



HAL
open science

Using Models of Baseline Gameplay to Design for Physical Rehabilitation

Antoine Loriette, Baptiste Caramiaux, Sebastian Stein, John Williamson

► **To cite this version:**

Antoine Loriette, Baptiste Caramiaux, Sebastian Stein, John Williamson. Using Models of Baseline Gameplay to Design for Physical Rehabilitation. [Research Report] University of Glasgow. 2022. hal-03800351

HAL Id: hal-03800351

<https://hal.science/hal-03800351v1>

Submitted on 6 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

USING MODELS OF BASELINE GAMEPLAY TO DESIGN FOR PHYSICAL REHABILITATION

A PREPRINT

A. Loriette, B. Caramiaux, S. Stein, J. H. Williamson

September 30, 2022

ABSTRACT

Modified digital games manage to drive motivation in repetitive exercises needed for motor rehabilitation, however designing modifications that satisfy both rehabilitation and engagement goals is challenging. We present a method wherein a statistical model of baseline gameplay identifies design configurations that emulate behaviours compatible with unmodified play. We illustrate this approach through a case study involving upper limb rehabilitation with a custom controller for a Pac-Man game. A participatory design workshop with occupational therapists defined two interaction parameters for gameplay and rehabilitation adjustments. The parameters' effect on the interaction was measured experimentally with 12 participants. We show that a low-latency model, using both user input behaviour and internal game state, identifies values for interaction parameters that reproduce baseline gameplay under degraded control. We discuss how this method can be applied to systematically balance gamification problems involving trade-offs between physical requirements and subjectively engaging experiences.

Keywords Game; Rehabilitation; Interaction Design; User modelling.

1 Introduction

The engaging nature of interactive systems, especially games, has the potential to improve rehabilitation sessions, as long as their interaction design aligns with patients' needs and capacities. One has to find and assess the relevant interaction design parameters that could be adapted to fit the rehabilitation requirements and specifications (Lopes and Bidarra, 2011).

Gamification (Raczkowski, 2014) seeks to incorporate “*video game elements in non-gaming systems to improve user experience and user engagement*” (Deterding et al., 2011). The use of games has been shown beneficial for patients in a wide variety of rehabilitation contexts. Focusing on the upper limb, Sietsema et al. (Sietsema et al., 1993) showed that an interaction with a Simon game produced significantly more range of motion than a standard rote exercise for patients with traumatic brain injury. However, when games are not initially intended for training purposes, the interaction between the game and the player must often be altered to meet rehabilitation goals. For example, *Half-life 2* (Valve Corporation, 2004) was partially controlled through a recumbent bicycle, by linking the pedalling speed to the move forward action of the playable character (Ketcheson et al., 2016), for purposes of player exertion. Then additional parameterisation of the interaction is typically needed to accommodate specific patients ability (Barrett et al., 2016). For instance, content from the game of *Fruit-Ninja* (Halfbrick Studios, 2010) was removed to cater for patients' reduced mobility during arm rehabilitation (Khademi et al., 2014). These modifications are often introduced using heuristics (Pirovano et al., 2016; Muñoz et al., 2018) or presets (Ketcheson et al., 2016; Chatta et al., 2015) but the extent to which they impact gameplay is difficult to predict.

In this paper, we propose to develop models of baseline gameplay, captured through game sessions from a control player group, to serve as a reference against which interaction modifications can be evaluated, alleviating the reliance on heuristics or presets.

Our contributions are twofold. First, we design a system for upper limb rehabilitation based on minimal game alterations and parametric interaction design. Second we validate a novel metric for predicting game difficulty using the designed system, based on models of baseline gameplay. These contributions benefit to the field of game design for rehabilitation by providing a new method for balancing game alterations compatible with a computational approach.

This research is motivated by field work with occupational therapists leading to the development of a system for movement-based rehabilitation using the game of Pac-Man. A series of workshop and testing sessions grounded our design choices: interaction adaptation is performed with the modifications to the game’s time rate and the optimisation of a newly designed input device physical properties. In addition, our contributions benefit to the field of Computational Interaction in HCI (Oulasvirta et al.), which relies on models to gain insight into interactive systems. We propose a model of baseline gameplay, constructed from low-level variables, that correlates with in-game score. We show that common measures of performance, such as in-game score, are prone to high latency and inter-user variability, and propose instead learning a statistical model of normative control behaviour.

The paper is structured as follows. In the next section we present previous work in modifying games for rehabilitation and related research to control input adaptation and user performance modelling. We then report the work carried out from an participatory workshops with occupational therapists and the resulting designed systems. An experimental study involving 12 participants follows wherein we investigate how a new input modality affects the user experience and users’ motions. Finally, to develop of a low-latency proxy for interaction difficulty, we detail a statistical modelling of collected data which is shown to correlate with game score.

2 Related Work

In line with our research objectives, we first review previous work in games for rehabilitation, and more precisely the modifications applied to the games to adapt to the rehabilitation specifications. Then we review related work on input gesture modelling in the context of gesture-based interaction.

2.1 Modifying Games for Rehabilitation

A large body of research has been dedicated to developing *Serious Games* (Michael and Chen, 2006; Susi et al., 2007) and design guidelines adapted to different patient’s needs have been established for their creation, such as those targeting stroke patients and upper arm rehabilitation (Burke et al., 2009; Barrett et al., 2016). More recently, repurposing commercial games has been proposed (Walther-Franks et al., 2013) with the intent of leveraging their established gameplay. The distinction between serious games and commercial games can sometimes be ambiguous. Commercial games intentionally designed for physical exercise, such as those present on the Wii Fit platform, or serious games based on well-known commercial games (Burke et al., 2009; Khademi et al., 2014) are conceptually very similar. The gameplay from commercial games is however unparalleled and Barret et al. (Barrett et al., 2016) found in their review that most appreciated serious games were those closely inspired by commercial ones. While likely more engaging than serious games, commercial games are usually less flexible in their potential for adaptation.

One commonality in the literature is the design of a new input control to adapt to different motor rehabilitation needs (Plaisant et al., 2000; Ketcheson et al., 2016; Khademi et al., 2014; Monedero et al., 2014; Chatta et al., 2015; Mandryk et al., 2013). Traditional hand-held controllers have been replaced by or augmented with other systems. These sense parts of the user’s body other than the hand, to solicit the motions required for physical rehabilitation while acting as input devices for control. Many studies have relied on the flexibility of optical technologies such as Microsoft Kinect (Bao et al., 2013; Chatta et al., 2015; Walther-franks et al., 2013), Leap Motion controller (Khademi et al., 2014) or motions capture systems (Muñoz et al., 2018). Accelerometer based sensors from the Wii system were also employed (Walther-franks et al., 2013; Monedero et al., 2014) while instrumented versions of exercise equipment, such as recumbent bicycles (Ketcheson et al., 2016), have also been used as input devices. Likewise, our work relies on a custom gamepad to recruit arm gestures in an interaction with a commercial game originally designed for finger-based control.

Then, the actual gameplay can be changed to fit the rehabilitation requirements. Walther-Franks et al. (Walther-Franks et al., 2013) proposed applying graphical overlays to unmodified games to fed back how well users performed rehabilitation-effective motions. Several systems have elaborated on this idea. Ketcheson et al. (Ketcheson et al., 2016) displayed the heart rate of the player and showed good results to support anti-sedentary levels of exertion, while Chatta et al. (Chatta et al., 2015) measured increased level of physical activity by employing graphical overlays to motivate players engaged with a commercial racing game. However, Chatta et al. also point out that this decouples exercise and in-game routine and, while it cleverly rely on “inconvenient interactions” (Rekimoto and Tsujita, 2014),

such methods can go against common recommendations for serious game design (Sinclair et al., 2009). These are examples of fixed modifications.

Strategies of automatic adaptation to the players' skill level and difficulties have also been investigated, notably in the field of serious gaming (Hendrix et al., 2019; Burke et al., 2009; Sinclair et al., 2009). These works highlight several challenges. Hendrix et al. (Hendrix et al., 2019) rightfully point out that games should not rely on presets, which require by nature a subjective player skill assessment, but adapt to an ideal challenge. Game knowledge can be useful for that matter (Sinclair et al., 2009). For instance, Burke et al. (Burke et al., 2009) mention that game difficulty can be tuned with the pace of the game, with a difficulty increasing along with the speed at which game elements move in the game. A common approach to tune these parameters is to use heuristic rules. Muñoz et al. (Muñoz et al., 2018) used heuristic rules to adapt both game difficulty and rehabilitation goals for a serious game they designed, inspired by Pong, targeting an optimal pulse rate. A set of rules linked to rehabilitation goals was also used by Pirovano et al. (Pirovano et al., 2016) to adapt the difficulty of custom games to specific users and a numerical objective was set to maintain an 80% success rate in the game. Similar game design elements are used in our system, and the method we propose is aiming at removing the needs for using presets to adapt for game difficulty. Instead the baseline gameplay model sets the default challenge.

2.2 Input control performance measure and modelling

Modifying a control input or swapping one for another purposefully designed is likely to have an impact on performance. A common measure of performance in games is the score, for example linked to the number of hits in First Person Shooter (FPS) games (Gerling et al., 2011). Isokoski and Martin (Isokoski and Martin, 2007) compared a keyboard with a standard mouse, a wheel mouse, a trackmouse and a Xbox360 gamepad. The gamepad was outperformed by all other devices in the task of FPS target acquisition. Another approach is to compute the information transmitted by the control inputs. Compared to a standard mouse, the Wii remote has been shown to perform poorly in terms of throughput, speed and error rate (Natapov et al., 2009), while an Xbox gamepad performed equally well in tracking a target's velocity (Klochek and MacKenzie, 2006). These differences can be explained by the differences in the limbs and muscles they recruit and is the topic of seminal research works. Card et al. (Card et al., 1991) compiled several Fitts' law experiments involving different limbs and showed a difference in available information throughput between neck, arm, wrist and finger – with four-fold decrease between finger and arm from 38bit/s to 9.5bit/s . Targeting more complex motions than pointing, Oulasvirta et al. (Oulasvirta et al., 2013) measured the information throughput in full body motions to provide researchers with a quantitative measure of control input capacity. Finally, to help designer in understanding the impact of new input methods on the user body, Bachynskyi et al. (Bachynskyi et al., 2015) used a clustering approach to summarise results about 3d pointing. In summary, newly designed input controls are likely sub-optimal when compared to reference devices and methods to predict performance beyond pointing, in particular (Oulasvirta et al., 2013; Bachynskyi et al., 2015), exist but are heavy to deploy in workshop settings.

Models for gameplay interaction have also been developed. For example, Smith et al. (Smith and Nayar, 2016) defined it as the player's raw input in contrast to methods that were only considering in-game states or event logs. In their work, they showed that models of gameplay can reliably identify player's identity based on their playstyle measured as "input words", i.e. strings of actions performed by the users. Patterns linking user identity to input behaviour was also found in the study from Dhakal et al. (Dhakal et al., 2018) in which they observed 136 millions keystroke of text input. In their analysis, several clusters identified users' typing behaviours. One of their performance measure included interkey interval, which is also a feature we rely on in this work.

Models of players have been used to replace humans in games by emulating their behaviour (Pfau et al., 2020). They have also been used for game design generation to optimise game worlds with specific outcome (Karavolos et al., 2018). When one has a direct access to player satisfaction, methods such as those proposed by Yannakakis et al. (Yannakakis and Hallam, 2009) can be employed, in which they trained a model to control the parameters of a custom game in real-time using recorded player satisfaction in pairwise A/B tests and sampled possible game designs through genetic optimisation. In contrast, the present work employs gameplay modelling for the purpose of identifying input configurations compatible with a baseline model.

2.3 Summary

The present work is, to the best of our knowledge, the first to develop baseline gameplay models of commercial games for adaptation to physical rehabilitation. In doing so, we tackle two challenges. The first challenge is finding parameters for the interaction loop that allows for rehabilitation and gameplay adaptation; this is addressed through co-design with occupational therapists. The second challenge is building a model of baseline gameplay from recorded

game sessions; we solve this by proposing a statistical model, based on data captured in an experimental study, with four parameters that leverages player typing behaviour and recorded effects on game state.

3 Motivation from a rehabilitation context

The present work is grounded by field work with a team of occupational therapists (OTs) from the spinal injury unit of Queen Elizabeth University Hospital (Loriette, 2019). They were interested in improving their process for upper limb reach rehabilitation, which has been less explored than lower limb treatment (Barrett et al., 2016).

3.1 Observations and Brainstorming workshops

The present collaboration started with a discussion related to the nature of OTs' patients and moved onto focus on their rehabilitation practises.

Their target group gathers patients with high level of spinal cord injuries (SCI), typically with spinal lesions at levels C5 and C6 (Jr et al., 1997), for whom the arm functions are severely compromised (Anderson, 2004). The induced partial or complete paralysis of specific arm muscles reduces patients' ability to perform coordinated motions. During the first stages of rehabilitation in the weeks following injury, when fine motor control of the fingers is absent, functional rehabilitation goals were limited to whole arm motions over tabletops. In routine rehabilitation, OTs used two physical artefacts. A skateboard, onto which one arm of the patient is strapped, was used to elicit linear back and forth motions while supporting the patient's arm, and facilitating movement through the reduced friction of the skateboard's wheels. The other artefact was a plastic tube rolled on its long dimension with patient's arm placed on top. A Velcro tape attached to the tube and to the table was sometimes used to increase friction as a mean to adapt task difficulty. For both devices, typical instructions from the OTs were for the patients to "move towards the window" or to "reach forward", involving four gestures in total: forward, backward, left, right on a tabletop. However, the simplistic, repetitive and artificial nature of these exercises led to low patient motivation, and they were described as "dull, tedious and boring" by the OTs.

Given the positive results reported in the literature regarding the use of games for rehabilitation (Barrett et al., 2016), this involvement of games was discussed between us (HCI practitioners) and the OTs during a series of meetings. These discussions were helpful in two ways. They helped in defining the main requirements by the OTs for a system which elicits the same types of movements as the ones exerted using the two artefacts introduced previously. And they helped in understanding the extent to which they can be parameterised in order to calibrate the level of difficulty. A two-hours participatory workshop was then organised at the hospital with five OTs and two of the authors. The workshop was split equally into two parts.

We first engaged in a brainstorming session to find mappings between game actions and rehabilitation movements. The list of rehabilitation movements was matched with a list of candidates for the games. When prompted for simple games they were familiar with, the participants mentioned Whack-a-Mole, Bubble Witch, Pong and Frogger to name a few, all of which were arcade games. The controls needed to implement their gameplay were then discussed. While some required four ways directional actions (e.g. Pac-Man), for aiming at a target for example, others only required button presses for triggering specific actions (e.g. Puzzle Bubble). To simplify the workshop complexity, it was decided that games with simple 4-ways controls should be focused on.

After the brainstorming, the second discussion addressed the parameterisation of specific exercises the OTs would like to see implemented by a gameplay. The movements that patients were producing when interacting with the skateboard, mentioned earlier, were used as a reference. A paper prototype of actionable controls was created for the interaction to be played out on the desk around which the workshop took place. The instructions from the OTs were to perform long reaching motions with the arm in contact with a planar surface. The goal they pursued was for their patients to reach always further. The flexibility of the paper prototype highlighted that the possibility of actionable controls whose locations on the surface could be easily modified was important for exercise adaptation.

The outcome of these workshops was then threefold:

- We identified that the game itself was not the focus for either parties (OTs and ourselves), as long as it is enjoyable for the patient and complies to the specifications given by OTs;
- We sketched/agreed on a control input with 4-ways actionable buttons to interface with various different games;
- We chose a rehabilitation parameter: the position of the gamepad's buttons.

3.2 System description

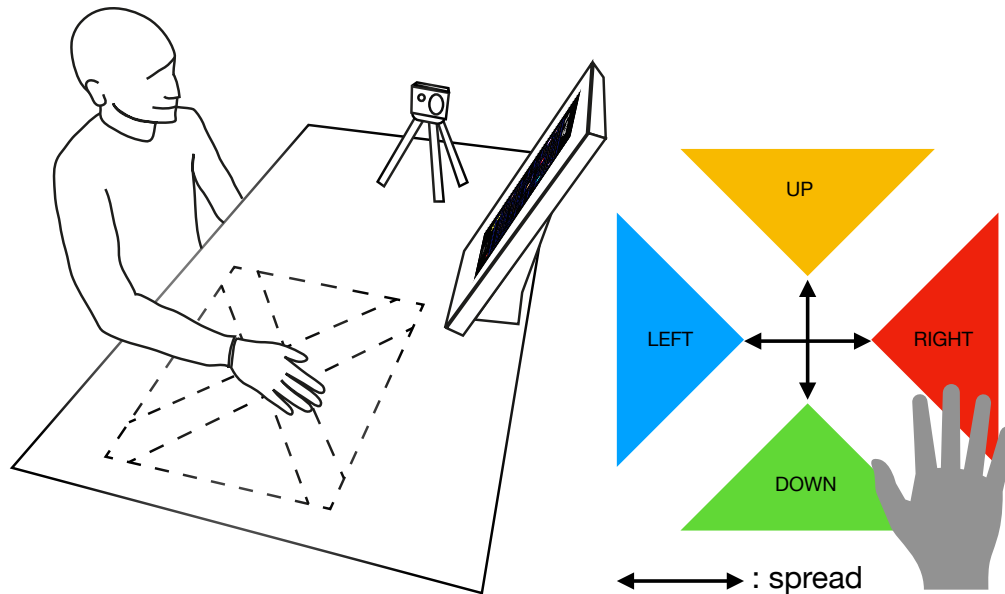


Figure 1: (left) Sketch of the proposed interaction. A user controls a digital game through four regions. Optical tracking tracks user’s hand and detects when it enters gamepad regions. (right) The virtual gamepad, parameterised with the *spread* variable, maps hand position to the commands UP, RIGHT, DOWN, LEFT.

The preliminary field work presented above has led to the prototype depicted in Figure 1 (left). A patient is seated by a tabletop on top of which reaching motions are performed with one arm. An optical device both tracks the patient’s hand and creates a virtual surface with four regions arranged in a cross shape. Each virtual button is activated when the hand of the user enters its respective region. This new input control is interfaced with a digital game running in front of the patient. The gestures required by the gameplay are meant to be similar to the one’s needed for the patient’s regular exercising.

As for the game, we chose Pac-Man (Namco, 1980) as test game. It is compatible with the four-way control and it is one of the most iconic arcade games where the rules are known by many. Regions of the virtual surface were mapped to directional commands (UP, RIGHT, DOWN, LEFT). Pac-Man always move forward at a given speed, the commands are only used to make directional changes.

The prototype was initially set up in the authors’ lab on a 78cm wide office desk. We used an Optitrack system for the tracking of the participants hand. A single Infrared (IR) reflective marker, placed on the user’s hand between the index and middle fingers, indicated the (x, y) location of the hand on the surface. A dedicated computer ran the Optitrack system whose output was transmitted to another machine running the game in a browser. We used an opensource version of Pac-Man¹. Translating (x, y) coordinates of the hand to game commands was done with custom software on the machine running the game. The origin of the controls was indicated by a protruding tactile marker. Directional commands were triggered when the tracked hand location entered the triangular control regions, which were not visible to the user. Audio feedback indicated when commands were triggered with a *bip*. Users would be equipped with headphones during use to ensure they could clearly hear this feedback and the in-game audio.

3.3 Interaction parameterisation

Based on the field work and the implemented system, interaction parameters have been identified in order to act upon the game difficulty and rehabilitation effectiveness.

A first parameter is the *spread* parameter, denoted r , defined as the distance between the inner tips of the triangular actionable regions (Figure 1 (right)). This parameter serves the purpose of varying task difficulty. This parameter was actually required by the OTs as a lever to adapt the game difficulty to the patient’s needs.

¹See <https://github.com/masonicGIT/pacman> for the source code and a discussion about the accuracy of this remake.

We conducted a first round of informal tests with members of our lab, revealing that the overall interaction difficulty was greatly increased due to the new input control modality (from button presses to reaching arm movements). We observed that the change in game difficulty incurred by the change of input control modality cannot be accommodated by the spread parameter alone. The latency and responsiveness of control is impacted by optical tracking delay and the effective reaching time necessary to move between regions.

To counterbalance this effect, we identified a second interaction parameter controlling the speed of the game, called *Time Rate*. This is similar to what games like Tetris use to set difficulty without altering other part of the gameplay mechanics and was for example recommended by Burke et al. (Burke et al., 2009) in their guidelines. It also has the benefit to limit game alterations to a unique and controllable parameter time rate T_r . By reducing the time rate, the number of actions per unit of time required for playing are reduced, facilitating the interaction.

The proposed interaction parameterisation can be represented as a block diagram (Figure 2). The perception loop includes the user, an input device and a game and produces engagement, depicted as a thick black line on the diagram. The interaction can be affected by two design parameters: the scale r , governing the distance that the participant has to travel with the arm in order to trigger a command; the time rate T_r , governing the speed of the game.

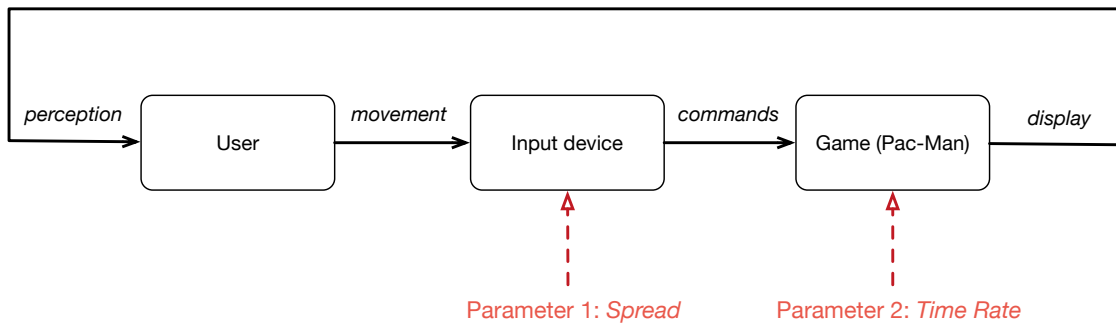


Figure 2: Conceptual block diagram for the interaction showing the primary control loop. A user interact with a game through an input device. Design parameters spread and frame rate are used accommodate for rehabilitation goals and gameplay engagement.

Both parameters relate to game difficulty, hence a significant impact on game engagement can be expected. As a matter of fact, providing a well-calibrated challenge is key to an engaging game experience. As phrased by Przybylski et al. (Przybylski et al., 2010), the “*mastery of controls plays an important role in game motivation, largely as a necessary, but not sufficient, condition for achieving psychologically need-satisfying play.*”

3.4 Summary

The co-design activities presented in this section led to the design of a rehabilitation system, based on a new gesture-based input device and the game of Pac-Man. We have conducted preliminary informal testing, including with participants having sustained a SCI.

In the remaining of the paper, we consider the problem of building a computational model of user performance while using the presented system. As introduced earlier, our motivation is to provide a more systematic method than heuristics in order to help adjusting the interaction parameters to define the game difficulty. Our approach relies on the estimation of the distance between a current performance and a baseline performance (reference). We split the problem in two:

1. Understanding the impact of the identified parameters on non-disabled participants’ performance while using the game. The fact to perform the study on non-disabled participants is motivated by the need to set a baseline performance and how parameters affect this baseline performance (as further developed in Section 4)
2. Proposing a behavioural model able to predict participants’ performance with respect to a baseline (Section 5).

4 Experimental Study

In this section we investigate the impact of the identified parameters on non-disabled participants’ performance while using the game. In addition, we collect usage data from the same population in order to set a baseline performance that will be further used in computational modelling (Section 5).

4.1 Apparatus

The system described above was used for the experiment. The game was instrumented to expose internal state variables: the commands received; the score; and the number of the steps taken (in multiples of 10). These were logged along with their frame number and wall-clock timestamp. The time rate control was implemented by fixing the duration of a game frame, equivalent to manipulating the frame rate. The movements made by players, tracked by the position of the motion capture markers, were also recorded alongside game state variables.

4.2 Task

We used the first level of Pac-Man as the task. In this game, the Pac-Man needs to consume all the pellets organised in a maze-like 2-dimensional rectangular map. The maze is populated with four ghosts (Blinky, Pinky, Inky and Clyde) which generally chase Pac-Man and kill him on contact. Four power-ups, placed at the four corners of the maze, grant invincibility to Pac-Man for a short period of time. The task was completed when all pellets were consumed or when all three lives had been used.

4.3 Design

We use a factorial design to measure potential effects and interactions of the independent variables gamepad control position (SPREAD) and time rate (T_RATE) on score (SCORE). SPREAD was evaluated with two levels (10cm and 40cm) and T_RATE was evaluated with three levels of slowdown ($T_r/3$, $2T_r/3$ and T_r). We opted for a within-subject design, as individual participants could be expected to have high variance. To counterbalance learning effects we used a balanced Latin square design. With independent variables SPREAD and T_RATE evaluated with respectively two and three levels, we have six testing conditions.

To establish a reference gameplay experience, a pre-test and a post-test were included where participants used ordinary keyboard (KEYBOARD) input to control the Pac-Man game. This provides the reference data we later use to build a statistical model of normative gameplay. The pre-test and post-test involved playing two rounds of Pac-Man.

Each block was designed with three repetitions of the task, provided the third repetition (e.g. game) was started before the 8 minute mark in that block. When it was not the case, only one or two games would be played to completion, this to ensure the experiment would last one hour maximum. The design of the experiment comported thus: $12 \text{ participants} \times (2 \text{ SPREAD} \times 3 \text{ T_RATE} \times 1 \text{ BLOCK} \times [1-3] \text{ Repetitions} + 2 \text{ PRE-TEST} + 2 \text{ POST-TEST}) = [120-264] \text{ total trials}$.

4.4 Procedure

Participants were welcomed and asked to provide written consent after reading an information sheet. Upon acceptance, they received £5 compensation. Participants were asked for demographic information and about their previous experience with *Pac-Man*. They were given a short explanation about the game mechanics. They were seated in front of a table, introduced to the tracking system, and equipped with a pair of headphones. An optical marker was attached with a stretchable band adaptable to different users' hand physical features. They were asked to acquire each of the gamepad controls once with the optical tracking system. The origin of the gamepad was chosen as a full forearm extension from the table edge adjacent to their trunk and, laterally, equidistant from both table edges. It was indicated on the tabletop with a protruding marker that taped in place for each participants. The participants were encouraged to produce their best possible score. To produce incentive for performance, we added an extra prize of £15 awarded to the participant with the highest score in any of the testing conditions. At the end of each testing conditions (a combination of r and T_r), participants were offered a break. Participants were responsible for advancing through the experiment by starting a new game. They were informed when they were changing testing condition but were not told about the current value of SPREAD or T_RATE. At the end of the experiment, participants were invited to provide qualitative feedback during a short debriefing discussion.

4.5 Participants

12 participants served in the experiment: four females and eight males, one of whom was left-handed, with a mean age of 29 ± 7 . The study was approved in advance by the University of X ethics committee. No participants presented with any relevant mental or physical disabilities.

4.6 Data Processing

For each testing conditions, we recorded the logs from the game of *Pac-Man*. From these, end-task SCORE was computed as the highest game score at trial completion. Pac-man scores are integers which increase as various on-screen items (pellets, powerpills, monsters) are consumed. Typical scores for one level of Pac-Man are in the range 3000-8000 points, with a theoretical maximum of 14800 points.

To produce a unit-less measure of score comparison between TRACKER and KEYBOARD conditions, the normalised value NSCORE was computed as the ratio between a raw score and the mean score from the keyboard reference games:

$$NSCORE = \frac{SCORE}{mean(SCORE_{keyboard})} \quad (1)$$

In other words, a mean value of 1 for NSCORE in the TRACKER condition means a level of performance equivalent to that produced with a keyboard.

4.7 Results

4.7.1 Learning Effects and Score Distribution

All participants reported having played Pac-Man, even if some reported that they could not remember when. We looked for learning effect between the reference PRE-TEST and POST-TEST conditions on KEYBOARD. An ANOVA on SCORE showed no significant difference between pre-test and post-test games. The mean score for all KEYBOARD games was 4390 (s.d. 950), see Table 1.

game #	PRE		POST		all
	1	2	3	4	
mean	3697	3992	4795	5077	4390
std	978	1438	1529	1705	950

Table 1: Mean value and standard deviation for SCORE over PRE-TEST and POST-TEST levels and averaged over games (all) in the last column.

The distribution of SCORE across participants was diverse (Figure 3). Some participants, such as 1, 8 or 11, produced scores with a very low variance while others, such as 6 and 10, had a much bigger range of scores.

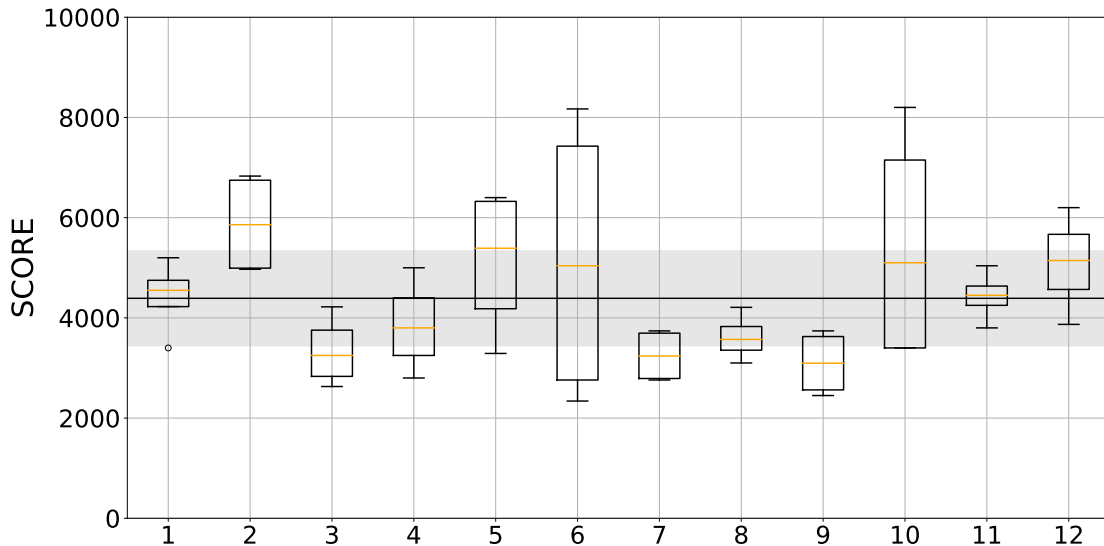


Figure 3: Distributions of SCORE across the 12 participants in the KEYBOARD reference games. The mean and standard deviation across participants are plotted with a black line and a black shade, respectively.

4.7.2 Effects on NSCORE

The results for NSCORE as function of T_RATE and SPREAD are plotted in Figure 4. A repeated measure two-way ANOVA on NSCORE with SPREAD and T_RATE as factors showed a significant main effect of T_RATE ($F_{1,47,16.15} = 22.11$, $ges = 0.41$, $p < 0.0001$), no significant main effect of SPREAD and no interaction T_RATE \times SPREAD on NSCORE. NSCORE was negatively correlated with T_RATE, decreasing from, on average, 96% of the reference keyboard SCORE at $T_r/3$ to 57% at T_r . We ran pairwise comparisons (adjusted with Holm-Bonferroni) on SCORE between all levels with the inclusion of KEYBOARD. It showed no differences of the NSCORE means between KEYBOARD and the condition where T_RATE equals to $T_r/3$.

Supporting the lack of significant effect of SPREAD on NSCORE, we found no differences between any SPREAD values for the same T_RATE. We also found no differences between levels ($T_r, 10cm$), ($2T_r/3, 10cm$) and ($2T_r/3, 40cm$). In other words, participants performed as well using tracker as with a keyboard, measured by NSCORE, provided T_RATE was reduced accordingly. The range of T_RATE values that produced a score equivalent to a keyboard score is however unknown, but most likely centred around a value of T_RATE close to $T_r/3$. This result validates T_RATE as a suitable parameter to balance task difficulty as induced by SPREAD with game performance as measured by NSCORE. For each value of SPREAD there exists a value for T_RATE such that NSCORE under this condition is statistically indistinguishable from reference gameplay.

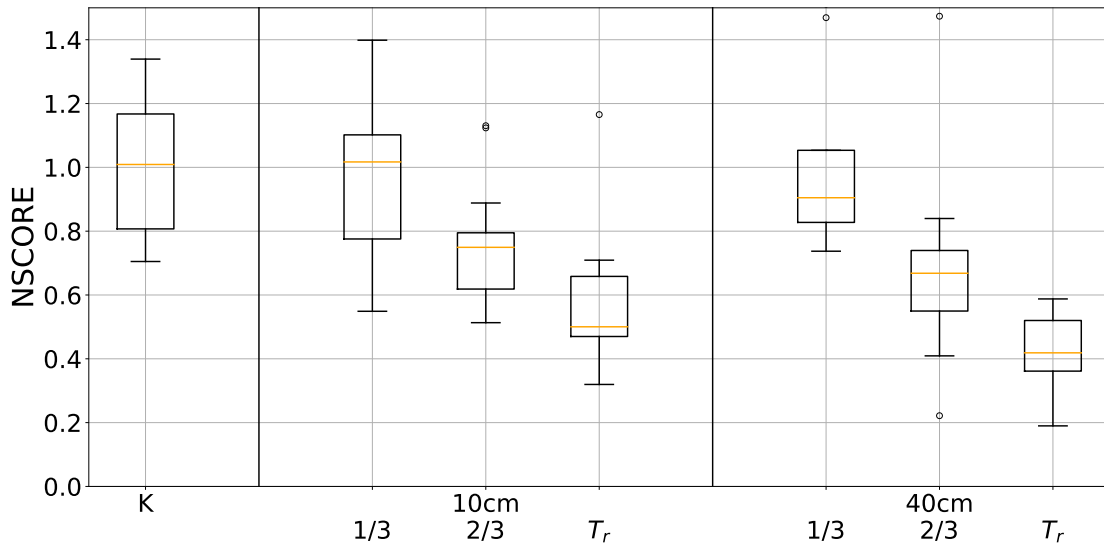


Figure 4: Effects of SPREAD and T_RATE on NSCORE for KEYBOARD (K) and TRACKER. The graph includes three parts, with KEYBOARD (K) on the leftmost subplot, and TRACKER with the six level combinations grouped by value of SPREAD on the middle and rightmost subplots. Values for SPREAD and T_RATE are reproduced on first row and second row, respectively.

4.8 Summary

This experimental study explored the impact of input modality and design parameters r and T_r on game scores. We showed that for low speed the scores are not affected by the change of input device. As pointed out in the preliminary testing, the interaction with our system is more difficult than with the reference input device. This result shows that it is possible to counteract this effect, as measured by the game score, by reducing the time rate parameter T_r . In addition, we found that the spread does not affect the game score.

With the objective of adapting the interaction such as ensuring a similar score than the one obtained using the baseline gameplay (reference input device and original speed), a metric must be defined to assess the performance between a given design and the reference one. Using the score as measured in this study is problematic. Since score is stochastic (people never play with the same exact performance twice), its reliable measurement requires a number of samples that increases with its variance. In addition, the value for score is sparse and delayed; available only once at the end of a game. Jointly, these factors limit the usability of score as a reliable, low-latency predictor of game performance due to its significant variation, high inter-subject variability and high latency.

In the next section we present a behavioural model and associated predictor for performance that is shown to be a good proxy for game score, to have lower variance than score and that can be computed with lower latency.

5 Modelling User Behaviour

We propose to investigate probabilistic modelling of observed behavioural characteristics during gameplay. The goal is, first, to fit a probabilistic model onto baseline behavioural data. Second, we propose to use the trained model to assess the likelihood of observed behavioural data, when changing the input modality or design parameters, under the model.

5.1 Behavioural Features

Prior to modelling, we need to specify which behavioural features to consider. Some domain knowledge can be needed to chose these features. Here, our reference model includes the actions the players produces, measured by the interkey interval (IKI), and the resulting effect in the game, measured by Pac-Man’s turning time (PTT). IKI is computed as the number of frames between two consecutive issued commands. PTT is computed as the number of frames in which Pac-Man stand motionless between two consecutive turns. While IKI is common in the literature related to typing (Dhakai et al., 2018), the choice of PTT has its roots from insights provided by game sessions with the prototype and backed by the experimental qualitative feedback: when poor control is afforded, producing rapid turns with Pac-Man becomes really hard. IKI and PTT features can be computed independently on the input modality considered (keyboard or tracker).

5.2 Gameplay Reference Model

The data needed to fit a reference model can be collected from baseline sessions. These baseline sessions have been recorded to that effect during the PRE-TEST and POST-TEST conditions on KEYBOARD in the previous experiment.

We made the following assumptions. Both IKI and PTT were modelled as continuous random variables. Even though they take discretised positive values measured in frames, they represent a time measurement. We also assumed that IKI and PTT generated independent events at some constant average rate.

Under these assumptions, we propose to model the variable X_{IKI} , representing a time difference between independent events, by a Gamma distribution. The probability density function (pdf) is given by:

$$Gamma(x; k, \mu, \gamma) = \frac{(x - \mu)^{k-1} e^{-(x-\mu)/\gamma}}{\Gamma(k)\gamma^k} \quad (2)$$

for $x \geq \mu$, $k > 0$, $\theta > 0$. $\Gamma(k)$ refers to the gamma function. We used maximum likelihood estimation (MLE) to estimate the distribution parameters. We fit the distribution on the IKI features from the dataset of all trials from PRE-TEST and POST-TEST (keyboard condition), across all participants. We found the following parameter values: $k = 2.19$, $\mu = -2.06$, $\gamma = 17.11$.

For the variable X_{PTT} , the shape of the histogram exhibits a sharp peak around 0 with a rapid decay, which is similar to an exponential distribution (Figure 5, right). Its probability density function (pdf) is given by:

$$Exp(x; \lambda) = \lambda \exp(-\lambda x), \quad x \geq 0 \quad (3)$$

Similar to previously, we estimated λ using MLE on the PRE-TEST and POST-TEST data, across all participants, and found that $\lambda = 2.2$ provides the best fit. Empirical distributions and the fitted models are shown in Figure 5. Histograms (yellow bars) represent the empirical distributions, with their associated pdf shown in black.

Our reference model $\theta_{ref} : (k_{IKI}, \mu_{IKI}, \gamma_{IKI}, \lambda_{PTT})$ is fully defined by this quadruplets of parameters. Given the observed play session of a user, this model can be used to estimate the similarity between the observed user behaviour and the baseline keyboard behaviour.

The observations of (X_{IKI}, X_{PTT}) over F frames are gathered into $\mathbb{X} : (x_t^{IKI}, x_t^{PTT})$. We propose to use the likelihood $p(\mathbb{X}|\theta_{ref})$ of the observed data given our reference model θ_{ref} (keyboard-based reference behaviour) as a measure for inferring how likely a user is to be behaving as if interacting in the baseline condition.

Assuming independence between X_{IKI} and X_{PTT} , we have:

$$\begin{aligned} p(\mathbb{X}|\theta_{ref}) &= p(X_{IKI}, X_{PTT}|\theta_{ref}) \\ &= p(X_{IKI}|\theta_{ref})p(X_{PTT}|\theta_{ref}) \end{aligned} \quad (4)$$

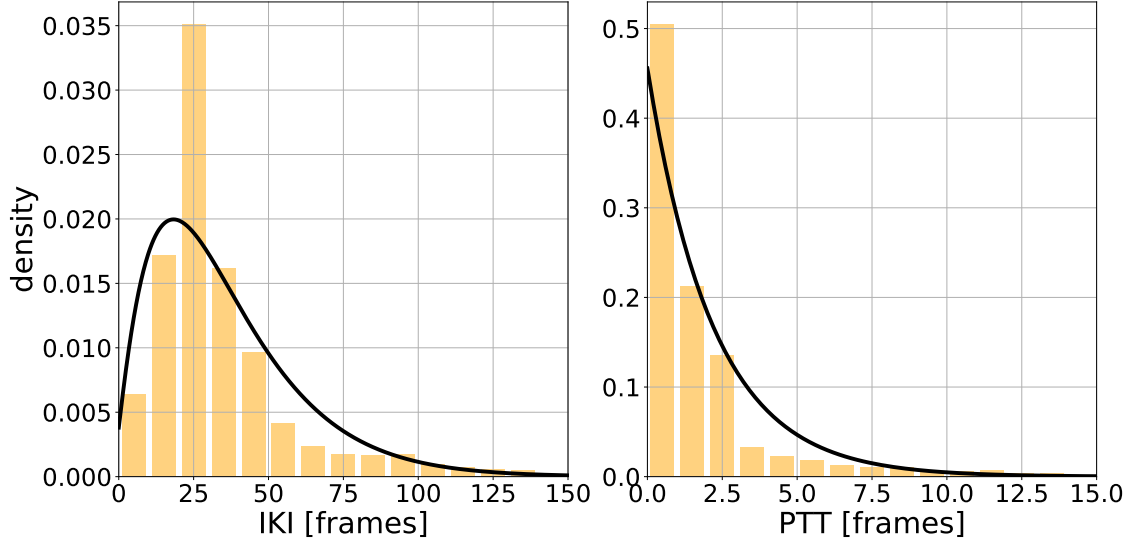


Figure 5: Empirical (yellow histogram) and fitted (black pdf) distributions for variables PTT and IKI on the left and right plots, respectively.

In practice we are taking the logarithm of this likelihood. We are interested in estimating the expected value of the log-likelihood, but we do not have access to the underlying $p(\mathbb{X})$ distribution. However, we can obtain samples from it and use the average value as an estimate:

$$\mathbb{E}[\log(p(\mathbb{X}|\theta_{ref}))] \approx 1/F \sum_t^F [\log(p(x_t^{IKI}|\theta_{ref})) + \log(p(x_t^{PTT}|\theta_{ref}))] \quad (5)$$

5.2.1 Normalisation

The values for the log likelihood (LL) were computed according to Equation 5. In the same fashion as SCORE, one value per game was obtained. We also computed a normalised value for LL, marked as NLL in the following, by taking the inverse of LL and multiplying it with its average value obtained over the PRE-TEST and POST-TEST levels, $NLL = mean(LL)/LL$, in a similar fashion to how NSCORE was computed.

5.3 Comparing NLL to NSCORE

We inspected the likelihood of IKI and PTT observations per participant from PRE-TEST and POST-TEST. Figure 6 reports the results. We observed a clear difference in standard deviations, within and across participants, between NSCORE and NLL with a value of 0.22 and 0.06, respectively. For NSCORE, five participants (2, 3, 5, 7 and 9) have an average NSCORE which lies further than one standard deviation from the overall mean, while for NLL only two participants (7 and 8) present the same deviation from the mean. This shows that NLL is less subject to inter-user variability than NSCORE.

Then we inspected the relationship between NLL and NSCORE. We computed the Pearson correlation through linear regression between NLL and NSCORE. The test revealed a significant correlation of 0.62 ($p < 0.001$, with $slope = 0.83$, $intercept = 0.14$, $stderr = 0.07$), leading to 39% of explained variance. A scatter plot (Figure 7) of their associated values for all measures except for the keyboard condition illustrates this relationship.

This correlation indicates that the statistical model based on the frequency of command inputs and their effect on game states is, to some extent, an indicator for user performance. The statistical model does not capture all the information necessary to predict score, such as the Pac-Man position in the maze or an understanding of the player’s tactics. However, it is logical that the user ability to issue commands and control the avatar, modelled by IKI and PTT, is a factor influencing success in the game.

In order to inspect the contribution of IKI and PTT in the correlation between NLL and NSCORES, we computed the correlation of the likelihood using only one of the variables IKI and PTT. We found that they had different linear relationship with NSCORE. A model with IKI only showed a statistically significant correlation of 0.55 ($p < 0.001$,

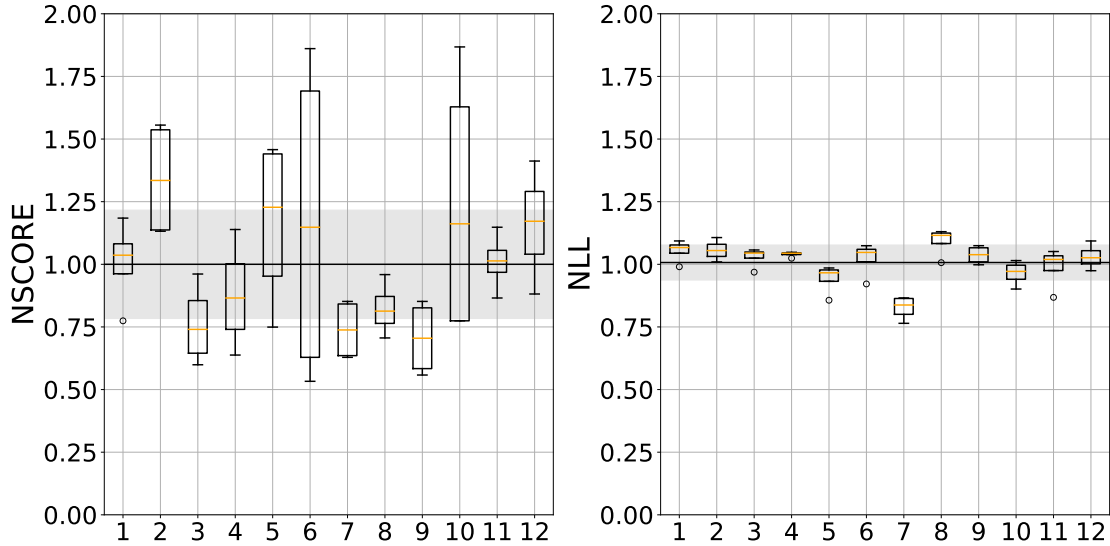


Figure 6: Distributions of NSCORE and NLL for all participants over the PRE-TEST and POST-TEST conditions on the left and right, respectively. The mean and standard deviation of their distributions across participant are represented by a black horizontal line and a shaded horizontal area.

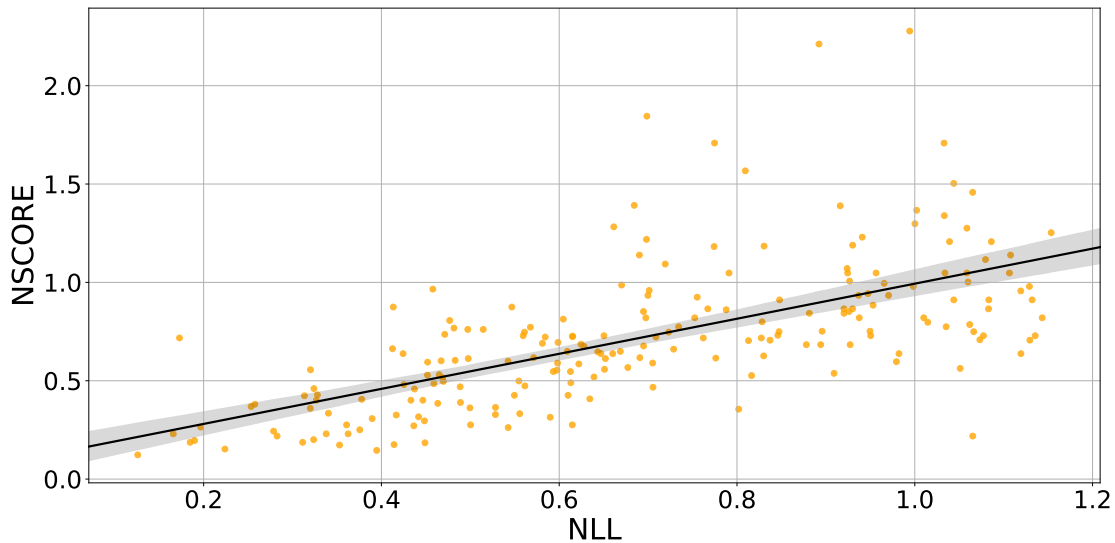


Figure 7: Linear relationship between NLL and NSCORE with a Pearson correlation coefficient of 0.62.

linear relationship with $slope = 1.85$, $intercept = -0.94$, $stderr = 0.19$) and a R^2 value of 0.30. A model with PTT only also showed a statistically significant correlation of 0.56 ($p < 0.001$, linear relationship with $slope = 0.50$, $intercept = 0.44$, $stderr = 0.07$) and a R^2 value of 0.31.

5.4 Effect of design parameters on NLL

Previous results showed that NLL exhibits a linear relationship with NSCORE. Here we inspected whether the design parameters (spread and time rate) impacts NLL. To do so, we computed IKI and PTT features from the data collected using both the keyboard input modality and the movement-based input modality.

Figure 8 depicts the statistics computed on NLL values for each condition. The figure reports the likelihoods computed using data from the movement-based input modality under the six testing conditions in the middle and rightmost columns.

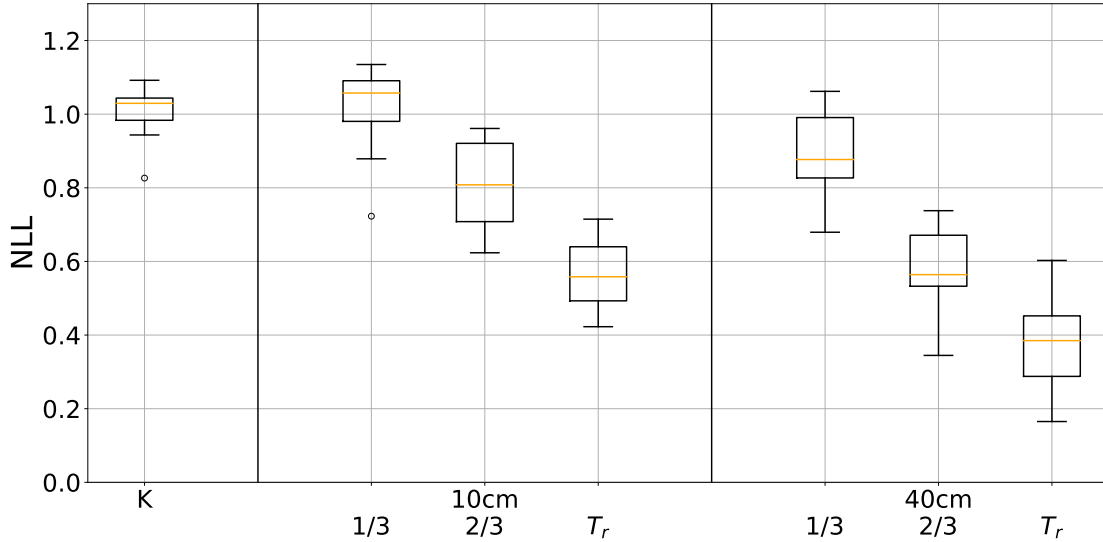


Figure 8: Effects of SPREAD and T_RATE on NLL for KEYBOARD (K) and TRACKER. The graph includes three parts, with KEYBOARD (K) on the leftmost subplot, and TRACKER with the six level combinations grouped by value of SPREAD on the middle and rightmost subplots. Values for SPREAD and T_RATE are reproduced on first row and second row, respectively.

We ran a repeated measure two-way ANOVA on NLL with SPREAD and T_RATE as factors. The analysis showed a significant main effect of T_RATE ($F_{1,96,21.58} = 83.34$, $ges = 0.76$, $p < 0.0001$). The analysis also showed a significant main effect of SPREAD ($F_{1,11} = 72.50$, $ges = 0.39$, $p < 0.0001$). We observed a negative impact of T_RATE and SPREAD on NLL with lower values with increasing values of T_RATE and SPREAD. We ran pairwise comparisons adjusted with Holm-Bonferroni on all levels with the inclusion of KEYBOARD. We found that only level ($T_r/3$, $10cm$) was similar to KEYBOARD. We also found some similarities between pairs ($T_r/3$, $10cm$) and ($T_r/3$, $40cm$) and pairs ($T_r/3$, $40cm$) and ($2T_r/3$, $10cm$) showing that similar effect on NLL can be achieved with different parameters pairs. Finally, the analysis found a small interaction T_RATE \times SPREAD ($F_{1,58,19.39} = 5.28$, $ges = 0.04$, $p = 0.02$). The reason seems to be the relatively high value of NLL for level (T_r , $40cm$). However, the amount of data collected in this condition is comparatively smaller than other conditions: the poor user performance had the result of shortened game sessions. This effect is thus likely an artefact.

These results should be compared with the effect of SPREAD and T_RATE on NSCORE (Figure 4). The goal of providing an experience similar to the keyboard reference translates to having parameters value (r and T_r) producing NSCORE and NLL equal to unity. The statistical analysis showed that the parameters value ($10cm$, $T_r/3$) is a solution for our design, whether NSCORE or NLL is considered. In other words, two models of user behaviour, based on widely different measurements (the outcome of the interaction for NSCORE and realisations of low-level variables for NLL), did identify the same value for T_r to produce an interaction similar to a keyboard. There are also some differences between both models. For example, the parameters value ($40cm$, $T_r/3$) is also a solution for the model based on NSCORE but not for NLL. The fact that NLL presents a much lower variability than SCORE could explain this difference.

5.5 Sampling Period

Finally, one of the reason for designing such model was the availability of many more samples for the variables IKI and PTT than observations of SCORE per unit of time. We measured from the TRACKER condition the time elapsed between samples for SCORE and LL (Figure 9). Obtaining one sample for SCORE took on average 2210 (s.d.1057) frames, while one sample for LL was available every 40 (s.d.31) frames on average. The measure of user behaviour based on low-level variable of gameplay provides 55 times more samples per frame than the counterpart model based on SCORE, thus exhibiting a lower latency.

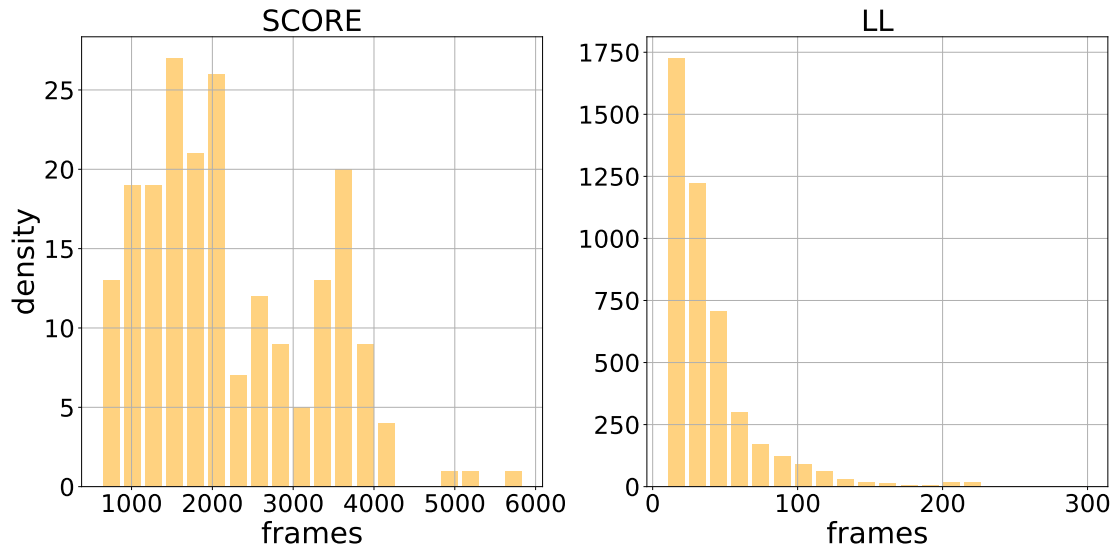


Figure 9: Distribution of time taken to measure a sample from SCORE and LL on the left and right, respectively.

6 Discussion

This paper contributes to research in interactive systems for motor rehabilitation. We presented two main contributions. First, we presented a gesture-based interactive system, interfacing with a commercial game, specifically designed such as complying with requirements from the context. These requirements were provided by occupational therapists involved in a co-design process. One of these requirements was to be able to adapt the level of the game difficulty through interaction parameters, that could be set by the OTs or automatically adapted. Our second contribution is to identify these parameters and propose a computational model estimating, in realtime, the input control performance as compared to a baseline performance, which could then be used as a proxy for game performance or optimisation function to adapt the parameters automatically. In this section, we discuss the different findings.

6.1 Design workshop and tools for modelling

The co-design activities (section 3) with occupational therapists were effective in exploring interaction possibilities and establishing requirements for a gamified interaction of the upper limb for patients with SCI.

The design of a new control modality was the principal outcome of the workshops with OTs. For motor rehabilitation, which targets specific motions from patients, the input control is the principal interface between the user’s body and the game. This is quite different from other types of rehabilitation targeting for example cognitive impairments (Mandryk et al., 2013). As a result, the game itself or the choice of the game did not emerge as a priority. Instead, the OTs mentioned the idea of reusing the proposed design with different games for satisfying specific patient preferences, a practice also described in the work by Hofmann et al. (Hofmann et al., 2019) as design reuse across clients.

Supported by rapid design iterations with a paper prototype, an adequate mapping from desired movement space to game control space (Walther-franks et al., 2013) was identified as well as a suitable parameterisation for adapting task difficulty (Lopes and Bidarra, 2011). The physical separation between controls placed on a tabletop effectively increased the reaching motions, in accordance to the rehabilitation requirements. This was helped by the natural connection between reaching motions and targeting motions but remains an issue for other types of activity. For more complex mapping, such as the one involving pedalling on a recumbent bicycle proposed by Ketcheson et al. (Ketcheson et al., 2016), an obvious solution to the mapping task is hard to obtain. We identify here potential opportunities for linking serious game design research with research related to HCI, such as (Bachynskyi et al., 2015; Oulasvirta et al., 2013). The breadth of possibilities afforded by the design of a new input control, and the effects these have on the user’s body are not yet fully understood.

As for the game adaptation, we chose not to use additional graphical overlays (Ketcheson et al., 2016). Instead, we relied on the flexibility of the control modality via the parameter spread and on the well-known game design pattern using time rate fluctuations. The OTs’ definition of the targeted motions was not very restrictive, on the contrary to

systems that target more precise motion executions, such as *YouMove* (Anderson et al., 2013) or *PhysioHome* (Tang et al., 2015), wherein the use of graphical feedback to convey how well the user is performing motions seems hardly avoidable. Put together, this affords a simple parameterisation, with the spread and time rate mainly controlling to rehabilitation goals and gameplay performance, respectively.

6.2 Interaction parameters adaptation

Game interaction with modified control input requires adaptation and usually results in (much) harder challenge for the player. Apart from controllers designed for performance, such as the one proposed by Kwak et al. (Kwak and Salem, 2009), new input controls can not afford users the same proficiency.

This effect was measured during the experimental study (section 4) by including a control condition and evaluation against a measure of in-game score. We showed that the parameter T_{rate} was successful in adapting the interaction difficulty: reducing T_{rate} from its original value decreased the level of difficulty, and with a preset level fixed around 1/3 the scores were indistinguishable. This is close to the 1/4 ratio in information throughput measured by Card et al. (Card et al., 1991) between the arm and the fingers. This also quantifies the significant change in users' control capabilities when input devices are swapped, which is thus an important effect to quantitatively measure when similar systems are proposed.

However, finding approximate preset values for the interaction parameters so that users could enjoy a performance similar to the keyboard reference is time consuming when score is used as a metric due to its sparse availability and high variability. In addition, any further modifications to the design of the control input would render previous presets unusable, requiring another round of measurements. Finally, since the data collected involved only unimpaired participants, it is unlikely that these preset values transfer to the final target group. As a result, we developed a low-latency metric of baseline gameplay that could replace the reliance on score for that matter.

6.3 Modelling Gameplay

We proposed modelling baseline gameplay (section 5) using game-specific features and user input, and using the normalised log-likelihood of observed gameplay under the reference model as a similarity measure. We showed that this measure is highly correlated with score, has lower variance and is available at lower latency.

We relied on the reasonable assumption that unmodified play is engaging. Our model of gameplay thus captures the behaviour of users, through their input frequency or IKI, and the effect they have on the game, through the specifically designed variable PTT. Even though PTT is hand crafted feature, it is simply an indirect measure of the avatar speed.

Then, we showed that this model correlates with score. This is evidence that maintaining gameplay, defined by IKI and PTT, is important to ensure game performance. The correlation between both metrics is however not perfect. The totality of the behaviours related to game outcome are not taken into account by the approach we proposed. Only the number of actions per unit of time and the speed of the Pac-Man in the game were considered. More complex models, such the one proposed by Smith et al. (Smith and Nayar, 2016) could be employed if the objective is to predict with accuracy the final score. Complex models are however hard to train. For instance, Smith et al. (Smith and Nayar, 2016) reported training duration in the order of days. The point here is that simple models, yet able to generalise, are valuable. The model we proposed relied on four parameters only, and was shown less prone to inter-user variability than in-game score.

Our results showed that the spread parameter does not impact the score. We can speculate that a higher spread may have impacted the score values, or that the spread considered would have impacted the performance of participants with impairments. However the spread does impact the log-likelihood values, suggesting that the proposed model may be more suitable to assess the input control performance than game score. The spread parameter was meant to steer motions towards those targeted by OTs for rehabilitation. Having a model that is capable of leveraging changes in spread is therefore important from a rehabilitation point of view.

Finally, a model linking engagement and performance is not easily established. Access to user engagement requires having specific measurements which often rely on questionnaires. Using the flow framework (Csikszentmihalyi, 2014), Limperos et al. (Limperos et al., 2011) compared the difference a Wii and a Playstation controller produced on the game experience, with the Wii controller being qualified as "natural" and requiring more involvement from the user body. They showed that the performance as defined as the end-game score was not alone explaining the enjoyment, and that the sense of control that users experienced was a rather salient indicator. Understanding what part of the gameplay makes a game engaging is an interesting research question. There is a need for models that are interpretable and that rely on quantifiable gameplay measures to further the understanding of user engagement.

6.4 Limitations and Future work

We acknowledge several limitations to this work. We made several assumptions with regards to the model. One was that samples were independent of each others, and IKI and PTT were also independent. These simplification assumptions are useful to build a simple model: IKI/PTT independence allows to model them separately, otherwise a joint distribution should have been learned which would have required a more complex learning procedure. The assumed independence of consecutive samples drawn from IKI and PTT is however useful to use simple probabilistic distributions with rather good generalisation capabilities. It is justified here with regards to the variables themselves: they describe short term processes that may not vary much with time.

The simplicity of the game we used is both an advantage and a disadvantage. For games using a similar control scheme, this approach is likely to be transferable. For game with more complex control commands, more work should be done on movement representation and probabilistic modelling. The distributions that were modelled in this work were simple (i.e. unimodal) and stationary which might not be always be the case.

The baseline model we trained is meant to be representative of a general population of players. This is one of the reason for its simplicity and the focus on command frequency: the simpler the model, the lesser the risk of overfitting. With a total population of 12 participants however, it is likely that certain category of players would be incorrectly modelled. In that case, it could be possible that players performing well with our system while behaving out of the norm would be qualified as unlikely by our model.

Finally, as future work we plan to test the model we developed to optimise online the interaction parameters. This extension to our computational approach is qualified as computational design (Oulasvirta et al.) and is a clear path forward for our research.

7 Conclusion

We have shown that a computational approach to design can be used as a mean to systematically control interaction performance in games where input device characteristics are varied. We argued that performance is an important side of the interaction to take into account when creating new control devices. We proposed a probabilistic model of low-level user behaviour that affords a measure of similarity against baseline gameplay, that could be used to online system adaptation. Through a complete methodology from field work to experimental and computational studies, this work brings a pragmatic account for designing interactive systems while maintaining user performance in a rehabilitation context.

8 Acknowledgements

We would like to warmly thank our contacts from the Spinal Unit of the QEU hospital (<https://www.spinalunit.scot.nhs.uk/>), in particular Michelle Rankin and Mariel Purcell, for the time they spent sharing their valuable expertise, helping with the design and testing of the prototype and their overall support for this research, which was funded through the MOREGRASP project (<http://www.moregrasp.eu/>).

References

- F. Anderson, T. Grossman, J. Matejka, and G. Fitzmaurice. YouMove. In *Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST '13*, pages 311–320, New York, New York, USA, 2013. ACM Press. ISBN 9781450322683. doi: 10.1145/2501988.2502045. URL <http://dl.acm.org/citation.cfm?doid=2501988.2502045>.
- K. D. Anderson. Targeting Recovery: Priorities of the Spinal Cord-Injured Population. *Journal of Neurotrauma*, 21(10):1371–1383, oct 2004. ISSN 0897-7151. doi: 10.1089/neu.2004.21.1371. URL <http://www.liebertonline.com/doi/abs/10.1089/neu.2004.21.1371>.
- M. Bachynskyi, G. Palmas, and T. Weinkauff. Informing the Design of Novel Input Methods with Muscle Coactivation Clustering. *Transactions on Computer-Human Interaction*, 21(6):1–25, jan 2015. ISSN 10730516. doi: 10.1145/2687921. URL <http://dx.doi.org/10.1145/2687921>.
- X. Bao, Y. R. Mao, Q. Lin, Y. H. Qiu, S. Z. Chen, L. Li, R. S. Cates, S. F. Zhou, and D. F. Huang. Mechanism of Kinect-based virtual reality training for motor functional recovery of upper limbs after subacute stroke. *Neural Regeneration Research*, 8(31):2904–2913, nov 2013. ISSN 16735374. doi: 10.3969/j.issn.1673-5374.2013.31.003. URL <https://www.ncbi.nlm.nih.gov/pubmed/25206611>.

- N. Barrett, I. Swain, C. Gatzidis, and C. Mecheraoui. The use and effect of video game design theory in the creation of game-based systems for upper limb stroke rehabilitation. *Journal of Rehabilitation and Assistive Technologies Engineering*, 3:205566831664364, jun 2016. ISSN 2055-6683. doi: 10.1177/2055668316643644. URL <http://journals.sagepub.com/doi/10.1177/2055668316643644>.
- J. W. Burke, M. D. J. McNeill, D. K. Charles, P. J. Morrow, J. H. Crosbie, and S. M. McDonough. Optimising engagement for stroke rehabilitation using serious games. *Visual Computer*, 25(12):1085–1099, dec 2009. ISSN 01782789. doi: 10.1007/s00371-009-0387-4. URL <http://link.springer.com/10.1007/s00371-009-0387-4>.
- S. K. Card, J. D. Mackinlay, and G. G. Robertson. A morphological analysis of the design space of input devices. *ACM Transactions on Information Systems*, 9(2):99–122, apr 1991. ISSN 10468188. doi: 10.1145/123078.128726. URL <http://dl.acm.org/citation.cfm?id=123078.128726>.
- A. Chatta, T. Hurst, G. Samaraweera, R. Guo, and J. Quarles. Get off the Couch. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15*, pages 47–56, New York, New York, USA, 2015. ACM Press. ISBN 9781450334662. doi: 10.1145/2793107.2793115. URL <http://dl.acm.org/citation.cfm?doid=2793107.2793115>.
- M. Csikszentmihalyi. Toward a Psychology of Optimal Experience. In *Flow Found. Posit. Psychol.*, pages 209–226. Springer Netherlands, Dordrecht, 2014. doi: 10.1007/978-94-017-9088-8_14. URL http://link.springer.com/10.1007/978-94-017-9088-8_14.
- S. Deterding, M. Sicart, L. Nacke, K. O’Hara, and D. Dixon. Gamification. using game-design elements in non-gaming contexts. In *Proc. 2011 Annu. Conf. Ext. Abstr. Hum. factors Comput. Syst. - CHI EA '11*, page 2425, New York, New York, USA, 2011. ACM Press. ISBN 9781450302685. doi: 10.1145/1979742.1979575. URL <https://dl.acm.org/citation.cfm?doid=1979742.1979575>.
- V. Dhakal, A. M. Feit, P. O. Kristensson, and A. Oulasvirta. Observations on typing from 136 million keystrokes. In *Conference on Human Factors in Computing Systems - Proceedings*, volume 2018-April, pages 1–12, New York, New York, USA, apr 2018. Association for Computing Machinery. ISBN 9781450356206. doi: 10.1145/3173574.3174220. URL <http://dl.acm.org/citation.cfm?doid=3173574.3174220>.
- K. M. Gerling, M. Klauser, and J. Niesenhaus. Measuring the impact of game controllers on player experience in FPS games. In *Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek '11*, page 83, New York, New York, USA, 2011. ACM Press. ISBN 9781450308168. doi: 10.1145/2181037.2181052. URL <http://dl.acm.org/citation.cfm?doid=2181037.2181052>.
- H. Halfbrick Studios. Fruit Ninja. Game [iOS], apr 2010.
- M. Hendrix, T. Bellamy-Wood, S. McKay, V. Bloom, and I. Dunwell. Implementing adaptive game difficulty balancing in serious games. *IEEE Transactions on Games*, 11(4):320–327, dec 2019. ISSN 24751510. doi: 10.1109/TG.2018.2791019.
- M. Hofmann, K. Williams, T. Kaplan, S. Valencia, G. Hann, S. E. Hudson, J. Mankoff, and P. Carrington. ”Occupational Therapy is Making”. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, pages 1–13, New York, New York, USA, 2019. ACM Press. ISBN 9781450359702. doi: 10.1145/3290605.3300544. URL <http://dl.acm.org/citation.cfm?doid=3290605.3300544>.
- P. Isokoski and B. Martin. Performance of input devices in FPS target acquisition. In *ACM International Conference Proceeding Series*, volume 203, pages 240–241, New York, New York, USA, 2007. ACM Press. ISBN 1595936408. doi: 10.1145/1255047.1255104. URL <http://portal.acm.org/citation.cfm?doid=1255047.1255104>.
- F. M. M. Jr, M. B. Bracken, G. Creasey, J. F. D. Jr, W. H. Donovan, T. B. Ducker, S. L. Garber, R. J. Marino, S. L. Stover, C. H. Tator, R. L. Waters, J. E. Wilberger, and W. Young. International Standards for Neurological and Functional Classification of Spinal Cord Injury. *Spinal Cord*, 35(5):266–274, 1997. ISSN 13624393. doi: 10.1038/sj.sc.3100432. URL <http://www.nature.com/doi/10.1038/sj.sc.3100432>.
- D. Karavolos, A. Liapis, and G. N. Yannakakis. Using a Surrogate Model of Gameplay for Automated Level Design. In *IEEE Conference on Computational Intelligence and Games, CIG*, volume 2018-Augus. IEEE Computer Society, oct 2018. ISBN 9781538643594. doi: 10.1109/CIG.2018.8490425.
- M. Ketcheson, L. Walker, and T. N. C. N. Graham. Thighrim and Calf-Life. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pages 2681–2692, New York, New York, USA, 2016. ACM Press. ISBN 9781450333627. doi: 10.1145/2858036.2858406. URL <http://dl.acm.org/citation.cfm?doid=2858036.2858406>.
- M. Khademi, H. Mousavi Hondori, A. McKenzie, L. Dodakian, C. V. Lopes, and S. C. Cramer. Free-hand interaction with leap motion controller for stroke rehabilitation. In *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems - CHI EA '14*, pages 1663–1668, New York, New York,

- USA, 2014. ACM Press. ISBN 9781450324748. doi: 10.1145/2559206.2581203. URL <http://dl.acm.org/citation.cfm?doid=2559206.2581203>.
- C. Kloczek and I. S. MacKenzie. Performance measures of game controllers in a three-dimensional environment. *Proceedings - Graphics Interface*, 2006:73–79, 2006. ISSN 07135424.
- M. Kwak and B. Salem. Designing a game controller for novice HALO3 players. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 5709 LNCS, pages 325–326. Springer, Berlin, Heidelberg, 2009. ISBN 3642040519. doi: 10.1007/978-3-642-04052-8_53. URL https://link.springer.com/chapter/10.1007/978-3-642-04052-8_53.
- A. M. Limperos, M. G. Schmierbach, A. D. Kegerise, and F. E. Dardis. Gaming Across Different Consoles: Exploring the Influence of Control Scheme on Game-Player Enjoyment. *Cyberpsychology, Behav. Soc. Netw.*, 14(6):345–350, jun 2011. ISSN 2152-2715. doi: 10.1089/cyber.2010.0146. URL <http://www.liebertpub.com/doi/10.1089/cyber.2010.0146>.
- R. Lopes and R. Bidarra. Adaptivity challenges in games and simulations: A survey. *IEEE Transactions on Computational Intelligence and AI in Games*, 3(2):85–99, jun 2011. ISSN 1943068X. doi: 10.1109/TCIAIG.2011.2152841.
- A. Lorientte. A computational approach to gestural interactions of the upper limb on planar surfaces. 2019.
- R. L. Mandryk, S. Dielschneider, M. R. Kalyn, C. P. Bertram, M. Gaetz, A. Doucette, B. A. Taylor, A. P. Orr, and K. Keiver. Games as neurofeedback training for children with FASD. In *Proc. 12th Int. Conf. Interact. Des. Child. - IDC '13*, pages 165–172, New York, New York, USA, 2013. ACM Press. ISBN 9781450319188. doi: 10.1145/2485760.2485762. URL <http://dl.acm.org/citation.cfm?doid=2485760.2485762>.
- D. D. R. Michael and S. L. Chen. *Serious games : games that educate, train and inform*. Thomson Course Technology, 2006. ISBN 1592006221. URL <https://dl.acm.org/citation.cfm?id=1051239>.
- J. Monedero, A. C. McDonnell, M. Keoghane, and D. J. O’Gorman. Modified Active Videogame Play Results in Moderate-Intensity Exercise. *Games Health J.*, 3(4):234–240, aug 2014. ISSN 2161-783X. doi: 10.1089/g4h.2013.0096. URL <http://www.liebertpub.com/doi/10.1089/g4h.2013.0096>.
- J. E. Muñoz, M. Cameirão, S. Bermúdez i Badia, and E. R. Gouveia. Closing the Loop in Exergaming - Health Benefits of Biocybernetic Adaptation in Senior Adults. In *The Annual Symposium on Computer-Human Interaction in Play Extended Abstracts - CHI PLAY '18*, pages 329–339, New York, New York, USA, 2018. ACM Press. ISBN 9781450356244. doi: 10.1145/3242671.3242673. URL <http://dl.acm.org/citation.cfm?doid=3242671.3242673>.
- Namco. Pac-Man. Game [Arcade], may 1980. URL <https://en.wikipedia.org/wiki/Pac-Man>.
- D. Natapov, S. J. Castellucci, and I. S. MacKenzie. ISO 9241-9 evaluation of video game controllers. In *Proceedings - Graphics Interface*, pages 223–230, 2009. ISBN 9781568814704. doi: 10.5555/1555880.1555930.
- A. Oulasvirta, P. O. Kristensson, X. Bi, and A. Howes. *Computational interaction*. ISBN 9780198799610.
- A. Oulasvirta, T. Roos, A. Modig, and L. Leppänen. Information capacity of full-body movements. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, page 1289, New York, New York, USA, apr 2013. ACM Press. ISBN 9781450318990. doi: 10.1145/2470654.2466169. URL <http://dl.acm.org/citation.cfm?id=2470654.2466169>.
- J. Pfau, J. D. Smeddinck, I. Bikas, and R. Malaka. Bot or not? User Perceptions of Player Substitution with Deep Player Behavior Models. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–10, New York, NY, USA, apr 2020. Association for Computing Machinery (ACM). doi: 10.1145/3313831.3376223. URL <https://dl.acm.org/doi/10.1145/3313831.3376223>.
- M. Pirovano, R. Mainetti, G. Baud-Bovy, P. L. Lanzi, and N. A. Borghese. Intelligent Game Engine for Rehabilitation (IGER). *IEEE Trans. Comput. Intell. AI Games*, 8(1):43–55, mar 2016. ISSN 1943-068X. doi: 10.1109/TCIAIG.2014.2368392. URL <http://ieeexplore.ieee.org/document/6967825/>.
- C. Plaisant, A. Druin, C. Lathan, K. Dakhane, K. Edwards, J. M. Vice, and J. Montemayor. A Storytelling Robot for Pediatric Rehabilitation. oct 2000. URL <https://drum.lib.umd.edu/handle/1903/1102>.
- A. K. Przybylski, C. S. Rigby, and R. M. Ryan. A Motivational Model of Video Game Engagement. *Review of General Psychology*, 14(2):154–166, 2010. ISSN 10892680. doi: 10.1037/a0019440.
- F. Raczkowski. Making points the point: Towards a history of ideas of gamification. 2014.
- J. Rekimoto and H. Tsujita. Inconvenient interactions: An alternative interaction design approach to enrich our daily activities. In *Proceedings of the Workshop on Advanced Visual Interfaces AVI*, pages 225–228, New York, New York, USA, 2014. Association for Computing Machinery. ISBN 9781450327756. doi: 10.1145/2598153.2598196. URL <http://dl.acm.org/citation.cfm?doid=2598153.2598196>.

- J. M. Sietsema, D. L. Nelson, R. M. Mulder, D. Mervau-Scheidel, and B. E. White. The use of a game to promote arm reach in persons with traumatic brain injury. *The American journal of occupational therapy*. : official publication of the American Occupational Therapy Association, 47(1):19–24, jan 1993. ISSN 02729490. doi: 10.5014/ajot.47.1.19. URL <https://ajot.aota.org/article.aspx?articleid=1872935>.
- J. Sinclair, P. Hingston, and M. Masek. Exergame development using the dual flow model. *Proceedings of the 6th Australasian Conference on Interactive Entertainment, IE 2009*, 2009. doi: 10.1145/1746050.1746061.
- B. A. Smith and S. K. Nayar. Mining controller inputs to understand gameplay. In *UIST 2016 - Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 157–168, New York, NY, USA, oct 2016. Association for Computing Machinery, Inc. ISBN 9781450345316. doi: 10.1145/2984511.2984543. URL <https://dl.acm.org/doi/10.1145/2984511.2984543>.
- T. Susi, M. Johannesson, and P. Backlund. Serious Games : An Overview. 2007. URL <http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A2416%7C&dswid=243>.
- R. Tang, X.-D. Yang, S. Bateman, J. Jorge, and A. Tang. Physio@Home. In *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst. - CHI '15*, pages 4123–4132, New York, New York, USA, 2015. ACM Press. ISBN 9781450331456. doi: 10.1145/2702123.2702401. URL <http://dl.acm.org/citation.cfm?doid=2702123.2702401>.
- Valve Corporation. Half Life 2, 2004.
- B. Walther-franks, D. Wenig, J. Smeddinck, and R. Malaka. Sportal : A First-Person Videogame Turned Exergame. In *Proceedings Mensch und Computer*, pages 539–542, 2013. URL <https://dl.gi.de/handle/20.500.12116/7665>.
- B. Walther-Franks, D. Wenig, J. Smeddinck, and R. Malaka. Exercise my game: Turning off-the-shelf games into exergames. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8215 LNCS:126–131, 2013. ISSN 16113349. doi: 10.1007/978-3-642-41106-9_15.
- G. N. Yannakakis and J. Hallam. Real-time game adaptation for optimizing player satisfaction. *IEEE Transactions on Computational Intelligence and AI in Games*, 1(2):121–133, jun 2009. ISSN 1943068X. doi: 10.1109/TCIAIG.2009.2024533. URL <http://ieeexplore.ieee.org/document/5067382/>.