but the hand was not covering the targets, and in the reflected condition. Importantly, we observed no significant difference between these conditions (mean difference 2ms; $t(5) = -0.199$, $p = 0.85$, $d = 0.024$), indicating that the alternative explanation we proposed cannot account for the findings of Experiment 1.

Next, to verify the results of the first experiment, we tested for the difference between the RTs to targets behind the hand and from the reflected condition. We observed approaching significance for these conditions (mean difference 14ms; $t(5) = 1.618$, $p = .083$, $d = 0.151$; Figure 9). This insignificance was, however, expected as there were only 6 subjects included in these analysis, and because of the smaller number of trials in the Hand-right condition compared to Experiment 1. We next compared the difference between the Hand-right and Hand-left condition and, despite the numerically slower RTs from the targets behind the hand, the data suggest no significant statistical difference (mean difference 16ms; $t(5) = 1.256$, $p = .132$, $d = .129$).

The alternative ex-Gaussian distribution analysis showed no difference between any of the pairs of conditions tested above for the mean of the normal component of the curve, variability of the normal part, or the skewness (tau) of the exponential part of the RTs curve.

3.2.3 Debriefing

The later interviewing revealed no subject who had guessed correctly the idea of the experiment. Some subjects even noted that the experiment was not about reaction times or the left-right movement itself, but about the coordination of the two tasks (multitasking). The RT measurement was several times noted as "arbitrary", or as one subject put it: "I found myself noticing some targets earlier or later, but the reaction was quite rhythmic". Together with the above results these comments indicate the subtle yet existing differences in the underlying processes of the sensorimotor system of the brain that the RT can distinguish.

Despite the rigorous training prior the experiment, when asked about whether they

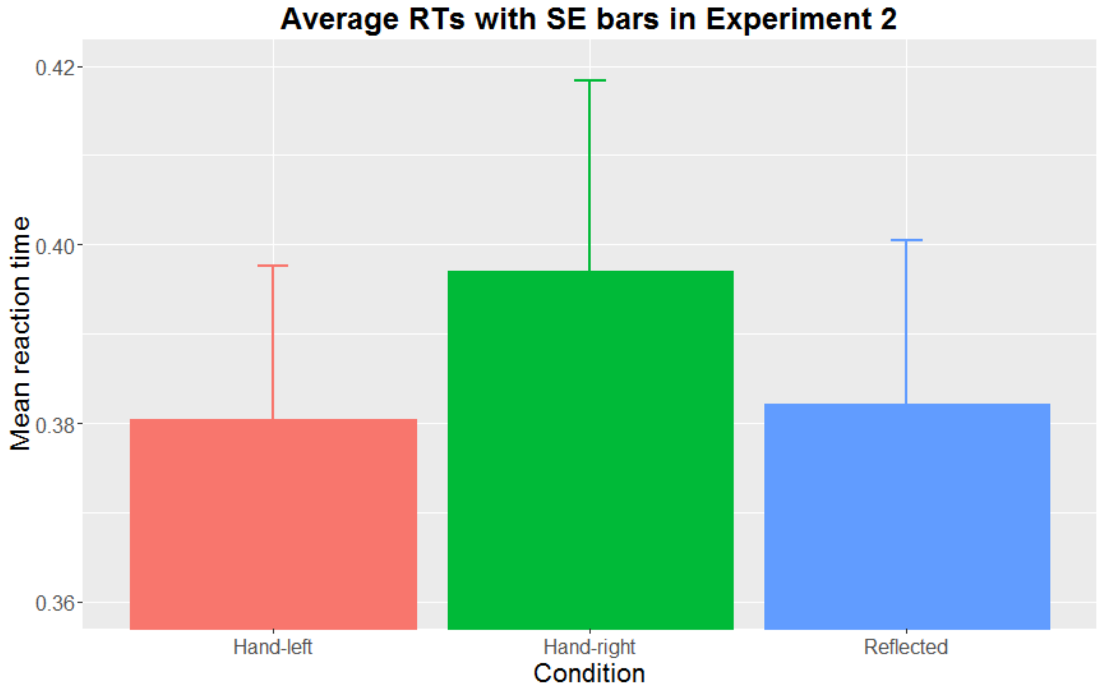

Figure 9: The mean RTs for all the non-random conditions measured. In the experimental condition, the hand was on the right side and the moving targets were presented behind the hand ("hand right"). In the additional control settings, the hand was shifted to the left and the target was not behind the hand ("hand left") or the targets were reflected to the right side of the visual field ("reflected").

concentrated on the hand movement or RT task, all subjects noted that the hand task took more effort in the beginning of the experiment. Most of the subjects also noted that they shifted part of their attention to the hand movement task for the next trial after getting a notification (i.e. "Did you forget to move your hand?").

3.3 Discussion

We conducted the second experiment in order to test whether the slower RTs in the condition where the targets were behind the hand could have been caused by the right hemisphere performing the hand movement task and processing of the target motion simultaneously. The data indicated no difference in the RTs to the two halfs of the visual field when the hand was not covering the movement of the targets. Hence, the results from the Experiment 2 support those of Experiment 1 and suggest that the slower RTs could have indeed been the effect of the hand movement being in front of the moving targets.

4 General discussion

4.1 Results of the experiments

Understanding the principles of how the human sensorimotor system works is essential for building intelligent humanoid machines that could learn by themselves to interact with complex environments. A part of this understanding are the algorithms and computations the brain uses to discriminate between the sensory signals of self-generated movement of e.g. the eyes, head, or limbs, and potentially important movement of other agents in the dynamically changing environment. As our everyday experience and the general theories about the brain (see Clark, 2016; Friston, 2010) suggest, the sensory system is constantly reducing or eliminating the perception of self-generated stimuli. In this study, we used novel VR and hand tracking devices that allowed absolute control over the visual scenery to test for the clear hypothesis that the brain also dampens the sensory consequences of the motion of the self-generated movement of the hand. We conducted two experiments where we hid the moving hand from VR environment while constantly monitoring the hand movement and presenting targets either behind the hand or away from it. This setup allowed us to measure the reaction times for moving targets, with the clear hypothesis that the RTs to the targets behind the hand are slower than in the control conditions.

In the first experiment, the experimental targets appeared behind the hand in the left field of view and the RTs were compared to the targets presented to the same spot in the right field of view. We observed significant difference between the mean RTs for the two conditions, namely, slower RTs to the targets behind the hand. However, as explained above in the discussion of the results of the experiment, one could argue that the faster RTs in the control condition could be caused by the different amount of workload of the respective hemispheres. Namely, while the left side of the brain is processing only the targets from the right field of view, the right hemisphere is controlling both the hand movement and processing the experimental targets in the left-half of the field of view.

To address this issue, we conducted the second experiment with an additional control condition where the hand was shifted to the left, but the targets were still presented in the left half of the visual field where it would appear otherwise behind the hand. We compared this extra control condition with the reflected condition similar to Experiment 1. Importantly, we found no difference between these two conditions. Hence, the slower RTs to the targets behind the hand indeed might be caused by the hand moving in front of the targets.

In the second experiment, we also tested the difference of the hand and reflected condition and found the difference only approaching statistical significance. However, the weaker power of these result was expected as data from only 6 subjects were included in the analysis. Moreover, there were also lower number of trials per condition (See Table 2 in Appendix II). Nevertheless, the second experiment was mainly conducted to verify that there is no difference between the hemispheres, not to repeat the results of the first experiment.

Taken together, we first showed that there is a clear difference between whether the moving targets were behind the moving hand or not. We verified these results by showing that the effect is not a mere consequence of one of the hemispheres being overloaded with the motor control of the hand movement. Therefore, these results provide, for the best of our knowledge, the first support for the hypothesis that the sensorimotor system of the human brain attenuates the well-predicted visual motion of self-generated movements.

4.2 From brain to machines

Accounting for the consequences of self-generated stimulation is relevant for building intelligent machines that could autonomously interact with complex environments. Interestingly, low-level (e.g. simple) versions of these systems are already around us today. For example, when having a Skype conversation, the systems that processes the input from the microphone has to eliminate the self-produced voice of the other speaker for not creating an endless reflection of the voice echoing from one end to the other. However, the more high-level machines are still lacking reliable systems for detecting self-generated movement, as, for example, the modern robots in industry are still hard-coded for the sensation of autogenous movements. Humans, on the other hand, are remarkably efficient in estimating the sensory consequences of their own bodily movement. Every time we lift our hands or move our eyes or head the brain could in principle be confused about what is causing the movement in the environment, but it is generally not. The brain has learned to anticipate and attenuate the sensory consequences of its own movement. In fact, there is quite some evidence to show that the breakdown of this self-monitoring mechanism could lead to symptoms distinctive to mental diseases such as schizophrenia (Frith, 1992; Wolpert & Ghahramani, 2000; Clark, 2016).

It is time for the developers and engineers of intelligent machines to stop using their own intuition of how the robots should make sense of the world. Biological brains have figured it out how to deal with the sensory consequences of self-generated movement and these intuitions could be used in designing intelligent machines. Early attempts using deep reinforcement learning have already been made, with brain-inspired systems controlling virtual robots only via visual sensory input of the movements (Lillicrap et al., 2015; Mnih et al., 2015). However, these systems are still in early phase of being able to accurately interact with the world similarly to humans (see e.g. Lake et al., 2016). Our results help to understand how evolution has solved this problem in the brain.

4.3 Experimental design

In both of the experiments we conducted, the targets in different conditions were presented symmetrically and with equal distance from the fixation point. At first, this might seem inessential for valid comparison of the conditions for two reasons. First, we designed the targets specifically in a way that allowed for a pop-out effect of the motion of the targets, that is, that the targets are automatically and clearly noticeable from the distracting horizontally moving objects. Second, although the highest foveal resolution is only in the 2 degree of the focus point, the peripheral sensitivity to motion is nearly equal to the foveal sensitivity (McKee & Nakayama, 1984; Leibowitz et al., 1972). However, these linear search times were affected by the narrow focal point radius of the lenses in the Oculus Rift headset. More specifically, the edges of the virtual field of view in the VR environment were more blurry with eccentricity function. That is, if two groups of conditions would have been even with the same spatial variability but with a small difference in the mean distances from the fixation point, the RTs to the targets closer to the fixation could have been faster. To control for the equality of the distances from the fixation point, we tested for the pairwise equality of mean distance from the fixation point and from the center of the clusters of the spatial positions of the targets for all conditions. Data for which these assumptions were not fulfilled were excluded from the analysis.

To obtain the same distribution of the targets in different conditions and preserve the equal distance from fixation point, we tested for several technical implementations for registering the motion of hand, raycasting, and presenting targets. Among the solutions used in the pilot study and the two experiments conducted, we also tested for the version where the hand was doing the same hand movement in the center of the visual field, and control targets were presented to the left, right, and top of the hand. The advantage of this method would have been that all the targets were chosen from the same central area around the fixation point and processed by both of the hemispheres (e.g. Neville & Lawson, 1987). However, there were many problems with this approach. For example, the hand position was hard to control to be in the exact center of the visual field of view, i.e. going "through" the vertical line of the fixation point. Also, the targets in the other condition were too sparsely located to allow for meaningful comparison. Notably, it is also not known to which extent the sensory dampening of the hand movement motion occurs, in other words, whether it is only in the area of the hand itself, or also in a certain buffer area around the hand. We opted for the solutions where both the targets of the experimental and control conditions were all chosen from the potential targets located behind the hand, with the variation of either reflecting the spot from the vertical line of the fixation point (Experiment 1), or moving the hand away from the spot itself (Experiment 2).

4.4 Technological considerations

Virtual reality technologies allow, for the first time in human history, a full mathematical control over the visual scenery. Wolpert & Ghahramani (2000) have said: "Having such control over the physics of the world with which subjects interact has allowed detailed tests of computational models of planning, control and learning." In this study, we used novel VR headset Oculus Rift, a VR headset that has been previously used in our lab to study change blindness (Vasser et al., 2015). The broad field of view and high-speed rendering of the Oculus Rift allowed us to conduct the visuomotor experiments that would have otherwise been impossible. However, there were two main drawbacks of the Oculus headset used. First, as described above, the high aberration in the edges of the simple lenses of the development headset causes the periphery of the visual field to be unnaturally blurry. Second, the developer-version Oculus also has relatively low quality displays with discernibly low number of pixels per inch. In comparison, the consumer level Oculus released later in spring 2016, and the competitive headset on the market, the HTC Vive (HTC Corporation Valve Corporation) have both high-quality lenses and high-resolution displays. Thus, the overall experience of the VR environment could be improved with the latest headsets.

Besides the VR headset allowing to generate immersive VR environments for scientific experiments, the experiments of this study could not have been possible without the novel Leap Motion hand-tracking device. With the ability to track the coordinates and velocity of the natural hand movement without showing the hand to the subject, we could test for the hypothesis of the brain dampening the visual sensation of the hand motion. However, as the tracking of the hands is based on image recognition software, the device has a few drawbacks which need to be accounted when designing such experiments. First, the hand has to be in the field of view of the Leap controller: if the hand out of the view, the hand model disappears. This was often one of the reasons the experimenter had to verbally assist subjects in the Experiment 2 to correct the hand movement. Second, although once the hand is recognized, the Leap uses internal models to estimate the hand position and gestures even if the hand is partly occluded for a short time or in a complex position, the first detection of the hand is dependent on the gesture of the hand. Hence, we specifically instructed the subject to hold the hand flat, fingers upwards and spread. Nevertheless, the experimenter constantly monitored the detectability of the hand by the Leap, and sometimes had to remind the subject to hold the hand correctly.

4.5 The perfect experiment

In the emergence of the novel virtual reality and hand-tracking technologies, many experiments that were unimaginable before could be conducted. In the future studies, the latest versions of the available technology should be used. For example, similar high-precision experiments should be conducted using high-resolution headsets such as the consumer version of Oculus Rift or the HTC Vive. The latter devices also deliver hand controllers that use electronic sensors and could be used to improve the detection of hands and allow for wider tracking space. An important improvement of the experiment would also be to train the subjects first as long as their hand movement is totally automatic. An interesting experiment to confirm the results of the current study would be to present the targets in the exact 3D spatial position the hand is in the physical environment, not just behind the (hidden) hand. Finally, different modalities such as color of the targets, and visual effects (e.g. tilted Gabor patches) should be tested besides the motion to test the generality of the sensory dampening accompanied with the self-generated movement.

Conclusions

In order to build intelligent systems capable of human-like understanding of the world, the mechanisms of the higher cognitive functions of the human brain itself have to be understood. We explored the important computational problem of how the brain processes self-generated movement. Specifically, we hypothesized that the brain actively attenuates the sensory perception of the motion of the self-generated hand movement, and conducted two virtual reality experiments with human subjects to verify this hypothesis. We first observed that the moving targets behind the moving hand were processed slower than other targets. In the second experiment, we verified these results by showing that the effect was non-specific to the side of the visual field the targets were presented to. All in all, these data indicate that the brain could indeed attenuate the motion of self-generated movements to discriminate the sensory consequences of its own body movement from the movement in the environment. These knowledge could be used by developers and engineers for building intelligent systems capable of interacting with complex and dynamic environment.

Acknowledgment

I express my sincere gratitude to Jaan Aru for his fatherly support, insightful comments, wisdom, and unending patience during my master studies and writing of this thesis. He has wisely guided me throughout my scientific journey so far, to which I am grateful for. Without him the thesis would not have been written.

I am thankful for the motivation and valuable comments of Raul Vicente who let me write my thesis in his lab. I also thank Madis Vasser for his the great 3D figures in this work, and all the Computational Neuroscience group members for supportive company the last 4 years.

Last but not least, I would like to thank my friends Kristjan Jansons and Jaanika Tammaru for being besides me this long spring in Tartu.

References

Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., ... Yoshida, C. (2009). Cognitive developmental robotics: a survey. *Autonomous Mental Development, IEEE Transactions on*, *1*(1), 12–34.

Asada, M., MacDorman, K. F., Ishiguro, H., & Kuniyoshi, Y. (2001). Cognitive developmental robotics as a new paradigm for the design of humanoid robots. *Robotics and Autonomous Systems*, *37*(2), 185–193.

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, *128*(1), 32.

Bell, C. C., Han, V., & Sawtell, N. B. (2008). Cerebellum-like structures and their implications for cerebellar function. *Annu. Rev. Neurosci.*, *31*, 1–24.

Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature neuroscience*, *1*(7), 635–640.

Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. *Vision research*, *15*(6), 719–722.

Brown, H., Adams, R. A., Parees, I., Edwards, M., & Friston, K. (2013). Active inference, sensory attenuation and illusions. *Cognitive processing*, *14*(4), 411–427.

Cardoso-Leite, P., Mamassian, P., Schütz-Bosbach, S., & Waszak, F. (2010). A new look at sensory attenuation action-effect anticipation affects sensitivity, not response bias. *Psychological science*.

Clark, A. (2016). Surfing uncertainty.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138.

Frith, C. (1992). The cognitive neuropsychology of schizophrenia.

Hervey, A. S., Epstein, J. N., Curry, J. F., Tonev, S., Eugene Arnold, L., Keith Conners, C., ... Hechtman, L. (2006). Reaction time distribution analysis of neuropsychological performance in an adhd sample. *Child Neuropsychology*, *12*(2), 125–140.

Holst, E., & Mittelstaedt, H. (1950). Das reafferenzprinzip. *Naturwissenschaften*, *37*(20), 464–476.

Kurzweil, R. (2005). *The singularity is near: When humans transcend biology.* Penguin.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2016). Building machines that learn and think like people. *arXiv preprint arXiv:1604.00289*.

Leap Motion unity plugin overview. (2016). Retrieved 2016-05-19, from \url{https://developer.leapmotion.com/documentation/unity/unity/Unity_Overview.html}

Leibowitz, H. W., Johnson, C. A., & Isabelle, E. (1972). Peripheral motion detection and refractive error. *Science*, *177*(4055), 1207–1208.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., ... Wierstra, D. (2015). Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*.

Luce, R. D. (1986). *Response times* (No. 8). Oxford University Press.

McKee, S. P., & Nakayama, K. (1984). The detection of motion in the peripheral visual field. *Vision research*, *24*(1), 25–32.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... others (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533.

Neville, H. J., & Lawson, D. (1987). Attention to central and peripheral visual space in a movement detection task: An event-related potential and behavioral study. i. normal hearing adults. *Brain research*, *405*(2), 253–267.

Oculus rift development kit 2 homepage. (2016). Retrieved 2016-05-19, from \url{https://www.oculus.com/en-us/dk2/}

Oculus Rift Documentation intoduction to best practices. (2016). Retrieved 2016-05-19, from \url{https://developer.oculus.com/documentation/intro-vr/latest/concepts/bp_intro/}

Oculus Rift Documentation overview. (2016). Retrieved 2016-05-19, from \url{https://developer.oculus.com/documentation/game-engines/latest/}

Oculus rift specs - dk1 vs dk2 comparison. (2016). Retrieved 2016-05-19, from \url{http://riftinfo.com/oculus-rift-specs-dk1-vs-dk2-comparison}

Poulet, J., & Hedwig, B. (2003). A corollary discharge mechanism modulates central auditory processing in singing crickets. *Journal of neurophysiology*, *89*(3), 1528–1540.

Poulet, J. F., & Hedwig, B. (2006). The cellular basis of a corollary discharge. *Science*, *311*(5760), 518–522.

Scripting in unity. (2016). Retrieved 2016-05-19, from \url{http://docs.unity3d.com/Manual/ScriptingSection.html}

Shergill, S. S., Bays, P. M., Frith, C. D., & Wolpert, D. M. (2003). Two eyes for an eye: the neuroscience of force escalation. *Science*, *301*(5630), 187–187.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... others (2016). Mastering the game of go with deep neural networks and tree search. *Nature*, *529*(7587), 484–489.

Unity Manual animator component. (2016). Retrieved 2016-05-19, from \url{http://docs.unity3d.com/Manual/class-Animator.html}

Unity Manual animator controller. (2016). Retrieved 2016-05-19, from http://docs.unity3d.com/Manual/class-AnimatorController.html

Unity Manual primitive and placeholder objects. (2016). Retrieved 2016-05-19, from \url{http://docs.unity3d.com/Manual/PrimitiveObjects.html}

Unity Manual vr overview. (2016). Retrieved 2016-05-19, from \url{http://docs.unity3d.com/Manual/VROverview.html}

Vasser, M., Kangsepp, M., Kilvits, K., Kivisik, T., & Aru, J. (2015). Virtual reality toolbox for experimental psychology—research demo. In *Virtual reality (vr), 2015 ieee* (pp. 361–362).

von Helmholtz, H. (1867). Treatise on physiological optics vol. iii.

Von Holst, E. (1954). Relations between the central nervous system and the peripheral organs. *The British Journal of Animal Behaviour*, *2*(3), 89–94.

Weichert, F., Bachmann, D., Rudak, B., & Fisseler, D. (2013). Analysis of the accuracy and robustness of the leap motion controller. *Sensors*, *13*(5), 6380–6393.

Wolpert, D. M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *nature neuroscience*, *3*, 1212–1217.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*(5232), 1880.

Appendix I: Figures

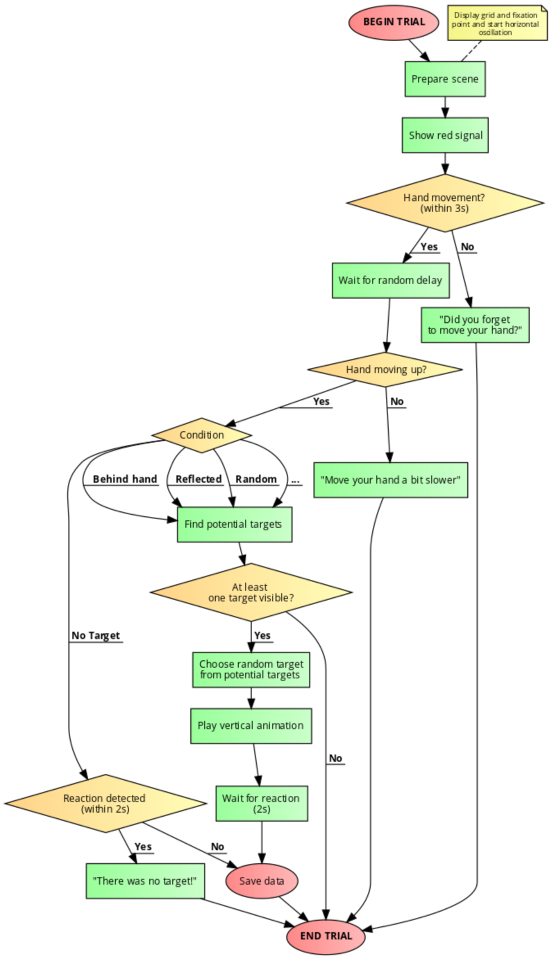

Figure 1: The flowchart of a trial of the Experiment 1

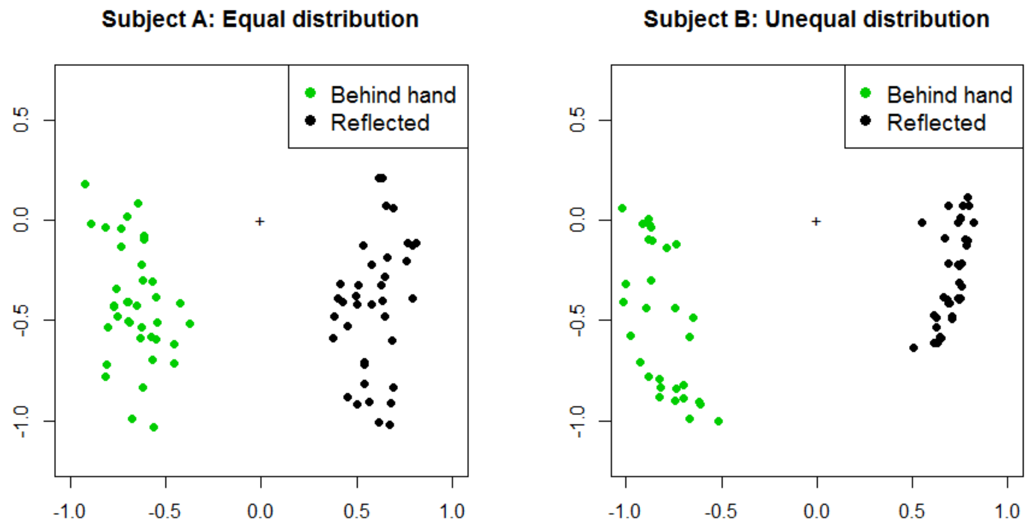

Figure 2: The spatial distribution of the targets in the Experiment 1 for two subjects (subject A and B, respectively). A. Targets are visually equally distributed in groups and from approximately similar distance from fixation point. B. Target groups are not equally distributed.

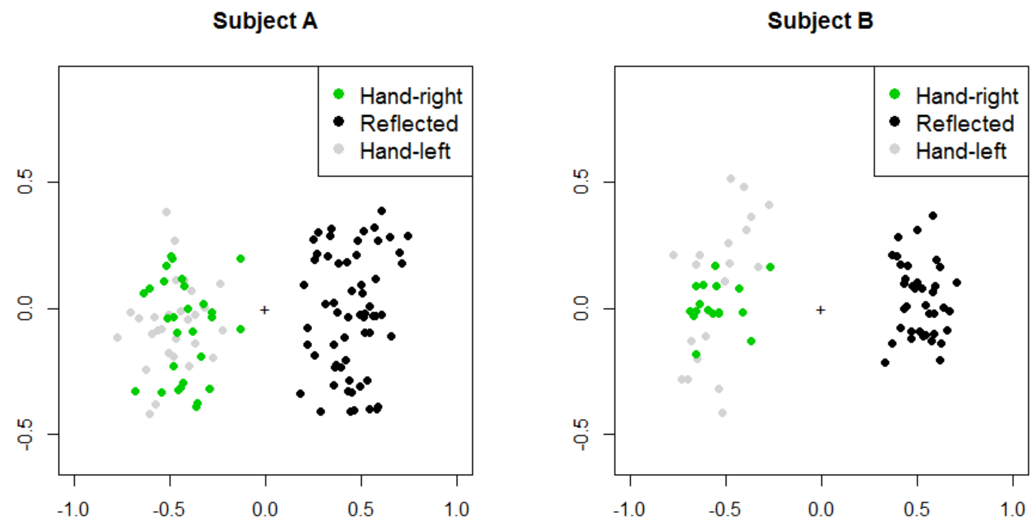

Figure 3: The spatial distribution of the targets in the Experiment 2 for two subjects whose data was rejected due to unequal mean distances from the centre of their respective clusters (subject A and B, respectively).

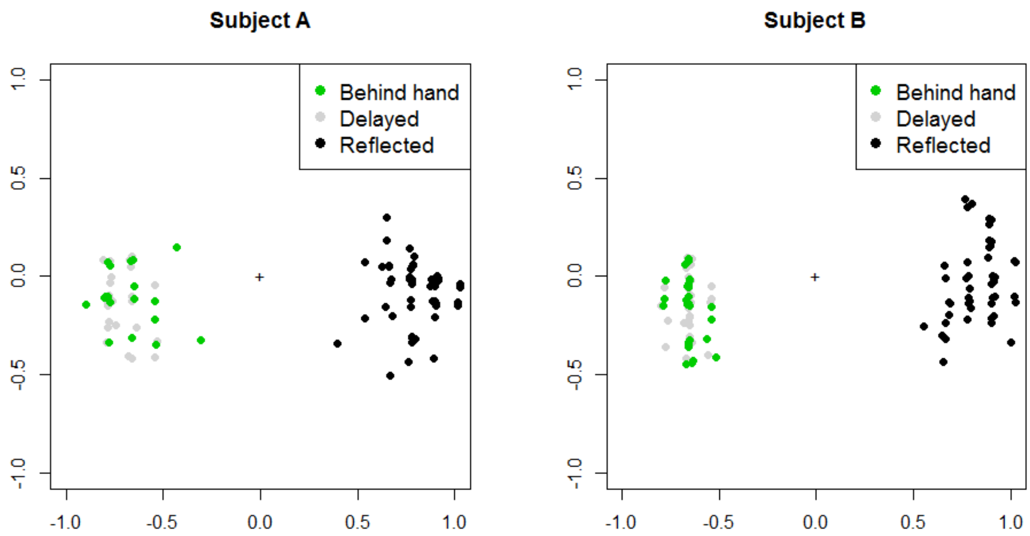

Figure 4: The spatial distribution of the targets in the pilot study for the two subjects whose data was rejected due to unequal mean distances from the fixation point for the hand and reflected condition (subject A and B, respectively). Although visually the groups of targets have similar mean spatial positions, these clusters were significantly different.

Appendix II: Tables

Table 1: Number of trials per subject per condition after preprocessing in the Experiment 1

	Condition	
Subject	Behind hand	Reflected
1	37	38
2	38	37
3	39	34
4	40	44
5	38	42
6	26	28
7	40	34
8	35	36

Table 2: The average RTs per subject per condition in the Experiment 1. Only data of subjects included in the analysis shown. Mean (SD)

	Condition	
Subject	Behind hand	Reflected
1	0.422 (0.065)	0.384 (0.068)
2	0.464 (0.128)	0.435 (0.110)
3	0.294 (0.078)	0.277 (0.064)
4	0.393 (0.082)	0.378 (0.103)
5	0.432 (0.103)	0.381 (0.075)
6	0.360 (0.052)	0.360 (0.035)
7	0.367 (0.042)	0.377 (0.064)
Mean(7)	0.391 (0.099)	0.372 (0.090)

Table 3: Number of trials per subject per condition after preprocessing in the Experiment 2

	Condition		
Subject	Hand-left	Hand-right	Reflected
1	29	24	46
2	21	35	60
3	19	21	45
4	32	24	48
5	29	28	60
6	19	17	38
7	21	24	40
8	31	26	57

Table 4: The average RTs per subject per condition in the Experiment 2. Only data of subjects included in the analysis shown. Mean (SD)

	Condition		
Subject	Hand-left	Hand-right	Reflected
1	0.362 (0.047)	0.399 (0.059)	0.357 (0.055)
2	0.359 (0.051)	0.359 (0.065)	0.368 (0.059)
3	0.388 (0.055)	0.374 (0.048)	0.356 (0.049)
4	0.457 (0.106)	0.463 (0.155)	0.459 (0.136)
5	0.381 (0.075)	0.454 (0.126)	0.413 (0.077)
6	0.335 (0.073)	0.334 (0.052)	0.34 (0.067)
Mean(6)	0.382 (0.084)	0.394 (0.102)	0.38 (0.088)

Table 5: Number of trials per subject per condition after preprocessing in the pilot study

	Condition		
Subject	Behind hand	Delayed	Reflected
1	18	23	50
2	18	20	48
3	34	34	62
4	22	26	53
5	19	20	40
6	25	28	58
7	27	30	48
8	22	27	58
9	26	25	57
10	20	23	55

Table 6: The average RTs per subject per condition in the pilot study. Only data of subjects included in the analysis shown. Mean (SD)

	Condition		
Subject	Behind hand	Delayed	Reflected
1	0.513 (0.095)	0.49 (0.051)	0.47 (0.076)
2	0.347 (0.044)	0.359 (0.049)	0.352 (0.048)
3	0.347 (0.095)	0.355 (0.088)	0.344 (0.066)
4	0.438 (0.106)	0.451 (0.083)	0.435 (0.112)
5	0.345 (0.047)	0.336 (0.072)	0.313 (0.061)
6	0.378 (0.053)	0.376 (0.072)	0.377 (0.072)
7	0.384 (0.085)	0.374 (0.082)	0.351 (0.067)
Mean(7)	0.388 (0.092)	0.388 (0.087)	0.375 (0.089)

Appendix III: Pilot study

Methods

Experimental design

The experiment followed the experimental paradigm and procedures of the Experiment 1, however, with two major differences. First, the control condition was a target in the same (left) side of the field of view as in the experiment condition. Second, the target was chosen in the downward motion of the hand, i.e. after the top turn of the movement was registered. More specifically, the control target was chosen from the same position where it would be behind the hand, but the target was presented after the tip of the middle finger passed the vertical position of the target plus the radius of the spherecast (see Figure 1 below). This allowed us to ensure that the target really is not behind the hand while keeping the spatial distribution of the targets in the two conditions approximately similar.

The targets were balanced between the following conditions: targets behind the hand, control condition from the same distribution as in the hand condition, but after the hand was moved out of the way; the reflected condition (same as in Experiment 1), and random trials. A total number of 200 trials were conducted for each subject.

The experimental algorithm for each trial was employed similar to Experiment 1. The only things different were the delay period that was started after the hand had reached to top height and started coming back down, the delay was chosen randomly from 150-275 ms, and the trial was rejected when the hand stopped before the target animation had played.

Notifications displayed

To control for the height and speed of the hand movement, many notifications were displayed for the subject when the hand did not follow the desired movement. For example, the subjects had to move the hand with the right speed and at least the length of a minimum distance during both the first 800 ms after the detection of the hand and after the top turn. Notifications were displayed for being both too slow and too fast. There were also notification for the height of the hand movement to ensure that the targets appear more in the center of the VFOV and

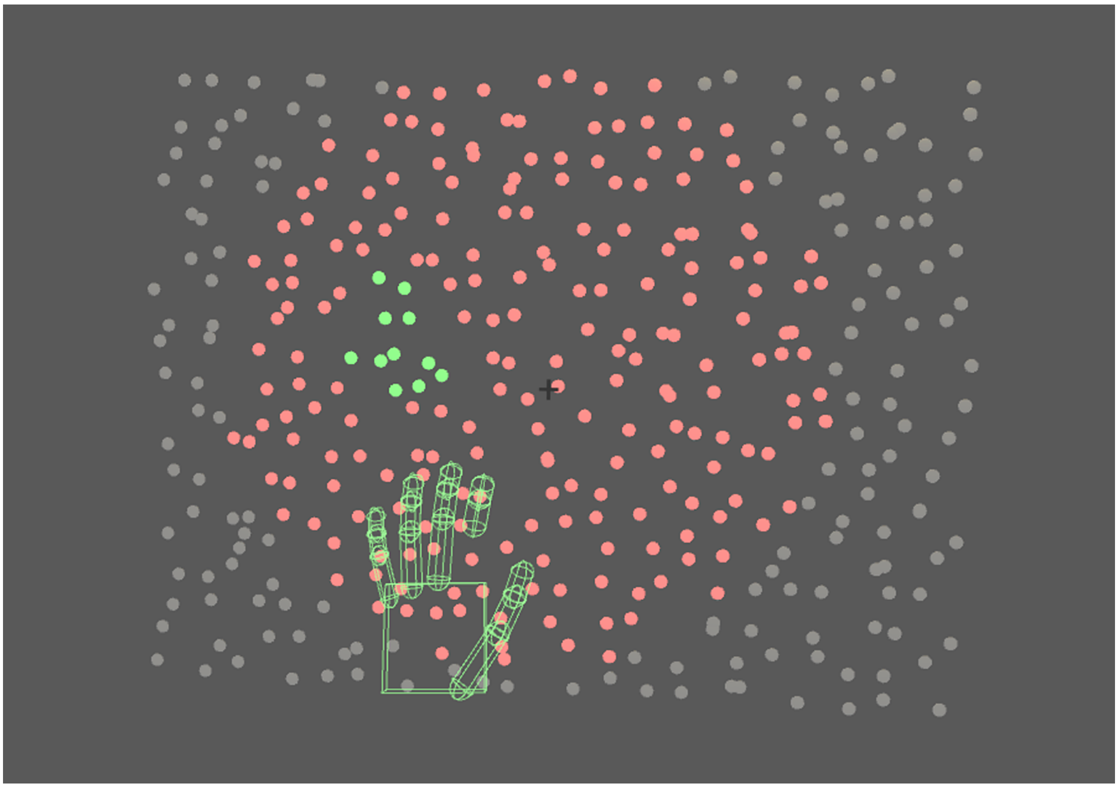

Figure 1: View from the camera for the left eye: the set of potential targets for the control condition in green, the nVFOV in red. In the control condition, the set of potential targets was exactly the same as when the target was shown behind the hand, except the target was shown when the hand had already moved away from that position. Specifically, target was presented when the middle finger had passed the position of the chosen target plus a diameter of the spherecast itself. Note that the figure is illustrative: the targets were not coloured and the hand was not visible to the subjects during the experiments.

while the hand velocity is the highest.

Results

Descriptive analysis

10 healthy subjects (5 female, 5 male; aged 21-27, mean 23.4 years) with normal or corrected to normal vision participated in the pilot experiment. A total number

of 1051 non-random trials were collected where the RT was measured. Further preprocessing removed 32 trials from the analysis (see Table 5 in Appendix II). Data from one subject was removed from the analysis as the subject reported not keeping focus on fixation point for a considerable amount of trials (For overall RTs for each subject, see Table 6 in Appendix II).

Equality of distances

These analysis were conducted similarly to Experiment 1. For two subjects, the mean distance from the fixation point was unequal for several pairs of conditions, thus, these data was excluded from further analysis (See Figure 4 in Appendix I).

Reaction time analysis

With the remaining 7 subjects, we tested for the difference of mean RTs in the following pairs of conditions: targets behind hand and delayed (control condition), targets behind hand and reflected, and targets delayed in the left and the reflected condition. We observed no effect between the RTs to targets behind hand and targets delayed ($t(6) = .360$, $p = .366$, $d = .001$, Figure 2 below). However, the RTs between the hand and reflected, and delayed and reflected conditions were significantly different ($t(6) = 2.188$, $p = .036$, $d = .138$, $t(6) = 4.254$, $p = .005$, $d = .139$, respectively). Note that the effect sizes for the latter comparisons are approximately equal. The ex-Gaussian measures showed no difference between the mean of the normal component of the curve, variability of the normal part, and the skewness (tau) of the exponential part of the RTs curve for the hand and delayed condition. However, the Gaussian mean was significant between the both the hand and reflected, and delayed and reflected condition ($t(6) = 2.442$, $p = .050$, $t(6) = 2.493$, $p = 0.047$, respectively).

Debriefing

All the subjects were found to be blind to the objectives of the experiment. Most of the subjects suggested testing of (i) the effect of the distance from the fixation point, (ii) coordination of the two tasks, (iii) the speed of the hand movement, or (iv) even the effect of the current trial on the next one. One subject also noted that we might study the RT of the left hand to the red "go" signal, and that the hemispheres might play a role in this. Notably, most subjects reported the hand

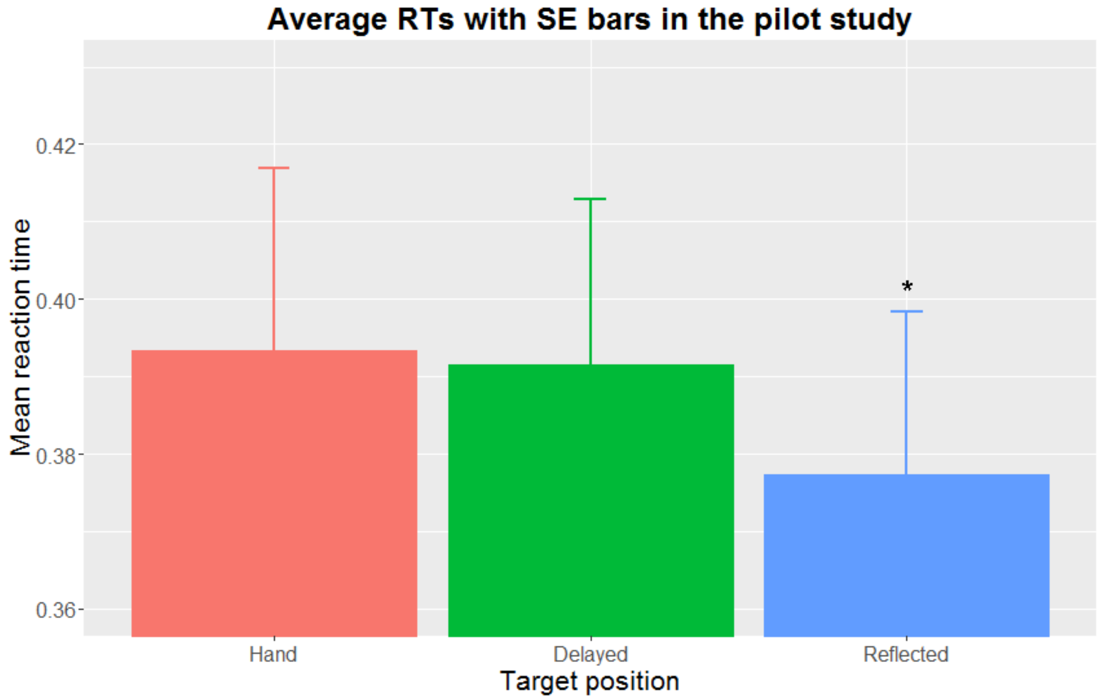

Figure 2: The average reaction times in the pilot study for the experimental (behind hand, red), delayed (green), and reflected condition (blue) with standard error (SE) bars, subject pooled.

movement taking a lot of attention. Moreover, many subjects verbally reported often slowing down the hand movement after the top turn in order to be more accurate in the RT task.

Discussion

The results of the pilot study indicated clear effect between the RTs to targets in the left side of the field of view (both hand and delayed condition) compared to the reflected targets. To add, there was no effect between the mean RTs of the hand and delayed condition. That could in the first observation indicate no effect of the hand movement on the sensory dampening of the predicted motion but rather an effect of the bilateral distribution of the targets between the left and right visual fields, that is , to the different processing in the left and right side of the brain. However, the pilot study had several critical experimental drawbacks.

First and most importantly, the subjects did not follow the instruction to move the hand in a constant speed. This was observed both by the experimenter visually during the experiments and the subjects themselves, as the debriefings revealed. This could have resulted in the delayed targets to be presented (a) near the top of the hand, and (b) when the hand velocity was significantly slow. Importantly, we do not know whether the sensory dampening appears right in the area of the visual field where the hand is located, or also in the near radius around the hand. Thus, both the targets behind the hand and delayed could have been under the influence of the same putative sensory attenuation of the well-predicted hand movement.

Second, the hand movement was in different phase during the hand and control condition. In other words, given an ideal hand movement, the targets in the hand condition appeared during the approximately fastest phase of the downward motion, but the delayed targets appeared when the hand was already slowing down or almost stopped. Hence, it is not clear whether the comparison of these conditions is completely valid.

Last, the full set of notifications were displayed to the subject when the hand movement did not follow the desired speed or accuracy. Moreover, the notification message often did not convey the real reason the notification was displayed. For example, the notification "Raise your hand higher" was often shown when the hand disappeared either because of getting too close to the Leap, or, paradoxically, getting too high above the Leap. This confused the subjects, leading them to pay considerable amount of attention on the hand movement rather than the RT task. Indeed, many subjects reported the experiment to be about the ability to coordinate between the two tasks.

Appendix IV: Code

For a detailed inspection of the Unity project including the scripts written by the author please fork the Github repository from `https://github.com/juliuslaak/msc-unity-project-laak`.

Licence

Non-exclusive licence to reproduce thesis and make thesis public

I, Kristjan-Julius Laak

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:

1.1.reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and

1.2.make available to the public via the university's web environment, including via the DSpace digital archives, as of 19.05.2017 until expiry of the term of validity of the copyright,

"From the brain to intelligent systems: The attenuation of sensation of self-generated movement",

supervised by Jaan Aru,

2. I am aware of the fact that the author retains these rights.

3. This is to certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, 19.05.2016