Kinect project: Endless Runner

Fritz Roth and Roman Seiler

Interaction Engineering

by Prof. Dr. Michael Kipp

winter semester 2020/21

03.02.2021

Augsburg University of Applied Sciences

Abstract

This research paper covers the procedure of developing the Kinect project: Endless Runner.

Picking up the people from their sitting desktop workspace and getting them to move and exercise playfully for a short period of time in order to stimulate the blood circulation, was the major goal.

We analyzed the different options of how to motivate our users to exercise and checked the overall available technology to the common home office employee.
Our decisions of the **implementation** were based on doability in the set timeframe, the pandemic restrictions and the spatial limitations.
So we had to combine these different demands on one mutual domain.

The **concept** helped us from square one on meeting our requirements in order to build the general structure of both hardware and software.

Three finalized sets were chosen by testing many different approaches of interaction techniques for the success of this application.
We outline how the operations are assembled, discuss setup methods and evaluate the finalized sets through direct benchmarks, so that we were able to gather information about possible future development.

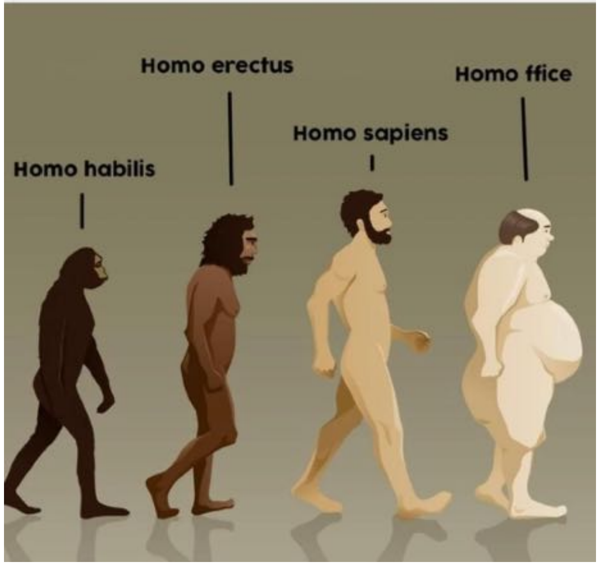
In the following we will describe the whole process in more detail.

Motivation

In times of a nation-wide lockdown where citizens are restricted to stay at home safe and sound, all professions earned the privilege of utilizing the home office for labour.
With neither the workplace nor a gym accessible, their daily movements got limited among bed, kitchen or desktop back and forth, we wanted to create a casual way to stay fit by **exercising** a bit even in a confined space.

Your first thought might be about virtual reality, to give them endless virtual space. But we see another physical constraint in wearing such a headset and controllers.We want the people to unleash themselves from being entangled in their peripherals.

Therefore we were looking for a possibility to interact with a computer completely **wireless** with nothing else but their own body.
A major device to provide such interaction techniques with enough knowhow available for us was an already ten years old camera with depth sensor, infrared emitters and skeleton tracking technology.

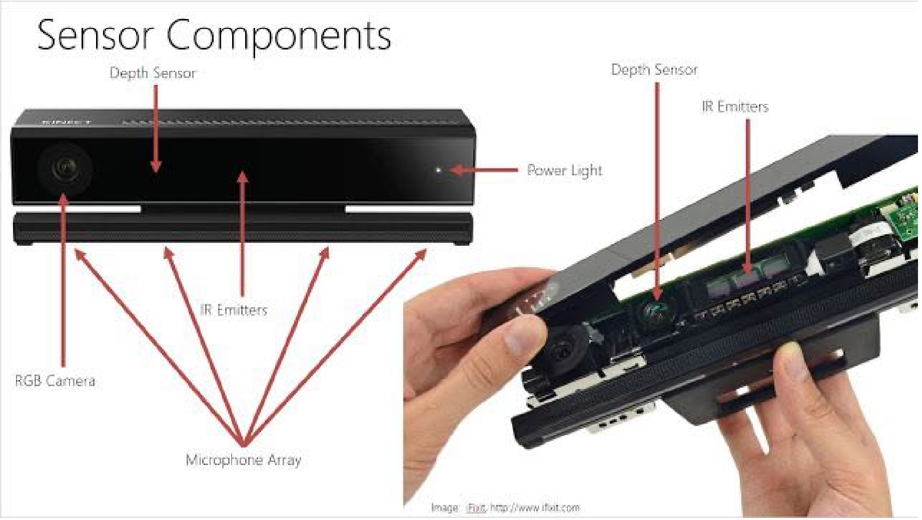


The **Kinect Gen 2**,
which refuses to be obsolete

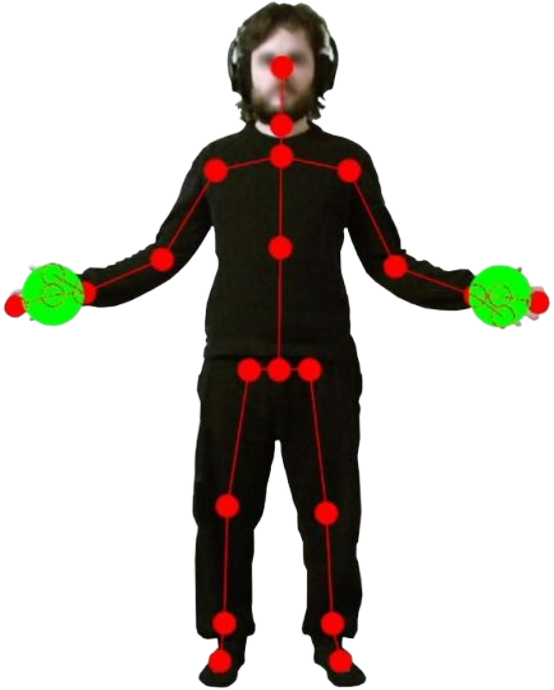
Concept - Technical aspects

The Kinect in its second Generation is equipped with the technology, that is interesting for our project:

- a Full HD RGB Camera
- a depth camera with a resolution of 512 x 424 pixels
- and infrared emitters, which create a grid to precisely **track** any movements regardless the lighting conditions



The Kinect is able to recognize multiple people and each of their skeletons to capture the **motions** simultaneously.
The ability to address all the joints, makes it possible to define conditions for various types of gestures.



Concept - Prototype idea

After deciding on the Kinect as our tool of choice, we had to take the next logical step and disconnect the people from their monitors.
Rather than bringing the player into the game, we **recreate** the interactable area beneath the player. For the purpose of drawing this idea into the physical space, an ordinary projector was sufficient.

Keeping the interactions intuitive for everybody was one of our main goals.
All interactions are executed by full body movements in an absolute scale.
Therefore we came up with an Endless Runner in which the player has to avoid obstacles, by simply dodging them left and right or **exercising** special interactions.

Acceptable area

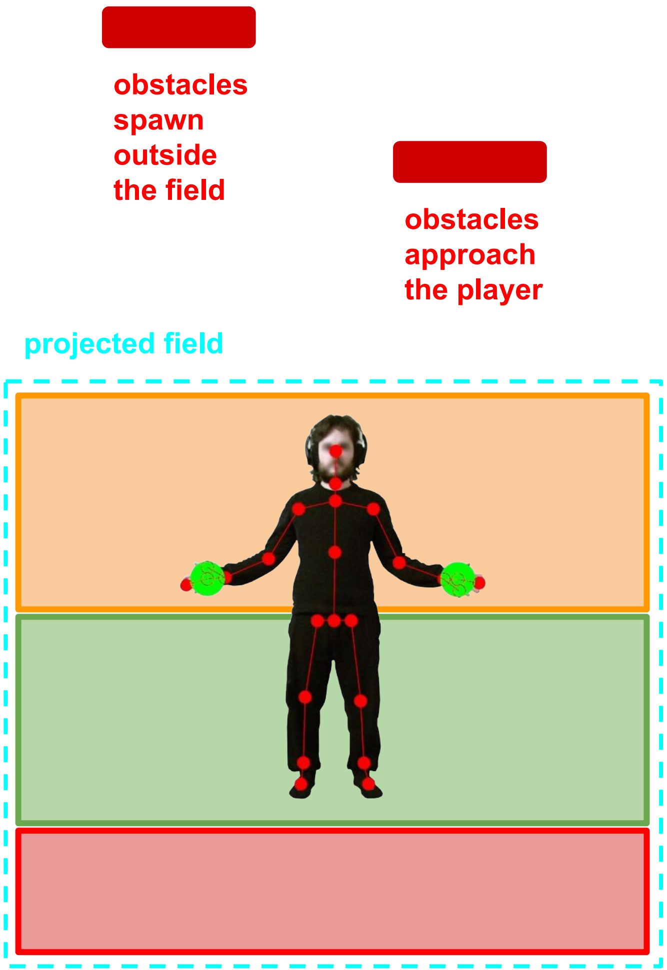
- player obstructs the displayed field behind him
- player has less time to react to obstacles

Optimal area

- player has enough time to react to obstacles
- movements get

Critical area

- tracking capabilities of the Kinect get worse with distance
- e.g. displaying the position gets shaky



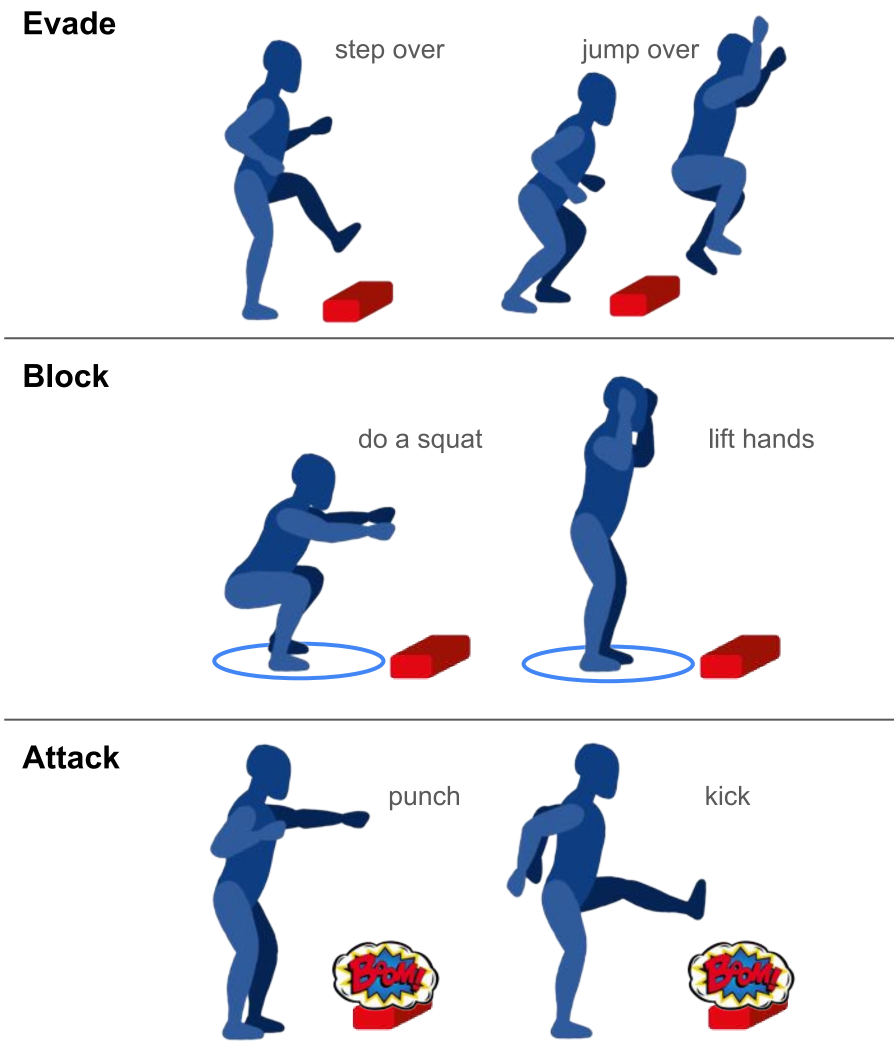
Interaction techniques

As stated in the concept, we motivate the player to move through an obstacle course.
We came up with three different sets of interactions. While each set consists of two movements, that are similar in their purpose, they vary in execution. All interactions **react** with audio and visual **feedback**.

Set 1: Evade
The player dodges approaching obstacles
- by either stepping over them
- or by jumping over them

Set 2: Block
The player protects himself from the obstacles
- while doing a squat and holding this position
- or while lifting one hand or both on head level

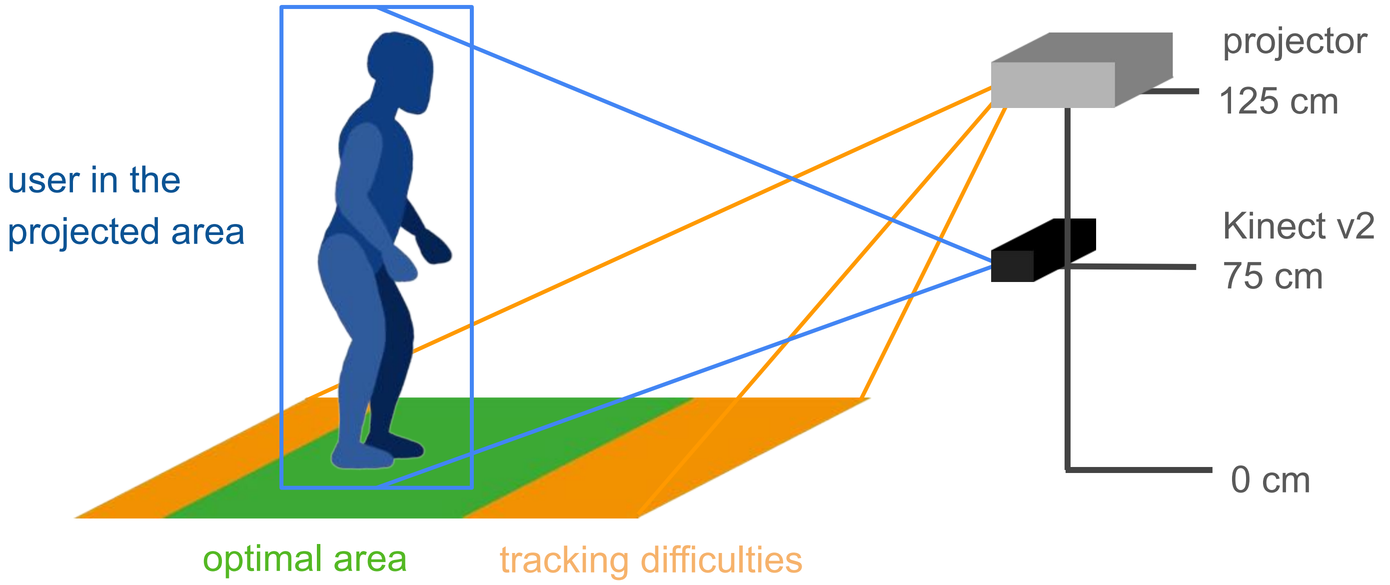
Set 3: Attack
The player destroys incoming obstacles
- by either kicking them
- or by punching them



Implementation - The setup

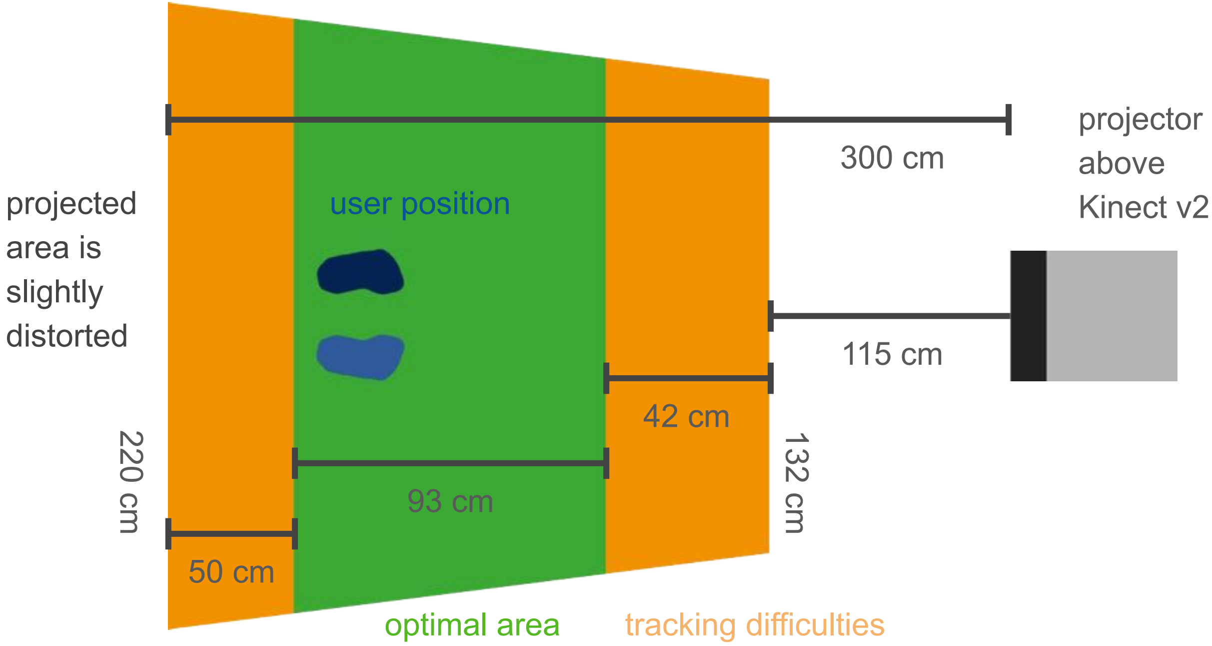
We replicated the desired conditions of a small enclosed apartment room as the testing grounds for our sets of basic exercises.
The projector is angled downwards in order to draw the obstacle course around the player. The projected image is flipped vertically, so that approaching obstacles and visual feedback are correctly aligned towards the player.
The Kinect tracks the skeletal joints of the player and transmits the **positional data** to the system, where an algorithm remaps it to the virtual two-dimensional playground.

Here is an image of the realized setup. The lense of the projector and the camera of the Kinect are arranged **perpendicular** to each other.
Therefore we did not have to add more calculations to the code to display the players position correctly.



Implementation - The setup

While the projector is angled downwards, it causes a slight distortion of the displayed field, which is negligible though.
Yet the distortion is taken into account for the position recognition and is being corrected inside the **remapping** algorithm.



Our anticipation about the acceptable, optimal and critical area were also confirmed by technical reasons.

In the **orange area** tracking difficulties occur due to the distance to the Kinect and its field of view. If the player stands too close, not all of the skeletal joints can be tracked.
If the player increases his distance to the camera, he starts to shrink in perspective while also rising in his captured height position. This causes issues with our defined heights, which are necessary for our interactions.
This will be explained in more detail on the next slide *Calibration*.

The **optimal area** describes the field, where the player can be tracked in full body size. This allows us to provide the best **experience** for the player without any inconvenience.

Implementation - Calibration

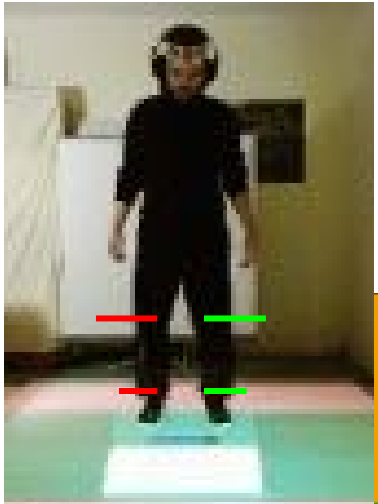
What did we calibrate and why?

The main reason for calibration was to gain a better distinction between the different interaction methods. Due to the perspective the further away a joint is to the Kinect, the more it appears to rise in height.

The **hovering** motion consists of one leg standing whilst the other is being lifted. So we are able to use the relative coordinates of the standing foot to compare them to the coordinates of the lifted one.

On the other hand we could not rely on any relative point while being **airborne**, as it consists of leaving the ground with both feet simultaneously.

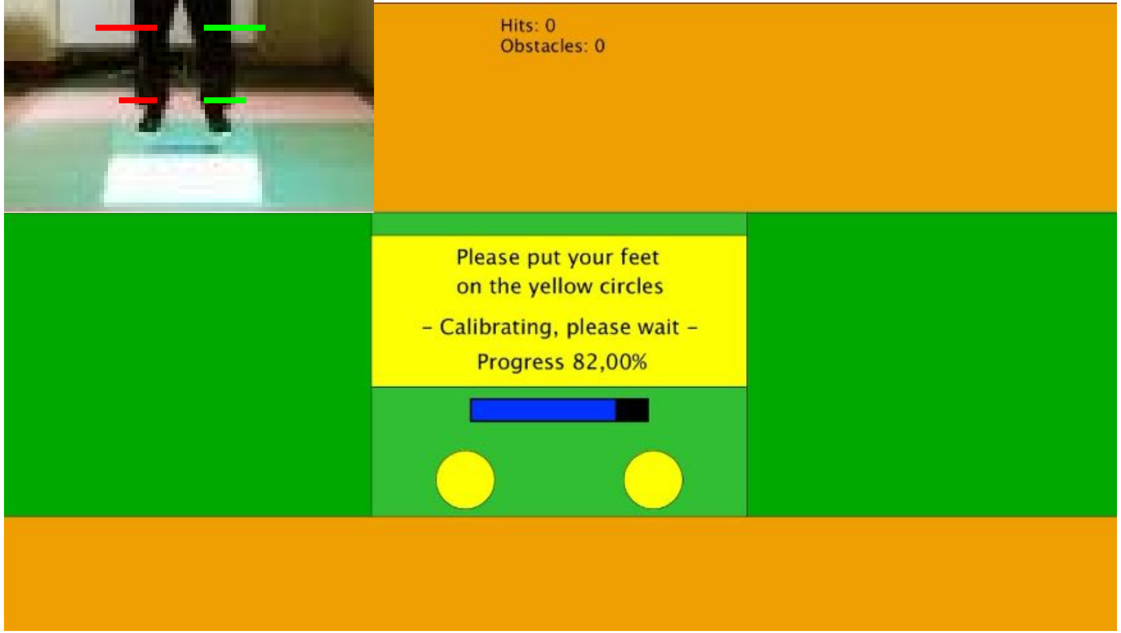
Therefore we were in need to create a virtual point to define an absolute floor height. The trigger-point to jump was an offset above the ground depending on the players knee height. This point had to be crossed with the lower ankle of both feet mid-air to trigger the jump interaction.



magnified image to illustrate

ankle and knee heights of

the right leg and left leg



Decision making

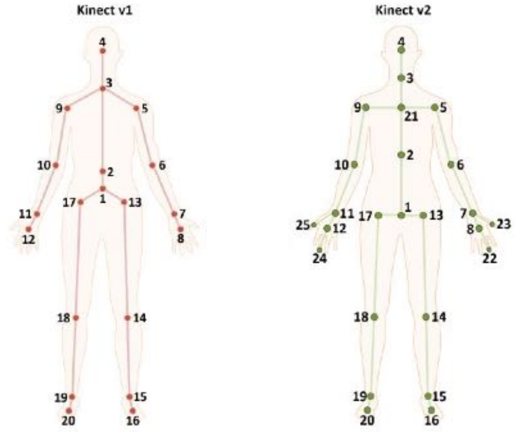
For reasons of optimal usage, we had to test every aspect of our available hardware in bipartite options.

The Kinect v2 had more advantages compared to the Kinect v1. With the second Generation we were able to use a **higher resolution RGB camera** and profit of the sixth fold amount of colored pixels to determine the body's position more precise.
The pixel count of the **depth camera** got almost tripled with the later version while also extended the overall **Field of View**.
Also the Kinect v2 provided us with **more skeletal joints** to track. The basis for starting the development were the open source Kinect libraries provided by **Thomas Sanchez Lengeling**.

In theory we speculated on the advantages of near-field projectors in the confined space that were our test grounds. Yet the distortion it caused on the displayed field, could not be used effectively.
A far more advanced picture was provided by an ordinary projector when we angled it downwards.

In the testing phase of the project we realized, that the implementation of modular object oriented coding might have simplified the process. Adding configuration files in order to handle the different circumstances in both of our workspaces at home, would have been an improvement in retrospect.

An automated calibration of a all the necessary criteria would have exceeded the time schedule and scope of our prototype.



Feature	Kinect for Windows 1	Kinect for Windows 2
Color Camera	640 x 480 @30 fps	1920 x 1080 @30 fps
Depth Camera	320 x 240	512 x 424
Max Depth Distance	~4.5 M	~4.5 M
Min Depth Distance	40 cm in near mode	50 cm
Horizontal Field of View	57 degrees	70 degrees
Vertical Field of View	43 degrees	60 degrees
Tilt Motor	yes	no
Skeleton Joints Defined	20 joints	26 joints
Full Skeletons Tracked	2	6
USB Standard	2.0	3.0
Supported OS	Win 7, Win 8	Win 8
Price	$299	TBD

Evaluation - User study

In a mixed team of Master and Bachelor students, the Master student had to perform the evaluation of the project.

In a user study I wanted to compare the two interaction methods in each of the sets. I developed a within-subject test, so every participant did both test runs with experiencing all the possible interaction techniques. Before the actual testing began we familiarized the participants with our project.

We split the runs into an A-test and a B-test to gather information about the exercise sets Evade, Block and Attack. In a direct comparison test we split up the interaction techniques of each set as the table on the right shows.
For the analysis the participants had to choose their **preferred option**.

Test A was designed to be more demanding in its execution than test B. Therefore we anticipated a higher user score for test B.

We also asked questions about the levels of complexity and exhaustion for each interaction individually and within the three sets.

Our participants rated the more technical aspects of our prototype in a range from -5 to 5, inspired by Likert scales.

direct comparison test	A	B
Evade	jump over	step over
Block	squat	lift hands
Attack	kick	punch

* master student only

Evaluation - Results

Due to the pandemic our number of participants was quite limited. The project only had 4 test-subjects in total, all adults, two between 20 to 40 and two between 40 to 60.
Considering the small amount of test subjects and the early stage of our project, the evaluation is rated as formative for possible work in the future.
The results of the direct comparison test are shown on the right.
All of our participants tend to prefer test B, yet tie in the set Evade.
So our anticipation got confirmed by the user score.

Overall test A was reported as more complex and exhausting than test B, but with an exception. The complexity to step over an obstacle was rated higher in its execution than to jump over it.

Questions without a semantic comparison came to mixed results:

- the **most exhausting** interaction technique:
  - 3 answered squat, 1 answered jump
- the **least exhausting** interaction technique:
  - 3 answered lift hands, 1 answered punch
- the personally **most difficult** interaction technique:
  - 1 answered step over
  - 1 answered jump over
  - 2 answered kick

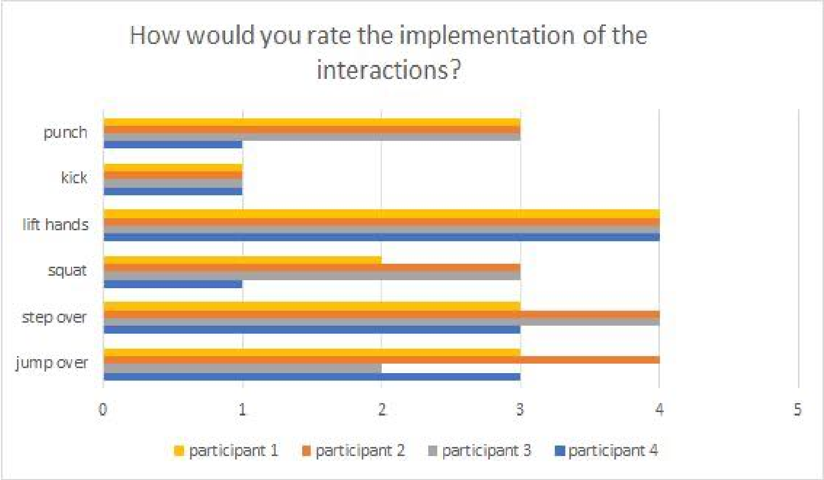
A	direct comparison, preference of the interaction techniques:		B
jump over	2	2	step over
squat	0	4	lift hands
kick	1	3	punch

* master student only

Evaluation - Additional information

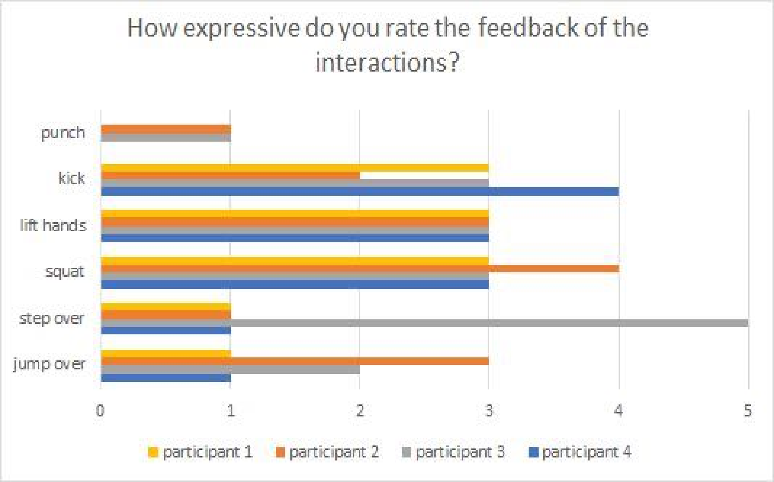
The meaning and range of the Likert scale rating system is shown in the figure on the right.
With no user score rating below zero the negative range of the scale was not necessary to be displayed in the diagrams.

insufficient					neutral					excellent
-5	-4	-3	-2	-1	0	1	2	3	4	5
O	O	O	O	O	O	O	O	O	O	O



In rating the implementation of the interactions

- The user score for “lift hands” is the highest: 4 in avg.
- The user score for “kick” is the lowest: 1 in avg.
- “squat” and “jump over” got inconsistent results



In rating the expressiveness of feedback of the interactions

- The user score for “squat is the highest: 3,25 in avg.
- The user score for “punch” is the lowest: 0,5 in avg.
- “kick” and “jump over” got inconsistent results

* master student only

Conclusion

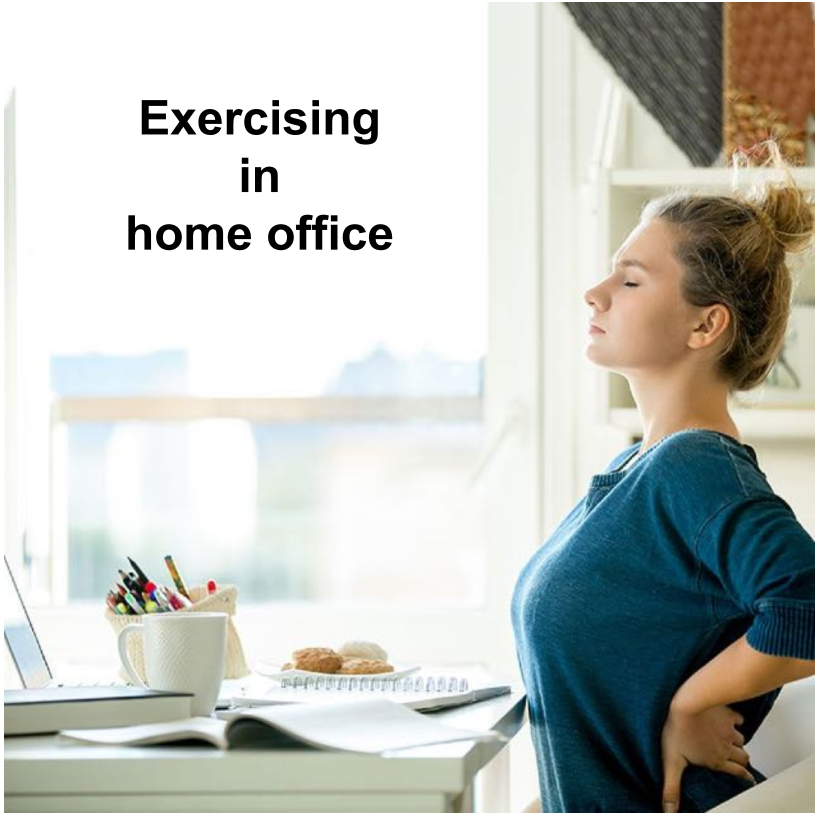
In the Kinect project: Endless Runner we developed three sets of various exercises, that are easy and intuitive to execute in order to overcome the approaching obstacle course and the own laziness. It has a vast accessibility, because of its minimal hardware requirements.

Our Endless Runner project is successful in motivating its players to move and exercise without any physical constraints. The activity of the players will boost their blood circulation and reduce their overall fatigue.

Even though people are stuck at home with their physical fitness slowly deteriorating, they don’t seem to be interested in doing exhausting exercises as the lower user score of test A shows. A more comfortable approach like the interaction techniques of test B is preferred.

A more modular approach in our coding architecture, would have improved the process and therefore benefited the progress of the project. Another future advantage might be saving the calibration settings of a user and the available spatial area in a configuration file. This file could be loaded at the programs initiation and spare the player from the recurring calibration time.
The gameplay of our application could be easily expanded and improved furthermore within a larger timespan.

“Endless Runner - End to Constraints”

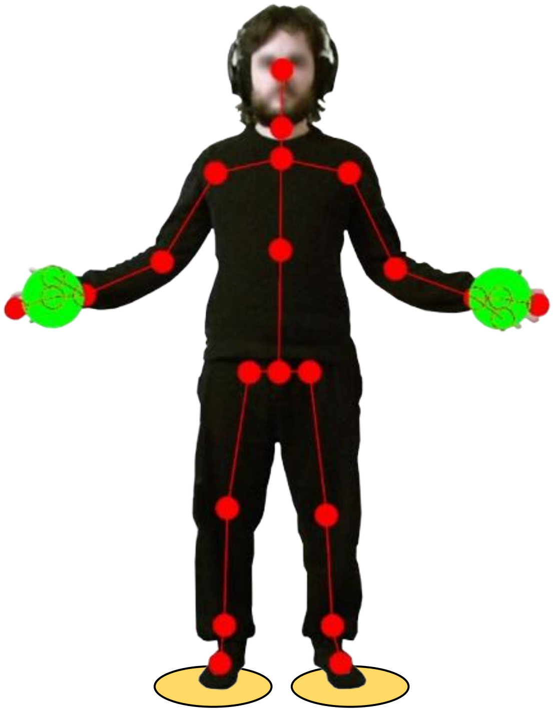


Possibilities in the future

Google recently started to implement skeleton tracking with ordinary **webcams** by enabling an Artificial Intelligence. This AI is capable of differentiating between a person and the background by splitting them, because modern webcams use advanced autofocus technology. Therefore the AI can locally calculate a person’s skeletal posture from the separated webcam input. For the full explanation visit Move Mirror:

https://experiments.withgoogle.com/collection/ai/move-mirror/view

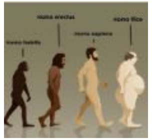
With this technology **advancing**, we could replicate our project onto these ordinary webcams in the future.



glossary

Boom.png

https://photobooth-props.co.uk/media/catalog/product/cache/1/thumbnail/600x600/9df78eab33525d08d6e5fb8d27136e95/b/o/boom_pop_art_coloured_photo_booth_prop_1.jpg



Homo office

João Paulo Rieg, https://jpreliquias.blogspot.com/2020/08/homo-office.html

Kinect v2

https://de.wikipedia.org/wiki/Kinect#/media/Datei:Xbox-One-Kinect.jpg

Kinect v2 technical

https://www.physio-pedia.com/File:Microsoft_Kinect.png

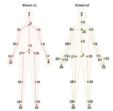
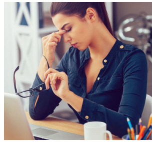
webcams

https://www.reviewgeek.com/p/uploads/2018/07/3a7c7f43.jpg

Snapshot from video of Move Mirror

https://winfuture.de/videos/Internet/Move-Mirror-Google-baut-Kinect-Funktion-einfach-ueber-Webcam-nach-19378.html

glossary

	Kinect v1 vs Kinect v2	http://zugara.com/wp-content/uploads/Kinect-1-vs-Kinect-2-Tech-Comparison.png
	joints	https://www.semanticscholar.org/paper/Kinect-v2-based-system-for-Parkinson's-disease-Rocha-Choupina/070b652524125ba30f5391fdef8feec441834227/figure/0
	home office	https://www.bauerfeind.de/fileadmin/user_upload/content-global/_blog/_sich-fit-halten/home-office-01.jpg
	fatigued employee	https://www.goodfon.com/download/fatigue-woman-office-worker/3840x2160/