



## Automated and unobtrusive measurement of physical activity in an interactive playground



Alejandro Moreno<sup>a,b</sup>, Ronald Poppe<sup>\*,c</sup>, Jenny L. Gibson<sup>d</sup>, Dirk Heylen<sup>a</sup>

<sup>a</sup> Human Media Interaction (HMI), University of Twente, The Netherlands

<sup>b</sup> Ambient Intelligence, Saxion University of Applied Sciences, The Netherlands

<sup>c</sup> Interaction Technology, Utrecht University, The Netherlands

<sup>d</sup> Centre for research on Play in Education, Development and Learning (PEDAL), University of Cambridge, United Kingdom

### ARTICLE INFO

#### Keywords:

Play  
Interactive playgrounds  
Automated behavior analysis  
Physical activity  
Exertion measurement  
Depth cameras

### ABSTRACT

Promoting physical activity is one of the main goals of interactive playgrounds. To validate whether this goal is met, we need to measure the amount of physical player activity. Traditional methods of measuring activity, such as observations or annotations of game sessions, require time and personnel. Others, such as heart rate monitors and accelerometers, need to be worn by the player. In this paper, we investigate whether physical activity can be measured unobtrusively by tracking players using depth cameras and applying computer vision algorithms. In a user study with 32 players, we measure the players' speed while playing a game of tag, and demonstrate that our measures correlate well with exertion measured using heart rate sensors. This makes the method an attractive alternative to either manual coding or the use of worn devices. We also compare our approach to other exertion measurement methods. Finally, we demonstrate and discuss its potential for automated, unobtrusive measurements and real-time game adaptation.

### 1. Introduction

Technology has become embedded into many aspects of children's lives, including children's play, and studies have suggested that it can limit its users to screen-based solitary interactions (Radesky and Christakis, 2016). A clear example of this is that children currently spend a significant amount of time consuming online digital media, and a considerable part is dedicated to digital gaming (Blumberg et al., 2013). Most young people play video games at least occasionally and many of them play daily (Desai et al., 2010; Ferguson and Olson, 2013). In doing so, the opportunities available for children to engage in full-body physical activity and in social interactions, both essential for their development, can be drastically reduced (Aggio et al., 2012; Carson et al., 2016; Plötner et al., 2015). Nonetheless, digital games can also be used to encourage positive aspects of play (Calvert et al., 2013). Exertion games or active video games (AVGs) provide the entertainment value of digital games while encouraging players to engage in physical activity (Müller et al., 2016; Peng et al., 2013). Interactive playgrounds are instrumented spaces where exertion games can be played, usually with small groups of players (Poppe et al., 2014). These playgrounds combine elements of traditional playgrounds with digital elements to promote key aspects of play, including physical activity (Moreno et al.,

2013). In general, these approaches are designed to put body movement as a core part of the gameplay in order to motivate players to exert themselves (e.g., Landry and Parés, 2014; Müller et al., 2012a).

This does not necessarily mean that players engage in appropriate levels of exertion (Peng et al., 2013). Players could move very little, or players might move too much and burn out quickly. Knowing beforehand how to stimulate players appropriately is difficult, and is likely to differ between individuals. One promising alternative to control the level of exertion is to adapt the stimulation of the players *in real-time*, based on measurements of the players' levels of physical activity (Altimira et al., 2017).

Traditional methods of measuring physical activity in play include the annotation of game sessions, interviews and self-reports (Hands and Larkin, 2006; Loprinzi and Cardinal, 2011). These validated methods provide varied information, but the outcome only becomes available after, not during, the game session. Annotation requires observers to categorize specific actions using annotation schemes, and it is typically performed on recorded game sessions (Bakker et al., 2008; Moreno et al., 2012). Questionnaires to evaluate physical activity are filled in after the game sessions since they ask players about their experiences. A different approach to measuring physical activity in games is to use sensors such as accelerometers, pedometers or heart rate monitors.

\* Corresponding author.

E-mail addresses: [a.m.morenocelleri@saxion.nl](mailto:a.m.morenocelleri@saxion.nl) (A. Moreno), [r.w.poppe@uu.nl](mailto:r.w.poppe@uu.nl) (R. Poppe), [jlg53@cam.ac.uk](mailto:jlg53@cam.ac.uk) (J.L. Gibson), [d.k.j.heylen@utwente.nl](mailto:d.k.j.heylen@utwente.nl) (D. Heylen).

These sensors are worn or carried and provide continuous in-game measurements so the data can be accessed directly. This presents an attractive alternative to manual annotation and opens up the possibility of in-game adaptation of gameplay based on sensor measurements. While these methods are suitable for the study of play in a laboratory setting, the requirement of fitting sensors and registering them to the game session hinders their use in everyday play. In the current paper, we present an approach that overcomes this limitation by measuring exertion in real-time and unobtrusively, using overhead cameras.

Our contributions are two-fold. First, we present a method to obtain in-game measurements of physical activity using a completely unobtrusive method. We track the players using cameras and computer vision algorithms and determine their level of activity by measuring the average movement speed. We compare this approach to a number of alternative sensor-based approaches and questionnaires. Second, we demonstrate that the level of physical activity can be influenced in real-time by adapting a single gameplay element in an interactive playground. To evaluate our approach, we conduct a user study with eight groups of four players. Together, these contributions demonstrate the potential of automatically and unobtrusively measuring and modulating physical activity in AVGs.

This paper is structured as follows. Section 2 presents an overview of how physical activity is currently measured and evaluated in active video games. In Sections 3 and 4 we describe the physical setup and the design of our user studies respectively. We then present and discuss the results of this study in Section 5, and conclude in Section 6 by discussing avenues for future work.

## 2. Measuring physical activity in active video games

AVGs are interactive games that stimulate users to be physically active. They are usually played in interactive playgrounds, physical installations that include actuators such as screens, projectors, or speakers, and sensors such as cameras and accelerometers. These sensors leverage body movements to drive system interactions. AVGs that promote full-body movement generate higher levels of exertion (Peng et al., 2013). Many of these systems are designed to promote specific types of movement and social interactions (Müller et al., 2017). For instance, Tetteroo et al. designed an open-ended game to stimulate physical activity for children (Tetteroo et al., 2011). In their playground, children wear hats with infrared reflectors that are tracked in the playing area. Shapes are projected on the floor and respond to nearby children, stimulating running. Similarly, Avontuur et al. designed a game where players chase and steal the “buzz” from players by approaching them (Avontuur et al., 2014). Besides running, other forms of movement can be promoted, as Mueller et al. show in “Hanging off a bar” (Müller et al., 2012a). In this game, a river is projected on the floor and players have to hang off a bar until a small raft drifts by and the player is allowed to jump onto it to rest. Once the raft drifts away, the player has to jump and hang onto the bar again, requiring endurance and arm strength for extended play. Another exergame designed to promote arm and upper body movement is “Rapid Recovery” by Shewaga et al. (2015). It is a kayaking simulation that requires players to paddle their canoe through a virtual course displayed in front of them, using an augmented baton.

Most AVG researchers use offline methods to measure physical activity such as the manual annotation of game recordings, or asking participants to assess their experience through questionnaires (Staiano and Calvert, 2011). These validated methods provide important information about a player’s experiences, feelings, opinions or behavioral characteristics but require additional time and personnel to carry out the interviews or annotations. In-game measurements are an attractive alternative because they provide a continuous stream of detailed data in real-time from sensors; therefore they present researchers with objective data. This comes at the cost of a more challenging interpretation of the data. Traditionally, sensors have been mostly used as

input for controlling games rather than as a measuring tool for physical activity. Body-mounted trackers have been used to sense upper body motion in games (Finkelstein et al., 2011), or gloves equipped with sensors to interact with projections on walls (Toprak et al., 2012).

With the advent of affordable and accurate sensors, in-game measurements in AVGs are becoming more common. The real-time data obtained through them can be used to react on-the-fly and adapt gameplay to steer behavior in positive directions (Ketcheson et al., 2015; van Delden et al., 2014). Typically, this proceeds through actuators such as screens and projectors. Considering that players’ in-game visual representation has been shown to affect the amount of exercise that players engage in Li et al. (2014), playing with these visual representations can be used to adapt the game play.

One goal of adaptation is to make the game more balanced for players with different physical fitness levels, such as in the “Heart Burn” racing game by Stach et al. (2009). Players have to speed virtual vehicles by pedaling on a stationary bicycle, while their heart rates are measured. Instead of using cycling speed, their measured heart rates determine the vehicle speeds. Consequently, people with better physical condition have to exert themselves more to compete against less fit players, effectively balancing the game. Similarly, Mueller et al. use heart rate to allow people at different locations to jog while “feeling together” in “Jogging over a distance” (Müller et al., 2012b). Participants are equipped with heart rate sensors and headsets for communication. The volume level of the joggers is adjusted based on their individual heart rates. When the heart rate of one participant is higher, his voice sounds as if he was ahead because of the additional effort he is putting in, encouraging the lagging participant to exert more.

Researchers have started to address the analysis and evaluation of a player’s movement and activity levels without the need of wearable sensors, making the measurement completely unobtrusive. Van Delden et al. measure player movement using Kinect depth cameras in “Hang in There” (van Delden et al., 2014). In the game, players hang from a climbing harness and move laterally to collect coins projected on a screen in front of them. Landry and Pares not only analyze movement, but use this to measure the physical activity of groups of children in the “Interactive Slide” installation (Landry and Parés, 2014). They project a game on a slide so that children have to climb up and slide down to interact with the game elements. Computer vision algorithms based on the difference between subsequent frames are used to measure the total amount of movement. This approach is fairly straightforward, and provides a collective measure of exertion. Despite the reported performance, the use of pixel difference is prone to issues of reliability due to the influence of clothing, body size, camera viewpoint and lighting conditions on the measurements.

We propose a pervasive approach for unobtrusive, in-game measurement of player activity during group play. By tracking the individual players, we overcome typical robustness issues when directly measuring exertion from sequences of video frames. Furthermore, we demonstrate how the level of activity can be influenced by adapting a gameplay element, effectively demonstrating the potential of real-time measurement of player exertion.

## 3. Measuring physical activity in an interactive playground

In this section, we detail the interactive playground in which we conduct our user study (Section 4). We then describe how we measure player exertion using depth cameras and computer vision algorithms.

### 3.1. Interactive tag playground

The Interactive Tag Playground (ITP) is an installation where players can play an enhanced version of the game of tag (Moreno et al., 2016). It provides a 6 m × 5 m space with floor projections. The location of each player is measured using four Kinects V1, located in the ceiling of the playground. Two projectors that are also mounted in the

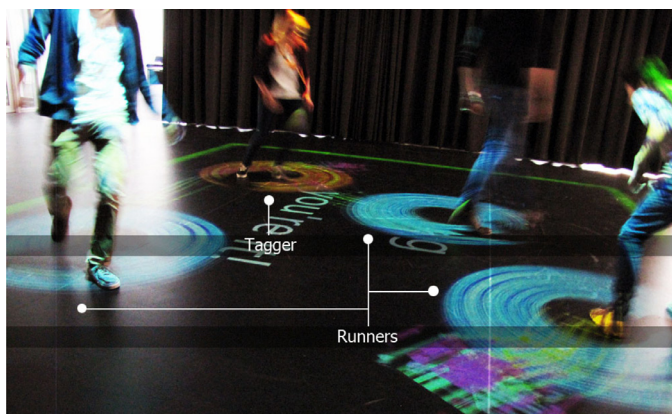


Fig. 1. People playing tag in the ITP.

ceiling, project circles around the estimated position of each player. The color of the circle represents the role of the player: orange for the tagger and blue for runners (see Fig. 1). To tag another player, the tagger has to make her/his circle touch that of a runner. Upon the tag, the roles of the two players switch. The new runner cannot be tagged for two seconds, to encourage the new tagger to look for another player.

The floor projections are based on the positions of the players, as discussed in the next subsection. As such, there is a feedback loop where the behavior of the players can be shaped by altering how the floor projections take into account the players' behaviors. In previous work, it has been demonstrated that careful adaptation of the floor projections could be used to increase proximity between players (van Delden et al., 2014) and stimulate risk taking (van Delden et al., 2017). In this paper, specifically, we experiment with different sizes of the circle that is projected around each player. Larger circles reduce the effort needed to tag another player. Consequently, we expect that the size of the projected circle affects the amount of exertion.

### 3.2. Unobtrusive measurement of exertion

We propose to measure the amount of activity in a completely unobtrusive manner through computer vision. We track players during the game and calculate their movement speeds. This allows us to estimate their physical activity levels.

To track players, we use depth images from Kinect sensors located above the play area and facing downwards. Since the depth images are obtained using an infrared sensor, the light from the overhead projectors is not an issue. Also, the sensor measurements are not sensitive to indoor lighting conditions, which is an additional advantage. Basic filtering operations are applied on the depth images to remove noise and enhance the head-shoulder region of the players. Effectively, this allows us to focus on those regions in the image that are closer to the camera. In a playground without physical props, these areas correspond to the players. We isolate these regions by looking for local peaks in the depth image. We then use a Kalman filter for each peak to track the corresponding player. Kalman filters predict the player's future location based on a motion model and live observations. The model is updated continuously based on the assigned detection's position.

Given proper calibration of the depth cameras, we can project pixel distances in the image to distances on the ground plane, in meters. By summing distances between subsequent measurements over one-second time intervals, we can nicely approximate the players' speed in meters per second. The tracker has been evaluated in Moreno et al. (2016) and performs well. Still, when players bump into each other, their tracks are occasionally swapped. While this is a rare event, we cannot guarantee that this never happens. From depth recordings, players cannot be easily differentiated or identified. While this is an advantage for player privacy, it prevents us from recovering from the track swaps. As a

consequence, we average the speeds of all players at a given moment. Our measure of exertion based on the speed of tracked players is therefore a group measure.

## 4. Experiment design

We conducted a user study in the ITP with two goals. Firstly, to demonstrate that players' speed can be used as a reliable measurement of physical activity in interactive playgrounds, we compare our estimates with a number of reported activity measurements including self-report and heart rate. Our second goal is to modulate physical activity in the ITP by changing players' circle sizes to manipulate the amount of effort it takes to tag other players. The experiment described here was approved by our university's ethical committee.

### 4.1. Participants

The participants were BSc, MSc and PhD students from the University of Twente. While these participants are not the intended target audience of interactive playgrounds, they are an easily accessible participant group. Moreover, for the aim of the present study, we focus on the automated measurement of exertion. Aspects such as specific interactions and enjoyment, which are arguably more age-dependent, are not the focus. Students were approached at the university and asked if they wanted to voluntarily participate in a 30min study in which they were going to play tag. If they agreed, they were taken to the playing area. In eight sessions with four players each, we had a total of 32 participants: 23 male, 9 female (age 19–28, mean 21.9, standard deviation 2.36).

### 4.2. Design

We ran a series of 8 playground tag sessions, each involving a group of four players. In each session, participants played three tag game sub-sessions with breaks in between. In each sub-session, the group played the tag game with different circle sizes (see Fig. 2). The first sub-session used the standard circle size and served to familiarise players with the game play. In the other two sub-sessions, which had a duration of 4 minutes each, players played 2 different versions of the game that were designed to test whether the size of the circles affected players' activity levels. In the High Exertion Condition (HEC), the circle size was smaller than the standard size while the circle was bigger than the standard size in the Low Exertion Condition (LEC). Specifically, the diameter of the circles was 102 cm in the standard condition, 66 cm in the HEC, and 149 cm in the LEC. The size of the circles in the HEC was set to resemble the average shoulder width of young adults, whereas the size in the LEC was set to approximate an arm's reach. Within each condition, the size of the circles did not change. To allow evaluation of possible order effects, we counterbalanced the conditions such that groups in the odd-numbered sessions played tag in the HEC before the LEC, and groups in the even-numbered sessions did the reverse. The study thus has a within-participants design with exertion condition (or circle size) as the independent variable.

### 4.3. Measurements

We used four exertion measures, including our novel, unobtrusive measure:

- **Speed of tracked players** is our newly proposed exertion measure. It is calculated from the player's tracked position, at 15 frames per second (see Section 3.2). The speed information is obtained by calculating the average track displacement per second, over all players. To eliminate the noise inherent to data collection, we applied a median filter with a window size of one third of a second on the position data, and we interpolated position values when tracks

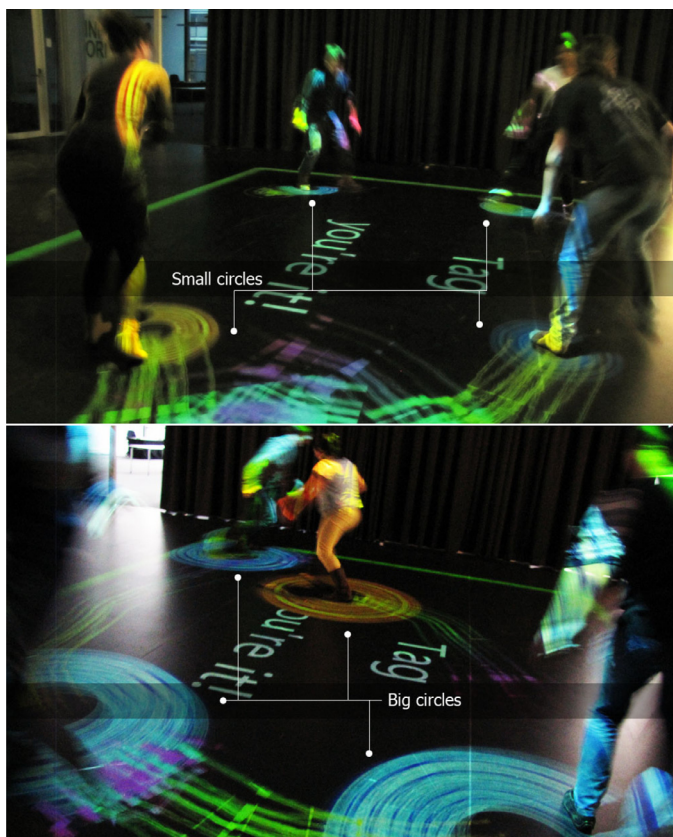


Fig. 2. Difference between circle sizes in the HEC(t) and LEC(b).

went missing.

- **Heart rate monitors** are widely used to determine exercise intensity (Achten and Jeukendrup, 2003), and provide ground truth measurements in this study. The unit used for the heart rate (HR) measurements is beats per minute (bpm). We used Scosche Rhythm Plus heart rate monitors, which were strapped to each player’s right forearm. The sensor is claimed to be accurate within 1 bpm. We measured HR every second for the entire duration (four minutes) of the condition sub-sessions.
- **Accelerometers** measure the amount of acceleration to which the sensor is subjected, a measurement related to the amount of movement of the user. Accelerometers have been shown to measure

exertion reliably in previous studies (Loprinzi and Cardinal, 2011; Sirard and Pate, 2001). We used YEI 3-Space wireless accelerometers for this study, which were placed in the right pockets of the players. We collect acceleration data 15 times per second. We applied a median filter with a window size of one third of a second to remove noise and interpolate missing values. The unit used for the acceleration values is  $g$ , the gravitational constant.

- **Pixel difference** is a computer vision method by which the difference between consecutive images is calculated. In the case of a video feed of people moving, the difference between consecutive frames yields an approximate measurement of how much movement is present in the sequence. Because we use floor projections and those can cause differences between frames even when the players do not move, we use the depth images from the Kinect instead of the traditional RGB images. To obtain the pixel difference value, we subtract consecutive depth images obtained from the Kinects 15 times per second. We applied basic morphological operations (dilation, erosion) and a median filter (window size of one third of a second) on the actual pixel difference values to remove noise.

For all measurements, we discarded the first two minutes of each recording because during this period the HR is rising (see Fig. 3). During the last two minutes, HR has stabilized and more accurately represents activity levels. We calculate, for each measure, the average value over the last two minutes. We furthermore average the measurements per group of players to be able to directly compare all measures, including our novel measure and pixel difference, which are both group measures. This means we are averaging individual differences. Also, this effectively reduces the number of observations, which makes statistical analyses conservative. Still, we believe that a group measure is appropriate because the behavior of all players is correlated.

In addition to the exertion measures, we used a questionnaire. It served two purposes. The first was to evaluate the perceived exertion of the players after a condition session. The second goal was to keep players occupied during the break so they would rest. Perceived exertion was measured using Borg’s Rating of Perceived Exertion (RPE) Scale (Borg, 1982). The Borg scale is a linear scale from 6 to 20. The range of the scale was designed to broadly represent the HR of healthy adults. A perceived exertion of 10 should coincide with a HR of 100 bpm. We also asked additional questions about players’ estimated fitness level, physical characteristics (height, weight) and preference of game elements. This information was not used for this study.

Below, we introduce, motivate and operationalize our hypotheses:

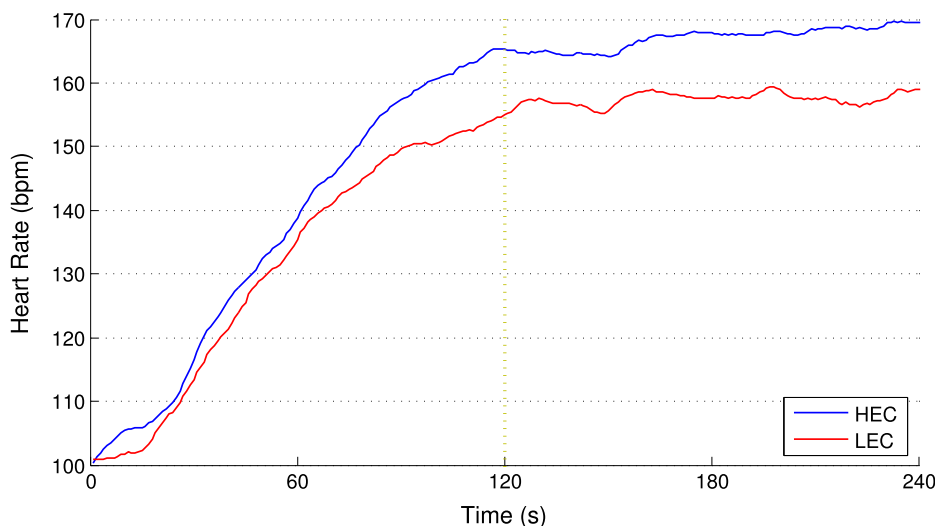


Fig. 3. Average HR measurements of all players in the LEC and HEC. Only the last two minutes of play are used for the data analysis.

- **Physical activity and circle size** With larger circles, less effort is required to tag others, potentially resulting in lower HR. Therefore, it should be possible to influence the amount of physical activity of players by varying circle sizes.

**H1:** The HR of players is higher in the HEC than in the LEC.

We compare, per group, the average HR of all players of the last two minutes in the LEC and HEC sub-sessions, respectively.

- **Average player speed and circle size** With larger circles, players should be able to run slower and still tag other players. Therefore, players' speed measurements through tracking should be lower in the LEC than in the HEC.

**H2:** The speed of the players is higher in the HEC than in the LEC. This hypothesis checks whether we can use player speed to measure physical activity. We compare the average speed of all players in a group for the last two minutes of the LEC and HEC sub-sessions.

- **Heart rate and player speed** The relation between heart rate and exertion has been well-studied (e.g., Boudet et al. (2004)). Exertion measurements not only depend on the amount of physical activity a player is undertaking, but also on his or her fitness level and body properties. This introduces a bias for different players. Still, players' speed should affect the amount of effort players are putting in when playing. Therefore, the speed at which players move should be related to their HR.

**H3:** The speed and HR of the players are positively correlated. This hypothesis checks whether we can replace HR measurements with players' speed measurements to measure physical activity. To test this hypothesis, we use the average speed and HR of all players in a group for the last two minutes of the LEC and HEC sub-sessions.

#### 4.4. Procedure

Before each session, all players were asked to read a consent form with a description of the game. After signing, players were taken to the playing area, were given an explanation of how the game works and left to play for one minute. Afterwards, they were asked to sit down for four minutes. During this time, the HRMs and accelerometers were fitted. The accelerometers were put in the players' trouser pockets, and the HRMs were secured to the upper part of the players' forearms. Once the break was over, players were asked to return to the playing area to play the second sub-session for four minutes. Then they filled in a short questionnaire while they rested. After the break, they were asked to play the last sub-session. Afterwards, they sat down again, filled in the last part of the questionnaire, and engaged in a short feedback session. Finally, we collected the sensors.

### 5. Experiment results

Our analyses are based on data from the questionnaires, the HRMs, the accelerometers and the depth cameras.

#### 5.1. Measuring physical activity using tracking

First, we verify whether the LEC and HEC conditions elicited different amounts of physical activity. Therefore, we check whether the HR measurements using the HRMs differed between conditions. In Fig. 3, the difference between conditions is noticeable almost from the beginning. To check whether the difference between conditions is statistically significant, we conducted a 2-tailed paired samples *t*-test. The test shows on average statistically higher HR in the HEC compared to the LEC ( $t(7) = 3.2, p < 0.05$ ), with a difference of 9.6 bpm. This confirms hypothesis **H1** that the HEC promotes more exertion than the LEC and that circle size can be used to influence the amount of activity.

With **H1** supported, we investigate whether the speed also differs between conditions. The speed of the players in both conditions can be seen in Fig. 4. We ran a 2-tailed paired samples *t*-test to evaluate whether speed differs significantly between conditions. The test shows

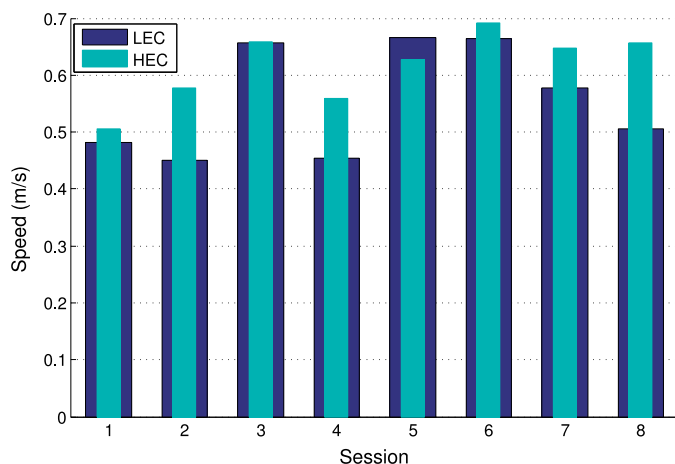


Fig. 4. Average group speed per condition, per group.

a statistically significantly higher average speed in the HEC compared to the LEC ( $t(7) = 2.5, p < 0.05$ ). This supports our hypothesis **H2** that players move faster in the HEC than in the LEC. The difference of 0.06 m/s between the conditions is small overall. In general, taggers run faster than runners in tag games (Moreno and Poppe, 2016), and we can investigate how circle size affects the speed for each role individually. We find that taggers are not affected much by the size of the circles (Table 1). Runners, on the other hand, have a significantly higher average speed in the HEC compared to the LEC (0.07 m/s difference,  $t(7) = 3.1, p < 0.05$ ). This difference hints at the possibility of balancing physical activity between roles in the ITP. In Session 5, the measurement is opposite to what we expected. This is discussed in more detail in Section 5.3.

Finally, given that both HR and speed significantly differ between conditions, we check whether the two variables are correlated. We used a 2-tailed bi-variate Pearson correlation test and found a statistically significant correlation ( $r = 0.72, p < 0.01$ ). Pearson's *r* measures the degree of linear relationship between two variables, the strength of their relationship. An *r* value of 0.72 is considered a strong relationship. Next, we check whether speed can be used as a predictor for heart rate. To this end, we use the Spearman rank correlator to estimate the magnitude of association between speed and HR. We found  $\rho = 0.74$ , which again indicates a strong relation between the two variables. Fig. 5(a) shows the regression line, with 95% confidence interval. We accept hypothesis **H3**. It is thus possible to use tracking and calculate average speed to measure physical activity as an alternative to HRMs.

#### 5.2. Comparison to other activity measurement methods

We compared our proposed approach to two other physical activity measurement methods: pixel difference and accelerometers. We also compare the actual exertion with the perceived exertion, obtained from the questionnaires.

##### 5.2.1. Pixel difference

Pixel difference has been used in interactive playgrounds to measure the physical activity of groups of children (Landry and Parés, 2014). An

Table 1  
Players' average speed values per role and condition.

		Speed (m/s)	
Role	Condition	Mean	SD
Runner	LEC	0.52	0.10
	HEC	0.59	0.06
Tagger	LEC	0.66	0.07
	HEC	0.68	0.07

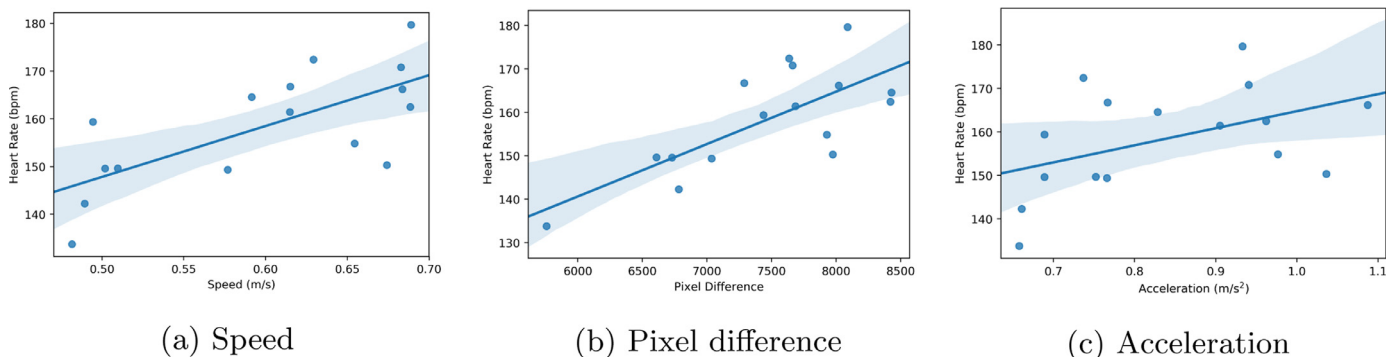


Fig. 5. Scatter plots showing relation of predictors speed (a), pixel difference (b) and acceleration (c) to heart rate. Regression line with 95% CI is shown.

issue with this approach is that it can be affected by elements not related to movement. Factors unrelated to exertion such as body size, clothing, viewpoint, occlusion can also affect the pixel count. Pixel difference is inherently a global feature, which means it considers the entire image for a single measurement. As a consequence, pixel difference can only be used as a group measure.

We found a statistically significantly higher pixel difference count of 0.97% in the HEC compared to the LEC ( $t(7) = 2.7, p < 0.05$ ). Spearman rank correlation resulted in  $\rho = 0.63$ , which indicates a slightly lower association between pixel difference and HR than between speed and HR (see also Fig. 5(b)). As we can see in Fig. 6, Session 3 did not show a higher number of changed pixels in the HEC. Sessions 1 and 5 also show a limited increase.

5.2.2. Accelerometer

Compared to tracking people and counting pixel differences, using accelerometers does not require having a computer vision system in place. Rather, accelerometers need to be attached to the body or clothing. This limits their practical value in autonomous public installations as additional personnel would be required to hand out and collect the sensors. Since accelerometers measure changes in speed, the measurements depend on where on the body they are attached (Rosenberger et al., 2013). To mitigate this issue, all players put the accelerometers in their right trousers pockets.

We found a statistically significantly higher average acceleration in the HEC compared to the LEC ( $t(7) = 2.5, p < 0.05$ , difference of 0.07g). Spearman rank correlation ( $\rho$ ) between acceleration and HR was 0.49, see also Fig. 5(c). From Fig. 7, it can be observed that acceleration also exhibits unexpected measurements in Session 5.

5.2.3. Perceived exertion analysis

Perceived exertion informs us whether players were aware they

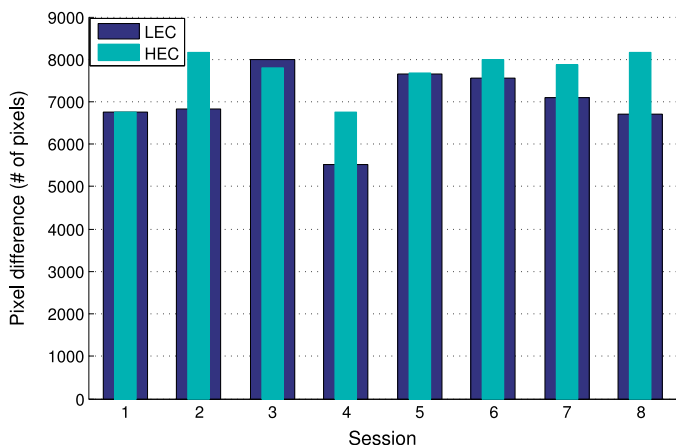


Fig. 6. Average group pixel difference per condition, per group.

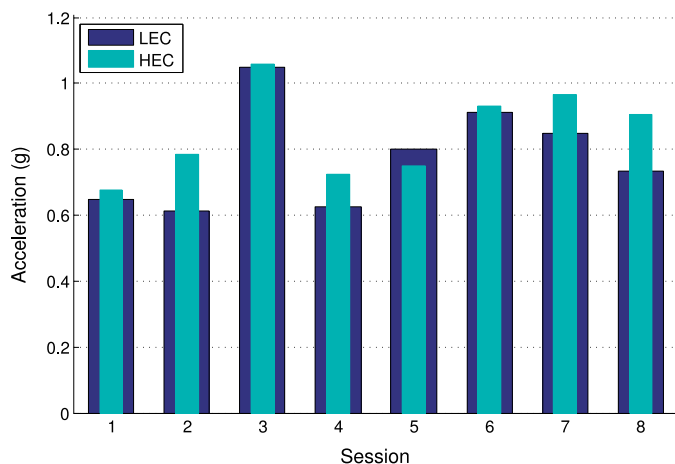


Fig. 7. Average group acceleration per condition, per group.

were exerting differently in the two conditions. This could be valuable in the design of future game interventions. Promoting high levels of exertion without players being aware of it could lead to longer play sessions. For this study, players rated their perceived exertion after each condition, resulting in two ratings per player.

To analyze the difference in perceived exertion, we conducted a 2-tailed paired samples  $t$ -test on the RPE questionnaire answers. This showed a statistically significantly higher perceived exertion in the HEC compared to the LEC on the Borg scale (difference of 1.88 corresponding to a 18.8 bpm) ( $t(31) = 6.2, p < 0.01$ ). This is roughly twice the difference measured using HRM (9.6 bpm). Players' higher exertion in the HEC is also consistent with their reports on the Borg scale. From the measured and perceived HR in Table 2, we observe that players underestimated their amount of exertion in both conditions, but especially in the LEC. This conveys that players thought they were exerting less than they actually were.

5.3. Discussion

The findings reported in this paper are important for a number of reasons. First, we have shown that completely unobtrusive measurements of physical activity are possible. Such measurements can alleviate the workload required to evaluate interactive playgrounds. In

Table 2  
Perceived and measured exertion (HR) for both conditions.

	HR (bpm)		
Condition	RPE	Perceived	Measured
LEC	12.75	127.50	157.53
HEC	14.63	146.30	167.08

**Table 3**  
Comparison of physical activity measurement methods by exertion measure.

Physical activity method	HRM	Tracking	Pix. diff.	Accel.
$\rho$ -HRM	1	0.74	0.63	0.49
Tracking		✓		
Contactless		✓	✓	
Individual	✓	Possible		✓

cases where the use of questionnaires or observational studies is not possible, our method can provide an estimation of the amount of physical activity. Second, these physical activity measurements can be used in-game to adapt gameplay and steer behavior.

Specifically in the ITP, we have shown that changing the size of the circles affects the speed at which players run. This could be useful in practice to ensure that players exert themselves as intended. For example, the circles could be enlarged or shrunk to modulate physical activity. Given the larger effect of circle size on the runners’ speed, we can also influence exertion levels through role assignment.

Table 3 shows a comparison of the investigated methods for physical activity measurements. Tracking players and measuring their speed is unobtrusive, has a strong correlation to HR, and could potentially be used to assess physical activity individually. Currently, estimating physical activity individually is only possible by manually correcting players’ tracks. However, our tracker can provide functionality beyond the assessment of group physical activity, such as analyzing movement patterns or estimating distances between players.

Pixel difference is also unobtrusive and has a moderate correlation to HR measurements. It does not require tracking but is affected by factors not related to physical effort such as the number of players, their size or their clothing. Also, without tracking, it is impossible to focus on individual players. Finally, accelerometers could measure physical activity differences between conditions, but their correlation to HR is limited. Also, even though accelerometers allow for measuring both group and individual activity, they are not unobtrusive. For interactive playgrounds, accelerometers are therefore not suitable to measure exertion.

We now turn to the different sessions. In Sessions 1, 3 and 5, measurements are contrary to what we expected. Since all odd-numbered sessions started with the HEC, we looked at the speed of the players based on whether they were played first or last, to see if there were any order effects (Table 4). Surprisingly, the speed at which players ran in the second condition was always higher. Due to exhaustion, we had predicted that it would be more likely for players to run slower in conditions played last, but the opposite was true. This might be because players were more willing to exert themselves in the last sub-session, having already invested significant effort in the previous one. It could also be that they were “warmed up” and gained confidence. This would explain why the expected difference between conditions was small or sometimes reversed in Sessions 1, 3 and 5. On average, the speed of the HEC when played first was only slightly higher than the LEC when played last. This effect is not very visible when the HEC was played last, probably because players are already

**Table 4**  
Speed values of players per condition and playing order using different time-windows.

	Playing order	Speed (m/s)		Speed (m/s)	
		(2 min)	(4 min)	(2 min)	(4 min)
LEC	First	Mean 0.52	SD 0.09	Mean 0.56	SD 0.10
	Last	Mean 0.60	SD 0.12	Mean 0.60	SD 0.12
HEC	First	Mean 0.61	SD 0.12	Mean 0.64	SD 0.12
	Last	Mean 0.62	SD 0.11	Mean 0.65	SD 0.11

running fast due to the intervention.

Since players ran faster in the sub-sessions that were played last, we looked at the speed of the players over time within each sub-session. In Fig. 8, we can see that player speed diminished over time within a session, probably due to exhaustion. The decrease in speed is not linear, but follows a sinusoidal-like pattern, with consecutive peaks and valleys of speed. This conveys that players had outbursts of high speed, followed by short recovery periods, and that all players adapted to this pattern.

This also provides some insight as to why player speeds in the LEC and HEC were very similar when the LEC was played last. Since we only used measurements obtained during the last two minutes of each sub-session, we only looked at the segments of play where exhaustion was kicking in. Since the HEC requires a higher level of effort, the onset of exhaustion should have been faster and more pronounced than in the LEC. Indeed, Table 4 shows that when calculating the speed of each condition based on the playing order, and taking into account the whole session, the difference between both conditions is more evident. When we calculate the average group speeds in both conditions using the whole session (Fig. 9), the speed in the HEC is higher in all sessions.

5.3.1. Limitations

These results show that the proposed approach has many merits. In its current state, there are also shortcomings. First, because the scope of our work is interactive playgrounds, our approach is designed to work in playing spaces of limited size. Increasing the playing area would require additional equipment.

A second limitation stems from the use of cameras to sense behavior. While we can measure player positions unobtrusively, there is information that cannot be easily measured visually, such as the age or the fitness level of the players. While this ensures privacy of the players, the lack of personal information might be an issue when studying properties of play. In such cases, questionnaires or interviews are necessary. Similarly, not all behavioral cues can be robustly measured visually. In these cases, observational approaches such as manually annotating behavior are needed.

Third, we have performed our user study with, mostly male, university students. Playgrounds, traditional and interactive, have predominantly younger audiences. While the play behavior between university students and primary school children is likely to be very different, the relation between exertion and movement on the playground might be more similar. More research is needed to understand how our findings relate to a younger audience.

Finally, we only showed that group speed measurements can be used to measure differences in group physical activity. When averaging speed and HR values for all players, we may be discarding relevant information that could lead to better insights on player behavior. Our work can be considered an important step towards individual-based measures. Future work should address the robustness of the tracking in terms of maintaining identity information for the tracks.

6. Conclusions and future work

We have investigated whether group physical activity can be measured unobtrusively by tracking players. To this end, we have conducted a user study in which we successfully manipulated the exertion levels of the players by varying one parameter of the Interactive Tag Playground. We showed that by tracking the players and measuring their speed, we could assess differences in group physical activity levels. Finally, we showed that speed measurements are strongly correlated to heart rate, which validates the usefulness of player tracking as a tool to measure physical activity.

We compared our approach to alternative physical activity measurement methods: using heart rate monitors, counting pixel differences and using accelerometers. Pixel difference also makes use of computer vision and is a suitable tool to assess physical activity, but only at a

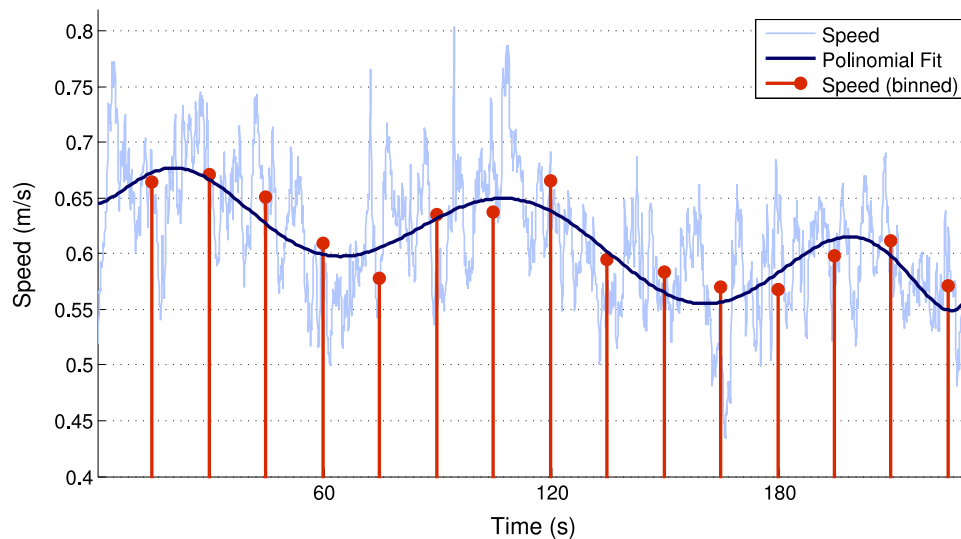


Fig. 8. Average player speed during a sub-session over all sessions.

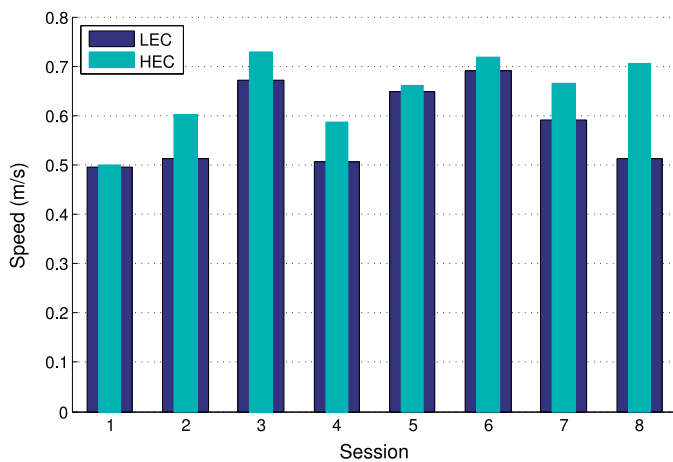


Fig. 9. Average group speeds for each group using the measurements of the entire session.

group level. Accelerometers showed a lower correlation compared to HR, and neither accelerometers and heart rate monitors are unobtrusive. Given that interactive playgrounds should run autonomously, with players being able to go in and out without hassle, we believe tracking players is an appropriate method. Moreover, such unobtrusive in-game measurements enable the system to respond to events on-the-fly.

Future work will be aimed at improving the tracker. If the tracker could reliably keep track of every player, we could measure physical activity of every player individually, which would provide better insight into their behavior. Additionally, we would like to measure how a player moves, instead of just how much. The quality of the movement can be further analyzed to understand play (Castañer et al., 2017). Another direction of research is to apply these findings in-game by setting specific exertion goals and automatically adapting the game to meet these. If we recognize that a particular player is not moving much, the system could assign her/him to the tagger role to encourage more movement. Changing circle size on an individual basis could also be an option. Even though adjusting circles sizes is inherently tied to the ITP, the automated measurement of physical activity using tracking can be easily used to trigger other interventions in different playgrounds.

Interactive playgrounds have value as entertainment tools, and we have shown that this is not their only use. They can be used to study play behavior, and to steer behavior to meet desired goals. Tracking is a

promising method that enhances this potential. We believe that game interventions based on tracked players open up exciting opportunities for game design and evaluation.

#### Acknowledgments

This publication was supported by the Dutch national program COMMIT.

#### References

- Achten, J., Jeukendrup, A.E., 2003. Heart rate monitoring. *Sports Med.* 33 (7), 517–538.
- Aggio, D., Ogunleye, A., Voss, C., Sandercock, G., 2012. Temporal relationships between screen-time and physical activity with cardiorespiratory fitness in english school-children: a 2-year longitudinal study. *Prev. Med.* 55 (1), 37–39.
- Altimira, D., Müller, F., Clarke, J., Lee, G.A., Billinghamurst, M., Bartneck, C., 2017. Enhancing player engagement through game balancing in digitally augmented physical games. *Int. J. Hum. Comput. Stud.* 103, 35–47.
- Avontuur, T., de Jong, R., Brink, E., Florack, Y., Soute, I., Markopoulos, P., 2014. Play it our way: customization of game rules in children's interactive outdoor games. *Proceedings of the Interaction Design and Children Conference*. pp. 95–104.
- Bakker, S., Markopoulos, P., de Kort, Y., 2008. OPOS: an observation scheme for evaluating head-up play. *Proceedings of the Nordic Conference on Human-Computer Interaction*. pp. 33–42.
- Blumberg, F., Blades, M., Oates, C., 2013. Youth and new media: the appeal and educational ramifications of digital game play for children and adolescents. *Zeitschrift für Psychologie* 221 (2), 67–71.
- Borg, G.A., 1982. Psychophysical bases of perceived exertion. *Med. Sci. Sports Exerc.* 14 (5), 377–381.
- Boudet, G., Albuissou, E., Bedu, M., Chamoux, A., 2004. Heart rate running speed relationships during exhaustive bouts in the laboratory. *Can. J. Appl. Physiol.* 29 (6), 731–742.
- Calvert, S., Staiano, A., Bond, B., 2013. Electronic gaming and the obesity crisis. *New Dir. Child Adolesc. Dev.* 2013 (139), 51–57.
- Carson, V., Hunter, S., Kuzik, N., Gray, C.E., Poitras, V.J., Chaput, J.-P., Saunders, T.J., Katzmarzyk, P.T., Okely, A.D., Connor Gorber, S., Kho, M.E., Sampson, M., Lee, H., Tremblay, M.S., 2016. Systematic review of sedentary behaviour and health indicators in school-aged children and youth: an update. *Appl. Physiol. Nutr. Metab.* 41 (6 (Suppl. 3)), S240–S265.
- Castañer, M., Camerino, O., Landry, P., Pares, N., 2017. Quality of physical activity of children in exergames: sequential body movement analysis and its implications for interaction design. *Int. J. Hum. Comput. Stud.* 96, 67–78.
- van Delden, R., Moreno, A., Poppe, R., Reidsma, D., Heylen, D., 2017. A thing of beauty: steering behavior in an interactive playground. *Proceedings of the International Conference on Human Factors in Computing Systems (CHI)*. pp. 2462–2472.
- van Delden, R., Moreno, A., Ramos, C., Carrasco, G., Reidsma, D., Poppe, R., 2014. Hang in There: A Novel Body-centric Interactive Playground. In: Rybarczyk, Y., Cardoso, T., Rosas, J., Camarinha-Matos, L. (Eds.), *Innovative and Creative Developments in Multimodal Interaction Systems*. Springer Berlin Heidelberg, pp. 160–178.
- Desai, R.A., Krishnan-Sarin, S., Cavallo, D., Potenza, M.N., 2010. Video-gaming among high school students: health correlates, gender differences, and problematic gaming. *Pediatrics* 126 (6), 1414–1424.
- Ferguson, C.J., Olson, C.K., 2013. Friends, fun, frustration and fantasy: child motivations for video game play. *Motiv. Emot.* 37 (1), 154–164.



- Finkelstein, S., Nickel, A., Lipps, Z., Barnes, T., Wartell, Z., Suma, E.A., 2011. Astrojumper: motivating exercise with an immersive virtual reality exergame. *Presence* 20 (1), 78–92.
- Hands, B., Larkin, D., 2006. Physical activity measurement methods for young children: a comparative study. *Meas. Phys. Educ. Exerc. Sci.* 10 (3), 203–214.
- Ketcheson, M., Ye, Z., Graham, T., 2015. Designing for exertion: How heart-rate power-ups increase physical activity in exergames. *Proceedings of the Symposium on Computer-Human Interaction in Play*. pp. 79–89.
- Landry, P., Parés, N., 2014. Controlling and modulating physical activity through interaction tempo in exergames: a quantitative empirical analysis. *J. Ambient Intell. Smart Environ.* 6 (3), 277–294.
- Li, B.J., Lwin, M.O., Jung, Y., 2014. Wii, myself, and size: the influence of proteus effect and stereotype threat on overweight children's exercise motivation and behavior in exergames. *Games Health* 3 (1), 40–48.
- Loprinzi, P.D., Cardinal, B.J., 2011. Measuring children's physical activity and sedentary behaviors. *J. Exer. Sci. Fitness* 9 (1), 15–23.
- Moreno, A., van Delden, R., Poppe, R., Reidsma, D., Heylen, D., 2016. Augmenting playspaces to enhance the game experience: a tag game case study. *Entertain. Comput.* 16, 67–79.
- Moreno, A., Poppe, R., 2016. Automatic behavior analysis in tag games: from traditional spaces to interactive playgrounds. *J. Multimodal User Interfaces* 10 (1), 63–75.
- Moreno, A., van Delden, R., Poppe, R., Reidsma, D., 2013. Socially aware interactive playgrounds. *IEEE Pervasive Comput.* 12 (3), 40–47.
- Moreno, A., van Delden, R., Reidsma, D., Poppe, R., Heylen, D., 2012. An annotation scheme for social interaction in digital playgrounds. *Proceedings of the International Conference on Entertainment Computing*. pp. 85–99.
- Müller, F., Gibbs, M.R., Vetere, F., Edge, D., 2017. Designing for bodily interplay in social exertion games. *ACM Trans. Comput.-Hum. Interact.* 24 (3), 24–41.
- Müller, F., Khot, R.A., Gerling, K., Mandryk, R.L., 2016. Exertion games. *Found. Trends Human-Comput. Interact.* 10 (1), 1–86.
- Müller, F., Toprak, C., Graether, E., Walminck, W., Bongers, B., van den Hoven, E., 2012. Hanging off a bar. *Extended Abstracts of the Conference on Human Factors in Computing Systems*. pp. 1055–1058.
- Müller, F., Vetere, F., Gibbs, M., Edge, D., Agamanolis, S., Sheridan, J., Heer, J., 2012. Balancing exertion experiences. *Proceedings of the Conference on Human Factors in Computing Systems*. pp. 1853–1862.
- Peng, W., Crouse, J.C., Lin, J.-H., 2013. Using active video games for physical activity promotion: a systematic review of the current state of research. *Health Educ. Behav.* 40 (2), 171–192.
- Plötner, M., Over, H., Carpenter, M., Tomasello, M., 2015. The effects of collaboration and minimal-group membership on children's prosocial behavior, liking, affiliation, and trust. *J. Exp. Child. Psychol.* 139, 161–173.
- Poppe, R., van Delden, R., Moreno, A., Reidsma, D., 2014. Interactive Playgrounds for Children. In: Nijholt, A. (Ed.), *Playful User Interfaces*. Springer Verlag, pp. 99–118.
- Radesky, J.S., Christakis, D.A., 2016. Increased screen time: implications for early childhood development and behavior. *Pediatr. Clin. North Am.* 63 (5), 827–839.
- Rosenberger, M.E., Haskell, W.L., Albinati, F., Mota, S., Nawyn, J., Intille, S., 2013. Estimating activity and sedentary behavior from an accelerometer on the hip or wrist. *Med. Sci. Sports Exerc.* 45 (5), 964–975.
- Shewaga, R., Rojas, D., Kapralos, B., Brennan, J., 2015. Alpha testing of the rapid recovery kayaking-based exergame. *Proceedings of the Games Entertainment Media Conference*. pp. 1–5.
- Sirard, J.R., Pate, R.R., 2001. Physical activity assessment in children and adolescents. *Sports Med.* 31 (6), 439–454.
- Stach, T., Graham, T.C.N., Yim, J., Rhodes, R.E., 2009. Heart rate control of exercise video games. *Proceedings of Graphics Interface*. pp. 125–132.
- Staiano, A.E., Calvert, S.L., 2011. The promise of exergames as tools to measure physical health. *Entertain. Comput.* 2 (1), 17–21.
- Tetteroo, D., Reidsma, D., van Dijk, E., Nijholt, A., 2011. Design of an interactive playground based on traditional children's play. *Proceedings of the International Conference on Intelligent Technologies for Interactive Entertainment*. pp. 129–138.
- Toprak, C.C., Platt, J., Müller, F., 2012. Bubble popper: considering body contact in games. *Proceedings of the International Conference on Fun and Games*. pp. 97–100.
- van Delden, R., Moreno, A., Reidsma, D., Poppe, R., Heylen, D., 2014. Steering gameplay behavior in the Interactive Tag Playground. *Proceedings of European Conference on Ambient Intelligence*. pp. 145–157.