

Assessing Dynamic Balance Performance during Exergaming based on Speed and Curvature of Body Movements

Venustiano Soancatl Aguilar, Jasper J. van de Gronde, Claudine J.C. Lamoth, Natasha M. Maurits, *Senior Member, IEEE* and Jos B.T.M. Roerdink, *Senior Member, IEEE*

Abstract—Improving balance performance among the elderly is of utmost importance because of the increasing number of injuries and fatalities caused by fall incidences. Digital games controlled by body movements (exergames) have been proposed as a way to improve balance among older people. However, the assessment of balance performance in real-time during exergaming remains a challenging task. This assessment could be used to provide instantaneous feedback and automatically adjust the exergame difficulty. Such features could potentially increase the motivation of the player, thus augmenting the effectiveness of exergames. As clear differences in balance performance have been identified between older and younger people, distinguishing between older and younger adults can help identifying measures of balance performance. We used generalized linear models to investigate whether the assessment of balance performance based on movement speed can be improved by incorporating curvature of the movement trajectory into the analysis. Indeed, our results indicated that curvature improves the performance of the models. Five-fold cross validation indicated that our method is promising for the assessment of balance performance in real-time by showing more than 90% classification accuracy. Finally, this method could be valuable not only for exergaming but also for real-time assessment of body movements in sports, rehabilitation and medicine.

Index Terms—Assessing dynamic balance performance, exergaming, speed and curvature, generalized linear models, Markov chain Monte Carlo estimation.

I. INTRODUCTION

THE assessment of the quality of body movements in real-time is of utmost importance in exergames, that is, digital games controlled by body movements. Older adults form a special target group for exergaming [1]. Because the number of injuries as well as the number of fatalities caused by fall incidences among older people have increased during the last decade [2], the ultimate goal of exergames is not only to provide fun and entertainment but also to improve postural balance performance. It is known that improving balance can reduce the incidence of falls among the older population [3]. Assessing balance during exergaming could be used to adapt the difficulty of the game as a function of the quality of body movements of the player, as well as to provide appropriate immediate feedback. This in turn could increase motivation of

Manuscript received ...; revised The project was financially supported by the Northern Netherlands Provinces Alliance, Course for the North.

Venustiano Soancatl Aguilar was supported by the Mexican National Council of Science and Technology (CONACYT) under scholarship number 313791 (correspondence e-mail: v.soancatl.aguilar@rug.nl)

the player and also increase the effectiveness of exergames as tools to improve balance [4].

Traditional methods to determine the effectiveness of exergames rely on the assessment of balance before and after intervention tests [1]. Real-time methods to assess balance are still scarce and a gold standard for dynamic balance assessment has not yet been established. Methods for real-time balance quantification of whole body movements as recorded by devices such as Microsoft Kinect are yet to be developed [5].

Common home exergame devices such as Microsoft Kinect 1 reliably capture body movements that could be used to assess balance performance during exergaming [5]. It is known that younger adults have better postural performance than older adults [6]. There is also evidence of significant deterioration of postural performance, as evaluated from the center of pressure (CoP), over 60 years of age [7]. Hence, distinguishing between older and younger adults on the basis of movement characteristics can help to identify measures of (dynamic) balance performance. One way to identify older and younger adults is by means of generalized linear models (GLMs) [8]. Assuming that the actual age of a group of people is unknown (or temporarily ignored), we use GLMs to predict their age based on curvature and speed of their body motion. The results of this model prediction are meaningful values that can be interpreted as *probabilities* of the people belonging to a younger or an older age class. In addition, these GLMs can be used in real time during exergaming and continuously provide insight into the behaviour of the player, as they would indicate whether this behaviour is similar to that of an older or a younger participant, representing weak or strong balance performance, respectively.

In this paper, we investigate curvature and speed of body movement as measures of interest for assessing balance performance in real time. Speed is a reliable measure of balance performance [9] and can be estimated instantaneously during exergaming. Although curvature is not commonly used, it has been identified as a promising measure for postural performance assessment in real-time [10], as it can provide additional information regarding the smoothness of body movements and can also be estimated instantaneously during exergaming. However, speed and curvature cannot be considered independent measures as they are related according to a power law [11]. We therefore also investigate whether linear regression parameters, derived from the speed-curvature

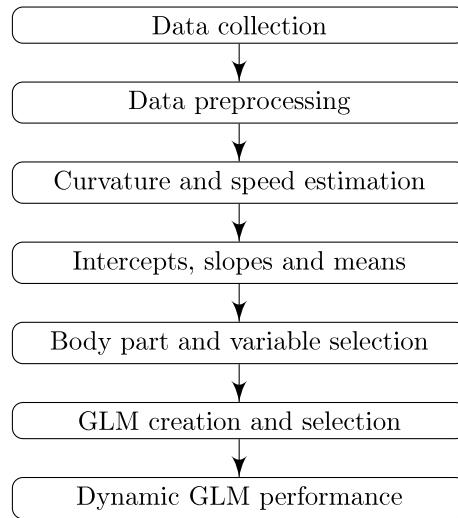


Fig. 1. Schematic showing the general steps followed in this study.

relationship on doubly logarithmic scale, can be used as compound measures of real-time balance performance. Fig. 1 shows our methodological approach in a flow chart.

II. METHODS

This study was performed in the context of the project *Exergaming for balance training of older adults at home* of the SPRINT research center [12] of the University Medical Center Groningen (UMCG). Part of the project was the development of a custom-made ice-skating exergame for unsupervised balance training of older adults. The study was executed in accordance with the ethical standards of the declaration of Helsinki and approved by the Medical Ethical Committee, and the Ethical Committee of the Center of Human Movement Sciences at the UMCG.

A. Participants

The data were collected from forty healthy participants in a previous study [13]. Twenty participants were older than 65 years (average 71.8, SD 4.0 years, twelve males). Twenty participants were younger than 61 years (average 36.9, SD 16.6 years, 9 males). Inclusion criteria were BMI < 30 and the ability to walk without aid (self-reported) for at least 15 minutes. Exclusion criteria were the use of medication affecting postural performance, musculoskeletal, visual, hearing, or neurological disorders that may affect balance performance, or an inability to understand the Dutch language.

B. Procedure and instrumentation

Each participant played the exergame ten times by swaying the center of body mass in lateral directions, resulting in 400 trials in total. On average, the trials lasted 44.3 seconds (SD 17.8 seconds). During game-play, 15 body parts were non-uniformly sampled at a frequency of about 30Hz using Microsoft Kinect version 1. Fig. 2 provides an example of the 15 point clouds recorded by Kinect during one minute of exergaming. For further details regarding the body positions captured by Kinect see [14, pp. 99].

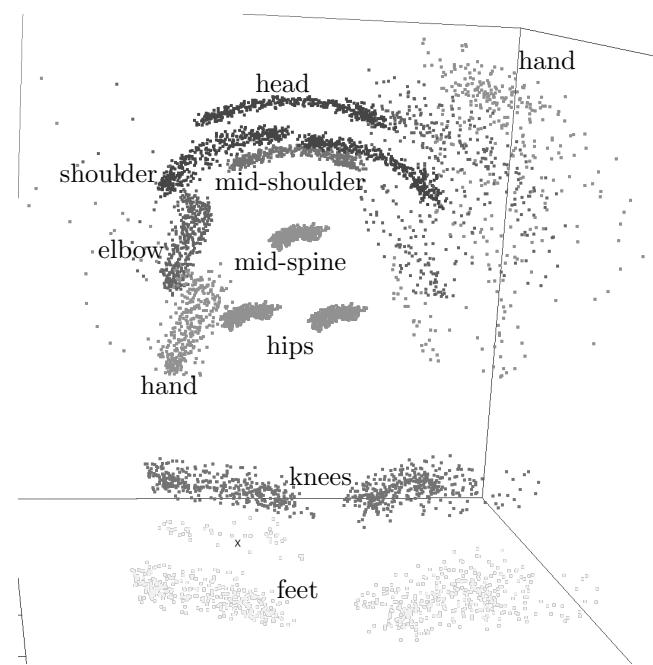


Fig. 2. Point clouds of 15 body parts tracked by Kinect during one minute of exergaming of a younger participant.

C. Data preprocessing

Kinect data were re-sampled at a fixed rate of 30Hz using cubic spline interpolation [15] to deal with possible sampling frequency deviations. For each trial, the first and last 5 seconds were removed to exclude motions that were not part of the swaying exercise. Feet, hands, and forearms were excluded from the analysis at this stage already, as they have the least reliably recorded body-part trajectories [5].

D. Curvature and speed estimation

For all the collected trials we estimated instantaneous speed (s) as the distance between two consecutive points divided by sample time. Instantaneous curvature (κ) was approximated by taking the inverse of the radius of a circle fitted to each three consecutive points [10].

E. Intercepts, slopes and means

Parameters from the power law relation between curvature and speed can be estimated [11] by fitting straight lines in doubly logarithmic space using the following linear model:

$$\log(s_{kj}(t_i)) = \gamma_{kj} + \beta_{kj} \cdot \log(\kappa_{kj}(t_i)) \quad (1)$$

where $k = 1 \dots 40$ is the index of the participants, $j = 1 \dots 9$ is a body part identifier, t is time, $i = 1 \dots n$, n is the number of samples per body part, γ_{kj} are the intercepts of the straight lines, and β_{kj} are the slopes. These parameters were estimated using [15]. For each participant and body part, average curvature $\bar{\kappa}_{kj}$ and speed \bar{s}_{kj} were also estimated.

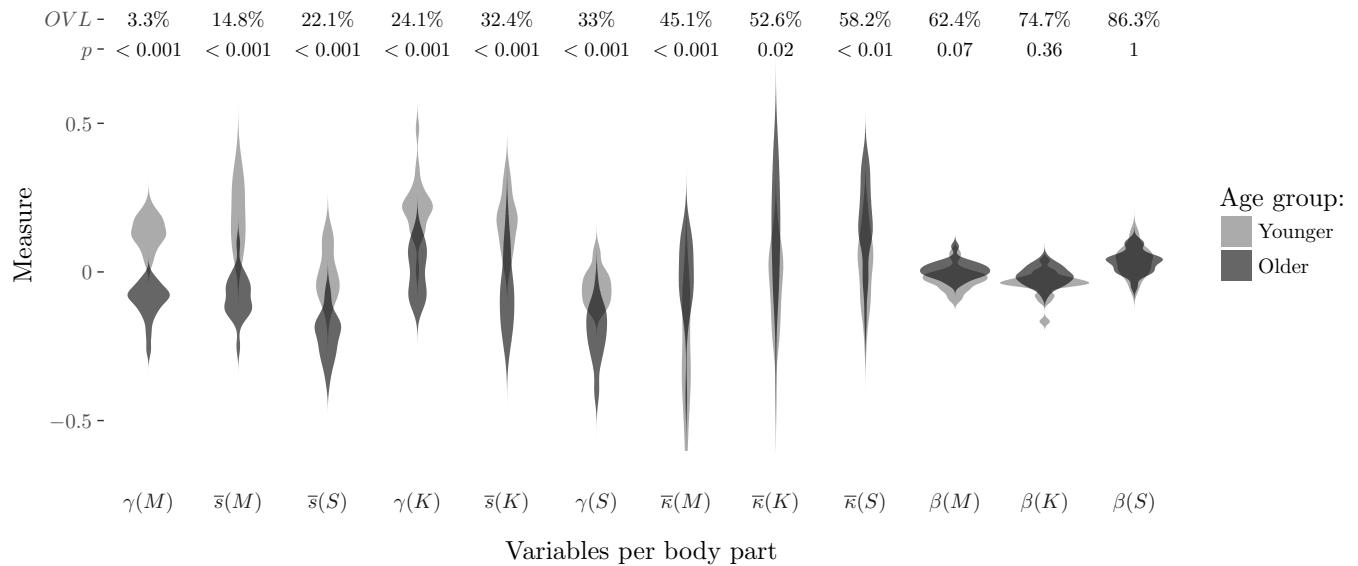


Fig. 3. Overlapping violin plots illustrating differences between older and younger participants. The variables (γ -intercept, β -slope, $\bar{\kappa}$ -mean curvature and \bar{s} -mean speed) derived from three body parts (K -knee, M -mid shoulder and S -mid spine) are shown on the horizontal axis arranged by the overlapping area (OVL) between distributions [17], and p is the statistical significance of the difference (Man-Whitney U-tests, Bonferroni corrected).

F. Body part and variable selection

The aim here is to reduce the complexity of the data as much as possible. For that purpose we estimated intercept correlations for all pairs of body part variables ($\gamma_j, \gamma_{j'}$), where $j' > j$. Highly correlated body parts ($r > 0.95$) were excluded. As a result only mid shoulder, mid spine and right knee body parts remained for analysis because head, mid shoulder and shoulders, mid spine and hips, and both knees turned out to be highly correlated. Using the remaining body parts we tested whether the variables speed (\bar{s}_j), curvature ($\bar{\kappa}_j$), and power law parameters (intercept γ_j and slope β_j) showed significant differences between older and younger participants, as assessed by Mann-Whitney U-tests [16] with Bonferroni correction for multiple comparisons. These differences provide insight into the importance of the variables to differentiate the two groups. Variables showing no differences between age groups were excluded from the analysis. As a result γ_j and \bar{s}_j derived from mid shoulder and mid spine were found to be the most important variables, $\bar{\kappa}_j$ also showed significant differences, but β_j did not. Therefore, β_j was excluded from the analysis (see Fig. 3).

G. GLM creation and selection

A GLM can be specified in three steps [8]. First, an assumption is made about the distribution of the outcome variable. In our case, we assumed that the response variable follows a Bernoulli distribution, that is, p_k is the probability that a participant is 61 years or older and $1 - p_k$ is the probability that a participant is younger than 61 years. Second, a linear model as a function of the explanatory variables is specified. Our explanatory variables are speed, curvature, and parameters derived from the power law relation between speed and curvature for mid shoulder, mid spine and right knee.

Third, the relationship between the mean value of the outcome variable and the linear model is specified.

Since in our case the outcome variable is binary (0 - young age class, 1 - old age class), we use *logistic* regression (or *logit* regression) which was invented by Cox [18], see also [19]. Logistic regression generates the coefficients of a formula to predict a *logit* transformation of the probability of a binary response based on one or more independent variables. The logit transformation involves the *logit* function, which transforms a probability value constrained between zero and one into the logarithm of the odds (or *log odds*),

$$\text{logit}(p_k) = \log(p_k/(1 - p_k)), \quad (2)$$

which can take any real value.

In this study, we defined GLMs to predict age category as follows.

- The age category of the k -th participant, $ageCat_k$, follows a Bernoulli distribution with probability p_k ;
- The logit transform of p_k is linearly regressed on the independent variables mean speed, mean curvature, and power law parameter (intercept) of the measured body parts, or a subset thereof.

Putting all this into a mathematical formula we get:

$$\begin{aligned} ageCat_k &\sim \text{Bernoulli}(p_k), \quad k = 1 \dots 40 \\ \text{logit}(p_k) &= \alpha + \sum_{l=1}^{|B|} \beta_l \cdot B_l, \\ \alpha &\sim \mathcal{N}(0, 10), \quad \beta_l \sim \mathcal{N}(0, 50) \end{aligned} \quad (3)$$

where $ageCat_k$ represents the age category *younger* (class 0) or *older* (class 1) of participant k , p_k is the probability that

participant k belongs to the older category, and B is a subset of the explanatory variables, i.e.,

$$B \subseteq \{\bar{s}(M_k), \bar{s}(S_k), \bar{s}(K_k), \bar{\kappa}(M_k), \bar{\kappa}(S_k), \bar{\kappa}(K_k), \gamma(M_k), \gamma(S_k), \gamma(K_k)\},$$

where mean speed \bar{s} , mean curvature $\bar{\kappa}$, and intercept γ are the measures derived for the body parts mid shoulder (M_k), mid spine (S_k), and right knee (K_k). The regression coefficients are the intercept α , and the coefficients β_l for each variable in B , where $|B|$ is the number of variables in B .

We selected the prior normal distributions \mathcal{N} for α and β_l in equation (3) with zero mean and standard deviation 10 and 50 respectively, based on the highest posterior density interval (HPDI) band, which is conceptually similar to the confidence interval band. We tested several priors using different standard deviation (SD) values, SD values of 10 for γ and 50 for β_l yielded the best fits (narrower bands are preferred). See [19, pp. 67] for further details regarding the HPDI.

Next, we investigated which combination of the variables would best predict age category. Evaluating all possible subsets (2^9 for a set of 9 variables) of the full model defined by Eq. (3) can be time consuming. Therefore, some specific models were considered based on the main aims of the study and the importance of the variables and body parts. Variables showing larger differences between older and younger participants were considered to be more important. In this sense, intercept (γ) and speed (\bar{s}) derived from mid shoulder (M) are some of the most important variables (see Fig. 3). We assumed that models without mid shoulder variables would perform worse than models that include these variables. We also wanted to investigate whether adding curvature to the models could improve the performance of models that use only speed as explanatory variable. Thus, in addition to the empty model

$$\text{logit}(p_k) = \alpha, \quad (m_0)$$

used only as reference [20], we defined 12 models falling within 3 classes: including only speed, including only intercepts, and models including speed and curvature as follows.

- **Model class I.** Only speed

$$\begin{aligned} \text{logit}(p_k) &= \alpha + \beta_1 \cdot \bar{s}(M_k), & (m_1) \\ \text{logit}(p_k) &= \alpha + \beta_1 \cdot \bar{s}(M_k) + \beta_2 \cdot \bar{s}(S_k), & (m_2) \\ \text{logit}(p_k) &= \alpha + \beta_1 \cdot \bar{s}(M_k) + \beta_2 \cdot \bar{s}(K_k), & (m_3) \\ \text{logit}(p_k) &= \alpha + \beta_1 \cdot \bar{s}(M_k) + \beta_2 \cdot \bar{s}(S_k) + \\ &\quad \beta_3 \cdot \bar{s}(K_k). & (m_4) \end{aligned}$$

- **Model class II** Only intercepts (here, curvature is used to estimate intercept values):

$$\begin{aligned} \text{logit}(p_k) &= \alpha + \beta_1 \cdot \gamma(M_k), & (m_5) \\ \text{logit}(p_k) &= \alpha + \beta_1 \cdot \gamma(M_k) + \beta_2 \cdot \gamma(S_k), & (m_6) \\ \text{logit}(p_k) &= \alpha + \beta_1 \cdot \gamma(M_k) + \beta_2 \cdot \gamma(K_k), & (m_7) \\ \text{logit}(p_k) &= \alpha + \beta_1 \cdot \gamma(M_k) + \beta_2 \cdot \gamma(S_k) + \\ &\quad \beta_3 \cdot \gamma(K_k). & (m_8) \end{aligned}$$

- **Model class III.** Curvature and speed:

$$\text{logit}(p_k) = \alpha + \beta_1 \cdot \bar{s}(M_k) + \beta_2 \cdot \bar{\kappa}(M_k) \quad (m_9)$$

$$\begin{aligned} \text{logit}(p_k) &= \alpha + \beta_1 \cdot \bar{s}(M_k) + \beta_2 \cdot \bar{\kappa}(M_k) + \\ &\quad \beta_3 \cdot \bar{s}(S_k) + \beta_4 \cdot \bar{\kappa}(S_k), \end{aligned} \quad (m_{10})$$

$$\begin{aligned} \text{logit}(p_k) &= \alpha + \beta_1 \cdot \bar{s}(M_k) + \beta_2 \cdot \bar{\kappa}(M_k) + \\ &\quad \beta_3 \cdot \bar{s}(K_k) + \beta_4 \cdot \bar{\kappa}(K_k), \end{aligned} \quad (m_{11})$$

$$\begin{aligned} \text{logit}(p_k) &= \alpha + \beta_1 \cdot \bar{s}(M_k) + \beta_2 \cdot \bar{\kappa}(M_k) + \\ &\quad \beta_3 \cdot \bar{s}(S_k) + \beta_4 \cdot \bar{\kappa}(S_k) + \\ &\quad \beta_5 \cdot \bar{s}(K_k) + \beta_6 \cdot \bar{\kappa}(K_k). \end{aligned} \quad (m_{12})$$

Fitting GLMs ($m_0 \dots m_{12}$) we encountered the *complete separation problem*, which happens mostly when the number of samples is small and a linear combination of the predictors perfectly or almost perfectly predicts the outcome [21], [22]. In such a case, the maximum likelihood fitting method provides implausible estimates. To circumvent this problem we used Markov chain Monte Carlo (MCMC) simulations to fit the GLMs [23]. We also used the probabilistic programming language Stan [24] and the *rethinking* [19] package as interface to fit the GLMs.

To compare the above models ($m_0 \dots m_{12}$) Watanabe-Akaike information criterion (WAIC) [25] scores were estimated. WAIC is a measure of the predictive accuracy of the models on new data [26]. Hence, this criterion provides a way of model selection. These scores were ordered from the lowest to highest (smaller values are preferred). For each model the *Akaike weight* was also estimated, which is a partition of the total weight of 1 among the models considered, so that the sum of the weights is always 1 (see Table I). These weights provide a more interpretable measure of the relative differences between the models because they are an estimate of the probability that a model will perform better on new data given the models considered. Further details can be found in [19, pp. 199, 207]. Note that for these comparisons the models were fitted using all the participants ($k = 1 \dots 40$). This is a preparatory step to select the models that will be assessed dynamically using five-fold cross-validation [27], [28] which we consider next.

H. Dynamic GLM performance

The main purpose of the defined GLMs $m_1 \dots m_{12}$ is to perform dynamical predictions on new data, that is, estimate balance performance continuously during exergaming. We restrict ourselves to a selection of four models to be tested dynamically using five-fold cross-validation, as this technique is computationally expensive. Based on the WAIC scores we selected GLMs with the three highest scores, that is, m_5 , m_7 , and m_{11} . We also selected the model with the lowest score m_3 because it is based only on speed values derived from mid shoulder and right knee body parts. This model is used as a benchmark to show that adding curvature and intercepts improves model performance.

Fig. 4 illustrates the first iteration of the five-fold cross-validation procedure. As part of the five-fold cross-validation procedure, the order of the participants was randomized and

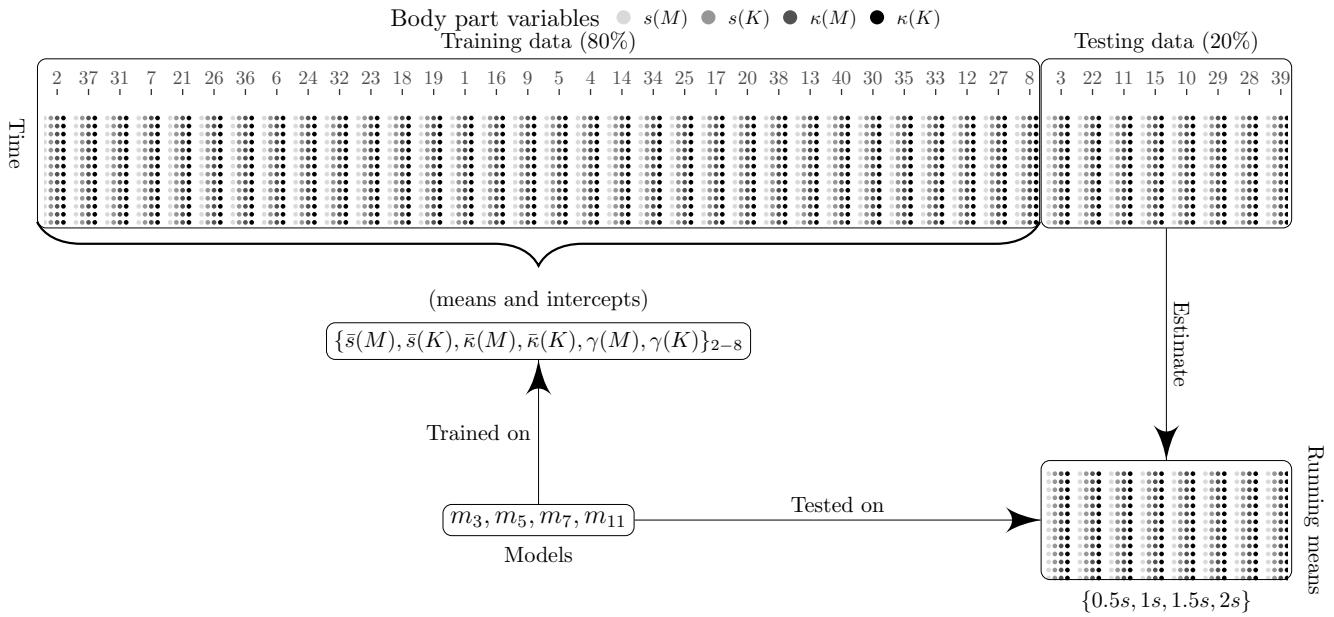


Fig. 4. Visualization of the training and testing disjoint subsets used to assess models m_3 , m_5 , m_7 , and m_{11} (only the first iteration of the five-fold cross validation procedure is illustrated). The order of the participants was randomized. Dots represent speed (s) and curvature (κ) estimations derived from two body parts, mid shoulder (M) and right knee (K), using Kinect recordings.

TABLE I

WAIC MODEL COMPARISON. SYMBOLS REPRESENT: M -MID SHOULDER, S -MID SPINE, AND K -RIGHT-KNEE, γ -INTERCEPT, $\bar{\kappa}$ -MEAN CURVATURE, AND \bar{s} -MEAN SPEED. COLUMN 'Measure': THE VARIABLES INCLUDED IN A MODEL; COLUMN 'Body part': THE INCLUDED BODY PARTS; COLUMN 'Weight': THE AKAIKE WEIGHT; COLUMN SE: THE STANDARD ERROR OF THE WAIC ESTIMATE.

Model	Body part	Measure	WAIC	Weight	SE
m_5	M	γ	2.4	0.15	1.08
m_7	M, K	γ	2.4	0.15	0.95
m_{11}	M, K	$\bar{\kappa}, \bar{s}$	2.6	0.14	0.85
m_6	M, S	γ	2.7	0.13	1.20
m_{12}	M, S, K	$\bar{\kappa}, \bar{s}$	2.8	0.13	0.94
m_8	M, S, K	γ	3.0	0.11	1.31
m_{10}	M, S	$\bar{\kappa}, \bar{s}$	3.6	0.08	1.37
m_9	M	$\bar{\kappa}, \bar{s}$	4.1	0.06	1.61
m_2	M, S	\bar{s}	5.9	0.03	2.61
m_4	M, S, K	\bar{s}	7.4	0.01	3.13
m_1	M	\bar{s}	14.2	0.00	7.74
m_3	M, K	\bar{s}	17.2	0.00	9.48
m_0			57.5	0.00	0.03

five training-testing disjoint subsets were created using 80% of the participants for training and 20% for testing. For each iteration of the cross-validation procedure, m_3 , m_5 , m_7 and m_{11} were fitted again (on means and intercepts) in the training phase. To test the trained models we estimated running means, from the 20% testing data (local curvature and instantaneous speed, see section II-D), using a moving window over the whole length of the trials in the testing set. To investigate the effect of the running window size, each model was tested using 0.5, 1, 1.5, 2 second running means.

The models were assessed using traditional metrics such as the F-measure, precision and recall [29]. Values of these metrics were estimated using the threshold at the point with the best sum of *sensitivity* and *specificity*, closest to the point (0,1) of the ROC curve. Sensitivity (also called "recall") is the proportion of correctly classified older participants. Specificity is the proportion of correctly classified younger participants. Precision is the number of correctly classified older participants divided by the total number of classified older participants. The F-measure is the harmonic mean of precision and recall. The threshold values were estimated using the pROC [30] R-package.

As a last step, we investigated whether more samples for training could improve model performance. For this purpose we redefined m_3 , m_5 , m_7 and m_{11} , to be trained on running means and running intercepts, as follows:

$$\text{logit}(p_{ik}) = \alpha + \beta_1 \cdot \bar{s}_i(M_k) + \beta_2 \cdot \bar{s}_i(K_k), \quad (m'_3)$$

$$\text{logit}(p_{ik}) = \alpha + \beta_1 \cdot \bar{\gamma}_i(M_k), \quad (m'_5)$$

$$\text{logit}(p_{ik}) = \alpha + \beta_1 \cdot \bar{\gamma}_i(M_k) + \beta_2 \cdot \bar{\gamma}_i(K_k), \quad (m'_7)$$

$$\begin{aligned} \text{logit}(p_{ik}) = \alpha + \beta_1 \cdot \bar{s}_i(M_k) + \beta_2 \cdot \bar{\kappa}_i(M_k) + \\ \beta_3 \cdot \bar{s}_i(K_k) + \beta_4 \cdot \bar{\kappa}_i(K_k), \end{aligned} \quad (m'_{11})$$

where \bar{s}_i and $\bar{\kappa}_i$ are the one-second running means of speed and curvature respectively at time i , and $\bar{\gamma}_i$ is the one-second running intercept. Note that training these variant models (indicated by a prime) was computationally expensive because we used the whole length of the trials (i is the same index as in equation 1). For each training set, these models were trained twice using 10% and 100% of the running means. The window size was selected after looking at the performance of the models $m_1 \dots m_{12}$. These models performed worse using

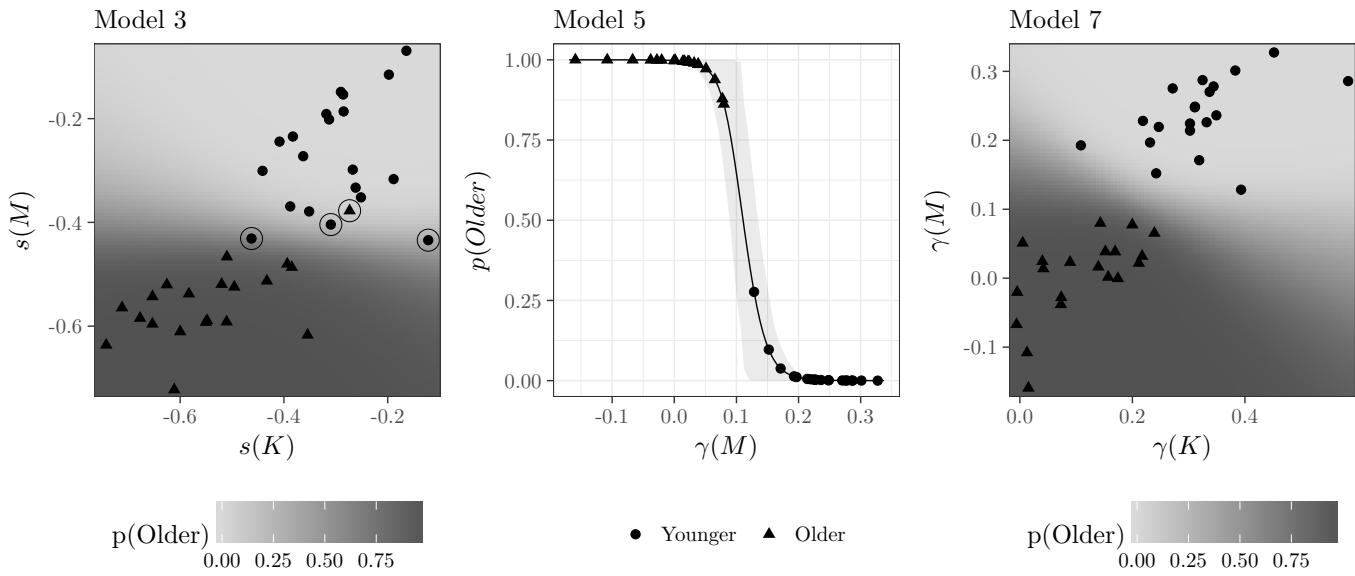


Fig. 5. Fits of the models m_3 , m_5 , and m_7 . The variables are intercept γ , speed s , curvature κ , whereas p represents the probability of belonging to the older age group. Dots and triangles represent mean values per participant in log – log scale. The red circles in *Model 3* indicate participants that are misclassified or not clearly classified as older or younger participants. In *Model 5*, the solid line represents the mean point estimates of the posterior distribution. The light-gray shaded area represents the 89% highest posterior density interval (HPDI) band for the means. The HPDI is the narrowest range containing the specified probability mass, similar to the common confidence interval.

a 0.5-second window size than using larger window-sizes, but there were no clear differences from 1 to 2 second window sizes. Hence we used a window size of one second.

III. RESULTS

Overlapping violin plots of mean speed and curvature, intercepts and slopes (of the speed-curvature relationship) per body part provide first insight into their potential to differentiate older and younger groups (Fig. 3). Smaller overlapping area (OVL) values suggest greater differences between groups. This figure shows that intercept and speed measures derived from mid shoulder movements, $\gamma(M)$ and $\bar{s}(M)$, differentiate better between older and younger groups than the other measures. Although curvature measures ($\bar{\kappa}$) from the three body parts show around 50% OVL, the difference between older and younger participants is still significant. Slope measures (β) from mid shoulder, mid spine, and knee show the least differences between older and younger participants. Indeed, slopes show both the highest OVL values (> 60%) and non-significant differences between older and younger groups and were therefore excluded from GLM fits.

Next, we first consider GLM model selection based on *global* means, where the averages are computed over the whole time interval based on models m_1-m_{12} in Section II-G. Subsequently, we investigate *dynamic* GLM performance, using the variant GLM models $m'_3, m'_5, m'_7, m'_{11}$, as defined in Section II-H.

A. GLM selection

Table I shows the model comparison results. It can be observed that models including either curvature and speed together or only intercept are among the “best” models. Also

notice that most of these models score similar Akaike weights (probability of making best predictions on new data, see Section II-G), suggesting similar model performance. Models including only speed measures score the lowest weights of the fitted models. This implies that including either curvature or intercept improves model performance on new data.

Fig. 5 shows how models m_3 , m_5 , and m_7 fit the data.

Model m_3 in this figure is based only on speed values derived from mid shoulder and right-knee body parts. It scores the lowest WAIC value in Table I. The visualization of the fit of m_3 provides insight into the performance of a model that does not include curvature as a predictor variable. It shows a clear, but not perfect, separation between older and younger participants.

Model m_5 in Fig. 5, is the “best” model according to WAIC criteria. It is the simplest model as it only uses “one” variable, the intercept γ from the speed-curvature relation, derived from mid shoulder movements. This visualization shows that including curvature in the model improves the separation between older and younger participants.

Model m_7 in Fig. 5 shows perfect separation between older and younger participants. In addition to intercept values γ derived from mid shoulder, γ values derived from knee movements are included. According the figure shows, this additional information helps to even better differentiate older and younger participants.

The fit of model m_{11} cannot be visualized in two dimensions because it involves four explanatory variables (speed and curvature means derived from mid shoulder and knee). However, we can use principal component analysis [31] to visualize a projection of the points in the four-dimensional space onto the two first principal components (PC). We can also use a

bi-plot to gain additional understanding of the contribution of the variables to the first two PCs. Fig. 6 shows the bi-plot and the projection of the four variables of model m_{11} . This figure also shows clear separation between older and younger participants in terms of probabilities, since all the triangles correspond to high probabilities, and all circles correspond to low probabilities. Note that in the projection, one older participant is located among the younger participants. Larger arrows in the bi-plot indicate stronger contribution of the variables to the PCs than shorter arrows, orthogonal vectors indicate weak correlation and opposite directions indicate strong negative correlation between variables [32]. Thus, curvature has a stronger contribution to the PCs, curvature and speed are strongly negatively correlated if they are derived from the same body part, but only weakly correlated if they are derived from different body parts.

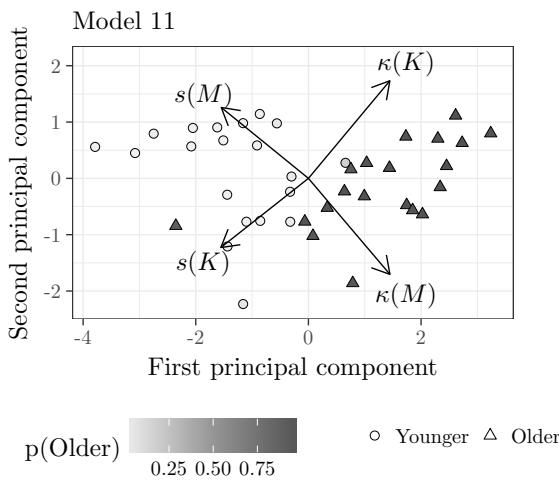


Fig. 6. Projection of the four variables of model m_{11} ($\bar{\kappa}$ and \bar{s} estimated from mid shoulder and knee) onto the first two PCs, p represents the probability of belonging to the older age group. The arrows represent the variable vectors with names at their end-points (M -mid_shoulder, K -right knee), see text for explanation. The first PC accounts for 78.63% of the variance and the second PC accounts for 16.81%.

In summary, Table I, and Figs. 5 and 6 show that including curvature in the models improves their predictive performance.

B. Dynamic GLM performance

Fig. 7 shows the performance of the models m_3 , m_5 , m_7 , and m_{11} and their variant versions (m') tested on the trials using 0.5, 1, 1.5, and 2 second running-means ($\bullet\bullet$, $\bullet\bullet$, $\bullet\bullet$, and $\bullet\bullet$). In contrast to WAIC comparison (Table I) where two γ -based models get the highest scores (m_5 and m_7), this figure shows that these models score the lowest when tested on 0.5-second running means. Moreover, m_5 , m_7 and their variants (m'_3 and m'_7) are among the models scoring the lowest accuracies, tested on 0.5-second and 1-second running means. Most of the models based only on speed values (m_3 and m'_3) scored better than models fitted on γ values. Clearly, the top 5 models are based on both speed and curvature values (m_{11} and m'_{11}). As for the size of the time window, although most of the models tested on 0.5-second running means have

the lowest scores, there are no clear differences between the models. Regarding the amount of data used to fit the models, surprisingly, there are no clear differences in performance between models fitted on means and intercepts, which include only 40 samples per variable, and models fitted on 10% or 100% (including about 30k and 300k samples) of the data see $\boxed{1}$, $\boxed{10}$ and $\boxed{100}$ labels on the left of the figure.

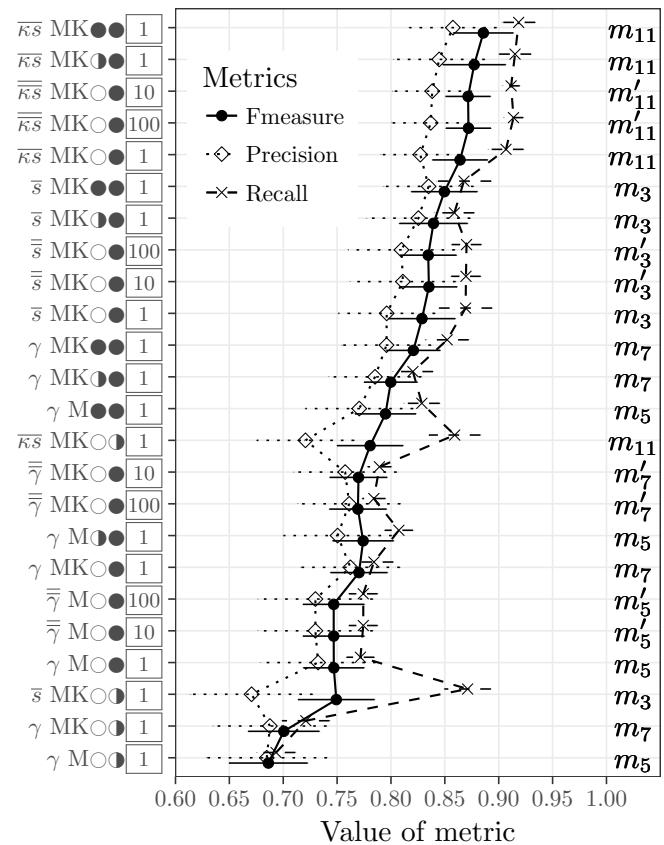


Fig. 7. Model performance as assessed by different metrics. The horizontal axis represents the value of the three metrics, precision, recall, and F-measure. The vertical axis represents the fitted models ordered by F-measure. The symbols represent γ -intercept derived from the power law, $\bar{\kappa}$ -curvature and \bar{s} -speed (one bar indicates means and double bar indicates running means), M -mid_shoulder, S -mid spine, and K -right knee; black filled circles represent time windows, \bullet -0.5 seconds and $\bullet\bullet$ -1 second; and boxes represent the amount of data used to fit the models: $\boxed{1}$ -means and intercepts, $\boxed{10}$ -10% of the running means, and $\boxed{100}$ -100% of the running means. In the plot, circles represent the average values of the metrics estimated by five-fold cross-validation, and horizontal bars represent standard errors.

F-measure, precision and recall, in Fig. 7, agree on the top-five models. Recall, that is the proportion of correctly classified older samples, shows that the top-five models correctly classify above 90% of the older samples. These models score between 0.8 and 0.86 on precision. The F-measure is commonly used as a measure between precision and recall. Among these measures, recall could be the best measure for our purposes because a) it shows the biggest gap between the top-five models and the rest of the models (see the distance between m_3 and m_{11}), b) it shows the highest accuracy in classifying older participants, which is the target population of exergames in this study, and d) it shows the smallest standard errors. All

in all, these results show that including curvature in the models indeed improve their performance.

Fig. 8 illustrates m_{11} , the top model, with dynamic predictions over time along the 400 trials. Even though the predictions for some participants are not clear cut (see participants, 19, 20, and 39), this figure shows a clear separation of older and younger participants over time, as on the left most of the probability values to belong to the older age group are low ("light gray") and on the right most of the values are high ("dark gray"). The values in this figure could be used to provide appropriate feedback during game-play, as the player could see his/her instantaneous performance as "older" or "younger" participant. These values could also be used to adjust the difficulty of exergames to meet the skills of the players.

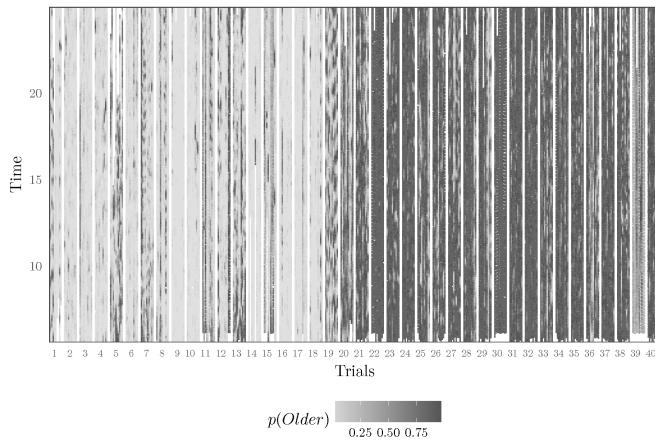


Fig. 8. Model m_{11} predictions from 400 trials using one-second running means. The horizontal axis represents 10 exergame trials for each participant. The vertical axis represents time. Each vertical line represents a trial. Participants are ordered by age, the first 20 are younger and the rest are older. For clarity, white lines are added to separate between trials per participant. Also, missing Kinect data are indicated by white lines.

In terms of balance performance our results suggest, as could be expected, that the younger participants performed better during the trials than the younger participants. This means that the older participants performed slower during the trials than the younger participants, illustrated in Figs. 3 and 5. Faster movements also indicate smoother movements (smaller curvature values) among the younger than among the older participants, because of the speed-curvature relationship [11].

IV. DISCUSSION

Our main goal in this study was to investigate whether balance performance as assessed by speed can be improved by adding curvature in the assessment, in two ways: a) using compound measures derived from the relation between curvature and speed, and b) using curvature and speed together. Our results suggested that including curvature can indeed improve postural performance assessment in real-time.

First, WAIC scores suggested that GLMs including both curvature and speed, and GLMs including intercepts would perform better on new data than GLMs including only speed (see Table I). Second, common classifier performance metrics

such as F-measure, precision and recall suggested that GLMs including curvature and speed will probably perform better on new data than GLMs fitted only on speed or only on intercepts (see Fig. 7).

Although obtaining high prediction accuracy was not the main goal of this study, it is a natural way to gain insight into the future usefulness of the models [33], in our case, the usefulness for assessing balance performance in real-time. In addition, the straightforward interpretation of the probability values estimated from logistic regression are useful to provide immediate and meaningful feedback and can be interpreted by common people. These probability values indicate how likely the quality of body movements is similar to that of an older person.

In a previous study, curvature and speed derived from force plate recordings were identified to be suitable measures of balance performance in real-time, because a) they show differences between older and younger participants, and b) they can be estimated instantaneously [10]. The results of the present study provide additional evidence of this suitability, but now using Kinect recordings. This relies on the ability to instantaneously characterize participants as "older" or "younger" in terms of probabilities. Even though the model predictions might be incorrect, the probability estimations could be useful because some older participants could behave as younger ones and vice-versa. In addition, curvature and speed provide a natural and intuitive interpretation of the results. For example, assuming that exergaming players are healthy it is natural to expect that fast and smooth movements reflect good postural performance. On the contrary, slow and non-smooth movements represent worse postural performance. As these features can be assessed by speed and curvature, higher speed and lower curvature values should reflect better postural performance.

In another study [13], balance performance was assessed using the same Kinect recordings. The authors used medial lateral movements (values of x and y coordinates, but not z) of nine body parts of the players to identify patterns using Self Organizing Maps. Then, they trained a kNN classifier using the identified patterns to discriminate older and younger participants. The accuracy of the classifier was 65.8%. Here, using two features derived from only two body parts we achieved more than 90% recall accuracy (see Fig. 7). These results suggest that features derived from body movement trajectories provide more information to differentiate older and younger participants than the coordinates of the trajectories.

In this study, all of the models showed a "clear" separation between older and younger participants. This may be because the number of participants is small. Thus, including more participants could be expected to give more overlap between older and younger participants. Still, the separation we found is consistent with the evident physical decline of people after 60 years of age. For example, older and younger participants were clearly differentiated using mean velocity during static tasks in [9]. Evident physical decline was also reported in walking speed and aerobic endurance for people in their 60s and 70s [34]. As exergaming is a physical activity, it may not be a surprise that we here observed similar differences between

older and younger participants.

One of the limitations of this study is that the lengths of the trials are short and the number of participants is small; to mitigate this, we applied 5-fold cross validation as an accepted method to gain insight into the performance of classifiers on new data. Further research is needed to investigate the usefulness of the presented method on long intervention studies. Another limitation is that the selection of body parts (mid shoulder, mid spine and right knee) is partly imposed by the limited accuracy of Kinect. More accurate devices may lead to a different selection of body parts for better classification performance.

The method presented here based on GLMs could be used to assess balance performance in different kinds of exergames using different kinds of tracking technology. That is because our method depends on features derived from the trajectory of body movements and not on the tracking device. Also, in addition to curvature and speed, other instantaneously estimated variables/features can also be included in the models such as torsion and their derivatives [35]. Once a GLM is fitted, the estimated parameters could be used to make predictions as shown in Fig. 8.

Studies such as [36] and [37] have suggested some desirable features for the development of adaptive exergames. Some of these features are: embracing age-related physical impairments, adapting individual differences in player range of motion, preventing overexertion by providing appropriate game pacing, and including automatic adjustment of difficulty. In this sense, the methods shown here form a firm step towards the development of adaptive exergames based on measures of balance performance in real time.

V. CONCLUSION

We have presented a promising method to assess dynamic balance performance during exergaming. Curvature derived from the trajectories of body movements can provide additional information to assess performance in real-time. GLMs provide a way to derive a single measure, as a function of curvature and speed, that represents the behaviour of participants as belonging to a younger or older age group. Given reliably captured body movement trajectories, this method could potentially be used to assess the quality of movements in real-time not only in exergaming but in other fields such as sports, rehabilitation, and medicine, offering instantaneous and appropriate feedback that could foster motivation and movement performance. Finally, in future work we plan to study the validity and sensitivity of our method to detect changes in balance performance, using trials recorded during an unsupervised six week exergaming training at home.

APPENDIX A COMPUTATIONAL COST

Fig. 9 shows the amount of time used to fit models m'_3 , m'_5 , m'_7 , and m'_{11} on 10% and 100% of the data, respectively. To fit these models we used a computer with two Intel Xeon E5-2630 processors of 2.3 GHz, each processor with 12 cores, and 64GB of main memory. For each model we ran

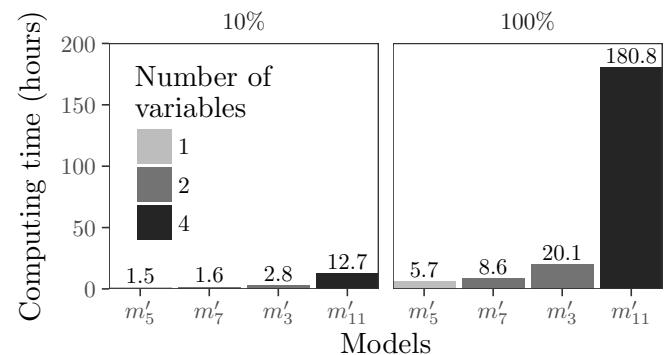


Fig. 9. Amount of time used to fit variant (m') models on 10% and 100% of the data, respectively. Note that as we used 5-fold cross validation, the bars represent the time to fit a particular model 5 times on different datasets.

24 simultaneous simulations, as there are 24 cores available, with 1000 warm-up samples and 3000 iterations, resulting in $24 \times (3000 - 1000) = 48000$ real samples in total. These models were fitted using only 1-second running means as these fits were computationally intensive. The figure clearly shows that the time increases exponentially as the number of variables and the number of samples increase, see for example model m'_{11} that was fitted on 4 variables with about 300k samples. This figure, together with the results shown in Fig. 7 suggest that models fitted on means and intercepts could be good enough to make reliable predictions on new data. In addition, the most “complex” model m_{11} tested using 5-fold cross validation based on means and intercepts required only about 2.5 minutes for fitting five different datasets.

ACKNOWLEDGMENT

The exergaming project has been performed on behalf of research center SPRINT of the UMCG and was supported by INCAS³ and 8D-Games. Data used in the present paper were collected by Mike van Diest in the realm of this project.

REFERENCES

- [1] M. van Diest, C. J. C. Lamoth, J. Stegenga, G. J. Verkerke, and K. Postema, “Exergaming for balance training of elderly: state of the art and future developments,” *Journal of Neuroengineering and Rehabilitation*, vol. 10, no. 1, p. 101, 2013.
- [2] CDC, “WISQARS (Web-based Injury Statistics Query and Reporting System)—Injury Center—CDC.” [Online]. Available: <https://www.cdc.gov/injury/wisqars/>
- [3] L. D. Gillespie, M. C. Robertson, W. J. Gillespie, C. Sherrington, S. Gates, L. M. Clemson, and S. E. Lamb, “Interventions for preventing falls in older people living in the community.” *The Cochrane database of systematic reviews*, vol. 9, p. CD007146, jan 2012.
- [4] C. Burgers, A. Eden, M. D. Van Engelenburg, and S. Buningh, “How feedback boosts motivation and play in a brain-training game,” *Computers in Human Behavior*, vol. 48, pp. 94–103, 2015.
- [5] M. van Diest, J. Stegenga, H. J. Wörtsche, K. Postema, G. J. Verkerke, and C. J. Lamoth, “Suitability of Kinect for measuring whole body movement patterns during exergaming,” *Journal of Biomechanics*, vol. 47, no. 12, pp. 2925–2932, sep 2014.
- [6] C. J. Lamoth and M. J. van Heuvelen, “Sports activities are reflected in the local stability and regularity of body sway: Older ice-skaters have better postural control than inactive elderly,” *Gait & Posture*, vol. 35, no. 3, pp. 489–493, 2012.
- [7] D. Abrahamová and F. Hlavacka, “Age-related changes of human balance during quiet stance.” *Physiological research / Academia Scientiarum Bohemoslovaca*, vol. 57, no. 6, pp. 957–964, 2008.

- [8] A. F. Zuur, E. N. Ieno, N. J. Walker, A. A. Saveliev, and G. M. Smith, *Mixed Effects Models and Extensions in Ecology with R*. New York: Springer Science + Business Media, 2009.
- [9] T. E. Prieto, J. B. Myklebust, R. G. Hoffmann, E. G. Lovett, and B. M. Myklebust, "Measures of postural steadiness differences between healthy young and elderly adults," *Transactions of Biomedical Engineering*, vol. 43, no. 9, pp. 965–966, 1996.
- [10] V. Soancatl Aguilar, J. J. van de Gronde, C. J. C. Lamoth, M. van Diest, N. M. Maurits, and J. B. T. M. Roerdink, "Visual Data Exploration for Balance Quantification in Real-Time During Exergaming," *PLOS ONE*, vol. 12, no. 1, p. e0170906, jan 2017.
- [11] P. L. Gribble and D. J. Ostry, "Origins of the power law relation between movement velocity and curvature: modeling the effects of muscle mechanics and limb dynamics." *Journal of neurophysiology*, vol. 76, no. 5, pp. 2853–2860, 1996.
- [12] "SPRINT Technologie die Ouderden Beweegt." [Online]. Available: <http://www.imdi-sprint.nl/>
- [13] M. van Diest, J. Stegenga, H. J. Wörtche, J. B. T. M. Roerdink, G. J. Verkerke, and C. J. C. Lamoth, "Quantifying Postural Control during Exergaming Using Multivariate Whole-Body Movement Data: A Self-Organizing Maps Approach," *Plos One*, vol. 10, no. 7, p. e0134350, 2015.
- [14] J. Webb and J. Ashley, *Beginning Kinect Programming with the Microsoft Kinect SDK*. Apress, 2012.
- [15] MATLAB, 8.6.0.267246 (R2015b), The Mathworks, Inc., Natick, Massachusetts, 2015.
- [16] H. B. Mann and D. R. Whitney, "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other," *The Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, 1947.
- [17] M. Ridout and M. Linkie, "Estimating overlap of daily activity patterns from camera trap data," *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 14, no. 50, pp. 322–337, 2009.
- [18] D. R. Cox, "The regression analysis of binary sequences (with discussion)," *J Roy Stat Soc B*, vol. 20, pp. 215–242, 1958.
- [19] R. McElreath, *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman and Hall/CRC, 2015.
- [20] T. A. Snijders and R. J. Bosker, *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, second edition. Sage Publications, 1999.
- [21] A. Albert and J. A. Anderson, "On the Existence of Maximum Likelihood Estimates in Logistic Regression Models," *Biometrika*, vol. 71, no. 1, pp. 1–10, 1984.
- [22] A. Gelman, A. Jakulin, M. G. Pittau, and Y. S. Su, "A weakly informative default prior distribution for logistic and other regression models," *Annals of Applied Statistics*, vol. 2, no. 4, pp. 1360–1383, 2008.
- [23] C. Rainey, "Dealing with Separation in Logistic Regression," *Political Analysis*, vol. 24, no. 3, pp. 339–355, 2016.
- [24] A. Gelman, D. Lee, and J. Guo, "Stan: A Probabilistic Programming Language for Bayesian Inference and Optimization," *Journal of Educational and Behavioral Statistics*, vol. 40, no. 5, pp. 530–543, 2015.
- [25] S. Watanabe, "Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory," *Journal of Machine Learning Research*, vol. 11, pp. 3571–3594, 2010.
- [26] A. Gelman, J. Hwang, and A. Vehtari, "Understanding predictive information criteria for Bayesian models," *Statistics and Computing*, vol. 24, no. 6, pp. 997–1016, 2014.
- [27] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," in *International Joint Conference on Artificial Intelligence*, 1995.
- [28] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. Springer, 2009.
- [29] T. Fawcett, "An introduction to ROC analysis Tom," *Pattern Recognition Letters*, vol. 27, pp. 861–874, 2006.
- [30] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, and M. Müller, "pROC: an open-source package for R and S+ to analyze and compare ROC curves," *BMC Bioinformatics*, vol. 12, no. 1, pp. 1–8, 2011.
- [31] I. T. Jolliffe, "Principal Component Analysis, Second Edition," *Springer Series in Statistics*, vol. 98, p. 487, 2002.
- [32] K. R. Gabriel, "The biplot-graphical display of matrices with applications to principal components analysis," *Biometrika*, vol. 58, pp. 453–467, 1971.
- [33] J. Piironen and A. Vehtari, "Comparison of Bayesian predictive methods for model selection," *Statistics and Computing*, vol. 27, no. 3, pp. 1–25, 2016.
- [34] K. S. Hall, H. J. Cohen, C. F. Pieper, G. G. Fillenbaum, W. E. Kraus, K. M. Huffman, M. A. Cornish, A. Shiloh, C. Flynn, R. Sloane, L. K. Newby, and M. C. Morey, "Physical Performance Across the Adult Life Span: Correlates With Age and Physical Activity," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 00, no. 00, pp. 1–7, 2016.
- [35] S. Wu and Y. Li, "Flexible signature descriptions for adaptive motion trajectory representation, perception and recognition," *Pattern Recognition*, vol. 42, no. 1, pp. 194–214, jan 2009.
- [36] K. Gerling, I. Livingston, L. Nacke, and R. Mandryk, "Full-body motion-based game interaction for older adults," in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, 2012, pp. 1873–1882.
- [37] A. Velazquez, A. I. Martínez-García, J. Favela, and S. F. Ochoa, "Adaptive exergames to support active aging: An action research study," *Pervasive and Mobile Computing*, 2016.

Venustiano Soancatl Aguilar received the B.Sc. degree in computer science from the Benemérita Universidad Autónoma de Puebla, México in 2001, and the MSc degree in computer science from the Instituto Nacional de Astrofísica, Óptica y Electrónica, Tonantzintla, Puebla, México in 2003. From 2006 to 2014, he was an Associate Professor in computer science at the Universidad del Istmo, Oaxaca, México. He is pursuing a PhD degree in computer science at the University of Groningen, The Netherlands. His research focuses on quantification of human movement in real-time for gaming technology and movement disorders. His interests include human-computer interaction, computer vision and machine learning.

Jasper J. van de Gronde studied computing science at the University of Groningen, the Netherlands, where he obtained his M.Sc. in 2011 and his Ph.D. in 2015 on a theoretical framework for mathematical morphology on colour and tensor-valued images. After a postdoctoral period at the same university he joined the ASML company in Veldhoven, The Netherlands, as Metrology Software Design Engineer in early 2017. His research interests include mathematical morphology, compressed sensing, and signal/image processing in general.

Claudine C.J. Lamoth is an Associate Professor and principal investigator at the Centre for Human Movement Sciences of the University Medical Centre Groningen (UMG), University of Groningen. Lamoth's research focus is in the area of motor control, motor learning, ageing and movement disorders, using concepts and tools from dynamical systems theory. She combines fundamental knowledge of motor control with sensor and game technologies and innovative data analysis in order to develop solutions for supporting independent mobility, preventing falls, and improving health.

Natasha M. Maurits received the MEng degree (1994) in applied mathematics, the MSc degree (1994) in numerical mathematics, and the PhD degree (1998) in chemistry, from the University of Groningen, The Netherlands. She is currently a full professor of Clinical Neuroengineering at the Department of Neurology at the University Medical Center Groningen, the Netherlands. She translates clinical neurological problems to physical-mathematical problems, tries to find a solution using state-of-the-art mathematical techniques and translates this solution back to neurology for application. Her interests include movement disorders, trauma, multimodal recordings, signal analysis and data visualization. She is a senior member of IEEE.

Jos B.T.M. Roerdink received a Ph.D. (1983) in theoretical physics from the University of Utrecht, the Netherlands. After a two-year position (1983-1985) as a Postdoctoral Fellow at the University of California, San Diego, he joined the Centre for Mathematics and Computer Science in Amsterdam, working on image processing and tomographic reconstruction. He was appointed associate professor (1992) and full professor (2003), respectively, at the Johann Bernoulli Institute for Mathematics and Computer Science of the University of Groningen, where he currently holds a chair in Scientific Visualization and Computer Graphics. His research interests include mathematical morphology, biomedical visualization, neuroimaging and bio/neuroinformatics.