MASTER THESIS

EFFECT OF GAMIFICATION ELEMENTS ON ENGAGEMENT AND PERFORMANCE IN A FORCE-TRACKING TASK

Wouter Schutte

FACULTY OF ENGINEERING TECHNOLOGY
DEPARTMENT OF BIOMECHANICAL ENGINEERING

EXAMINATION COMMITTEE
dr. E.H.F. van Asseldonk
A. Ruiz Rodríguez MSc
R.E. Ornelas Kobayashi MSc

DOCUMENT NUMBER
BE - <NUMBER>

UNIVERSITY OF TWENTE.

09/08/2024

Acknowledgements

Starting in 2011, I would never have thought that finishing what I started would take thirteen years. Though it should not come as a surprise anymore, finishing this master thesis also took a while longer than expected. The fact that it is now finished however, cannot be attributed to me alone. Many people have helped me through these last two years, so I would like to extend my gratitude to those people.

Edwin for the feedback and his relaxed attitude, letting me deliver a 54-page document only a couple of days before he needs to review it. Rafael and Aurora for being not only nice supervisors, but great people as well. Thanks for all the hot chocolate! Also, thanks for really making me make plannings, even though I rarely managed to stick to them. I have tried my best to randomize the order in which I addressed you in emails as much as possible, so neither of you would feel second-rank. We never scheduled that boardgame night we talked about in the beginning! Maarten, for all the help with Unity. Internet support is only suited for people that have basic knowledge, and I was decidedly below that point. Cynthia for some brainstorming and help with the Biodex. It is a clunky machine at times. I am just glad it never broke down while I was using it! All volunteers in the experiment, for spending their free time to provide me with results. I hope the chocolate bars made it worth your time. Maxime, for being the best wife I could ever have hoped for. I am sorry you had to suffer from more stress about my deadlines than I ever did. And finally, my parents, for showing support and urging me to not give up. I would probably not have completed my master's without you.

Abstract

Movement results from the interaction between the central nervous system and the musculoskeletal system. Damage to the central nervous system can result in motor impairment. For people suffering from motor impairment, assistive technologies have been developed to support locomotion or limb movement. These technologies often make use of electromyogram (EMG) recordings. However, while EMG can be used to predict muscle force using musculoskeletal (MSK) models, it remains challenging to retrieve the neural code from EMG. To obtain more detailed information about the translation of intent of movement to force and vice-versa, MSK models can be expanded to account for individual motor unit (MU) activations. Decomposition of high-density EMG can give insight in MU activations. However, to perform decomposition, a constrained motor task must be developed. Due to how the characteristics of movement influence rate coding and recruitment strategies, good performance of the motor task is of equal importance. Motor performance of subjects can be influenced by different factors, including engagement. However, in current research, cognitive factors involved in motor performance are typically not taken into consideration. Since engagement can be influenced by gamification, this study investigated whether engagement could be increased using gamification elements, and whether increased engagement led to improved motor performance.

In this study, a feedback interface for a motor task was developed. Performance and engagement of two groups performing a plantar flexion trapezoidal ramp motor task were compared: a control group (n=11) using a standard version of the feedback interface that shows target and output torque, and a gamified group (n=11) using the same interface but with added gamification elements such as score tracking and motivational messages. Force tracking errors were significantly ($p<.05$) larger for the gamified group in most of the experimental conditions. However, coefficient of variation of measured torque was significantly ($p<.05$) lower for the gamified group in half of the conditions. No significant difference in engagement, as measured using a modified version of the Game Experience Questionnaire, was found between the two groups. However, Content Analysis of the question “what did you think of the interface?” showed a higher frequency of words from the Fun category in the gamified group. As no significant difference in engagement was found, the difference in performance cannot be attributed to engagement. In the current state, the usage of the gamified interface over the standard interface is not recommended. Future work is needed to improve the gamified interface.

Table of Contents

1. Introduction

Movement results from the interaction between the central nervous system and the musculoskeletal system. Alpha motor neurons innervate muscle fibres of skeletal muscles. An alpha motor neuron and the muscle fibres it innervates constitute a motor unit (MU) [1]. The activation of the alpha motor neuron causes the innervated muscle fibres to contract, resulting in muscle force [2]. The central nervous system cannot control MU's independently. Instead, it supplies pools of MU's with a neural drive [3], which results in the recruitment of MU's largely according to Henneman's size principle [4].

Motor impairment is caused by damage to the central nervous system, which can result in loss of voluntary motor control. Physical trauma such as spinal cord injury or lesion, or neurological conditions such as stroke, brain injury or migraine, are the most prevalent causes of motor impairment [5], [6]. Of these, stroke is the most common [6]. In the Netherlands, the incidence of stroke is about 30.000 people each year. 41% of survivors experience difficulty in the performance of activities of daily living: a collection of skills necessary for selfcare, like locomotion, eating and washing oneself [7], [8].

For people suffering from various forms of motor impairment, physical therapy has been shown to potentially improve the ability of carrying out activities of daily living, and/or stop deterioration of the ability to perform these activities [9]. In addition, wearable technologies have been developed to support limb movement and locomotion [10], [11]. These devices often make use of the measurement of electromyograms (EMG) [10], [12], [13]. While needle electrodes can be used [14], the non-invasive nature of surface EMG (sEMG) makes it a preferable mode of data acquisition in wearable technologies.

Using only sEMG for driving wearable technologies has limitations however, including that the translation of EMG to force is non-linear [15]. While musculoskeletal (MSK) models can translate the EMG data into muscle forces, e.g. using a Hill-type model [16], which can then be translated into joint torques using forward dynamics [17], [18], they are still constrained to EMG limitations. Namely, that EMG is the summation of action potentials of multiple MU's. Retrieval of the neural code embedded in the EMG proves challenging [19]. Neural drive can be estimated from the EMG amplitude, but only crudely [20]. Also, because of the mentioned simplifications that MSK models employ, deriving neuromechanical information and the description of internal muscle dynamics from the MSK model remains limited [16].

To obtain more detailed information about the translation of intent of movement to force and vice-versa, MSK models can be expanded to incorporate motor unit (MU) activations. Because action potentials produced by the motor neurons are transmitted linearly to the muscle fibres they innervate [21], measurement of MU action potentials allows for interfacing with the central nervous system. Thus, the analysis of individual MU action potentials gives insight in muscle recruitment strategies and provides a more accurate estimation of how the central nervous system directs movement, including in pathological situations [22], [23]. Decomposition of High-Density sEMG (HDsEMG) has been studied in stroke survivors and the results may indicate a difference in MU recruitment threshold, MU recruitment range, and motor neuron firing rate, compared to healthy individuals [24], [25]. By transforming motor neuron spike trains into motor unit forces, the force of the whole muscle can accurately be predicted [16]. Hence, particularly for motor impaired individuals, the calibration of personalized MSK models that account for these differences is key in accurate force prediction.

The MU action potentials required for this can non-invasively be obtained via decomposition of HDsEMG [26]-[31]. Different techniques exist to decompose the EMG signal back into separate MU activations. These techniques assume that the summation of the activity of multiple MU's together composes the EMG signal [32] and that therefore, inversely, by decomposition of the EMG signal, the sources (i.e. MU's) can be decoded. EMG decomposition is typically performed with isometric contractions, due to the underlying assumptions of the technique [33]. Limitations of sEMG decomposition include a lower number of identifiable MU's during high contraction forces due to superimposition [34]; a lower number of identifiable MU's in subjects with a thicker subcutaneous layer due to spatial attenuation [34]; a lower number of identifiable MU's during higher contraction speeds [35]; a possibility of overlap between two or more MU action potentials, causing superposition which complicates classification [36] and increasingly occurs with higher contractions [37]; MU synchronization, which is increased in stroke survivors and makes separating the signal into their constituents challenging [38]; and finally, the MU action potentials are sometimes attributed to duplicate MU's, while in reality the MU action potentials stem from a single MU. These duplicates should be merged, but this has to be done manually [32].

As mentioned, motor control strategies depend on the characteristics of the motor task [3]. Both rate coding and MU recruitment influence muscle forces, however, their relative influence varies depending on the motor task because neural strategies differ depending on the force requirement [39]. During gradually increasing isometric forces, MU discharge rate typically increases. In contrast, during rapid (i.e. ballistic) contractions, initial high-frequency discharge rates followed by successive decline in discharge rate is observed [39]. As for recruitment, isometric contractions typically follow an exponential curve in the number of MU's that are recruited [40]. However, depending on the characteristics of the force task, different muscle fibre types may be recruited [14]. Combined with the assumptions of the decomposition technique and the difference in how muscle fibre types constitute different muscles in different proportions, this shows the importance of designing a suitable motor task. The motor task needs to be designed specifically for the muscle of interest, and constrained in such a way that the expectation of the contributions of rate coding and MU recruitment, as well the neural strategies, match the design of the experiment.

A correct performance of the motor task by the subject is likewise important, for the same reasons. A subject's motor performance can be influenced by multiple factors, including experience and age [41], cognitive impairment [42], physical training [43], and engagement [44], [45]. Measurement of MU activity typically involves showing the output and target force or joint torque to the subject using a real-time feedback interface. The subject is verbally encouraged to follow the target to the best of its ability [46]-[48]. However, cognitive factors involved in motor performance are typically not taken into consideration. Thus, the potential influence of engagement on motor performance specifically in the context of motor tasks, and consequently the assessment of MU activity, remains uncertain.

To summarize, engagement potentially improves motor performance. Motor performance in a force task influences the results of EMG decomposition. EMG decomposition gives insight in the subject-specific motor control strategies and MU activations, which can be accounted for in personalized MSK models to improve the accuracy of force prediction. Thus, understanding the impact of engagement on performance in this context can prove valuable.

Engagement is defined by O'Brian and Toms as:

> *"A quality of user experience characterized by attributes of challenge, positive affect, endurability, aesthetic and sensory appeal, attention, feedback, variety/novelty, interactivity, and perceived user control."* [49]

Thus, engagement can be influenced by gamification [50]-[54]. Gamification is the practice of applying game mechanics to non-game contexts [55]. It has been used to influence subject behaviour [56] in various relevant fields, such as therapeutics and rehabilitation. Gamification has also been studied in the context of physical exercise [45], [53], 54]. Studies show increased motivation [45], [53] and higher exercise performance compared to a control group [45].

Therefore, the aims of this study were two-fold: Firstly, to determine whether the addition of gamification features improves engagement in the context of a motor task. Secondly, to test whether engagement influences performance in a motor task.

To accomplish this, a digital interface for a motor task was developed. The interface had two versions, one that only shows the target and output force (i.e. not gamified), and one including gamification features such as a scoring system and motivational messages. Engagement and performance between users of the two interfaces was compared under the hypothesis that the addition of gamification elements would lead to increased engagement, and thus improved motor performance.

The study found significantly larger tracking errors, but significantly lower coefficient of variation in subjects using the gamified interface. However, the study found no significant difference in user engagement between interfaces. Because of this, the difference in performance cannot be attributed to engagement. In the current state, the usage of the gamified interface over the interface without added gamification elements is not recommended. Improvements to the gamified interface are suggested however, which can potentially increase engagement and reduce tracking errors. If future versions show a significant improvement in performance, the addition of gamification elements can lead to a increased performance on the motor task. This results in a more accurate estimation of MU-activations, which in turn leads to a better estimation of motor control strategies. This can improve personalized MSK models, so that force can more accurately be predicted from EMG, especially for people suffering from motor impairment.

2. Methods

To measure a potential difference in engagement and performance between a gamified and a non-gamified environment, a motor task was presented to two groups. The test group was presented with a feedback interface with added gamification elements, while the control group was presented with only the target and current ankle torque. To accomplish this, an interface was developed in which gamification features can be turned on or off. Then, a single-blind study on healthy participants was performed. Finally, performance and engagement were measured and compared between the two groups. This section thus consists of three parts. First, the design and development of the software will be described, as well as the implementation and evaluation. Afterwards, the study population and test protocol will be detailed. Finally, the measurement and definition of performance and engagement will be discussed, as well as the learning effect.

2.1 Interface design and gamification elements

The design methodology for the software was based on work by Morschheuser *et al.* [57]. It provides a structured approach for engineering gamified software. This methodology consists of seven phases: (i) project preparation, (ii) context analysis, (iii) user analysis, (iv) ideation, (v) design, (vi) implementation, and (vi) evaluation. Not all elements of the process-deliverable diagram from [57] were used, either due to time constraints or because elements were not applicable to this research.

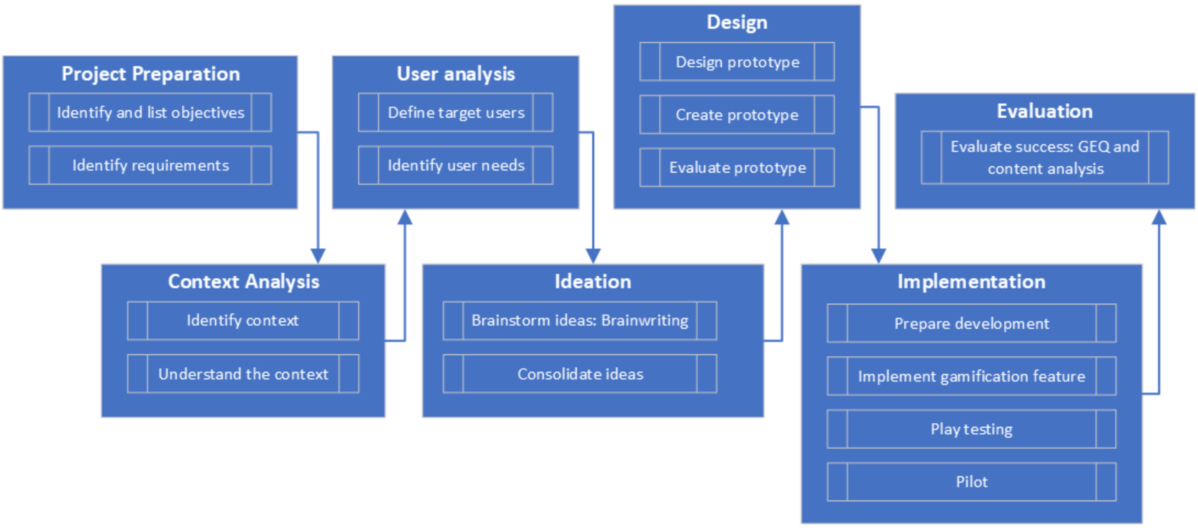


Figure 1: Design flowchart and processes.

The design of the interface followed the phases and processes shown in Figure 1. It was not possible to conduct a comprehensive user analysis. A simplified version of this phase was conducted instead, which defines the target users and their needs. For the ideation phase, a brainstorm was conducted with two software engineers, two biomechanical engineers, and two people with experience in the design of therapeutical games. The brainstorm used a variant of the Brainwriting method [58], in which participants spent the first four minutes for each question writing down their ideas in Google Jamboard [59], and were allowed to add to already existing ideas, but were not allowed to discuss them. Afterwards, the ideas were discussed. This was done for three questions: (i) *how would you increase engagement of the user*, (ii) *what should the* *feedback interface be like*, and (iii) *how do we adapt to needs and preferences*. The ideas were then bundled and consolidated into a concept.

The evaluation phase includes quantitative and qualitative results. Quantitative data was collected in the form of performance and engagement measurements, which are detailed in sections 2.6 and 2.7, but also in the form of the delay between the target and the measured torque. Qualitative data was collected in the form of answers to two open questions: (i) *what did you think of the test,* and (ii) *what did you think of the interface*. These answers were analysed using the Content Analysis method [60].

2.2 Interface implementation

The interface was built in two parts. One part was built in Mathwork's Simulink [61] and took care of the acquisition of data from the dynamometer, storage of raw data, normalization and filtering. The other part was built in Unity Technologies' Unity3D [62] and took care of the feedback interface.

2.2.1 Goal of the interface

The goal of the interface was to provide the subject with visual feedback regarding their applied torque and the target torque, and to persuade the subject to match the applied torque and the target to their best ability. To accomplish this, the following gamification elements were chosen in accordance with the brainstorm session: Challenge, Collection, Feedback, Level, Point, and Single-Player [63]. These together cover all four of player types in the Bartle typology [64]. The user interface should not distract the user from the goal and should be usable by a broad target audience.

2.2.2 Signal processing

The torque signal was filtered using a 100 ms moving average window, because a low-pass filter introduced delays larger than 0.5 s between measurement and display of the output torque. While larger moving average windows increase the smoothness of the signal, they also increase the delay between input and interface response. The 100 ms moving average window was chosen as a compromise between signal smoothness and interface responsiveness. For inter-subject comparison, the measured torque was normalized with respect to the respective subject's MVC. Communication between the recording interface and the feedback interface was done using User Datagram Protocol which ensures low-latency data transfer and that the latest sample is read by the interface. User Datagram Protocol has the added benefit that there is no buffer buildup of packages that are yet to be received, which improves responsiveness of the interface. Simulink also provided the Unity-based interface with the trial parameters (target %MVC, RoFD, plateau duration, resting time between repetitions, and number of repetitions) and a trigger signal at the start of the measurement so that synchronization of measurement data with the displayed target could be guaranteed in post-processing.

2.2.3 System specifications

While running the data acquisition in Simulink and the interface at the same time, the Unity inspector showed a potential framerate higher than 400 Hz, which surpasses the 30 Hz typically used in videos [65]. Leaving continuous data acquisition in Simulink running for thirty minutes did not noticeably impact the performance of the interface or the measurements, at which point the data acquisition was terminated. This far surpasses the five minutes needed for measuring the longest trial.

2.2.4 Accessibility

Where applicable, guidelines for application design were used [66], as well as ISO 9241-125 as a standard for the presentation of visual information [67]. The study was carried out with healthy volunteers. However, to make the interface suitable for testing a broader audience, several accessibility features were implemented. Colour vision deficiency, hearing impairment and illiteracy were considered. Design characteristics for these impairments were selected because they could be implemented without drastically impacting the design, and because the impairments are common. Colour vision deficiency is common in males (5.26% to 11.36%) [68], so the elements of the interface were designed with this in mind. Combined with ISO 9241-125, this meant that contrast between elements had to be maximized. Colour differences were never the only way to differentiate between different interface items or meanings. It is suspected that globally, about 20% of people suffer from hearing impairment [69]. Sound was not used in the interface so that users did not have to rely on hearing to perform the experiment. Finally, while illiteracy is rare relative to colour vision deficiency or hearing impairment, occurring in about 1% of adults in the USA [70], the interface was made to be language-agnostic because it can also be beneficial for children and people with cognitive impairment.

2.2.5 Latency measurement

To quantify the delay between the target torque and the measured torque, the cross-correlation of the two signals was computed. This gives the displacement between the two signals, which is a measure of the latency caused by the system. Because two different interfaces were used between the groups, it is necessary to eliminate any latency caused by the system, as latency causes inaccuracies in the calculation of the performance metrics.

2.3 Interface evaluation

The responses to the two open questions were evaluated using the Content Analysis method. This method was chosen because it is a flexible tool that can be manually performed with the amount of data in this study [60]. Content Analysis identifies one or more conceptual categories and then records the frequency of occurrence of words or phrases related to them. In this analysis three pairs of conceptual categories were used. Within each, the frequency of related words was recorded. Frequencies were then compared between the two questions and between test groups. The categories are shown in Table 1.

Table 1: Categories used for the Content Analysis. Three pairs were construed, each with a positive and a negative category.

Positive category	**Negative category**
Fun	Unfun
Engaging	Boring
Clear	Confusing

2.4 Study population

The study population consisted of twenty-two healthy volunteers (18-64 years old, average age 27.4, 11 females). Before recruitment, additional demographic data was collected on topics that are suspected to influence the performance and/or engagement measurements, including experience and physical training [41], [43]. Subjects were asked whether they play videogames, and if so, how often, chosen from a 5-point Likert scale [71] consisting of 'Very rarely', 'Rarely', 'Occasionally', 'Frequently', and 'Very frequently'. These were given a score from 1 to 5, respectively. A score of 0 was given to subjects that did not play videogames. The questions regarding physical activity are based on the Global Physical Activity Questionnaire (GPAQ) [72], a standardized tool developed by the World Health Organization as a way to quantify physical activity. Subjects were asked whether they participate in sports, and if so, which and for how many hours in a typical week. High-intensity sports, as defined by the GPAQ, had the number of hours per week doubled for the purpose of a better group division and population mean to get a better comparison. Subjects were then asked how many times in a typical week they walk or use a bicycle for at least 10 minutes continuously to get to and from places. Finally, subjects were asked whether they are familiar with force tracking measurements. These demographics can be seen in Table 2.

Healthy individuals aged 18 to 64 were invited to participate, with the following exclusion criteria:

- Impaired ankle movement (e.g., due to sprain history, neurophysiological disorders, or injury).
- Severe visual impairment.

Subjects were recruited on a voluntary basis without a promise of compensation. All subjects gave explicit written consent to volunteer in the study and permission for their data to be processed by signing an informed consent form. See also Appendix 1.

Subjects were divided into two groups: One gamified group that performed the force tracking task using the gamified interface, and one control group, that performed the same task but with the non-gamified interface. Subjects were divided into groups such that the demographics of each group were as similar as possible, which can be seen in Table 2. Priority was given in this order: Familiarity with force tracking tasks, frequency of video gaming, physical activity as described by sports frequency and movement by walking or by bicycle, and finally age range and gender. For the calculation of mean age and the standard deviation, the middle of the reported range was used.

Table 2: Demographics of the control and gamified group. Items are in order of priority. Presented scores of the bottom four demographics are the group mean (± standard deviation).

	Control group	**Gamified group**
Group size [-]	11	11
Number of subjects familiar with force tracking tasks [-]	1	1
Frequency of video gaming [-]	2.64 (±1.63)	2.64 (±1.29)
Participation in sports [hour/week]	3.10 (±2.93)	3.59 (±2.91)
Travel by walking or bicycle [week^-1]	12.82 (±6.76)	11.27 (±6.52)
Age [year]	26.09 (±7.74)	28.64 (±11.33)

2.5 Test protocol

Each subject participated in six measurement trials of five repetitions each. Decomposition experiments typically record HDsEMG during isometric contractions at varying target forces and contraction speeds [33], [34], [73]. To keep the chosen force task as close to the intended use case as possible, this experiment mimicked a typical force task used in decomposition. Each trial was a combination of one of three target percentages of MVC (%MVC), namely 15%, 30% and 45%, and one of two rates of force development (RoFD), namely 5%/s and 20%/s. These parameters are similar to those used in HDsEMG measurement for the decomposition of isometric contractions of the *musculus tibialis anterior* [35], [74]. These parameters were constant within a trial. A repetition consisted of four parts, as illustrated in Figure 2.

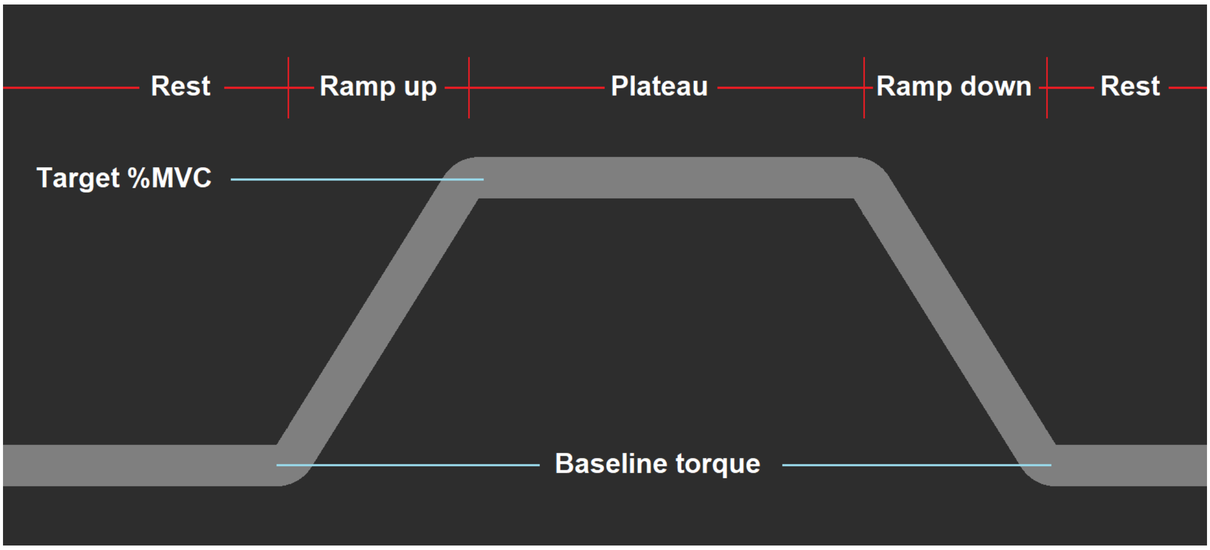


Figure 2: The different segments of the trapezoidal ramp protocol, and indications of baseline torque and target %MVC. A repetition consists of a Rest, Ramp Up, Plateau, and Ramp Down section, and then the next repetition starts again with a Rest segment. The light grey area shows the target including the tolerance.

Trials started with a Rest segment. The resting times increased with increasing target %MVC, to avoid fatigue affecting the performance, as can be seen in Table 3. After the Rest segment came a Ramp Up segment, with an incline rate equal to the RoFD, up until the target %MVC. Then a Plateau segment at the target %MVC for 5 seconds, and finally a Ramp Down segment with a decline rate equal to the RoFD, down to 0% MVC. To eliminate the impact of muscle fatigue on performance, after each 15% MVC trial subjects had a resting time of one minute, and after each

30% and 45% MVC trial subjects had a resting time of two minutes [74]. The six test conditions are summarized in Table 3. Subjects were asked to not drink coffee up to two hours before the measurements, because caffeine increases cortical neuronal excitability [75], which could potentially influence results. Upon arrival, subjects were asked if they had any more questions, and were informed that they had to fill in a form regarding their experience at the end of the experiment. They were then introduced to the experiment and the setup. Subjects were seated in front of the dynamometer in an upright position, with their right foot secured with Velcro straps to the footplate of the dynamometer, which can be seen in Figure 3. The chair and footplate position were adjusted so that the knee angle was approximately 120 degrees, and the ankle angle was 90 degrees. The footplate was fixed, so that ankle torque due to isometric contraction of the *musculus triceps surae* could be studied. The screen with the visual feedback interface was placed 1.5 m in front of the subject, at eye level. The researcher then explained the measurement protocol and the force task to the subject. Before starting the measurements, ankle torque was normalized by measuring the baseline torque with the ankle joint relaxed. The subject was then asked to perform maximum plantarflexion for a few seconds, after which they could relax the ankle again. The acquired baseline torque and MVC torque were then used to normalize the measurements during the trial and to normalize the torque for the data analysis.

Table 3: List of trials with their respective experimental conditions.

Trial name	Target %MVC	RoFD	Rest time
15-5	15%	5% MVC/s	8 s
30-5	30%	5% MVC/s	15 s
45-5	45%	5% MVC/s	22 s
15-20	15%	20% MVC/s	8 s
30-20	30%	20% MVC/s	15 s
45-20	45%	20% MVC/s	22 s

Figure 3: Setup showing the subject seated in the chair in front of the dynamometer with the right foot fixed on the footplate and the screen in front of the subject.

For the control group, the control interface was shown on screen, for the gamified group, the gamified interface was shown on screen. The differences between the interfaces can be seen in Figure 4. See also section 3.1. Subjects were familiarized with the interface by asking them to press against the footplate so they could see the cursor of the interface on the screen in front of them move up. They were then told that they had to follow the centre of the target line as best as they could, and that each trial consists of five repetitions. Subjects were informed that in total there are six trials, with three different target %MVC and two different RoFD. They were also informed of the different resting times between repetitions and trials. Subjects were presented the trials in randomized order and were informed before each trial of the target %MVC and RoFD. After the final trial, subjects were asked to fill in the modified GEQ (see also Appendix 2) and answer the two open questions mentioned in section 2.7. Finally, subjects were thanked for their participation and received a chocolate bar of their choice.

2.6 Measurement of performance

Torque was measured using Biodex Medical Systems' System 4 Pro dynamometer [76] at a sampling frequency of 1 kHz and imported using MathWorks' Simulink [61]. The raw data was stored for post-analysis. For the data analysis, the raw torque data was filtered using a second order 20 Hz low-pass Butterworth filter. The data was then normalized with regards to MVC. The latency of the system was determined using the cross correlation of the target and the measured torque. This latency was used to align the signals. Root-mean-square error (RMSE) and mean absolute error (MAE) between the target and measured torque was used as a measurement of performance. While related, both RMSE and MAE were included, because RMSE is more sensitive to outliers than MAE. MAE, conversely, shows a better estimate of the true magnitude of the error. These two metrics together can give a better insight in the nature of the error. RMSE and MAE were calculated separately for each segment of the trapezoidal isometric force tracking task, as illustrated in Figure 2. The feedback interface showed a line with a tolerance area of 6.25% MVC, i.e. 3.125% to either side. RMSE and MAE were calculated while accounting for the tolerance area, by first subtracting 3.125 from the absolute error, and setting all negative values to zero. Coefficient of variation (CV) was measured for each subject of a group, for each repetition, but separated by trial, as a measure of dispersion from the mean torque. Because the torque profile is a trapezoid, the target was first subtracted from the measured torque.

2.7 Measurement of engagement

Subject engagement was measured using a modified version of the Core Module of the Game Experience Questionnaire (GEQ) [77]. The GEQ is a self-reported questionnaire that was administered immediately after the test finished. The modified version of the GEQ can be found in Appendix 2. The GEQ was chosen for this because its factors are related to the definition of engagement, given in the introduction. The questionnaire is freely available and is commonly used to assess experiences with (gamified) software [78], [79]. The GEQ originally assesses game experience on seven items, though validation studies show the originally used 7-factor design is not stable [78], [79]. Johnson *et al.* [78] propose a 5-factor structure instead, combining Negative Affect, Challenge, and Tension/Annoyance into a single factor, and dropping some items that either load too much on two factors, or do not load sufficiently on any factor. Of these five, four factors were used: Flow, Competence, Positive Affect, Negativity. The Immersion factor was removed as it was not relevant for this type of software. The only other adaptation was changing item 21 "I was fast at reaching the game's targets" to "I was good at reaching the game's targets", as the set speed of the protocol does not allow for the subject to progress faster or slower. This new wording is in line with the Competence factor it represents. The modified GEQ consisted of 27 items, which subjects responded to on a 5-point Likert scale: 'Not at all', 'Slightly', 'Moderately', 'Fairly', or 'Extremely' [71]. These answers respectively corresponded to the numbers 0 through 4.

2.8 Learning effect

To determine whether the gamified interface results in a higher learning rate, one question was added at the end of the GEQ questionnaire as a self-reported assessment of learning effect: "28. I felt it got easier over time." This item used the same 5-point Likert scale as the other questions. In addition, the median RMSE was taken from each repetition for each group, so that a potential change in performance over the course of several repetitions could be shown.

2.9 Data analysis

To assess a potential performance difference between the gamified group and the control group, RMSE, MAE and CV were compared between groups. This was done on a per trial, per repetition basis, using the mean of the group. For RMSE and MAE, this was done separately for each of the four trial segments. For CV, this was done considering entire repetitions. The mean value was then taken from each subject's repetitions, separately for each segment in case of RMSE and MAE. Statistical analysis was done using the Shapiro-Wilk test to test for normality, and in case of normally or not normally distributed data either an unpaired T-test or a Mann-Whitney U test (also known as a two-sided Wilcoxon rank sum test) was performed, respectively.

To assess whether there was an engagement difference between the gamified group and the control group, the responses on the GEQ were compared. This was done on a per item basis, as well as per factor. In case of comparing factors, the median of all items of the factor was taken for each subject, due to the ordinal nature of the data. Statistical analysis was done using a Mann-Whitney U-test, because the outcomes of the GEQ are ordinal. In addition, group mean rank was calculated for each item and factor.

3. Results

This section consists of two parts. The first part consists of an evaluation of the implementation and final design of the interface, as well as signal delay measurements and the qualitative analysis of the results from the user responses using the Content Analysis method. The second part consists of the results of the performance and engagement measurements, as well as the results of the learning effect measurements.

3.1 Interface design

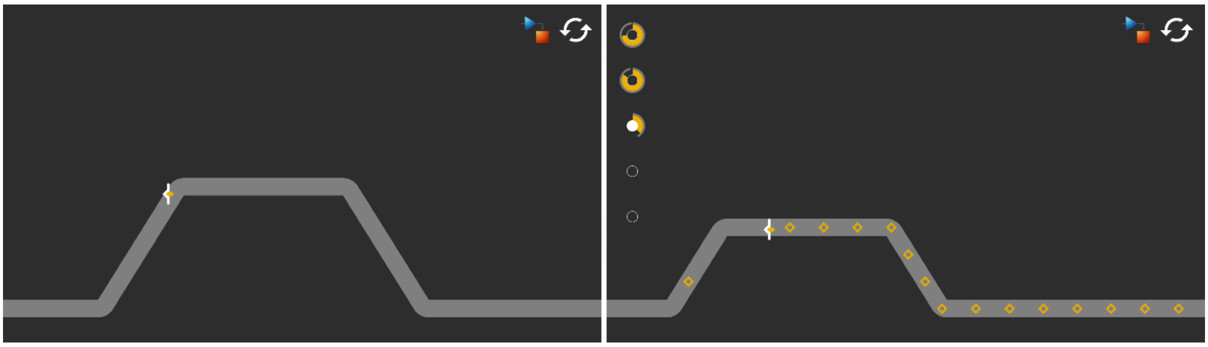
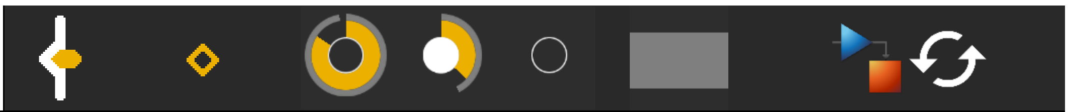


Figure 4: Comparison of the interface for the Control group (left) and the Gamified group (right). Target MVC corresponded with the height of the target. The light grey area shows the target and the tolerance. The x-position of the white and yellow cursor was constant; the y-position corresponded to the measured torque.

Figure 4 shows the design of the interface for the control group and the gamified group. The left image shows a trial with a target %MVC of 45% and a RoFD of 20 %MVC/s; the right image shows a trial with a target %MVC of 30% and a RoFD of 20 %MVC/s. The relation between the y-position of the cursor and the measured torque was equal across all trials. This made the required force to reach the target more predictable across trials. During a trial, the target line shifted towards the left side of the screen, comparable to how side-scrolling games function. This is illustrated in Figure 5. This interface took inspiration from existing interfaces, most notably Technogym's Biostrength [80] and Flappy Bird [81], for the side-scrolling design. The scoring system and style of motivational messages were inspired by Ubisoft's Just Dance [82]. Table 4 lists all the different interface items and shows in which version of the UI they were present. Apart from the user interface buttons, which were only operated by the researcher, each of these elements is discussed in the subsections below.

Table 4: Elements in the Control version (C) and Gamified version (G) of the interface.

Element					
Name	Cursor	Collectible	Score circles	Line	User interface buttons
Shown in	C, G	G	G	C, G	C, G

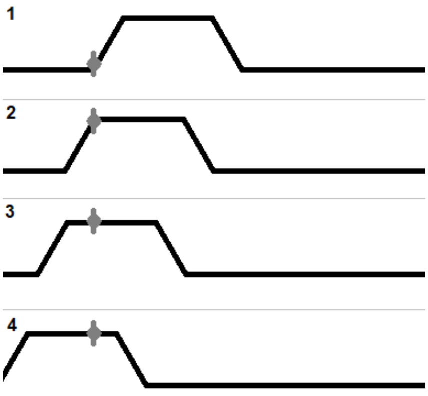


Figure 5: Illustration of the side-scrolling mechanic. As time passed (indicated by the number), the target line (black) slided to the left, while the cursor (grey) remained in the same x-position. Subjects needed to keep the cursor in the centre of the line.

3.1.1 Cursor

The cursor contrasted with the background and the target line. Its shape helped the user to pinpoint the centre of the cursor. The tip was the same colour as the collectibles. In the gamified version, this strengthened the idea that the collectibles needed to be collected with the centre of the cursor. The cursor's x-position was more to the left of the screen. This allowed the user to see changes in the target line more in advance.

3.1.2 Collectible

In the gamified version, diamond-shaped collectibles were used to guide the user to the centre of the target line. When the centre of the cursor collided with these collectibles, the collectible disappeared, and the yellow circle corresponding to the current repetition filled up more. Because of the prediction that users tend to have larger errors in the ramp sections, the ramps had a higher density of collectibles. This was done to persuade users to stay on target especially in these segments. Spacing of the collectibles was one per second on horizontal segments, and two per second on ramp segments. In some subjects in the gamified group, nearly missing collectibles resulted in overcompensation behaviour. An example of this can be seen in Figure 6.

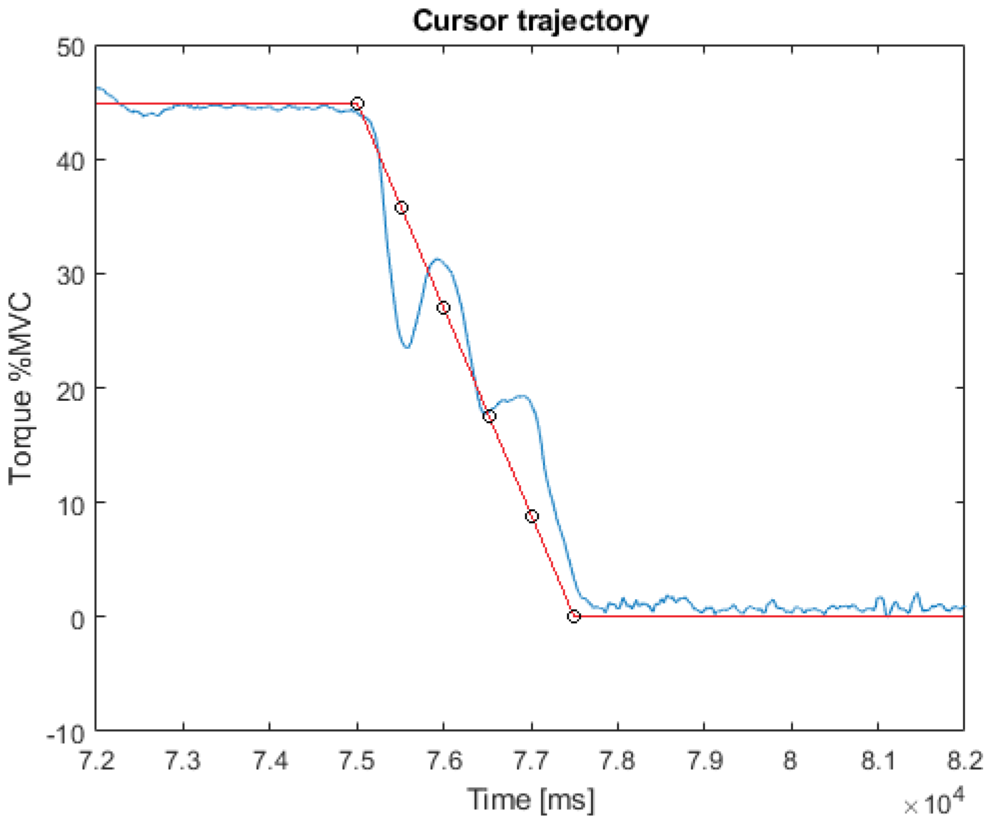


Figure 6: Cursor trajectory in the feedback interface is shown in blue, target line is shown in red, collectibles are marked with black circles. This image was reconstructed using the measured Torque but uses the same processing methods as was used with the data sent to the feedback interface. The data was normalized with respect to MVC and filtered with a 100 ms moving average window. This data is from subject 16's second repetition of the 45% MVC trial with RoFD of 20 %MVC/s. Subject 8 was part of the gamified group.

3.1.3 Score circles

In the gamified version, positioned on the left side of the screen was a number of circles equal to the number of repetitions. Initially, these circles were empty. The circle of the current repetition was filled with white, and after the repetition the circle became empty again. This gave the user an indication of progression, as well as an indication of the current and total number of repetitions. If the cursor remained within the tolerance boundaries, the grey outer circle filled up. If collectibles were collected, the yellow inner circle filled up. A corresponding circle could be filled up fully if the user stayed on the target line for the entire repetition or collected all collectibles. This gave the user of the gamified version an indication of the amount of time the cursor remained between the tolerance boundaries as a percentage of the maximum, and the percentage of collectibles in each past repetition that had been collected. At the end of a repetition, a motivational message was shown on screen. The wording of the message was based on the number of collectibles collected and the time the cursor remained between the tolerance boundaries, with higher aggregated score resulting in a more positive message, as can be seen in Table 6. The aggregated score was the average of the percentage of time the cursor remained between the tolerance boundaries and the percentage of collectibles collected in the repetition. Due to a rounding error, users were not able to get the motivational message corresponding to an aggregated score of 100% in the test. The circles were positioned on the left of the screen, so that they did not obstruct vision of the target line. Other than gradually filling up, these circles had no animation and did not emit particle effects. This was done to not distract the user from the goal.

Table 5: Motivational messages and their required aggregated scores.

Aggregated score	Message
100%	Flawless!
≥87.5%	Amazing!
≥75%	Wonderful!
≥50%	Great!
≥25%	Not bad!
<25%	Keep it up!

3.1.4 Line

The target torque was displayed using the feedback interface, which was displayed on a monitor in front of the subject. The target torque was displayed with a tolerance boundary of 6.25% MVC (3.125% in both directions). This made the target torque more visible at a distance and reduced the chance of overcompensation errors. Additionally, this boundary purveyed an indication of allowed deviation from the target, to which a point scoring system for the gamified interface was attached.

Table 6: Presence of gamification elements in the two different versions of the interface.

Element	Control	Gamified
Challenge	✓	✓
Collection		✓
Feedback	✓	✓
Level	✓	✓
Point		✓
Single-Player	✓	✓

3.2 Gamification elements

As can be seen in Table 5, both the Control and Gamified version of the interface contained gamification elements. Differences between the two interfaces are in the Collection, Feedback, and Point elements. The Control version of the interface had no collectibles and the only feedback given to the user was a visualization of the output and target torque. No points were awarded in the Control version of the interface. Conversely, in the Gamified version of the interface, collectibles could be collected, which awarded the user points.

3.3 Interface implementation

3.3.1 Normalization and filtering

The output of the dynamometer is a noisy signal. The noise amplitude exceeded the tolerance boundary by a factor of five. It was thus necessary to filter the torque before displaying it in the interface. Despite this, the cursor trajectory showed jittery behaviour as can be seen in Figure 7. This was commented on by four subjects. The data from Figure 7 was taken from one of the subjects who commented on the jitteriness.

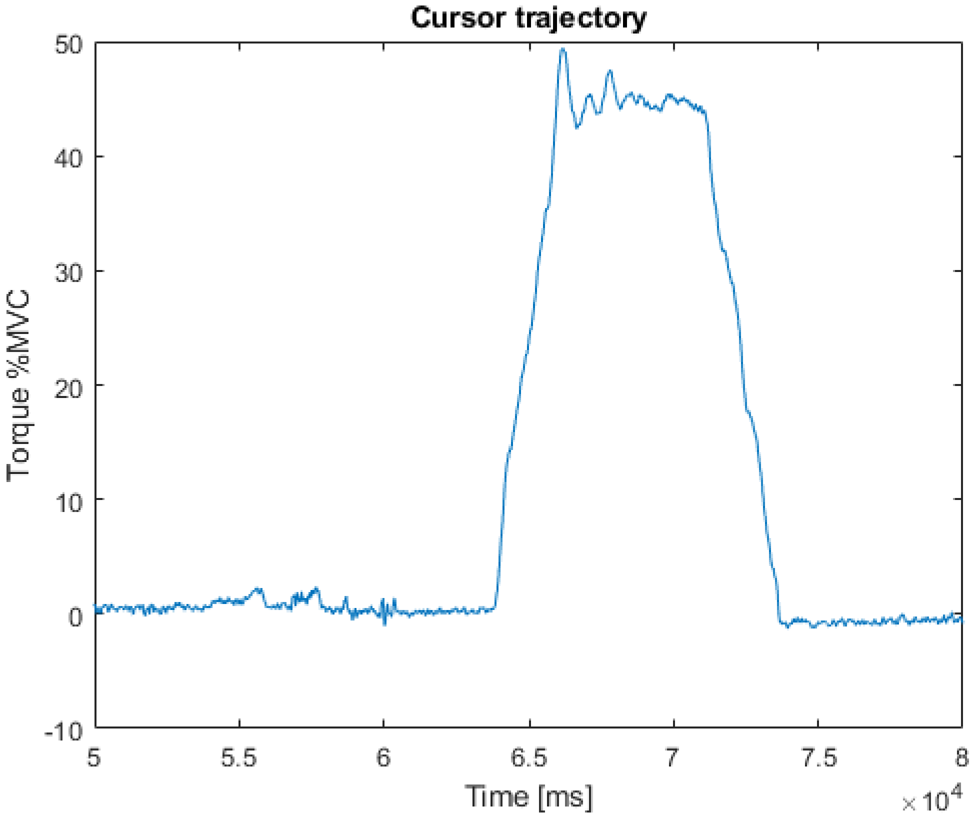


Figure 7: Cursor trajectory in the feedback interface. This image was reconstructed using the measured Torque but uses the same processing methods as was done with the data sent to the feedback interface. The data was normalized with respect to MVC and filtered with a 100 ms moving average window. This data is from subject 8's second repetition of the 45% MVC trial with RoFD of 20 %MVC/s. Subject 8 was part of the control group.

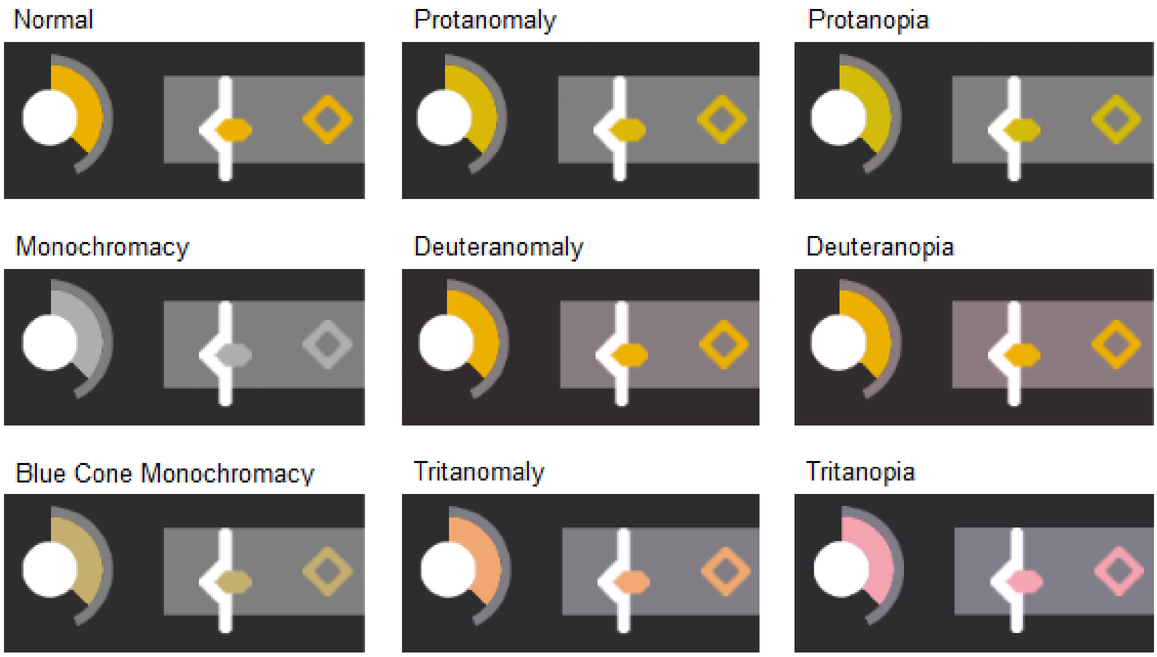


Figure 8: Comparison between normal trichromatic view and eight forms of colour vision deficiency, simulated using Coblis [83].

3.3.2 Accessibility results

Coblis [83] was used to test whether the different interface elements are discernible for people with colour vision deficiency. A comparison of eight versions of colour vision deficiency to the trichromatic view can be seen in Figure 8. Apart from the start and stop messages, there was no language involved in the interface. No sounds were used in the interface, making it equally usable for hearing impaired users.

Table 7: Median latency with Q1 and Q3 intervals [Q1, Q3] in ms between the target and the torque signal. Median latency is given for all trials, only 5% MVC trials, and only 20% MVC trials. Latency was compared between the Control and Gamified groups. A positive latency means the measured torque lags behind the target. p-values were determined by the Wilcoxon Rank sum test. One, two, or three asterisks (*) denote a p-value below 0.05, 0.01, 0.001, respectively. The p-values in the bottom row compare the medians of 5% MVC and 20% MVC data of the same group.

TRIAL	CONTROL MEDIAN	GAMIFIED MEDIAN	p	
All	**139** [82, 172]	**196** [152, 270]	<.001	***
5% MVC	**142** [81.75, 197.25]	**248** [156, 279.75]	<.001	***
20% MVC	**125** [78.75, 158.5]	**193** [151.75, 229.75]	<.001	***
p	.214	.105		

3.3.3 Latency measurement

The latency between the target and measured torque, as determined using the cross-correlation of the two signals, can be seen in Table 7. The control group showed a significantly smaller (p<.001) latency between the signal and the target, compared to the gamified group. This was also the case when only 5% MVC trials or only 20% MVC trials were compared to those of the other group. Within a group, there was no significantly different delay when comparing between their 5% MVC trials and their 20% MVC trials.

3.4 Content Analysis

To code each category, the words that were counted in each of the three category pairs are shown in Table 8. Twenty responses were given in Dutch and two in English. Words used for the coding of each category have been translated to English for the sake of clarity. The original responses can be seen in Appendix 3.

Table 8: Categories used in the Content Analysis. The word pool shows words that were counted, including conjugations of those words (e.g. fun, funny).

Category	Word pool
Fun	Fun, fine, nice, good
Unfun	Unfun, annoying, difficult
Engaging	Engaging, focused, challenging, not distracted, interesting, drive to perform
Boring	Boring, monotonous, simple, distracted, barebone
Clear	Clear, intuitive
Confusing	Confusing, unclear, uncertainty

For each subject, the answers to the two open questions (Q1: "What did you think of the test?" and Q2: "What did you think of the interface?") were coded using the word pools. The results are shown in Table 9. In certain cases, both categories of a pair were scored in a single response. This occured, for example, when a subject described a certain aspect of the interface as fun, but another aspect as unfun. Subtracting the result of the negative category from the result of the positive category results in the aggregated category pair score. The aggregate serves as a score for the category pair. Positive scores indicate that words from the positive category occur more often, while negative scores indicate that words from the negative category occur more often. Notable aggregate score differences (≥3) between groups were found in the Fun-Unfun pair, for both Q1 and Q2, and in the Engaging-Boring pair for Q1. Responses contained words from the Fun category more often in the Gamified group than in the Control group and contained words from the Unfun category more often in the Control group, but this difference was larger for Q1 than for Q2. Responses contained words from the Boring category more often in the Gamified group, but only for Q1. Responses contained words from the Clear category more often in the Control group, but only for Q2.

Table 9: Content Analysis results for each category, for each question, compared between the two groups. 'C' denotes Control group, 'G' denotes Gamified group. 'Occurrences' columns show the frequency of words from a category occurring in the responses for that group. 'Number of subjects' columns show number of subjects whose answers contained words from the corresponding category. The 'Aggregate' rows subtract the result of the negative category from the result of the positive category.

Q1: What did you think of the test?

	C Occurrences	G Occurrences	C Number of subjects	G Number of subjects
Fun	14	17	8	11
Unfun	5	2	3	1
Aggregate	9	15	5	10
Engaging	5	4	4	3
Boring	0	3	0	2
Aggregate	5	1	4	1
Clear	1	0	1	0
Confusing	0	0	0	0
Aggregate	1	0	1	0

Q2: What did you think of the interface?

	C Occurrences	G Occurrences	C Number of subjects	G Number of subjects
Fun	5	8	4	6
Unfun	1	0	1	0
Aggregate	4	8	3	6
Engaging	3	1	2	1
Boring	9	7	6	6
Aggregate	-6	-6	-4	-5
Clear	8	4	6	3
Confusing	2	1	2	1
Aggregate	6	3	4	2

3.5 Measurement of performance

Figure 9 shows that force-tracking errors seemed similar between the control and gamified groups. Force-tracking errors seemed larger around points where a slope transitioned into a horizontal segment, or vice versa. Force-tracking errors also seemed larger during slope sections than during the plateau section. Measured torque for the other conditions can be seen in Appendix 4.

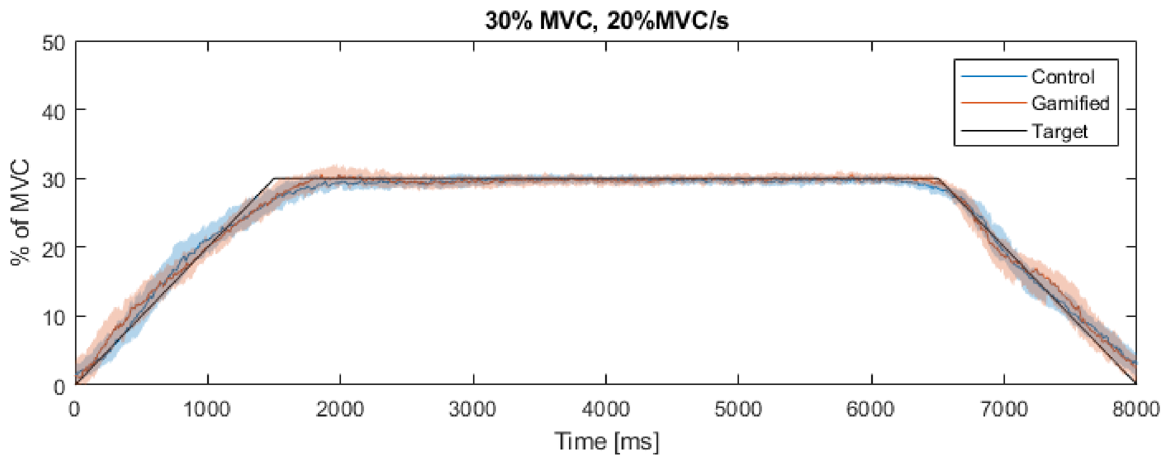


Figure 9: Median and IQR of measured torque on all repetitions of the 30% MVC at 20% MVC/s trial. The torque was filtered and normalized with respect to MVC, as well as shifted to eliminate latency. Coloured lines show median and shaded area shows IQR. Only the Ramp Up, Plateau, and Ramp Down sections are shown. Results of other trials can be found in Appendix 4.

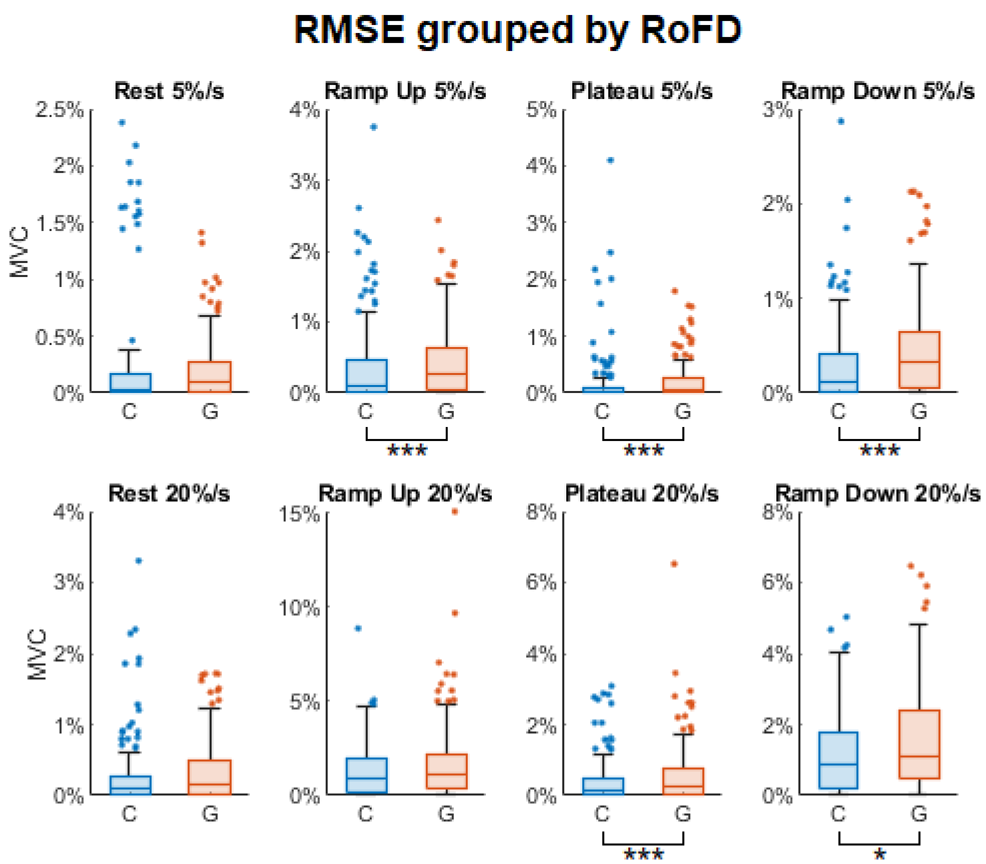


Figure 10: Boxplots of the RMSE in terms of percentage of MVC of the four trial segments, for the two RoFD conditions. C is the Control group; G is the Gamified group. The line inside the box shows the median RMSE. The bottom and top edges of the box show Q1 and Q3, respectively. Whiskers denote maximum and minimum non-outlier values. Outliers are shown as dots and randomly displaced horizontally for visibility. Outliers are defined as values more than 1.5 IQR away from top or bottom of the box. p-values are determined by the Wilcoxon Rank sum test. One, two, or three asterisks (*) denote a p-value below 0.05, 0.01, 0.001, respectively.

3.5.1 Root mean square error

Figure 10 shows the median and interquartile ranges, as well as outliers, of the RMSE between the measured torque and the target, including the 3.125% tolerance area. Each box plot shows the RMSE grouped by RoFD. Each box plot contains five repetitions of three target %MVC's from eleven subjects in the given group (165 measurements). Differences in RMSE between the control and gamified group were significant in both Plateau and Ramp Down sections, but the RMSE in the Ramp Up section only differed significantly between the groups for the 5% MVC/s condition. For all conditions where a significant difference was found, the median RMSE was larger in the gamified group. Errors seemed larger in Ramp (Up and Down) sections than in horizontal (Rest and Plateau) sections. Errors also seemed larger for the 20% MVC/s condition than for the 5% MVC/s condition.

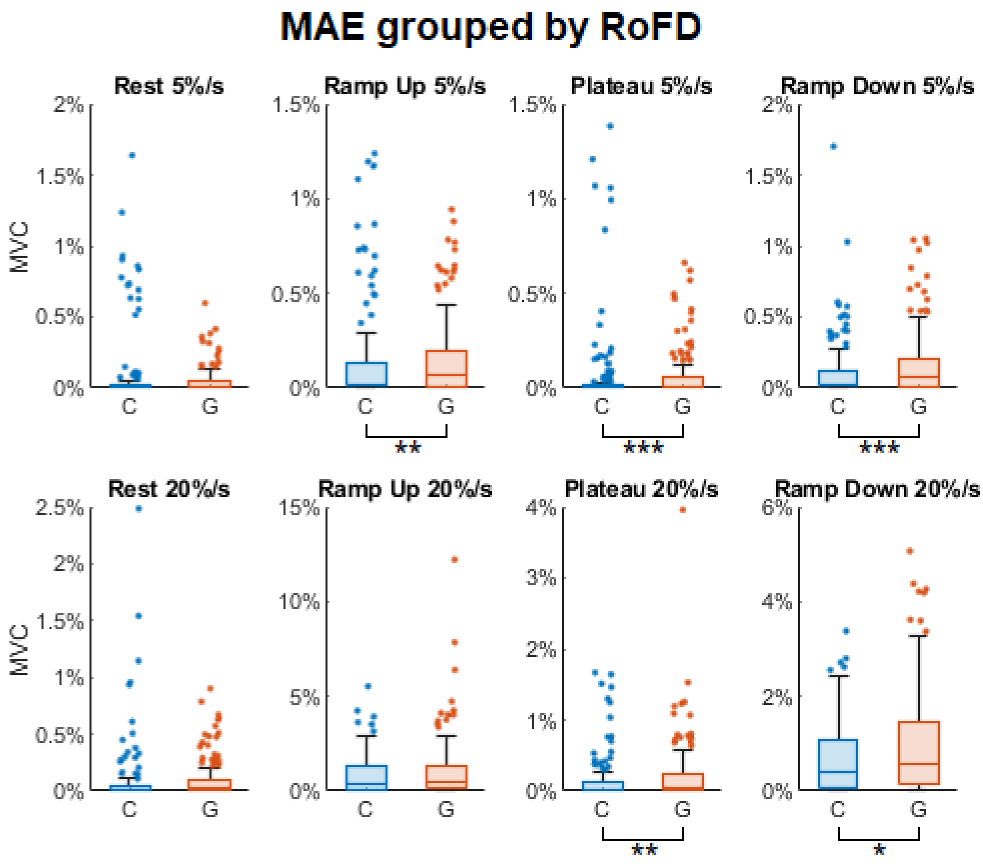


Figure 11: Boxplots of the MAE in terms of percentage of MVC of the four trial segments, for the two RoFD conditions. C is the Control group; G is the Gamified group. The line inside the box shows the median MAE. The bottom and top edges of the box show Q1 and Q3, respectively. Whiskers denote maximum and minimum non-outlier values. Outliers are shown as dots and randomly displaced horizontally for visibility. Outliers are defined as values more than 1.5 IQR away from top or bottom of the box. p-values are determined by the Wilcoxon Rank sum test. One, two, or three asterisks (*) denote a p-value below 0.05, 0.01, 0.001, respectively.

3.5.2 Mean absolute error

Figure 11 shows the median and interquartile ranges, as well as outliers, of the MAE between the measured torque and the target, including the 3.125% tolerance area. Each box plot shows the

MAE grouped by RoFD. Each box plot contains five repetitions of three target %MVC's from eleven subjects in the given group (165 measurements). Differences in MAE between the control and gamified group were significant in both Plateau and Ramp Down sections, but the MAE in the Ramp Up section only differed significantly between the groups for the 5% MVC/s condition. For all conditions where a significant difference was found, the median MAE was larger in the gamified group. Errors seemed larger in Ramp (Up and Down) sections than in horizontal (Rest and Plateau) sections. Errors also seemed larger for the 20% MVC/s condition than for the 5% MVC/s condition.

3.5.3 Coefficient of variation

Figure 12 shows the CV for each trial. For each repetition by each subject in the group, the CV was calculated. Each box pot shows the CV grouped by trial. Each box plot contains five repetitions of each subject in the given group (55 measurements). CV considers whole repetitions instead of segments, so that it shows a measure of torque output variability over an entire trial. In all the 5% MVC/s trials, CV was significantly closer to zero for the gamified group. In the 20% MVC trials, the differences were not significant.

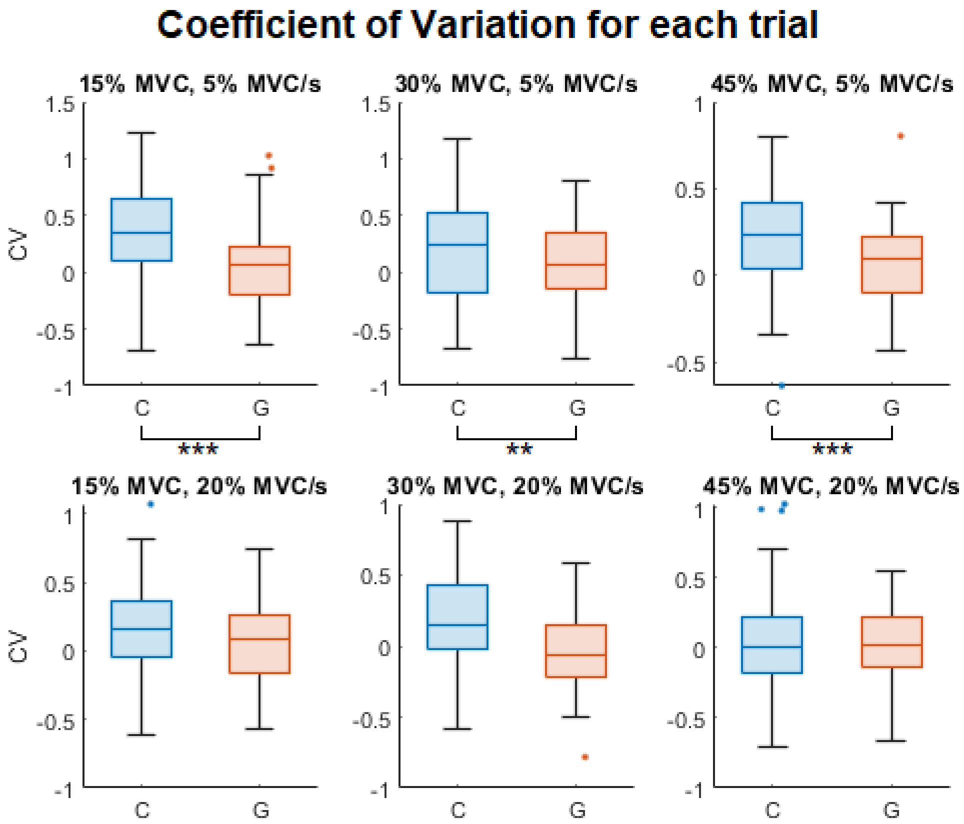


Figure 12: Boxplots of CV for each trial. C is the Control group; G is the Gamified group. The line inside the box shows the median CV. The bottom and top edges of the box show Q1 and Q3, respectively. Whiskers denote maximum and minimum non-outlier values. Outliers are shown as dots and randomly displaced horizontally for visibility. Outliers are defined as values more than 1.5 IQR away from top or bottom of the box. p-values are determined by the Wilcoxon Rank sum test. One, two, or three asterisks (*) denote a p-value below 0.05, 0.01, 0.001, respectively.

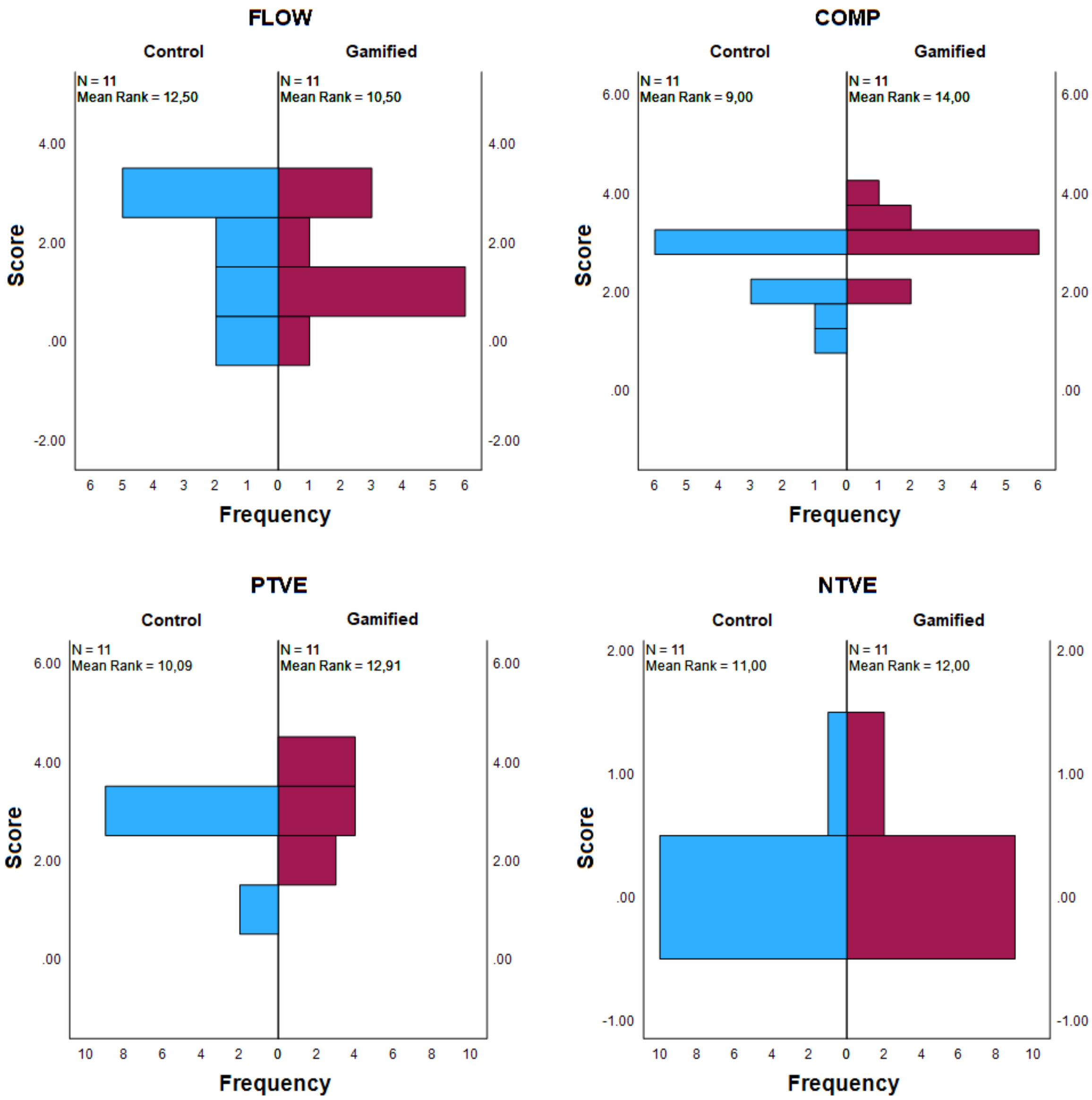


Figure 13: Histograms and mean rank of each GEQ factor, compared between the two groups. Frequency is the number of subjects in the given group that gave the indicated score on the corresponding factor. Score corresponds to the responses (0-4) on the 5-point Likert scale (e.g. on the GEQ items that constitute the Positivity factor, eight subjects in the Control group had a median score of 3 which corresponds to "Fairly" on the Likert scale).

3.6 Measurement of engagement

Figure 13 shows the histograms of the scores for each factor of the GEQ, compared between the groups. Table 10 shows the mean ranks of each separate item and factor from the GEQ. A higher Mean Rank indicates higher agreement to the statement (or factor) for that group. None of these items or factors showed a significantly different score between groups. The Competence factor showed the largest difference between groups (C = 9.00; T = 14.00) and the Negativity factor showed the smallest difference between groups (C = 11.00; T = 12.00). The item with the largest mean rank difference was item C09: “I felt competent” (C = 8.82; T = 14.18). All factors except for Flow showed higher mean ranks for the Gamified group.

Table 10: Mean rank and Mann-Whitney U test significance of comparing the median responses of GEQ items and factors between Control and Gamified groups. Item numbers correspond to the items on the GEQ, see also Appendix 2: Modified Game Experience Questionnaire.

	Mean rank		
	Control	Gamified	p
FLOW	12.50	10.50	.478
F04	11.32	11.68	.898
F11	12.27	10.73	.606
F21	12.41	10.59	.519
F23	11.73	11.27	.898
F25	11.55	11.45	1.000
COMP	9.00	14.00	.076
C02	9.36	13.64	.133
C09	8.82	14.18	.056
C13	10.14	12.86	.332
C17	9.50	13.50	.151
NTVE	11.00	12.00	.748
N06	11.50	11.50	1.000
N18	11.05	11.95	.748
N19	10.64	12.36	.562
N20	11.00	12.00	.748
N24	11.50	11.50	1.000
N26	12.50	10.50	.478
N27	12.36	10.64	.562
PTVE	10.09	12.91	.332
P01	9.86	13.14	.243
P03	10.86	12.14	.652
P05	10.50	12.50	.478
P12	10.59	12.41	.519
P16	10.50	12.50	.478

3.7 Learning effect

Figure 14 shows the histogram of responses to the statement ‘I felt it got easier over time’. While the Mean Rank was higher for the Gamified group, these results were not statistically significant. In contrast to the GEQ factors, the learning effect consists of a single item. Figure 15 shows that in Ramp sections, errors tended to decrease as the repetition number increases. The IQR got closer to the median as the repetition number increased as well. No improvement was found in the horizontal sections over time. While both RMSE and MAE seemed to show smaller errors for the Control group across almost all datapoints, no difference was found in the extent to which these errors decreased over time between the Control and Gamified group.

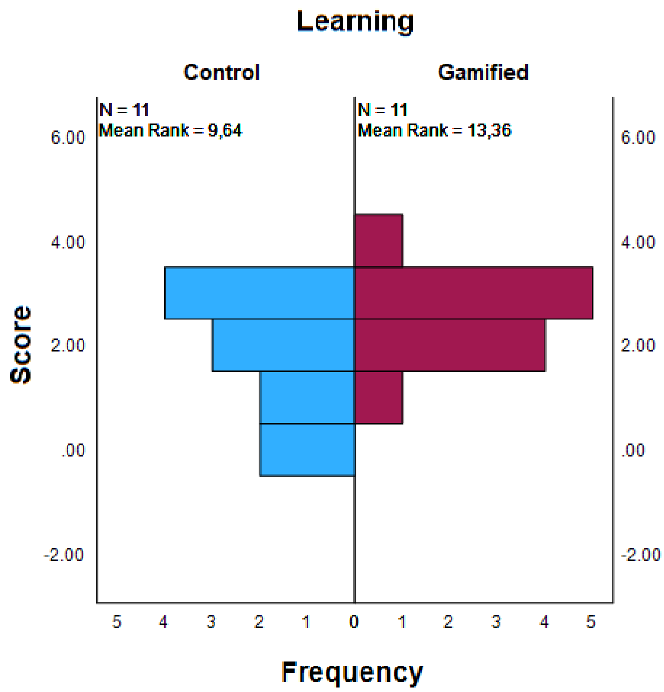


Figure 14: Histogram of the responses to the question 'I felt it got easier over time' from the Control and Gamified (Test) groups. Scores correspond to the 5-point Likert scale used in the GEQ. Frequency is the number of subjects in the given group that had the indicated score. Score corresponds to the responses (0-4) on the 5-point Likert scale.

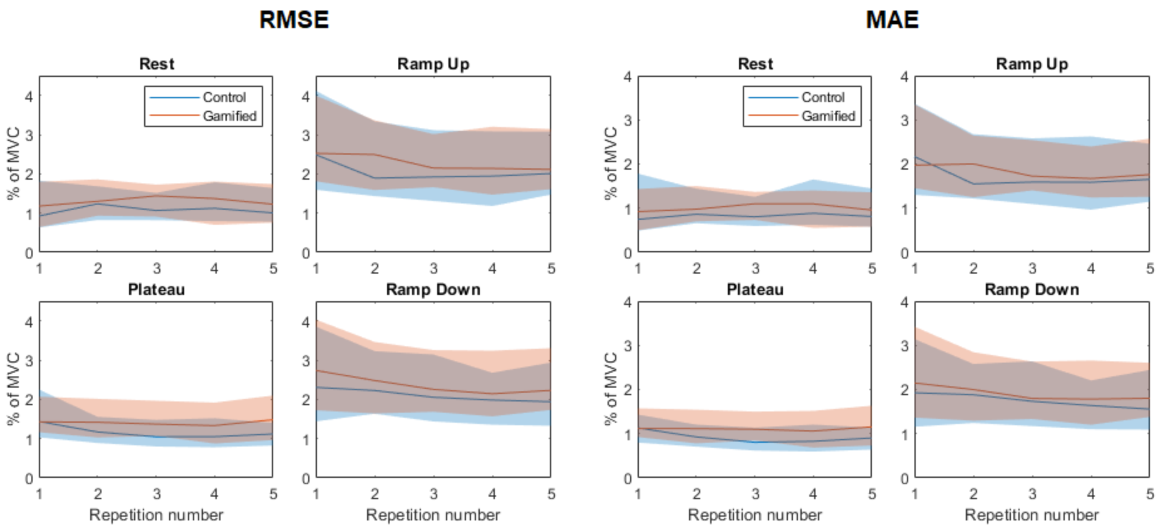


Figure 15: Median and IQR of RMSE and MAE in terms of percentage of MVC of the four trial segments, for all conditions combined. The line shows the median RMSE for each repetition for each group. IQR is shown by the shaded area.

4. Discussion

The goal of this study was to assess whether gamification improves subject engagement, and subsequently whether improved engagement leads to better performance on a motor task. An improvement in engagement and performance in the group using the gamification interface would indicate that the application of gamification elements improves the quality of the measured parameters. This would be relevant in the context of EMG measurement and would lead to improved MSK modelling.

In this study an interface was designed and developed for the measurement of torque during a trapezoidal ramp motor task. The feedback interface had two versions, one with and one without additional gamification elements. This was done to determine whether gamification elements increase a subject's engagement, and whether engagement influences performance on the motor task. A single-blind study was performed to test the hypotheses that gamification increases engagement, and that increased engagement improves performance. The gamified group showed a larger RMSE and MAE in both Plateau and Ramp Down sections, as well as in the 5% MVC/s Ramp Up sections. CV was significantly lower for the gamified group in the 5% MVC/s trials. Finally, more words from the Fun category were scored in the qualitative data from the gamified group. Other metrics showed no significant differences between the two groups.

The gamified group showed higher RMSE and MAE, which is likely influenced by overcompensation to avoid missing a collectible, as can be seen in Figure 5. While this behaviour increases a subject's score in the interface, it is detrimental for extracting MU activity. Contraction velocity influences MU recruitment strategies [84]. Therefore, sudden overcompensation can lead to unwanted rate coding or recruitment order of MU's. Future iterations of the gamified interface should avoid rewarding overcompensation.

The moving average window was chosen as a filtering technique because low-pass filtering in Simulink introduced a noticeable delay between measurements and the display of the cursor on the screen. As shown in Figure 7, cursor trajectory in the feedback interface shows jitter, especially in the Rest segments. Jitter was reported during the open questions by four subjects, of which three from the control group. In two subjects in the control group, their answers to the open questions contained words from the Unfun and/or Uncertainty categories in the Content Analysis, specifically related to the jitter. It is not known how this experience influences their results on the GEQ, or whether more subjects' experiences were influenced by this phenomenon. Biquadratic filtering was not tested but could be a solution to this problem [85]. It provides a method to filter digital signals using only a handful of timesteps, instead of the 100 used in this study for the moving average window. Recommendation for future improvement of the system is to rework the filtering of data that is sent to the Unity interface.

Table 7 shows that the latency between the target and the measured torque was significantly larger for the gamified group. This difference is also significant when comparing 5% or 20% MVC/s trials between groups, but not when comparing 5% MVC/s trials with 20% MVC/s trials within a group. This suggests that the latency difference is not caused by the test protocol. Removing the latency in post processing using the cross-correlation coefficient thus was justifiable. In doing so however, any potential delay caused by the subject's performance was removed as well, which could influence the performance results. With the current measurement results, it was not possible to differentiate between human error and latency due to software or hardware. Future versions of the interface should add a synchronization trigger between the target and the measurements, so that the latency caused by the system can be accounted for without removing human errors as well.

Table 8 shows that the gamified group scored better on the Fun-Unfun category in the Content Analysis regarding the interface (Q2: "What did you think of the interface?"). Other category pairs for Q2 did not show the same differences between the groups. This indicates that the gamified group described the interface as more fun, compared to the control group. While the results of the GEQ are not statistically significant, a tendency towards higher Positive Affect can be seen in the gamified group, which would support the results from the Content Analysis.

As can be seen in Figures 10 and 11, a significant increase in RMSE and MAE was measured in the gamified group in the Plateau and Ramp Down segments of 5% and 20% MVC/s conditions, as well as in the Ramp Up segments of the 5% MVC/s conditions. However, Figure 12 shows that the CV was lower for the Gamified group in the 5% MVC/s trials. Tracking errors, as described by RMSE and MAE influence the results of EMG decomposition. In isometric contraction of the *musculus tibialis anterior*, a recruitment of 50% of the MU pool results in about 10% of maximal voluntary contraction (MVC) [40]. Especially for smaller contraction forces (<20% MVC), errors between the produced and target force result in a relatively large discrepancy in MU activation [40]. In addition, a contraction force larger than the target force leads to the recruitment of additional motor units, potentially resulting in different MU firing patterns during the rest of the task due to activation hysteresis caused by the de-recruitment thresholds [86]. Hence, these results suggest that the used gamified environment lowers motor performance from the standpoint of MU activation analysis. From a neuromechanical standpoint, these findings contrast with a previous study, which investigated the performance of test subjects between a gamified and control group, in this case in robot-assisted upper extremity training [87]. That study showed that a gamified environment led to higher accuracy, higher smoothness, but slower speed in stroke patients, even in the first training session. Whether this is due to the different test population, motor task, or performance metrics cannot be concluded from this study alone. Gamification has also been shown to positively influence engagement and performance in the context of rehabilitation [88] and motor coordination [89]. However, the conducted studies concern long-term improvement whereas this study focused on the immediate difference between the two groups. To the best of author's knowledge, no other systematic study compared the immediate effects of gamification elements on motor task performance. Future research should further explore these aspects, as gamification may be beneficial for the calibration of personalized MSK models.

As shown in section 3.6, no significant difference in engagement was found between groups. This indicates that the addition of the score trackers, motivational messages and collectibles did not influence the subject's engagement, or that it was not captured by the GEQ. The measurement of engagement is discussed later in this section. The use of motivational messages as a reward for good performance has been shown to increase task efficacy, even though positive feedback does not influence self-report behaviour [90], [91]. This can explain not finding a significant difference in the Positive Affect factor of the GEQ between the gamified and control groups. Conversely, a study done on the impact of positive feedback on behaviour in the context of customer service shows that positive feedback lowers performance of the subject [92]. Studies thus show different results regarding the influence of positive feedback on performance in different contexts. However, in this study, the control group shows significantly lower RMSE and MAE errors, which could indicate that positive feedback has an adverse effect on performance on a motor task. These findings are not conclusive, because positive feedback is not the only variable between the two interfaces used in this study. Additionally, while RMSE and MAE were lower, CV was higher for the control group. To further test this, a study would need to test the motor performance of two groups on the same motor task, with the test group receiving positive feedback, and the control group not receiving positive feedback. The results could shine a light on the effect of positive feedback in the context of motor tasks.

As shown in section 3.6, the Competence factor of the GEQ showed the largest (though not significant) difference between the groups. However, as shown in section 3.5, the performance by the Gamified group was significantly lower than that of the control group. The 5-point Likert scale is a subjective scale, and answers tend to converge to the middle response [71]. Given the positive feedback in the gamified interface, this could explain the difference in self-reported Competence, as positive feedback might influence a subject's opinion on their performance [90], [91]. Conversely, the reduced feedback of the control interface could lead to results closer to the average score of the Likert scale (2 – Moderately), because subjects in the control group interpret their performance using only the tolerance boundary.

Also, as shown in section 3.6, the control group scored higher on the Flow factor of the GEQ, though this difference was not significant. A literature review shows that the impact of gamification on flow experience is inconsistent [93]. Another study shows that flow experience does not necessarily change with a more personalized set of gamification elements, that is suited to the subject's player type [94]. To increase flow experience, more research is needed to provide a more complete personalization of the system, and to determine which elements affect flow experience for which player types.

Finally, on the topic of engagement measurement, the GEQ measures the experience of a user regarding a game [77]. The reason the GEQ was used for this study, is because given the definition of engagement posed in the introduction, this measure of experience is closely related to engagement. Another reason for using the GEQ is because is free availability and its ease of use. Currently, it is inconclusive whether there was no difference in engagement between the two groups, or if there was a difference in engagement that the GEQ could not measure. Engagement is most commonly measured using self-report scales [95], but can also be measured using other methods, including interviews, Content Analysis, eye-tracking, and physiological sensors. In future research, is recommended to look for alternative options to measure engagement, to see whether the GEQ is sufficient for measuring engagement in this context.

Another potential reason for not finding a significant difference in the GEQ factors between the two interfaces could be that the interfaces were not different enough to result in a significant difference in GEQ results between them. The creation of a third environment, more akin to a full-blown video game was considered, but ultimately dropped due to time constraints. The addition of more gamification elements in future versions of the interface may provide a definitive answer. However, it remains important to test whether a potential difference in performance between groups is due to the added engagement of a more gamified environment, or simply due to a different way of visualizing the feedback. The larger the difference between the compared environments, the larger the risk of confounding errors.

As described in section 3.7, no significant differences were found between the gamified group and the control group in rate of learning. This was the case for both the measured and self-reported results. This indicates that the used gamified environment does not increase the rate of learning. This is in accordance with the previously mentioned study done by Ozgur *et al.* [87] which also shows no difference in learning rate between the control and gamified groups.

To summarize, this study did not find significant differences in engagement between groups. The latency and performance differences indicate that further work is needed to improve the gamified interface. While CV was lower for the gamified group, RMSE and MAE were higher. While the gamified group reported more words from the Fun category in the Content Analysis, this was not reflected in the results of the GEQ. Given the current results, it is not known whether there was no significant difference in engagement between groups, or if the GEQ did not capture it. Recommendations for expanding the measurement of engagement have been given. Possibilities for improvement of the interface have also been given, which may influence future results. Additionally, while this study addresses the impact of gamification on engagement and performance in the context of improving MSK modelling, no EMG was recorded. If a future version of the gamified interface succeeds in consistently positively influencing engagement and performance, it is recommended to perform a study to compare the quality and accuracy of parameters derived from EMG decomposition between groups, as those are ultimately needed for the improvement of MSK models.

5. Conclusion

In this part first the two research questions are answered, afterwards a general conclusion is given, which places the findings of this study into the perspective given into the introduction.

5.1 Does gamification improve engagement?

No difference in GEQ results was found between the two groups. None of the four factors (Flow, Competence, Positive Affect, Negativity) showed a significant difference between the control and gamified group. None of the GEQ items showed a significant difference between the control and gamified group, but a tendency towards higher Competence results for the gamified group could be seen. It is not known whether a difference in engagement was not found because there was no difference between the groups, of if the GEQ did not measure it. Content Analysis results of Q2: "What did you think of the interface?" show similarities with the results from the GEQ. Namely, in the Fun-Unfun category pair, the gamified group scores better compared to the control group, as seems to be the case with the Positive Affect factor.

5.2 Does engagement improve performance?

The used gamified interface increased RMSE and MAE but lowered CV, compared to the interface used by the control group. RMSE was lower for the control group in the Ramp Up section for the 5% MVC/s trials ($p < .05$), and in the Plateau and Ramp Down sections in both 5% MVC/s and 20% MVC/s trials ($p < .001$ and $p < 0.05$ respectively). Rest segments and Ramp Up at 20% MVC/s showed no significant differences between groups. Though significant in the mentioned cases, median error differences were smaller than 0.4% MVC. MAE was lower for the control group in the Plateau and Ramp Down sections in both 5% MVC/s and 20% MVC/s trials ($p < .001$ and $p < .05$ respectively). Rest and Ramp Up segments did not show a significant difference between groups. Though significant in the mentioned cases, median error differences were smaller than 0.3% MVC. CV was significantly lower for the gamified group for all 5% MVC/s trials ($p <.01$).

The used gamified interface did not improve the learning effect when compared to the interface used by the control group. The responses to the item 'I felt it got easier over time' did not show a significant difference between the groups. Comparing RMSE and MAE between repetitions did not show a difference in the decrease in error between the two groups in case of the Ramp Up and Ramp Down sections and did not show improvement in the Rest and Plateau sections for both the control and gamified groups.

5.3 Conclusion

This study did not find a consistent effect of gamification on engagement. Because no significant difference was found in engagement between the two interfaces, it could not be concluded that the differences in performance were due to a difference in engagement. In this study, the addition of gamification elements had a negative impact on error magnitude, which is significant in the Plateau and Ramp Down segments of the force task, and in the Ramp Up section only for the 5% MVC/s condition. This was the case for both RMSE and MAE. Conversely, CV was significantly lower for the gamified group in the 5% MVC/s trials, but not in the 20% MVC trials. With the recommendations given in the Discussion, potential improvements to the interface can be made. In addition, not finding a difference in engagement may have been caused by the relative similarity of the two interfaces, or due to the measurement of engagement using the GEQ. Due to the dearth of studies on the immediate effect of gamification on motor task performance, further research is recommended, because the addition of gamification elements has shown beneficial results in many different relevant contexts.

References

[1] C. J. Heckman and R. M. Enoka, “Motor Unit,” *Comprehensive Physiology*, vol. 2, no. 4, Oct. 2012, doi: https://doi.org/10.1002/cphy.c100087.

[2] G. Valli, P. Ritsche, A. Casolo, F. Negro, and Giuseppe De Vito, “Tutorial: Analysis of central and peripheral motor unit properties from decomposed High-Density surface EMG signals with openhdemg,” *Journal of electromyography and kinesiology*, vol. 74, pp. 102850–102850, Feb. 2024, doi: https://doi.org/10.1016/j.jelekin.2023.102850.

[3] N. J. Marshall *et al.*, “Flexible neural control of motor units,” *Nature Neuroscience*, vol. 25, no. 11, pp. 1492–1504, Oct. 2022, doi: https://doi.org/10.1038/s41593-022-01165-8.

[4] E. Henneman, “Relation between Size of Neurons and Their Susceptibility to Discharge,” *Science*, vol. 126, no. 3287, pp. 1345–1347, Dec. 1957, doi: https://doi.org/10.1126/science.126.3287.1345.

[5] World Health Organization, “Spinal cord injury,” *Who.int*, Apr. 16, 2024. https://www.who.int/news-room/fact-sheets/detail/spinal-cord-injury (accessed Jul. 31, 2024).

[6] WHO, “Over 1 in 3 people affected by neurological conditions, the leading cause of illness and disability worldwide,” *www.who.int*, Mar. 14, 2024. https://www.who.int/news/item/14-03-2024-over-1-in-3-people-affected-by-neurological-conditions--the-leading-cause-of-illness-and-disability-worldwide (accessed Jul. 31, 2024).

[7] S. M. A. A. Evers, J. N. Struijs, A. J. H. A. Ament, M. M. L. van Genugten, J. C. Jager, and G. A. M. van den Bos, “The disease impact, health care management and cost of stroke in the Netherlands | RIVM,” *www.rivm.nl*, 2002. https://www.rivm.nl/publicaties/disease-impact-health-care-management-and-cost-of-stroke-in-netherlands (accessed Jul. 12, 2024).

[8] P. F. Edemekong, D. L. Bomgaars, and S. B. Levy, “Activities of Daily Living (ADLs),” *NCBI*, Jun. 26, 2023. https://www.ncbi.nlm.nih.gov/books/NBK470404/ (accessed Jul. 05, 2024).

[9] L. A. Legg, S. R. Lewis, O. J. Schofield-Robinson, A. Drummond, and P. Langhorne, “Occupational therapy for adults with problems in activities of daily living after stroke,” *Cochrane Database of Systematic Reviews*, vol. 7, no. 7, Jul. 2017, doi: https://doi.org/10.1002/14651858.cd003585.pub3.

[10] N. Mizukami *et al.*, “Effect of the Synchronization-Based Control of a Wearable Robot Having a Non-Exoskeletal Structure on the Hemiplegic Gait of Stroke Patients,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 5, pp. 1011–1016, Mar. 2018, doi: https://doi.org/10.1109/tnsre.2018.2817647.

[11] E. Anselmino, A. Mazzoni, and Silvestro Micera, “EMG-based prediction of step direction for a better control of lower limb wearable devices,” *Computer methods and programs in biomedicine*, vol. 254, pp. 108305–108305, Sep. 2024, doi: https://doi.org/10.1016/j.cmpb.2024.108305.

[12] G. Durandau, W. F. Rampeltshammer, H. van der Kooij, and M. Sartori, “Neuromechanical Model-Based Adaptive Control of Bilateral Ankle Exoskeletons: Biological Joint Torque and Electromyogram Reduction Across Walking Conditions,”

IEEE Transactions on Robotics, vol. 38, no. 3, pp. 1380–1394, Jun. 2022, doi: https://doi.org/10.1109/TRO.2022.3170239.

[13] Y. Celik, S. Stuart, W. L. Woo, and A. Godfrey, “Gait analysis in neurological populations: Progression in the use of wearables,” *Medical Engineering & Physics*, vol. 87, pp. 9–29, Jan. 2021, doi: https://doi.org/10.1016/j.medengphy.2020.11.005.

[14] A. J. Fuglevand, D. A. Winter, and A. E. Patla, “Models of recruitment and rate coding organization in motor-unit pools,” *Journal of Neurophysiology*, vol. 70, no. 6, pp. 2470–2488, Dec. 1993, doi: https://doi.org/10.1152/jn.1993.70.6.2470.

[15] H. Tankisi, “Standards of instrumentation of EMG,” *Clinical Neurophysiology*, vol. 131, no. 1, pp. 243–258, Jan. 2020, doi: https://doi.org/10.1016/j.clinph.2019.07.025.

[16] A. H. Caillet, Andrew, D. Farina, and Luca Modenese, “Motoneuron-driven computational muscle modelling with motor unit resolution and subject-specific musculoskeletal anatomy,” *PLoS Computational Biology*, vol. 19, no. 12, pp. e1011606–e1011606, Dec. 2023, doi: https://doi.org/10.1371/journal.pcbi.1011606.

[17] J.-J. Luh, G.-C. Chang, C.-K. Cheng, J.-S. Lai, and T.-S. Kuo, “Isokinetic elbow joint torques estimation from surface EMG and joint kinematic data: using an artificial neural network model,” *Journal of Electromyography and Kinesiology*, vol. 9, no. 3, pp. 173–183, Apr. 1999, doi: https://doi.org/10.1016/s1050-6411(98)00030-3.

[18] K. Min, D. Shin, J. Lee, and Shinji Kakei, “Muscle synergy control model-tuned EMG driven torque estimation system with a musculo-skeletal model,” *PubMed*, Jul. 2013, doi: https://doi.org/10.1109/embc.2013.6610289.

[19] D. Farina, R. Merletti, and R. M. Enoka, “The extraction of neural strategies from the surface EMG: an update,” *Journal of Applied Physiology*, vol. 117, no. 11, pp. 1215–1230, Dec. 2014, doi: https://doi.org/10.1152/japplphysiol.00162.2014.

[20] D. Farina, R. Merletti, and R. M. Enoka, “The extraction of neural strategies from the surface EMG,” *Journal of Applied Physiology*, vol. 96, no. 4, pp. 1486–1495, Apr. 2004, doi: https://doi.org/10.1152/japplphysiol.01070.2003.

[21] J. Duchateau and R. M. Enoka, “Human motor unit recordings: Origins and insight into the integrated motor system,” *Brain Research*, vol. 1409, pp. 42–61, Aug. 2011, doi: https://doi.org/10.1016/j.brainres.2011.06.011.

[22] J. Á. Gallego *et al.*, “Influence of common synaptic input to motor neurons on the neural drive to muscle in essential tremor,” *Journal of Neurophysiology*, vol. 113, no. 1, pp. 182–191, Jan. 2015, doi: https://doi.org/10.1152/jn.00531.2014.

[23] G. Drost, “Propagation disturbance of motor unit action potentials during transient paresis in generalized myotonia: A high-density surface EMG study,” *Brain*, vol. 124, no. 2, pp. 352–360, Feb. 2001, doi: https://doi.org/10.1093/brain/124.2.352.

[24] N. L. Suresh, X. Li, P. Zhou, and W. Z. Rymer, “Examination of motor unit control properties in stroke survivors using surface EMG decomposition: A preliminary report,” *PubMed*, Aug. 2011, doi: https://doi.org/10.1109/iembs.2011.6092032.

[25] L. C. Miller, C. K. Thompson, F. Negro, C. J. Heckman, D. Farina, and J. P. A. Dewald, “High-density surface EMG decomposition allows for recording of motor unit discharge from proximal and distal flexion synergy muscles simultaneously in individuals with stroke,” *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, vol. 2014, pp. 5340–5344, 2014, doi: https://doi.org/10.1109/EMBC.2014.6944832.

[26] J. Fang, G. C. Agarwal, and B. T. Shahani, “Decomposition of EMG signal by wavelet spectrum matching,” *Proceedings of the 19th Annual International Conference of the*

IEEE Engineering in Medicine and Biology Society. "Magnificent Milestones and Emerging Opportunities in Medical Engineering" (Cat. No.97CH36136), vol. 3, Oct. 1997, doi: https://doi.org/10.1109/iembs.1997.756598.

[27] D. Farina, F. Negro, S. Muceli, and R. M. Enoka, "Principles of Motor Unit Physiology Evolve With Advances in Technology," *Physiology*, vol. 31, no. 2, pp. 83–94, Mar. 2016, doi: https://doi.org/10.1152/physiol.00040.2015.

[28] A. Holobar, D. Farina, M. Gazzoni, R. Merletti, and D. Zazula, "Estimating motor unit discharge patterns from high-density surface electromyogram," *Clinical Neurophysiology*, vol. 120, no. 3, pp. 551–562, Mar. 2009, doi: https://doi.org/10.1016/j.clinph.2008.10.160.

[29] R. Merletti, A. Holobar, and D. Farina, "Analysis of motor units with high-density surface electromyography," *Journal of Electromyography and Kinesiology*, vol. 18, no. 6, pp. 879–890, Dec. 2008, doi: https://doi.org/10.1016/j.jelekin.2008.09.002.

[30] S. Muceli *et al.*, "Accurate and representative decoding of the neural drive to muscles in humans with multi-channel intramuscular thin-film electrodes," *The Journal of Physiology*, vol. 593, no. 17, pp. 3789–3804, Sep. 2015, doi: https://doi.org/10.1113/jp270902.

[31] Shun Kunugi, Aleš Holobar, Tsutomu Kodera, H. Toyoda, and K. Watanabe, "Motor unit firing patterns of lower leg muscles during isometric plantar flexion with flexed knee joint position," *Journal of electromyography and kinesiology*, vol. 67, pp. 102720–102720, Dec. 2022, doi: https://doi.org/10.1016/j.jelekin.2022.102720.

[32] D. Stashuk, "EMG signal decomposition: how can it be accomplished and used?," *Journal of Electromyography and Kinesiology*, vol. 11, no. 3, pp. 151–173, Jun. 2001, doi: https://doi.org/10.1016/s1050-6411(00)00050-x.

[33] A. Gallina *et al.*, "Consensus for experimental design in electromyography (CEDE) project: High-density surface electromyography matrix," *Journal of Electromyography and Kinesiology*, vol. 64, p. 102656, Jun. 2022, doi: https://doi.org/10.1016/j.jelekin.2022.102656.

[34] A. Del Vecchio, A. Holobar, D. Falla, F. Felici, R. M. Enoka, and D. Farina, "Tutorial: Analysis of motor unit discharge characteristics from high-density surface EMG signals," *Journal of Electromyography and Kinesiology*, vol. 53, p. 102426, Aug. 2020, doi: https://doi.org/10.1016/j.jelekin.2020.102426.

[35] A. S. Hassan *et al.*, "Properties of Motor Units of Elbow and Ankle Muscles Decomposed Using High-Density Surface EMG," *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 2019, doi: https://doi.org/10.1109/embc.2019.8857475.

[36] D. W. Stashuk, "Decomposition and quantitative analysis of clinical electromyographic signals," *Medical Engineering & Physics*, vol. 21, no. 6–7, pp. 389–404, Jul. 1999, doi: https://doi.org/10.1016/s1350-4533(99)00064-8.

[37] I. Rodriguez-Carreno, L. Gila-Useros, and A. Malanda-Trigueros, "Motor Unit Action Potential Duration: Measurement and Significance," *Advances in Clinical Neurophysiology*, Oct. 2012, doi: https://doi.org/10.5772/50265.

[38] C. Dai, N. L. Suresh, A. K. Suresh, W. Z. Rymer, and X. Hu, "Altered Motor Unit Discharge Coherence in Paretic Muscles of Stroke Survivors," *Frontiers in Neurology*, vol. 8, May 2017, doi: https://doi.org/10.3389/fneur.2017.00202.

[39] R. M. Enoka and J. Duchateau, "Rate Coding and the Control of Muscle Force," *Cold Spring Harbor Perspectives in Medicine*, vol. 7, no. 10, p. a029702, Mar. 2017, doi: https://doi.org/10.1101/cshperspect.a029702.

[40] J. Duchateau and R. M. Enoka, “Distribution of motor unit properties across human muscles,” *Journal of Applied Physiology*, vol. 132, no. 1, Oct. 2021, doi: https://doi.org/10.1152/japplphysiol.00290.2021.

[41] J. J. Thompson, M. R. Blair, and A. J. Henrey, “Over the Hill at 24: Persistent Age-Related Cognitive-Motor Decline in Reaction Times in an Ecologically Valid Video Game Task Begins in Early Adulthood,” *PLoS ONE*, vol. 9, no. 4, p. e94215, Apr. 2014, doi: https://doi.org/10.1371/journal.pone.0094215.

[42] S. Y. Schaefer, A. Hooyman, and K. Duff, “Using a Timed Motor Task to Predict One-Year Functional Decline in Amnestic Mild Cognitive Impairment,” *Journal of Alzheimer’s Disease*, vol. 77, no. 1, pp. 53–58, Sep. 2020, doi: https://doi.org/10.3233/jad-200518.

[43] S. Bruhn, N. Kullman, and A. Gollhofer, “The Effects of a Sensorimotor Training and a Strength Training on Postural Stabilisation, Maximum Isometric Contraction and Jump Performance,” *International Journal of Sports Medicine*, vol. 25, no. 1, pp. 56–60, Jan. 2004, doi: https://doi.org/10.1055/s-2003-45228.

[44] J. Penalver-Andres, K. A. Buetler, T. Koenig, R. M. Müri, and L. Marchal-Crespo, “Providing Task Instructions During Motor Training Enhances Performance and Modulates Attentional Brain Networks,” *Frontiers in Neuroscience*, vol. 15, Dec. 2021, doi: https://doi.org/10.3389/fnins.2021.755721.

[45] J. Mora-Gonzalez, I. J. Pérez-López, I. Esteban-Cornejo, and M. Delgado-Fernández, “A Gamification-Based Intervention Program that Encourages Physical Activity Improves Cardiorespiratory Fitness of College Students: ‘The Matrix rEFvolution Program,’” *International Journal of Environmental Research and Public Health*, vol. 17, no. 3, p. 877, Jan. 2020, doi: https://doi.org/10.3390/ijerph17030877.

[46] X. Hu, W. Z. Rymer, and N. L. Suresh, “Accuracy assessment of a surface electromyogram decomposition system in human first dorsal interosseus muscle,” *Journal of Neural Engineering*, vol. 11, no. 2, p. 026007, Feb. 2014, doi: https://doi.org/10.1088/1741-2560/11/2/026007.

[47] S. Bao, Y. Wang, D. L. Wright, J. J. Buchanan, and Y. Lei, “Differences in motor unit recruitment patterns and low frequency oscillation of discharge rates between unilateral and bilateral isometric muscle contractions,” *Human Movement Science*, vol. 83, p. 102952, Jun. 2022, doi: https://doi.org/10.1016/j.humov.2022.102952.

[48] A. M. Zero, E. A. Kirk, K. Hali, and C. L. Rice, “Firing rate trajectories of human motor units during isometric ramp contractions to 10, 25 and 50% of maximal voluntary contraction,” *Neuroscience Letters*, vol. 762, p. 136118, Sep. 2021, doi: https://doi.org/10.1016/j.neulet.2021.136118.

[49] H. L. O’Brien and E. G. Toms, “What is user engagement? A conceptual framework for defining user engagement with technology,” *Journal of the American Society for Information Science and Technology*, vol. 59, no. 6, pp. 938–955, 2008, doi: https://doi.org/10.1002/asi.20801.

[50] F. S. Rahayu, L. E. Nugroho, and R. Ferdiana, “E-Learning Gamification Model using Video Game Elements that Influence Engagement or Addiction,” *2022 International Seminar on Intelligent Technology and Its Applications: Advanced Innovations of Electrical Systems for Humanity, ISITIA 2022 - Proceeding*, Jul. 2022, doi: https://doi.org/10.1109/isitia56226.2022.9855202.

[51] S. Adlakha, D. Chhabra, and P. Shukla, “Effectiveness of gamification for the rehabilitation of neurodegenerative disorders,” *Chaos, Solitons & Fractals*, vol. 140, p. 110192, Nov. 2020, doi: https://doi.org/10.1016/j.chaos.2020.110192.

[52] L. Sardi, A. Idri, and J. L. Fernández-Alemán, “A systematic review of gamification in e-Health,” *Journal of Biomedical Informatics*, vol. 71, pp. 31–48, Jul. 2017, doi: https://doi.org/10.1016/j.jbi.2017.05.011.

[53] T. Kari, J. Piippo, L. Frank, M. Makkonen, and P. Moilanen, “To Gamify or Not to Gamify? : Gamification in Exercise Applications and Its Role in Impacting Exercise Motivation,” *Bled eConference*, p. 20, Jan. 2016.

[54] Z. Zhao, S. Ali Etemad, and A. Arya, “Gamification of Exercise and Fitness using Wearable Activity Trackers,” *Proceedings of the 10th International Symposium on Computer Science in Sports (ISCSS)*, pp. 233–240, Nov. 2015, doi: https://doi.org/10.1007/978-3-319-24560-7_30.

[55] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, “From game design elements to gamefulness,” *Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek ’11*, pp. 9–15, Sep. 2011.

[56] J. Hamari, J. Koivisto, and H. Sarsa, “Does Gamification Work? -- a Literature Review of Empirical Studies on Gamification,” *2014 47th Hawaii International Conference on System Sciences*, pp. 3025–3034, Jan. 2014, doi: https://doi.org/10.1109/hicss.2014.377.

[57] B. Morschheuser, L. Hassan, K. Werder, and J. Hamari, “How to design gamification? A method for engineering gamified software,” *Information and Software Technology*, vol. 95, pp. 219–237, Mar. 2018, doi: https://doi.org/10.1016/j.infsof.2017.10.015.

[58] A. B. VanGundy, “BRAIN WRITING FOR NEW PRODUCT IDEAS: AN ALTERNATIVE TO BRAINSTORMING,” *Journal of Consumer Marketing*, vol. 1, no. 2, pp. 67–74, Feb. 1984, doi: https://doi.org/10.1108/eb008097.

[59] “Google Jamboard: Collaborative Digital Whiteboard | Google Workspace for Education,” *Google for Education*. https://edu.google.com/intl/ALL_en/jamboard/ (accessed May 31, 2024).

[60] A. Luo, “What is content analysis and how can you use it in your research?,” *Scribbr*, Jul. 18, 2019. https://www.scribbr.com/methodology/content-analysis/ (accessed Feb. 20, 2024).

[61] “Simulink - Simulation and Model-Based Design,” *Mathworks.com*, 2019. https://www.mathworks.com/products/simulink.html

[62] Unity Technologies, “Unity,” *Unity*, 2019. https://unity.com/ (accessed Jul. 13, 2024).

[63] A. C. T. Klock, I. Gasparini, M. S. Pimenta, and J. Hamari, “Tailored gamification: A review of literature,” *International Journal of Human-Computer Studies*, vol. 144, p. 102495, Jun. 2020, doi: https://doi.org/10.1016/j.ijhcs.2020.102495.

[64] R. Bartle, “Richard A. Bartle: Players Who Suit MUDs,” *mud.co.uk*, 1996. https://mud.co.uk/richard/hcds.htm

[65] L. Milazzo, “A production expert’s guide to frame rates and FPS,” *Vimeo Blog*, Aug. 16, 2021. https://vimeo.com/blog/post/fps-for-live-streaming/ (accessed Aug. 17, 2024).

[66] M. Ballantyne, A. Jha, A. Jacobsen, J. S. Hawker, and Y. N. El-Glaly, “Study of Accessibility Guidelines of Mobile Applications,” *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia - MUM 2018*, 2018, doi: https://doi.org/10.1145/3282894.3282921.

[67] International Organization for Standardization, *ISO 9241 Ergonomics of human-system interaction - Part 125: Guidance on visual presentation of information*. Koninklijk Nederlands Normalisatie-instituut, 2017.

[68] M. Fareed, M. A. Anwar, and M. Afzal, “Prevalence and gene frequency of color vision impairments among children of six populations from North Indian region,” *Genes &*

Diseases, vol. 2, no. 2, pp. 211–218, Jun. 2015, doi: https://doi.org/10.1016/j.gendis.2015.02.006.
[69] L. M. Haile *et al.*, “Hearing loss prevalence and years lived with disability, 1990–2019: findings from the Global Burden of Disease Study 2019,” *The Lancet*, vol. 397, no. 10278, pp. 996–1009, Mar. 2021, doi: https://doi.org/10.1016/S0140-6736(21)00516-X.
[70] S. P. Katiyar, “Adult Illiteracy: A Global Social Problem,” *The Palgrave Handbook of Global Social Problems*, pp. 1–20, Nov. 2021, doi: https://doi.org/10.1007/978-3-030-68127-2_160-1.
[71] P. Bhandari and K. Nikolopoulou, “What is a likert scale? | guide & examples,” *Scribbr*, 2023. https://www.scribbr.com/methodology/likert-scale/ (accessed Aug. 08, 2024).
[72] T. Armstrong and F. Bull, “Development of the World Health Organization Global Physical Activity Questionnaire (GPAQ),” *Journal of Public Health*, vol. 14, no. 2, pp. 66–70, Mar. 2006, doi: https://doi.org/10.1007/s10389-006-0024-x.
[73] C. J. De Luca, A. Adam, R. Wotiz, L. D. Gilmore, and S. H. Nawab, “Decomposition of Surface EMG Signals,” *Journal of Neurophysiology*, vol. 96, no. 3, pp. 1646–1657, Sep. 2006, doi: https://doi.org/10.1152/jn.00009.2006.
[74] A. M. Zero, E. A. Kirk, K. Hali, and C. L. Rice, “Firing rate trajectories of human motor units during isometric ramp contractions to 10, 25 and 50% of maximal voluntary contraction,” *Neuroscience Letters*, vol. 762, p. 136118, Sep. 2021, doi: https://doi.org/10.1016/j.neulet.2021.136118.
[75] V. Cerqueira, A. de Mendonça, A. Minez, A. R. Dias, and M. de Carvalho, “Does caffeine modify corticomotor excitability?,” *Neurophysiologie Clinique/Clinical Neurophysiology*, vol. 36, no. 4, pp. 219–226, Jul. 2006, doi: https://doi.org/10.1016/j.neucli.2006.08.005.
[76] “System 4 Pro™,” *Biodex*. https://biodexrehab.com/products/system-4-pro/ (accessed Dec. 12, 2023).
[77] W. A. IJsselsteijn, Y. A. W. de Kort, and K. Poels, “The Game Experience Questionnaire,” Technische Universiteit Eindhoven, Eindhoven, 2013. Accessed: Dec. 12, 2023. [Online]. Available: https://research.tue.nl/en/publications/the-game-experience-questionnaire
[78] D. Johnson, M. J. Gardner, and R. Perry, “Validation of two game experience scales: The Player Experience of Need Satisfaction (PENS) and Game Experience Questionnaire (GEQ),” *International Journal of Human-Computer Studies*, vol. 118, pp. 38–46, Oct. 2018, doi: https://doi.org/10.1016/j.ijhcs.2018.05.003.
[79] E. L.-C. . Law, F. Brühlmann, and E. D. Mekler, “Systematic Review and Validation of the Game Experience Questionnaire (GEQ) - Implications for Citation and Reporting Practice,” *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, Oct. 2018, doi: https://doi.org/10.1145/3242671.3242683.
[80] “Gym equipment | gym equipment for home | fitness solutions | Technogym,” *www.technogym.com*. https://www.technogym.com (accessed Aug. 17, 2024).
[81] “Play Flappy Bird - Official Online Game - Submit Your Highscore!,” *Flappy Bird*. https://flappybird.net/ (accessed Dec. 12, 2023).
[82] “Just Dance 2024 Edition: Nintendo Switch™, PlayStation 5, Xbox Series X|S,” *www.ubisoft.com*. https://www.ubisoft.com/en-us/game/just-dance/2024 (accessed Dec. 12, 2023).
[83] Colblindor, “Coblis — Color Blindness Simulator – Colblindor,” *Color-blindness.com*, 2016. https://www.color-blindness.com/coblis-color-blindness-simulator/ (accessed Apr. 22, 2024).

[84] J. M. Wakeling, K. Uehli, and A. I. Rozitis, “Muscle fibre recruitment can respond to the mechanics of the muscle contraction,” *Journal of the Royal Society Interface*, vol. 3, no. 9, pp. 533–544, Aug. 2006, doi: https://doi.org/10.1098/rsif.2006.0113.
[85] D. J. Shpak, “Analytical Design of Biquadratic Filter Sections for Parametric Filters,” *Journal of the Audio Engineering Society*, vol. 40, no. 11, pp. 876–885, Nov. 1992.
[86] T. Oya, S. Riek, and A. G. Cresswell, “Recruitment and rate coding organisation for soleus motor units across entire range of voluntary isometric plantar flexions,” *The Journal of Physiology*, vol. 587, no. 19, pp. 4737–4748, Sep. 2009, doi: https://doi.org/10.1113/jphysiol.2009.175695.
[87] A. G. Ozgur *et al.*, “The effect of gamified robot-enhanced training on motor performance in chronic stroke survivors,” *Heliyon*, vol. 8, no. 11, pp. e11764–e11764, Nov. 2022, doi: https://doi.org/10.1016/j.heliyon.2022.e11764.
[88] S. Adlakha, D. Chhabra, and P. Shukla, “Effectiveness of gamification for the rehabilitation of neurodegenerative disorders,” *Chaos, Solitons & Fractals*, vol. 140, p. 110192, Nov. 2020, doi: https://doi.org/10.1016/j.chaos.2020.110192.
[89] J. M. Cenizo-Benjumea, F. J. Vázquez-Ramos, S. Ferreras-Mencía, and J. Gálvez-González, “Effect of a gamified program on physical fitness and motor coordination,” *Cultura, Ciencia y Deporte*, vol. 17, no. 52, Jun. 2022, doi: https://doi.org/10.12800/ccd.v17i52.1786.
[90] R. P. Brooks, M. T. Jones, M. W. Hale, T. Lunau, N. Dragano, and B. J. Wright, “Positive verbal feedback about task performance is related with adaptive physiological responses: An experimental study of the effort-reward imbalance stress model,” *International Journal of Psychophysiology*, vol. 135, pp. 55–62, Jan. 2019, doi: https://doi.org/10.1016/j.ijpsycho.2018.11.007.
[91] E. Faulconer, J. Griffith, and A. Gruss, “The impact of positive feedback on student outcomes and perceptions,” *Assessment & Evaluation in Higher Education*, vol. 47, no. 2, pp. 1–10, Apr. 2021, doi: https://doi.org/10.1080/02602938.2021.1910140.
[92] R. Waldersee and F. Luthans, “The impact of positive and corrective feedback on customer service performance,” *Journal of Organizational Behavior*, vol. 15, no. 1, pp. 83–95, Jan. 1994, doi: https://doi.org/10.1002/job.4030150109.
[93] W. Oliveira *et al.*, “Does gamification affect flow experience? A systematic literature review,” *arXiv.org*, Jun. 18, 2021. https://arxiv.org/abs/2106.09942 (accessed Nov. 15, 2023).
[94] W. Oliveira *et al.*, “The effects of personalized gamification on students’ flow experience, motivation, and enjoyment,” *Smart Learning Environments*, vol. 9, no. 1, Mar. 2022, doi: https://doi.org/10.1186/s40561-022-00194-x.
[95] S. Lİ, “Measuring Cognitive Engagement: An Overview of Measurement Instruments and Techniques,” *International Journal of Psychology and Education Studies*, vol. 8, no. 3, pp. 63–76, Jul. 2021, doi: https://doi.org/10.52380/ijpes.2021.8.3.239.

Appendices

Appendix 1: Informed Consent Form

Informed consent form

- I have read the information letter, know that I could ask questions, and any questions have been answered sufficiently.
- I agree that research data gathered for the study may be published or made available provided my name or other identifying information is not used.
- I know that participation is voluntary, and that I can decide at any time to not participate in the research or to stop, without stating reasons.
- I agree that the data obtained from this study will be stored for as long as two years and will be destroyed after that.

Please, tick yes or no below

I agree to participate in this research Yes ☐ No ☐

Please, fill in your full name

____

Signature: Date:

Appendix 2: Modified Game Experience Questionnaire

Game Experience Questionnaire **Subject nr:**

Please indicate how you felt after you finished completing all the trials for each of the items, on the following scale:

	Not at all	Slightly	Mode-rately	Fairly	Ex-tremely
	0	1	2	3	4
1. I felt content	☐	☐	☐	☐	☐
2. I felt skillful	☐	☐	☐	☐	☐
3. I thought it was fun	☐	☐	☐	☐	☐
4. I was fully occupied with the game	☐	☐	☐	☐	☐
5. I felt happy	☐	☐	☐	☐	☐
6. It gave me a bad mood	☐	☐	☐	☐	☐
7. I thought about other things	☐	☐	☐	☐	☐
8. I found it tiresome	☐	☐	☐	☐	☐
9. I felt competent	☐	☐	☐	☐	☐
10. I thought it was hard	☐	☐	☐	☐	☐
11. I forgot everything around me	☐	☐	☐	☐	☐
12. I felt good	☐	☐	☐	☐	☐
13. I was good at it	☐	☐	☐	☐	☐
14. I felt bored	☐	☐	☐	☐	☐
15. I felt successful	☐	☐	☐	☐	☐
16. I enjoyed it	☐	☐	☐	☐	☐
17. I was good at reaching the game's targets	☐	☐	☐	☐	☐
18. I felt annoyed	☐	☐	☐	☐	☐
19. I felt pressured	☐	☐	☐	☐	☐
20. I felt irritable	☐	☐	☐	☐	☐
21. I lost track of time	☐	☐	☐	☐	☐
22. I felt challenged	☐	☐	☐	☐	☐
23. I was deeply concentrated on the game	☐	☐	☐	☐	☐
24. I felt frustrated	☐	☐	☐	☐	☐
25. I lost connection with the outside world	☐	☐	☐	☐	☐
26. I felt time pressure	☐	☐	☐	☐	☐
27. I had to put a lot of effort into it	☐	☐	☐	☐	☐
28. I felt it got easier over time	☐	☐	☐	☐	☐

Appendix 3: Qualitative data

Subj 01 G
What did you think of the test?
"Wel leuk."
What did you think of the interface?
"Goed afgestemd op het pedaal, responsive."

Subj 02 C
What did you think of the test?
"Wel prima, wel leuk om te doen. Voelt been wel. Vooral door de positie, maar ook wel door inspanning."
What did you think of the interface?
"Simpel, prima, duidelijk. Smiley aan het eind was leuk. Niet helemaal duidelijk wat binnen de lijn moest blijven, maar ik focuste op het gele puntje."

Subj 03 C
What did you think of the test?
"Leuk om te doen, licht uitdagend, niet te complex. Binnen het spel ruimte voor optimalisatie. Kracht -> Respons. Als je dat eenmaal weet, loopt het daarna soepeler."
What did you think of the interface?
"Duidelijk."

Subj 04 C
What did you think of the test?
"Ja, prima."
What did you think of the interface?
"Interface vrij onaangekleed. Cursor was heel lijp."

Subj 05 C
What did you think of the test?
"Hogere threshold voelt als goede training. Lagere set voel je minder van. Niet als training, ook niet als moeilijk. 45%MVC maakt latere trials interessanter qua spiergebruik."
What did you think of the interface?
"Fijn dat je een range hebt. Marge geeft rust (niet helemaal verkeerd). Minder overcorrectie. Trillen van de indicator is vervelend, geeft onzekerheid (is het de Biodex of de game). Voetsteun zit wat los."

Subj 06 G
What did you think of the test?
"Fun, not quite difficult. User friendly."
What did you think of the interface?
"No issues or in-between problems. Simple and good."

Subj 07 G
What did you think of the test?
"Grappig om te doen, vond het wel leuk. Goed te doen. Paar keer goede score, soms iets minder, maar was wel leuk."
What did you think of the interface?

"Zag er wel strak uit, niet teveel afleiding."

Subj 08 C
What did you think of the test?
"Leuk om te doen, interessant hoe het werkt met de spieren. Soms ook moeilijk, met name de 45% en de minder steile met name naar beneden, in trapjes. Wel interessant om te merken."
What did you think of the interface?
"Duidelijk aangegeven wat je moet doen. Soms trilt de cursor. Verder duidelijk pad. Cursor is duidelijk waar je precies zit tussen de lijn."

Subj 09 G
What did you think of the test?
"Wel grappig, meer krachtmeting. Vrij simpel."
What did you think of the interface?
"Ziet er simpel uit, beetje retro-game-achtig."

Subj 10 G
What did you think of the test?
"Op zich wel grappig, beetje eentonig, weinig variatie en verdere uitdaging. Prima te doen."
What did you think of the interface?
"Doet wat 'ie moet doen. Prima voor het behalen van resultaat. Vraag is of het de aandacht vasthoudt."

Subj 11 C
What did you think of the test?
"Niet zo'n sterke mening over. Duidelijk waar het voor dient, verder niet veel bijzonders."
What did you think of the interface?
"Simpel, duidelijk, saai. Muziek zou zorgen voor meer focus."

Subj 12 C
What did you think of the test?
"Goed te doen, prima test. Niet echt een specifieke mening."
What did you think of the interface?
"Simpel, super simpel. Best wel degelijk."

Subj 13 C
What did you think of the test?
"Uitdagend, ook wel lastig om het goed te doen. Bewust van de kracht in mijn been. Normaal beweeg je gewoon, maar nu zie je hoeveel je beweegt. 45% was wel lastig om vast te houden. Betere timing naarmate de game vorderde. Loslaten is lastig om niet in vrije val te doen."
What did you think of the interface?
"Zorgt niet dat je afgeleid raakt. Zou helpen als je een rood licht krijgt als je buiten de lijn zit en groen als je erin zit. Gaat je in lange stukken focussen op het voorbijschuiven van de lijn. Qua kleur prima. Duidelijk wat je moet doen."

Subj 14 G
What did you think of the test?

“Wat moet ik ervan vinden? Soms irritant dat de cursor op en neer ging. Vervelend om collectibles te missen, drive om rondjes te vullen. Verder goed te doen. Goede kracht leveren word je snel beter in.”
What did you think of the interface?
“Erg simpel, beetje saai.”

Subj 15 G
What did you think of the test?
“Goed te doen, niet te moeilijk maar ook niet zo makkelijk dat ik er geen zin meer in had.”
What did you think of the interface?
“Simpel maar duidelijk.”

Subj 16 G
What did you think of the test?
“Leuk, grappig. Deed me denken aan snake (een beetje). Ook wel uitdagend, leidt tot de wens om het beter te doen.”
What did you think of the interface?
“Prima, duidelijk. Helder wat je moest doen, waar je moest kijken.”

Subj 17 C
What did you think of the test?
“Opstelling goed, been ondersteunen is goed. Prima voor de rest.”
What did you think of the interface?
“Wel interessant. Bij de lange stukken rust soms afgeleid, voor de rest prima.”

Subj 18 G
What did you think of the test?
“Curious about the results, difference between left and right. Waiting time was a bit long maybe. Fun to do, just a bit slow, but that’s my life.”
What did you think of the interface?
“Looking nice, simple. Serving the purpose, it seems. Screen size is good, imagine or color blindness it’s also good. At first, I thought the dots were about the height. Maybe put them on top?”

Subj 19 G
What did you think of the test?
“Spanning bilspier maakt dat je wat krampachtig zit. Verder vond ik het leuk en nuttig; belangrijk dat ik het heb kunnen doen.”
What did you think of the interface?
“Prima, niks op aan te merken. Zou niet weten wat je daarna kan verbeteren.”

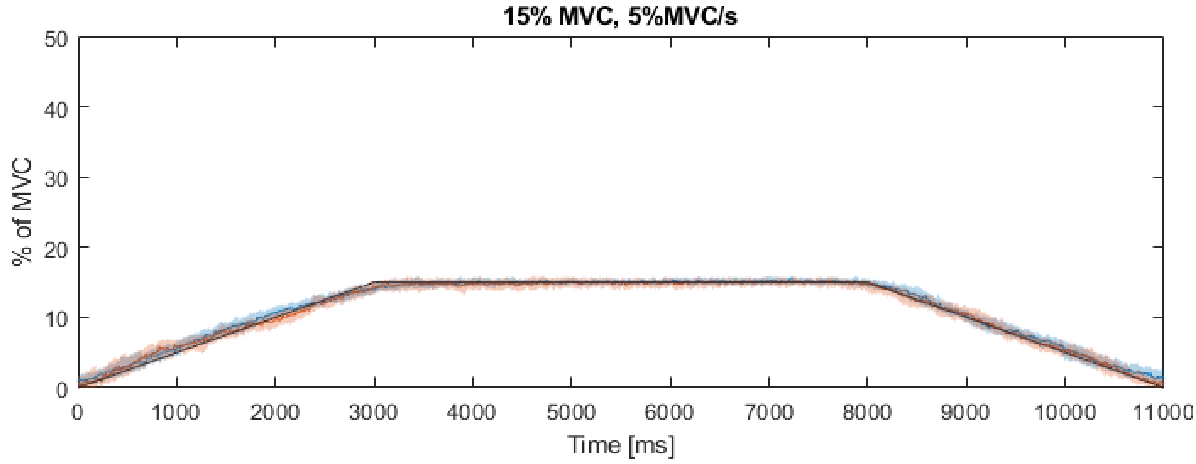
Subj 20 C
What did you think of the test?
“Leuk om te doen. Leuk om mee te helpen aan onderzoek. Leuk om wat van mee te krijgen.”
What did you think of the interface?
“Saai-ig. Vast met opzet. Had het wel leuk gevonden als het poppetje bewoog, meer als een spelletje. Je merkt dat je aandacht wat verslapt omdat er niet zoveel gebeurt op het scherm. Voor training wellicht goed om het interessanter te maken.”

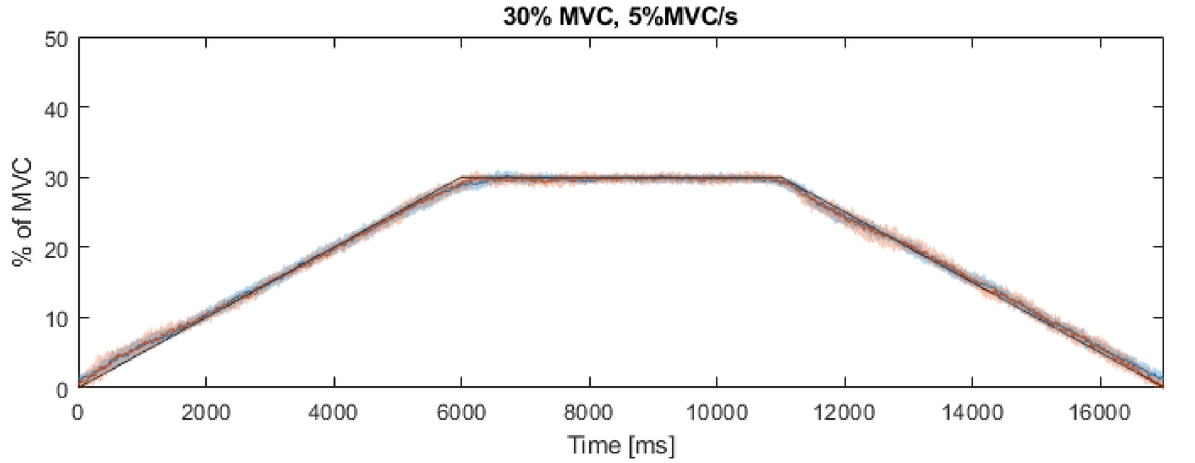
Subj 21 C

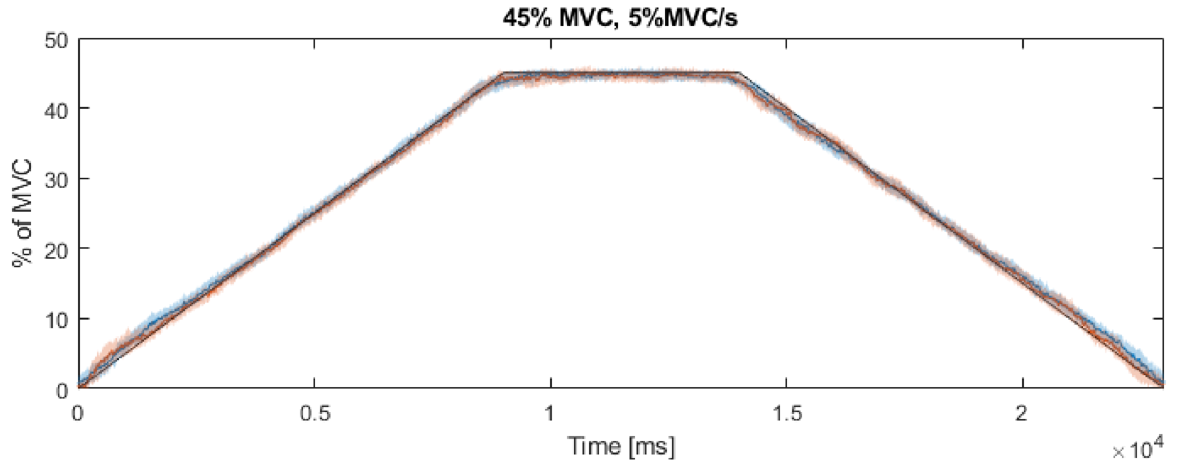
What did you think of the test?
"Leuk om te doen, kon nog aardig binnen de lijntjes blijven. Lastig om doel te beoordelen als je dat nog niet weet, maar leuk om een keer te doen."
What did you think of the interface?
"Heel clean, niks te moeilijk. Duidelijk dat geel het midden is en wit de marge. Ging voor mijn gevoel heel vloeiend."

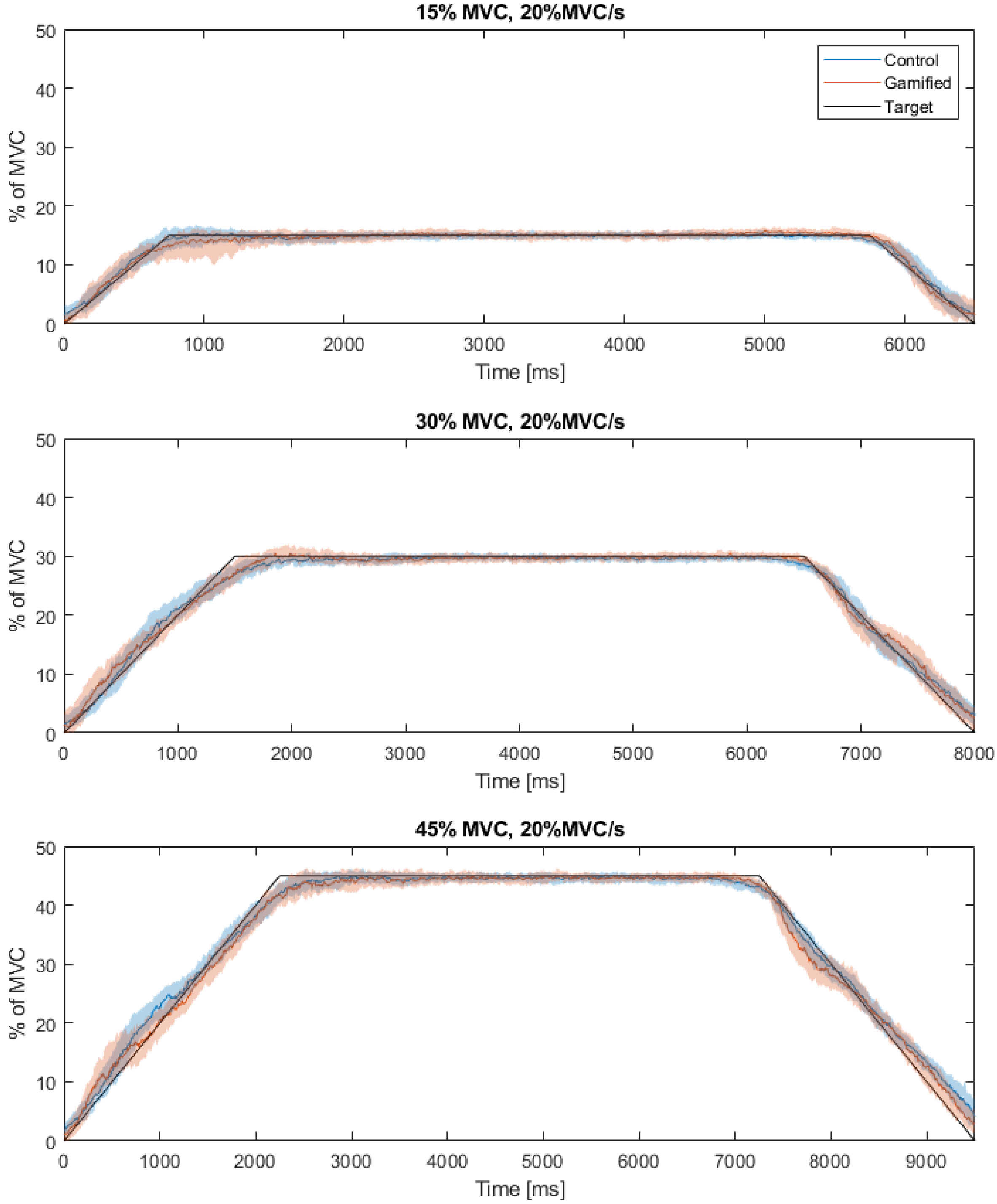
Subj 22 G
What did you think of the test?
"Prima om te doen. Was aan het nadenken wat er precies getest zou worden. Niet iets wat ik vaak doe, machine nooit eerder gezien, nooit die spieren gebruikt voor een spel."
What did you think of the interface?
"Intuïtief. Snapte de bedoeling. Zag er simpel uit, gewoon wat je nodig had. Grijze lijn had niet gehoeven, hoewel wel esthetisch. Trail van collectibles is voldoende. Knoppen geen aandacht aan besteed."

Appendix 4: Median torque and IQR








15% MVC, 20%MVC/s
Control
Gamified
Target
% of MVC
Time [ms]
30% MVC, 20%MVC/s
% of MVC
Time [ms]
45% MVC, 20%MVC/s
% of MVC
Time [ms]