Antonio Iyda Paganelli

Reliability of Wii Balance Board and Microsoft Kinect for capturing posturographic information during balance tests

Dissertação de Mestrado

Dissertation presented to the Programa de Pós-graduação em Informática of PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Informática.

Advisor: Prof. Alberto Barbosa Raposo

Rio de Janeiro
September 2018

Antonio Iyda Paganelli

Reliability of Wii Balance Board and Microsoft Kinect for capturing posturographic information during balance tests

Dissertation presented to the Programa de Pós-graduação em Informática of PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Informática. Approved by the undersigned Examination Committee.

Prof. Alberto Barbosa Raposo
Advisor
Departamento de Informática – PUC-Rio

Profª. Greis Francy Mireya Silva
Pesquisador Autônomo

Prof. Moacir Marocolo Júnior
UFJF

Prof. Márcio da Silveira Carvalho
Vice Dean of Graduate Studies
Centro Técnico Científico – PUC - Rio

Rio de Janeiro, September. 21st, 2018

Antonio Iyda Paganelli

Graduated in Informatics from Pontifical Catholic University, Rio de Janeiro (PUC-Rio) in 1994 and in Physical Education from Estácio de Sá University in Rio de Janeiro in 2015. He got a Master in International Business from Florida International University, USA in 2008, and joined the Master in Informatics at PUC-Rio in 2016.

Bibliographic data

Acknowledgments

To my family, Adriana, Gabriel, Diana, and Maria for giving unconditional support in my decisions and challenges. To my Mom, Tomoko Iyda Paganelli, for guiding and inspiring me the whole life. To my Dad, Ernesto Antonio Jorge Paganelli, wherever you are, surely you have been looking at me. My brothers, Kenzo and Guerno, helping me whenever it was necessary.

To Professor Carol Tolomini and the municipal school Paulo Freire in Magé, for opening the space, and the students of that school, for participating as volunteers. All the support given from the team of Laboratório da Performance Humana, especially doctors Fabrício Braga, Flávia Brasil and Christiane Prado for giving me the freedom to use hospital facilities and calling volunteers for my research. My colleague and physiologist Gabriel Espinosa who helped me during the whole research process.

To my advisor Professor Alberto Raposo, for his dedication during the development of this work. I am also very thankful to Professor Moacir Marocolo and the department of Physiology of Federal University of Juiz de Fora in Minas Gerais. Professor Marocolo kindly participated in the discussion of the project proposal and gave valuable advice.

Thanks to my colleagues and Professors from Department of Informatics for all the support. Thanks to Tecgraf team, in providing the devices, and sharing their incredible experience with similar projects.

Abstract

Paganelli, Antonio Iyda; Raposo, Alberto Barbosa (Advisor). **Reliability of Wii Balance Board and Microsoft Kinect for capturing posturographic information during balance tests.** Rio de Janeiro, 2018. 148p. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Body balance is an important physical skill and it is fundamental for elderly´s health, considering that falls are a major cause of unintentional injuries leading to the loss of autonomy and death in this group. Growth of aging in world population and being balance impairment one of the major causes of physiotherapeutic attendance, simple, affordable, portable, and reliable devices for evaluating body balance are of great relevance. Several studies have been examining concurrent validity and reliability of Microsoft Kinect (Kinect) and Nintendo Wii Balance Board (WBB) during balance tests. The majority of these studies suggested that those devices could be used as reliable and valid tools for assessing balance in semi-static positions. Based on that, this study investigated test-retest reliability using Kinect and WBB, concurrently, in three standing positions, and analyzed variables related to center of pressure (CoP) and center of gravity (CoG), in static manikins and in 70 healthy subjects. Each participant performed the set of tests twice in the same day. Our solution demonstrated sensibility to identify different body sway patterns. Tests showed that the most reliable variables were average speed and total path length in all directions and tasks. Despite tests with static manikin signalized excellent reliability, tests with individuals were considered poor to good. However, variables of consolidated data based on different tasks achieved excellent scores. CoP properties outperformed those related to CoG, suggesting that WBB was superior when compared to Kinect in providing more reliable body sway information. This study reinforced that these devices may provide reliable quantitative information that enhances qualitative body balance assessments.

Keywords

Body sway; balance assessment; posturography; center of pressure; center of mass; Wii; Kinect; reliability; test-retest.

Resumo

Paganelli, Antonio Iyda; Raposo, Alberto Barbosa. **Confiabilidade do Wii Balance Board e do Microsoft Kinect na captura de informações posturográficas durante testes de equilíbrio**. Rio de Janeiro, 2018. 148p. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

O equilíbrio corporal é uma importante habilidade física e um aspecto fundamental para a saúde dos idosos, sendo as quedas a principal causa de lesões não intencionais que levam a perda de autonomia e ao óbito neste grupo. Com o envelhecimento da população mundial e sendo os déficits de equilíbrio uma das maiores causas de atendimentos fisioterapêuticos, o uso de instrumentos portáveis de baixo custo e confiáveis para avaliação do equilíbrio são de extrema relevância. Diversos estudos avaliaram a validade concorrente e/ou confiabilidade do Microsoft Kinect (Kinect) e do Nintendo Wii Balance Board (WBB) durante testes de equilíbrio. Estes estudos foram favoráveis ao uso destes equipamentos nestas avaliações em posições semi-estáticas. Este estudo examinou a confiabilidade teste-reteste com o uso concorrente do Kinect e do WBB durante um teste de equilíbrio em três posições semi-estáticas, analisando variáveis do centro de pressão (CoP) e do centro de gravidade (CoG) em manequins e em 70 indivíduos saudáveis. Cada participante executou dois testes na mesma sessão. Os equipamentos demonstraram sensibilidade para identificar diferentes padrões de oscilação corporal. As variáveis mais confiáveis foram a velocidade média e o percurso total em todas as direções e tarefas. A confiabilidade foi considerada de fraca a boa nos testes com pessoas e nos testes com manequins, excelente. Porém, os resultados com as variáveis das pessoas baseadas em dados consolidados das três tarefas alcançaram confiabilidade excelente. Propriedades do CoP demonstraram ser mais confiáveis do que as do CoG, sugerindo que o WBB seja superior ao Kinect nesta tarefa. O presente trabalho corrobora estudos anteriores, podendo, estes dispositivos, prover informações quantitativas confiáveis, aprimorando avaliações qualitativas do equilíbrio.

Palavras-chave

Oscilação corporal; teste de equilíbrio; posturografia; centro de pressão; centro de massa; Wii; Kinect; confiabilidade; teste-reteste.

Table of Contents

List of Figures

List of Tables

1
Introduction

Body balance is an important skill for performing daily physical activities and it is fundamental for elderly´s health, considering that falls are a major cause of unintentional injuries leading to the loss of autonomy and death in this group (Burns & Kakara, 2018; Gale *et al*., 2018). Growth of aging in world population (WHO, 2007) and being balance impairment one of the major causes of physiotherapeutic attendance (Horak *et al.*, 2009), simple, reliable, affordable and portable devices for evaluating body balance are of great relevance.

Furthermore, balance assessment is an important tool for rehabilitation of stroke (Tyson *et al*., 2006), Parkinson disease (Palacios-Navarro *et al.*, 2015) and other clinical cases that affect the physiological systems involved in balance control (Knippenberg *et al*., 2017), such as the vestibular, optical and proprioceptive systems (Silverthorn, 2003. pp. 306).

The use of sensors and computational technology has been advanced drastically in recent years increasing their reliability and reducing their costs (Hondori & Khademi, 2014). Microsoft Kinect (Kinect) is a low cost device that provides 3D spatial information of body segments with high precision on frontal plane and for changes in positions from a stable posture (Yeung *et al*., 2014). It has been object of research in the field of body balance and postural assessments (Yeung *et al*., 2014; Clark *et al*., 2015; Lim *et al*., 2015; Díaz-Centeno, 2016; González *et al.*, 2018), not only in healthy individuals, but also in patients with Parkinson disease and other clinical cases (Procházka *et al.*, 2015; Díaz-Centeno, 2016). Kinematic evaluation of body movements and estimation of center of gravity (CoG) displacements beyond the analysis of asymmetries are usually supported by body markers and 3D cameras (González *et al.*, 2012).

Likewise, Nintendo Wii Fit Balance Board (WBB) has been used as an alternative for costly force platforms in neurorehabilitation and balance training because it may provide center of pressure information and infer data of horizontal displacements of center of mass (Gobble *et al*., 2014). Posturographic analysis

related to semi-static balance[1] tests can be performed using WBB through the identification of trajectory, path length and velocity of the center of pressure (CoP) among other characteristics associated to body oscillation. In the same way as Kinect, WBB has been used in balance tests in healthy young and elderly adults (Clark *et al.*, 2010; Chang *et al.*, 2013) and unhealthy individuals with multiple sclerosis (Castelli *et al*., 2015; Severini *et al*., 2017), visual impairment (Jeter *et al*., 2015), or chronic stroke (Kaewkaen *et al*., 2018). There are evidences that support the use of Kinect and WBB to acquire reliable and valid posturographic data (Ruff *et al.,* 2015).

Additionally, the advantage of automating a balance test with instrumented testing devices is to enhance its validity and reliability (repeatability and reproducibility). Normally, variations in range and frequency of CoP oscillations are impossible to be noticed by an observer. Therefore, using quantitative data collected by sensors and computing devices improves analysis and interpretation of balance variables, independently of observer level of training and judgment. In some specific cases, observer judgment would be affected by environmental conditions, such as when the tests are executed at high altitude. Furthermore, the high cost, the burden of setup and maintenance of force platforms or of high precision 3D cameras, make the current alternative for using quantitative analysis during balance tests very expensive, what reduces its access for the majority of the population.

There is a vast literature about reliability of WBB and Kinect showing Intraclass Correlation Coefficients (ICC) scores from .27 to .997 (Clark *et al*., 2018) for WBB, and ICC = -0.03 to 0.90 for Kinect depending on the evaluated joint and plane direction (Clark *et al*., 2015). ICC is one of most common index of reliability used in standing tests that utilizes WBB as a platform (Clark *et al*., 2018). This coefficient indicates that values close to zero mean no reliability, while values close to 1, a perfect reliability. Scores are considered poor when <.40 and excellent when >.75 for evaluating reliability of balance tests (Clark *et al.*, 2010; Bower *et* al., 2014; Severini *et al.* 2017). Additionally, some authors (Bartlett *et al*. 2014; Clark *et al*., 2018) not only recommended the use of Kinect and WBB for balance assessment tests when lower accuracy and precision is acceptable, but also tested the integrated

---

[1] Even in a quiet standing position, the body center of gravity is oscillating, then it is more precise the term semi-static position instead of static position.

solution on balance tests in old subjects (Yang *et al*., 2016). Thereby, it is desirable to compare the reliability of Kinect and WBB when used concurrently for assessing body balance. As far as we know, we could not find studies that assess test-retest reliability performing a balance test using both devices concurrently or using Kinect Xbox One for estimating center of mass through a kinematic method during three different leg support balance tasks. Moreover, it is necessary to verify the reliability of the devices simulating real proposed conditions, taking the equipment to different places and testing individuals in a variety of semi-static positions.

The Balance Error Scoring System test (BESS) is a qualitative balance test and served as a reference for our experiment because it evaluates balance in three different leg stances: double, single and tandem. BESS has been used broadly and had demonstrated moderate to good reliability, correlated with laboratory-instrumented measures for criterion-related, and construct validity (Bell *et al*., 2011; Chang *et al.*, 2014).

Therefore, the main purpose of this study was to investigate the test-retest reliability of the most important variables, related to CoP and CoG, examined during balance tests, and obtained from WBB and Kinect concurrently. Additionally, the devices are known to be portable (Yang *et al.*, 2016; Chakravarty *et al*., 2017). Then, this study also aimed at exploring the portability characteristics of the instruments. Therefore, the tests were performed in different locations with the same devices. Tests took place at hospital rooms, at university laboratory, at school classroom, and at homes. Test-retest reliability was checked under such conditions.

Another objectives were to examine the signal filtering process and the statistical tests aimed at verifying repeated measures reliability. Furthermore, it was investigated which device provided more reliable data and which variables presented the highest reliability, not only in dimensionless indexes, but also in absolute values. It is important to highlight that this study did not investigate the validity of measures captured by these devices, since theirs validities were already tested in other studies and considered sufficient for performing balance tests (Clark *et al.,* 2018, Ruff *et al*., 2015).

Statistical tests and procedures were commented out across the text, showing their strengths and weaknesses for evaluating test-retest reliability as well as the difficulty to compare results with other studies when those procedures are

not fully reported. The study investigated relative reliability using the Intraclass Correlation Coefficient (ICC) and Lin´s Concordance Correlate Coefficient (CCC) tests and the absolute reliability through the standard error of measurement (SEM), minimal detectable change (MDC), repeatability coefficient and Bland-Altman graph analysis. Along the statistical analysis, some other tests were proposed comparing coefficients of variation, differences of distributions and relative size of differences. Relative and absolute reliability were not only compared between test-retest conditions, but between CoP and CoG properties as well.

Balance tests were performed twice with a very short time interval between test and retest condition on static manikins and on individuals. On manikins, the short interval (almost immediately) was aimed at isolating devices and procedures variability, and in the same way, on individuals (around 1 minute), for the examination of within-subject variability captured by those devices.

A customized software was built up in order to manage test procedures and collect information from participants, test conditions, and from devices. Figure 1 gives an overview of the performed balance tests. Seventy apparently healthy participants with no reported balance impairment were selected to participate in this study. Figure 2 shows an example of a balance test performed with a participant in a classroom.
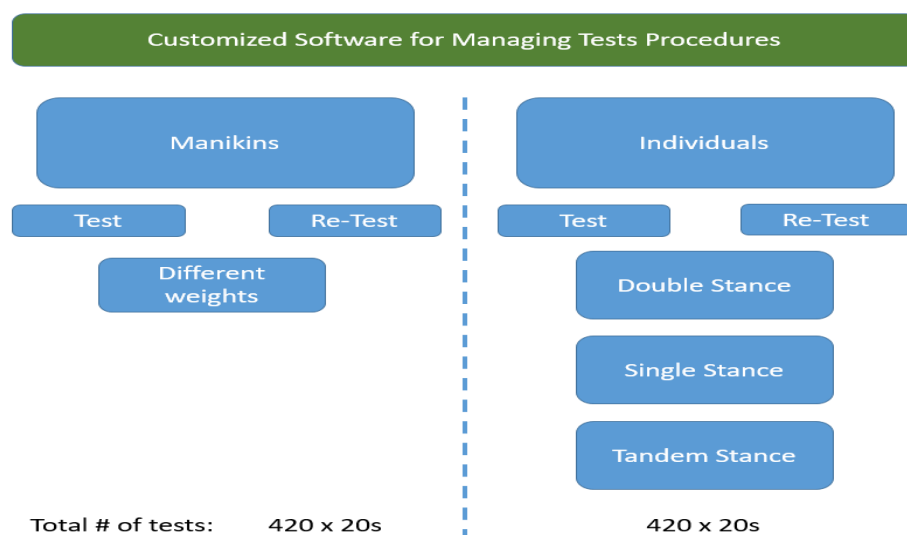
Figure 1 Test procedures applied in manikins and individuals following the same procedures and managed by the same customized software.

This project started with a literature exploration about the main concepts related to body balance and statistical approaches for reliability tests. Followed by

investigating the use of Kinect and WBB aimed at body balance training and assessment, examining studies that verified devices concurrent validity with gold standard devices, and all sorts of reliability tests.

A prototype was built with the objectives of integrating the obtained data from Kinect and WBB, and of organizing the procedures during data collection. The customized software solution logged raw data and inferred CoG and CoP displacements and their characteristics during the tests. Then, a vast set of test and retest with static manikins simulating a human body and with healthy individuals were conducted. Following, data preparation tests with different cut-off frequencies of low-pass filter were executed. A series of statistical tests were implemented to reach results that were analyzed in our study and compared to published literature. Finally, conclusions and limitations of this work were drawn based on raised information.

Figure 2 Example of an actual balance test performed at a classroom during this study.

This chapter introduced the motivation of this work, the related topics addressed for the development of this research, and a general view of the adopted procedures and results. Chapter 2 reviews important concepts of body balance, posturography, the behind technology of WBB and Kinect as well as, statistical tests for reliability examination. Chapter 3 reviews literature covering balance tests and reliability studies using Kinect or WBB. Chapter 4 describes in details the methodology for data acquisition and analysis. Chapter 5 gives an overview of the implemented routines. Chapter 6 presents the results describing manikins and individuals data. In Chapter 7 results are interpreted and compared to the literature, limitations are commented and suggestions are given for future work. Finally, in Chapter 8, main findings are highlighted as well as the contributions of this work.

2
Concepts

The objective of this chapter is to review some concepts in order to provide a better understanding of the context of body balance assessments technology and statistical procedures for verifying test-retest reliability.

2.1 Body balance

Body balance can be interpreted as the capacity to control oscillations and keep the alignment against gravity, whilst stability can be defined as the resistance to lose balance (Hall, 2009). Body balance is a result of the integration of sensory-motor system. Visual, vestibular and proprioceptive stimuli provide complex information processed by our nervous system that coordinates the motor responses in order to keep body stability (Silverthorn, 2003). Roughly, vision is involved with planning and anticipating actions. The vestibular system deals with linear and angular accelerations, and finally the proprioceptive system handles the sense of position and velocity of all corporal segments (Winter, 1995).

Stability maintenance is an important ability for performing daily physical activities and its impairment may entail loss of autonomy and death, mainly in the elderly population (Burns & Kakara, 2018). The study of body stability is also important for athletic performance and in many cases is associated with body mobility. The higher the stability the lower the mobility. Mechanically speaking, five factors define the level of stability and mobility: size of the body base of support, height of the center of gravity, location of center of gravity projection, body mass, and friction (Whiting & Rugg, 2006).

Furthermore, balance assessments aim at screening for balance problems, predicting fall risk, and evaluating interventions. Although these assessments are not directed to guide treatment, there are some studies that investigate underlying causes of balance loss (Horak *et al*., 2018).

Besides that, assessment of balance is part of the field of studies related to postural control. A posture could be interpreted as an arrangement of body joints in a time instant, and it may be associated to actions, such as walking, running, throwing or jumping, for example. Additionally, a posture may simply represent a quiet-standing position (Duarte & Freitas, 2010). A standstill position is a result of the dynamics of postural control system to maintain balance (Prieto *et al.*, 1996). Therefore, balance analysis is applied by assessing quiet or dynamic tasks. The former is evaluated with the individual staying as static as possible in standing positions. The later, some controlled perturbation is applied during the test and the responses to that perturbation are analyzed (Duarte & Freitas, 2010).

Tests can be performed with eyes open or closed. With eyes open, it is recommended that the individual look at a specific point within a standard distance. Then, all sensory-motor systems involved with balance will be evaluated acting together. With eyes closed, body balance tends to be worse without visual feedback relying mainly in proprioceptive sensors and in small extent in the vestibular system during semi-static tests.

Methods for evaluating balance can be qualitative or quantitative. In qualitative methods, the subject is observed during specific tasks and utilizes standardized procedures and rate scales. Quantitative methods, in general, utilize instrumented-devices such as a force platform or 3D high precision cameras that provide information estimating CoP and CoG displacements (Figure 3). Understanding the compromises of balance control associated with aging or of a vast set of diseases and syndromes that affect postural control may help to enhance interventions and fall risk prevention (Prieto *et al.*, 1996).

Figure 3 Force platform (left) [2] and a set of high-precision 3D cameras (right). Both instruments are used for estimating center of pressure and center of mass displacements.

[2] Image source: Bastos *et al.*, (2005).

Considering that human bodies suffer actions of internal and external forces all the time, such as gravity, ground reaction force, cardiac beats, blood circulation, respiratory dynamics or even one´s own muscular movements, the body is always oscillating and it would be imprecise to define static posture. In an empirical sense, the lower the variation of body sway, the higher the steady state (Duarte & Freitas, 2010). Mechanically, the equilibrium occurs when the sum of all forces and moments of force acting over a body is zero. It means, in practice, that it only happens in short instants of time. The name of the technique utilized for measuring body oscillation is posturography (Duarte & Freitas, 2010).

2.2 Posturography

Center of mass (CoM) is a theoretical point equivalent of total body mass, to where applying a force on it, does not cause rotation of the body. It is the point where all body segments are equally distributed. Yet, the center of gravity (CoG) is the center of mass under the effect of gravity force. They may differ because CoG refers only to the vertical direction of the force (Hamil, Knutzen and Derrick, 2015). The estimation of the whole body CoG can be obtained by the kinematic method, that is, the positions of body segments are acquired in a certain instant, for each segment it is used inertial parameters of the segment mass related to total body mass, and the position of the CoG of the segment. Then, it is calculated the body segment moment about each plane based on the acquired parameters. The sum of all moments of each segment will be equal to the total body moment, which defines the CoG. Many studies (Yang *et al.*, 2014, Yeung *et al.*, 2014) consider a model with 15 segments of the human body: head, upper trunk (thorax and abdomen), lower trunk (pelvis), two upper arms, two forearms, two hands, two thighs, two shanks and two feet.

The capacity to keep the vertical projection of body´s CoG into the body´s support base, may define postural balance, it can encompass stay upright or recover balance after an external dynamic perturbation (Juras *et al*., 2008). In its turn, the support base is a polygon area delimited by the most external edges of the body in contact with the support surface (Hall, 2009). Consequently, when the CoG vertical

projection is within the support base, there is stability for moving upper body´s segments as exemplified in Figure 4. Furthermore, the limit of stability is known as an area within the support area that limits corporal stability and entails in the functional support base (Duarte & Freitas, 2010). In such conditions, like during drunkenness, aging or high altitude, the functional support area will be smaller, although the support area remains equal, for example.
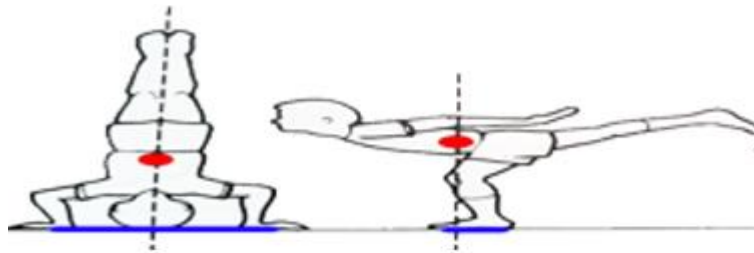
Figure 4 - Center of gravity and its projection within the base of support.

Nonetheless, the most important variable for posturography is the center of pressure (CoP) which is the resulting point of applied vertical force over a support area (Pietro *et al.*, 1996). If only one foot is on the ground, the CoP is under that foot, if both feet are on the ground, the location of CoP will be between them (Winter, 1995). The coordinates on anterior-posterior (A-P) and medial-lateral (M-L) planes often describe CoP. Frequently a force platform is used to analyze CoP behavior (Palmieri *et al*., 2002). A force platform consists in having four force sensors (load or piezoelectric cells) distributed in order to capture the three directions of the force components.

It is important to notice that CoP represents the localization of the ground reaction force resultant vector that is equal and opposite to the weighted mean of all forces acting over the force platform, such as body weight and internal forces of the body. Therefore, CoP is a measure of the displacement that depends on CoG movements. Thus, CoP represents in great extent the neuromuscular response of CoG displacement and position (Duarte & Freitas, 2010).

Normally, the trajectory of CoP is analyzed by two charts named statocinesigram and stabilogram. The first is a map of A-P versus M-L directions and the second plots planar trajectory of CoP along time (Pietro *et al*., 1996). It is possible to perform a global analysis of CoP variation based on the variation patterns in the time and frequency domains. The variables commonly analyzed in time domain are related to CoP excursion, speed, and oscillation area, because they demonstrated good reliability (Duarte & Freitas, 2010).

When analyzing a semi-static standing position, the vertical variation of CoG is minimal and much smaller than its horizontal variation. For this reason, posturography mainly focuses on the horizontal variation. This, in part, justifies why force platforms can be very simple, measuring only the vertical component of ground reaction force and the two moments on A-P and M-L axes.

A limiting factor for the use of force platforms is the great variability of CoP signal. Several repetitions of the same tasks would minimize this limiting aspect. On one hand, repetition can cause a learning effect. On the other hand, many or long repetitions of a task may cause fatigue. Both cases may influence test results. There is no consensus in the scientific literature about the number of trials that a test should apply for obtaining the best precision and accuracy of a specific position. In general, it is well-accepted two-four repetitions and a duration between 30 seconds up to few minutes (Duarte & Freitas, 2010).

Precautions with other factors are also important during a postural control analysis such as feet position, illumination, noises and any other external stimulus that may disturb the assessed subject. Another important point is that corporal stability is inversely related to the height of CoG and a common way to normalize the CoP data variation is divided it by the height of the analyzed subject.

Although there has been a vast discussion about measuring body sway, and which properties best represent it, capturing CoP and CoG is the easier part of the problem. The most challenging task is interpreting their meaning.

2.3 Signals

Natural phenomena can be expressed by signals, which represent a set of data of one or more independent variables defined by a mathematical function. Periodic functions, that is, functions that repeat themselves over time, carry the properties of amplitude, frequency, period and phase. The shortest period that the function repeats itself is the fundamental period and its inverse, the fundamental frequency. The periodic function can be decomposed into its AC (alternate current) and DC (direct current) components. The DC is the average value of the function while the AC is the oscillatory part of the function. The independent variable that

characterizes the signal may vary continuously, but capturing it by a computer device will discretized it. The name of this reduction is sampling. The frequency that the sampling occurs is the sampling rate and its unit is the Hertz (Hz). Following the Nyquist-Shannon sampling theorem the sampling rate should be at least two times the highest frequency component of the observed phenomena, if this rule is violated, the discrete signal will not correctly represent the continuous information of the phenomena. As digital signals are always discrete, then its amplitude can be specified within a resolution that dependents on the number of bits that represent it. Two bits of data can represent only four levels while four bits, sixteen, and so on (Biomechanics and Motor Control [BMC], n.d.).

A mathematical function can represent only deterministic signals. A random behavior could not be directly translated into the model. Statistical probabilities can be used in those cases with the addition of some amount of error. Most of the observed natural phenomena contain deterministic and random signals. Whenever there is a translation of a phenomenon into a signal, there is a part of the signal that is what we want to measure, and another part, that is noise. Noise can happen by random factors as well as for external interference during signal capture and discretization. The signal to noise ratio is a measure of the proportion of noise content into some data. It is very difficult to distinguish what part is the actual signal and what one is noise (Biomechanics and Motor Control [BMC], n.d.). Some techniques have been proposed to identify it.

Scientific devices and game devices have different specifications for capturing and handling signals of observed phenomena. Sampling rates, number of independent variables measured, signal to noise ratio and quantization resolutions will interfere more or less in the quality of obtained data. This interference may significantly increase variability and error, reducing the accuracy and precision of measures.

The acquisition frequency of the CoP or CoG signals is an important feature to be defined. Kapteyn *et al.* (1983) suggested that the highest valuable component in stabilometric signal would be at 10Hz. Therefore, a low-pass filter should be applied to reduce disturbance effect of higher frequencies. Then, for the quiet standing position, following the Nyquist theorem, the sampling frequency of 20Hz would be sufficient (Kapteyn *et al.*, 1983; Duarte & Freitas, 2010).

2.4 Wii balance board technology

Nintendo launched WBB in December 2007 aimed at being an interface for computer games. The platform (Figure 5) was associated to the application Wii Fit that encourages players to perform physical exercises and to keep track of their level of fitness. Additionally, it may provide information about body mass and using the data provided from each of its load cells, CoP displacement may be estimated.
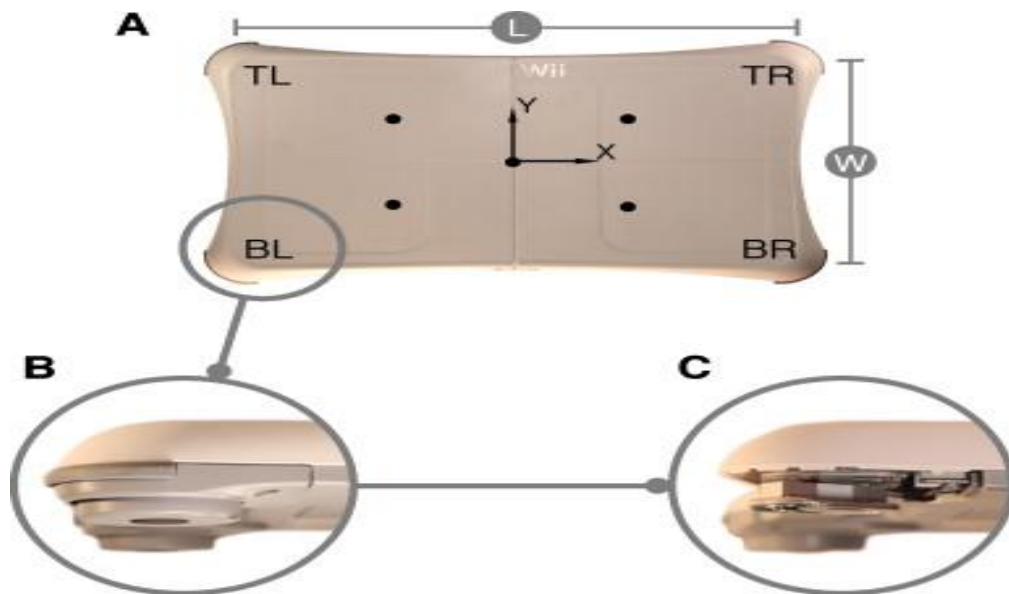
Figure 5 Top view of the Wii Balance board (A) showing the position of each load each in the top left (TL), top right (TR), bottom left (BL) and bottom right (BR) positions. In the detail (B and C), it is highlighted the load cell[3].

WBB has some similarities to devices used in clinical diagnosis such as force plates and baropodometers. Beyond that, it has some advantages when compared to commercial-grade force plates such as being much cheaper, portable and easier to configure. These facilities when applied to body assessments may become the tests more accessible for the population.

The device has four axial transductors located in each edged of the platform that are used for analyzing the force distribution exerted over it (Bartlett *et al*., 2014). These sensors can verify the vertical force applied on them and calculate M-L and A-P asymmetries, similarly to scientific force platforms. Each transductor is a load cell that is a metal ligament linked to a tension measurer. The conversion of the force applied over the load cell to a voltage value is digitalized. Then, the

---

[3] Source: Bartlet *et al.*, (2014).

obtained information is transmitted via Bluetooth to other devices. WBB sampling rate may reach 100Hz and accurately measures up to 150 Kg. The dimensions of the platform are about 45 x 26.5 cm (Wii brew Website).

There are some open-source libraries that facilitate the establishment of Bluetooth connection and data acquisition from WBB. A general specification of the data structure can be found at (Wii brew Website) and customized software solutions are available for free download. However, WBB presents many limitations when compared to commercial or scientific force platforms. WBB plastic surface suffers more deformation when compared to metal surfaces of plate platforms if heavy loads are applied over it (Leach *et al.*, 2014). Additionally, force plates provide tri-axial information about forces and moments while WBB provides only data of vertical forces limiting the studies about horizontal and shear force components.

The use of WBB during balance tests and its characteristics for capturing center of pressure information will be explored in section 3.1 on page 38.

2.5   Microsoft Kinect technology

Microsoft Kinect is a human-computer interaction device that captures images and sounds through an RGB camera, infrared sensors and a set of microphones (Figure 6). It allows the interaction of users and a computing system through body movements and/or voice commands. It was aimed at enhancing users' experience with computer games, controlling avatars or other menu options.

Kinect was launched in November 2010 in a version known as Kinect for Xbox 360 (Kinect v1). In 2012, a second version was released under the name of Kinect for Xbox One (Kinect v2). It was estimated that 35 million of units were sold by October 2017 (Wikipedia Kinect).

Microsoft provided open access to software development libraries and documentation focused on developers of games and other applications in order to propagate the use of Kinect technology. Therefore, this low cost device opened new possibilities in several fields, like virtual and augmented reality, as well as in applications for physical rehabilitation, assessment, and identification of motor patterns.

Kinect v2 has a high-resolution RGB camera (1920 x 1080 pixels) and infrared sensors with depth resolution of 512 x 424 pixels. The depth accuracy is relatively constant at about 2mm in the central view cone, increasing to 2-4mm up to 3.5m distance (Wang *et al.* 2015). Differently from Kinect v1, v2 requires a USB 3.0 connection.

Figure 6 Microsoft Kinect fox Xbox One.

The infrared sensors are used to calculate the distance of objects from the center of the camera using a technique known as Time of Flight. It interpolates three different infrared signal frequencies and calculates the time between the emission and capture of those signals back in relation to the sensor lens center. Using the time and the speed of light it estimates the distance of the objects in its field of view (Figure 7). It also considers the focal distance and the distance of a reference point to calculate the real distance of points (Valgma, 2016). Following Wasenmüller and Stricker (2017), the position mean error would be about 18 mm in a distance up to 3 meters from the camera and the variability would increase according to the distance from the device.
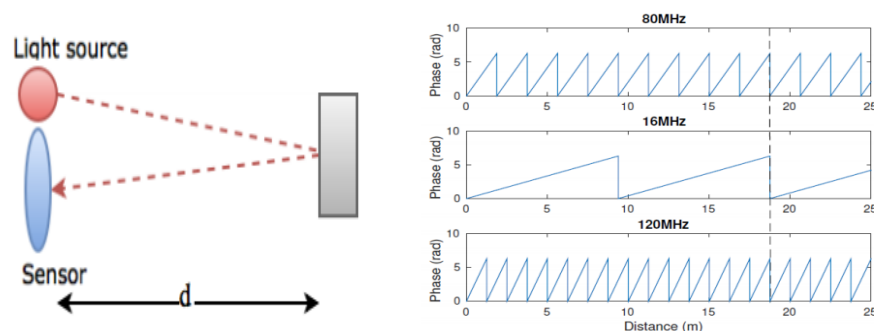
Figure 7 Depth sensor 3D reconstruction using Kinect v2 camera.[4]

Kinect v2 is able to identify up to six persons in its field of vision, but it keeps track only of two of them. It is able to estimate the 3D position of twenty-

---

[4] Source: Valgma, (2016)

five points representing the main joints of the human body at 30 Hz, frame by frame, based on a deep randomized decision forest algorithm. Thousands of synthetized images were used for training each decision tree of this forest (Shotton *et al*., 2001). After inferring the joint positions, it is possible, using simple algorithms, to identify static and dynamic poses and postures. As well as WBB, Kinect has the same advantages when compared to commercial 3D motion capture cameras, it is affordable, portable and easy to use, which may amplify its adoption when used for evaluating body balance. The use of Kinect during balance tests will be investigated in section 3.2 on page 41.

2.6   Reliability tests

When using instruments for measuring a specific property of an object it is important to know the accuracy and precision of these instruments. Accuracy is determined by the degree of agreement between the observed value and the real value. It is related to the validity of the measurement. Validity is checked, in general, comparing new methods to a gold standard reference. Yet, precision concerns measuring the same object many times and the agreement between repeated observed values. It is related to the reliability of the measurement. Reliability encompasses repeatability of a measure, which means two measurements on the same samples, and reproducibility that means two different observers performing the same procedures on identical samples (Watson & Petrie, 2010). Reliability is important in health and sports studies, mainly to understand within-subject variability as well as an effect or lack of it on analyzed variables due to interventions.

Different statistical methods have been proposed to estimate the degree of agreement between measurements. This section reviews basic concepts about agreement and the most important concordance methods used with continuous scale data in the literature.

2.6.1. Variability and error

It is common to reach inequalities in results when repeating a measurement of a variable on different individuals or on the same individual. This happens

because there is a variation between individuals, and a variation within the same individual when tested in different occasions. These variations are known as measurement variability. This concept differs from error measurement that refers to the differences between the observed value and the true value. The true value of any variable is very difficult to obtain and sometimes impossible. However, the literature accepts, in most cases, the term true value when referring to a measurement obtained by a gold standard method.

Error can be random or systematic. In a random error, the observed value has an aleatory behavior, being higher or lower than the true value. The systematic error has a bias or tendency of the observed value in comparison to the true value (Watson & Petrie, 2010). Knowing the variability and error of a method is fundamental to interpret correctly the measured values of any phenomena.

Generally, reliability is expressed as the ratio of (Koo & Li, 2016):

$$\frac{True\ Variance}{True\ Variance\ +\ Error\ Variance}$$

A reliability rate close to zero represents no reliability. In this case, error variance is very high when compared to true variance. On the other extreme when reliability rate is near to one, reliability is very strong and error variance is negligible.

2.6.2. Repeatability

As mentioned before, repeatability and reproducibility are measurements of reliability. They are concerned to the consistency of measurements. Since this study uses only numerical data, our approach in this section will not cover categorical data and our focus will be on repeatability tests.

Some studies demonstrated that there is no advantage in using more than two or three measures per subject, and that is better to compare pairs of results on a greater number of individuals than many results on fewer subjects (Walter, Eliasziw, & Donner, 1998).

The most basic test for checking repeatability is aimed at verifying systematic differences between the two measures. Depending on normality of data a paired T-Test or Wilcoxon signed ranks tests may be applied. The null hypothesis of these tests indicates that the mean differences between the two measured values are zero, meaning that there is no evidence of a systematic or bias difference.

However, a significant result of those tests may suggest that there is a systematic difference in the measures. It may be related to a constant and/or to random factors.

The linear correlation between repeated measures is another way to check repeatability. A high correlation coefficient indicates that correspondent pairs follow the same behavior and a regression line estimates the mean differences between measures. However, correlation tests such as Pearson or Spearman are not ideal for checking agreement, because they do not find bias between measures (Chen & Barnhart, 2008). Paired data can correlate perfectly, but with no agreement as shown in Figure 8. On the top of the graph, the red points represent a perfect correlation (r = 1.0), the best fit linear model perfectly represents the measures. However, there is a clear bias where all values of method 1 have higher values than those of method 2. Yet, the blue points have a poor correlation (r = 0.5) and a poor agreement, none of the measures coincided. However on average, the methods agreed, the best fit linear model showed no bias, in a perfect 45º line (Watson & Petrie, 2010).
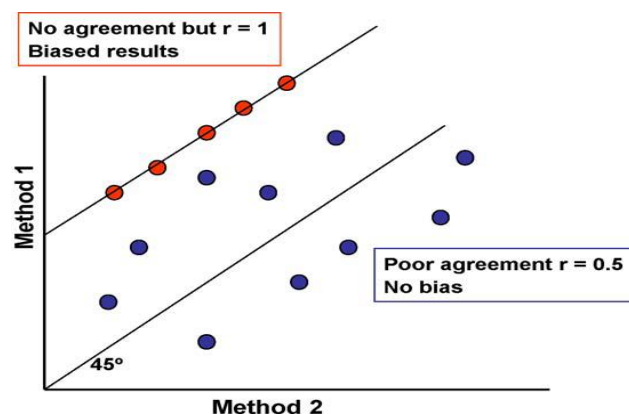
Figure 8 Correlations of measures of different methods. In the red point example, correlation coefficient (r) is perfect, but there is a bias on results. In the blue points example, correlation and agreement are poor, but on mean, there is no bias between the two methods measures.[5]

2.6.3. Lin's Concordance Correlation Coefficient

A concordance coefficient (CCC) was proposed by Lin (1989) for comparing the agreement between two readings from the same sample when measures were in a continuous scale, focused on reproducibility of results. When results from a new method are plotted against those of gold standard in a scatterplot graph and a best-fit line represents the linear model of the relationship between the

[5] Image source: Watson & Petrie, 2010.

two methods, Pearson correlation coefficient will estimate the distance of the points to the best-fit line. Lin´s coefficient beyond of assessing how close the data is about the best-fit line, it also evaluates how far that line is from a 45-degree line passing through the origin (Watson & Petrie, 2010). Not only the linearity, but also the additivity relationship between values measured by different methods should be considered to attest reliability.

Lin´s coefficient can be calculated using the formula:

$$r_c = \frac{2\, r\, S_x\, S_y}{S_x^2 + S_y^2 + (\bar{x} - \bar{y})^2}$$

r = Pearson´s coefficient, $S_x{}^2$, $S_y{}^2$ = data variances, $\bar{x}$ $\bar{y}$ = data series means.

Then, CCC has been proposed as a more trustworthy index to estimate the degree of concordance in reliability studies. In general, this index is not used to compare repeated measures. It is based on Pearson's correlation test coefficient that supposes two independent variables. In order to reach independent repeated measures, the subject should perform the repeated trial without any information about the first one, what is quite difficult because people self-evaluate their performance or may feel the effort effect of the first trial.

However, this basic formula was expanded to better handle outliers, and treat more than two observers, as well as, to deal with repeated measures, as proposed by (King, Chinchilli & Carrasco, 2007). In that study, repeated measurements variance was estimated through U-statistics methodology performing well with many or few repeated measurements.

2.6.4. Intraclass Correlation Coefficient

Intraclass Correlation Coefficient (ICC) index is widely used in the literature and there are different forms of calculating it (Koo and Li, 2006). The correct ICC formula to be used depends on the type of study and its design peculiarities. When measuring reliability there are, at least, three different scenarios: interrater, test-retest, and intrarater reliability. The first one occurs when there are two or more raters evaluating the same group of individuals. The second reflects the variation of measures taken by an instrument on the same individuals

under the same conditions. Finally, the last one reflects the variation of measures taken by one rater across two or more trials (Koo and Li, 2016).

Following Shrout and Fleiss (1979), each ICC model can be labelled: 1, 2 or 3. The label 1 designates study designs when each subject is evaluated by a different set of raters. The second, when the subjects are evaluated by the same set of raters, and these raters were selected randomly from raters' population. Finally, the third model is used when the subjects are evaluated by the same group of raters, and these raters are the only group of interest. Beyond that, a second label designates if it is going to be used only one score, then 1, or the mean of several scores, using the letter 'k' to represent it (Weir, 2005).

ICC is based on analysis of variance (ANOVA) and three sources of variability are specified, subjects, trials and error. If systematic and random errors variability is treated uniquely, then it is called a 1-way model. If they are treated separately, then it is called 2-way model (Weir, 2005). An ICC score of zero means no reliability and one a perfect reliability.

Another way to classify the type of ICC to be used was proposed by McGraw and Wong (1996). In this case, for example, our study that was interested in the reliability of the same instrument measuring the same individuals on the same conditions, it would be classified as a two-way mixed effects, with absolute agreement, and single rater / measurement, an ICC (A, 1) equivalent of Shrout and Fleiss (1979) convention to $ICC_{2, 1}$.

In test-retest design study, it is recommended to use absolute agreement instead of consistency, because the repeated measures should reflect the degree of agreement (Koo & Li, 2016). Then, the formula for calculating this type of ICC is:

$$ p = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)} $$

MS_R = Subjects Mean Square (rows), MS_E = Mean Square Error, MS_C = Trial Mean Square (columns). n = number of subject, k = number of measurements.

Clarifying the formula, MS_R indicates the estimate variance of between subjects, the inter-subject variability, while MS_E indicates the estimate the error term of the within-subject variability (the interaction between subjects and trials,

random error), and MSc is the within-subject variability, or trials effect (systematic error). Then, the formula is a variance of the between-subject variability, representing the true value, over then between-subject variability plus the error variability, representing the systemic and random errors.

ICC and CCC results are very similar to each other when applied to the same study design. However, ICC assumes that ANOVA assumptions hold.

Shoukri, Asyali and Donner (2004) reached a conclusion that for continuous measures a small number of replicates would minimize the variance of estimated ICC in reliability studies. Moreover, for clinical investigations where an ICC coefficient of at least 0.60 is required, the authors recommended to apply two or three replications per subject. Additionally, Duarte and Freitas (2010) revision on posturographic methods suggested the use of two to four samples of CoP. Koo and Li (2016) recommended to report not only the ICC coefficient, but also its 95% confidence interval and the utilized significant value.

There is some criticism on indexes of reliability since they do not provide to decision-making actors, in a practical sense, the relevance of a measure, because they are dimensionless scores (Hopkins, 2010, Weir, 2005). On the other hand, with dimensionless variables are easier to compare results among different studies.

2.6.5. Standard error of measurement and minimal detectable change

The indexes of concordance, ICC and CCC are expressions of relative reliability and the standard error of measurement and minimal detectable change represent absolute reliability. Relative reliability informs the consistency in the positions of scores from the same individual within a group during repeated measures. Absolute reliability will estimate the variability between repeated measures. A high absolute reliability informs that the variability is low in repeated measures and vice-versa (Ries *et al.*, 2009).

A simple way to measure within-subject variation is the standard error of measurement (SEM) (Hopkins, 2000). It tries to provide an indication of the dispersion of measurement errors between observed and true score (Brown, 1999), and it determines absolute reliability instead of relative reliability. SEM can be calculated using the formula:

$$SEM = S * \sqrt{1 - r_{xy}}.$$

Where S is the standard deviation of the test, r_{xy} is the estimated reliability coefficient of the test. Commonly, it is used the ICC or CCC as the reliability coefficient, but it may be used a correlation coefficient as well. Moreover, the standard deviation (SD) can be chosen from the first measure, the mean SD of both measures, or the SD of the differences (Cook *et al.*, 2014). Since, the intention is to measure the within-subject variation that is part of the method variation, not a real effect of an intervention, and interpreted as an error, it was selected in our study the within-subject standard deviation as defined by Bland and Altman (1999). Being the square root of the mean square of residuals, following the formula:

$$S_w = \sqrt{\frac{\sum_1^n (y_i - x_i)^2}{n}}$$

n = sample size, y_i = retest observation, x_i = test observation.

Based on SEM, it is possible to estimate the minimal detectable change (MDC), which is the amount of change within a statistical confidence level that is beyond the measurement error. MDC is considered a general measure of reliability and normally it is used the 95% confidence level, following the formula:

$$MDC = 1.96 * \sqrt{2} * SEM$$

Although, some authors have argued that 95% may be too conservative in some applications, for example, in high performance sports where very small changes in performance are very significant. MDC and SEM can be represented as percentages of absolute values of observed variable, allowing comparisons among different studies' designs as well.

2.6.6. Limits of Agreement

Bland and Altman chart (B-A) analysis is another method for analyzing agreement between repeated measures (Weir, 2005). Although, the method has been used vastly to compare the differences between two methods, the authors proposed, a variation for the assessing repeated measures (Bland and Altman, 1999).

A scatter plot with the means of pairs plotted in the horizontal axis and the correspondent differences in the vertical axis shows the limits of agreement (LOA) based on standard deviations of pairs' differences. Two lines, representing mean +/- 1.96 standard deviations, are printed in the charts. Assuming that the differences

are normally distributed, points within those lines represent 95% of the found differences. Moreover, it is plotted a line expressing the mean difference between the measures, which represents the systematic error. Systematic error includes constant error and bias (Weir, 2005). Bias can be analyzed if the differences vary depending on the magnitude of values, a funnel form may be seem on the plot. A correlation test may check if exists any monotonic relationship between the pairs and their differences (Hirata & Camey, 2009). The specific application and objective of the evaluation will determine if the found differences and their distribution would be reasonable or not. Then, B-A is a method for organizing and facilitating the identification of the differences and their behavior.

Finally, for analyzing repeated measures, Bland and Altman (1999) proposed to calculate a variable named reliability coefficient (RC) that is calculated using the traditional 95% confidence interval, following the equation: $1.96 * \sqrt{2} * S_w$. Where, S_w is the within-subjects standard deviation. Square root of 2 was used because this variation is based on two trials, the first observation and the second one.

3
Related Work

The objective of this chapter is to review relevant studies of reliability using the WBB or Kinect, especially during balance tests.

3.1   Reliability of Wii Balance Board in posturographic studies

Recent studies have been using WBB as a tool for analysing body balance and body oscillation variables (Bartlett *et al*., 2014; Sgrò *et al.*, 2014, Castelli *et al*., 2015; Severini *et al*., 2017). Reliability of WBB was tested in healthy young and older subjects, stroke survivors, visual impaired and multiple sclerosis patients (Clark *et al.*, 2018).

Sgrò *et al.* (2013) suggest that Computing Science researchers should work with human movement professionals in order to develop customized game applications based on experiences of physical activities. Sgrò *et al.* (2014) compared results obtained from WBB and a high-quality force platform during two semi-static balance tests, with open and closed eyes. They evaluated CoP total path length (TPL) showing high and significant correlation coefficients between the devices. ICC achieved excellent levels and chart analysis using Bland-Altman plots demonstrated good validity of TPL. Additionally, similarity T-tests were consistent with above mentioned results not finding significant differences between the two instruments. The authors concluded that the agreement of values obtained from both devices confirm the role of WBB as a low-cost, portable and friendly-use device for balance assessment, as well as the need of a customized software for data collection.

On the other hand, Pagnacco *et al*. (2011) highlighted the low resolution (0.5 mm), low and inconsistent sample rate (time jitter), low signal to noise ratio and glitches in the WBB data. The authors suggested that WBB signal noise might be caused due unshielded cables, electronics components incapable of reducing noise, and unsynchronized sampling across the four sensors.

Bartlett *et al.* (2014) verified the precision of measures of force and CoP coordinates among nine different WBBs, three of them lightly used and six heavily used, and comparing to a scientific-grade platform. Uncertainty and reliability of WBB was tested across varying sway conditions. The authors found total uncertainty of CoP measurement to be within 4.1mm across the nine-tested WBB, much higher than that recommended for posturography (0.1 mm), and when using the same WBB, it decreased to 1.5 mm. Yet, the authors indicated that WBB behave linearly with statistically significant increase in error from center to corner and from light to heavy static loads. WBB has a linear behaviour across the surface of the platform only increasing in 1 mm the measured error in different localizations with growing loads. Besides that, there was no effect of wear on mean CoP measurement error. Table 1 summarizes findings in Bartlett *et al.* (2014) study.

Table 1. Force and CoP coordinates differences using distinct Wii Balance Boards and the same platform related to a high precision force platform.

Test conditions	Force variation	Center of Pressure variation
Distinct WBBs	+/- 9.1 N	+/- 4.1 mm
Same WBB	+/- 4.5 N	+/- 1.5 mm

Additionally, that study also compared internal calibration values with those calculated experimentally by them. The authors declared that extra calibration on WBB was not necessary because the mean error between fabric and customized methods was lower than the uncertainty degree of the device. In addition, WBB would be sufficient precise for detecting postural variation differences greater than 10mm, what, following the authors, may distinguish healthy individuals from the ones with balance impairment. Finally, the authors concluded that WBB could provide useful information about force and CoP displacements, in situations that low accuracy and low precision were acceptable.

Leach *et al.* (2014) in contrast to Bartlett´s *et al.* (2014) study used controlled dynamic sways instead of semi-static sways for assessing WBB reliability. They utilized a customized mechanical system of an inverted pendulum to evaluate the dynamic load. It was controlled the pendulum's displacement angle and load height in order to modulate CoP displacement and to analyse the frequency, amplitude and direction (A-P vs M-L) across 12 different WBB. The authors had two main objectives, measure CoP error under controlled dynamic conditions comparing simultaneous results from WBB and a laboratory-grade force

plate, as well as to determine inter-device variability. The tests were performed in 30s trials and the weight over the platform was 58.3 Kg. They found higher error levels on M-L plane compared to A-P plane with WBB. The average CoP measured error across all sway amplitudes and WBBs was about -10% and -11% for all A-P and M-L time-domain measures, respectively. WBB overestimated force-plate measures. Applying a proposed formula for calibrating the results, the error fell to -0.01 and -0.18%. Moreover, after calibration, similarity tests showed no statistical differences between WBB and the scientific-grade force plate. The authors concluded that WBB was an affordable, portable, and easily accessible device that would be used where low-accuracy and low precision of CoP displacements were plausible. The linear calibration used in their study was recommended to provide more reliable time-domain CoP measures with WBB.

Goble *et al*. (2014) reviewed the use of WBB as a tool for neurorehabilitation and concluded that WBB provides valid and reliable data of CoP after the analysis of several scientific work.

Kaewkaen and colleagues (2016) assessed reliability of WBB in thirty stroke patients during a double stance task with eyes open and closed for 30s. Participants performed three trials with 10-minute intervals. Sway path, velocity and area were calculated. All analysed variables reached excellent $ICC_{3,1}$ scores with 95% confidence interval scores >.82, suggesting that WBB could be used for evaluating balance in this population.

In a broad revision made by Clark *et al*. (2018) in six scientific bases, twenty-five studies were selected from 2078 distinct articles. The authors assessed balance researches that used WBB in relation to reliability intra-rater, test-retest and concurrent validity that compared results to gold-standard force platforms. The majority of studies supported the validity and reliability of WBB when compared to scientific force platforms in assessments of healthy and unhealthy subjects. Among the selected articles, twelve of them investigated reliability, twenty-one, concurrent validity and eight examined both properties. Some evidences were conflicting, WBB presented several limitations, such as inconsistent sampling rate, noise-signal rate inadequate and inefficiency for detecting shear forces and their moments. Table 2 shows some of the parameters used in balance tests consolidated in this revision.

Table 2. Most used parameters in balance tests with Wii Balance Board.[6]

Sampling rate	25 – 100 Hz
Low-pass filter cut-off frequencies	6 – 12 Hz
Test procedures	Double/single stance, open/closed eyes, rigid/foam support, with/without visual stimuli.
Duration	10s – 50s (more often 30s, 56% of studies)
Number of tries	3 (60% of studies)
Analysed variables	CoP total path length or velocity (92% of studies)

Yet, this review reported that reliability of WBB compared to gold-standard platforms was considered moderate to excellent in line with previous reports in literature. The CoP path length variable consistently showed excellent reliability (ICC > 0.75). Most of studies reported excellent concurrent validity and the authors concluded that WBB could provide data with concurrent validity compared to commercial force platforms. Moreover, WBB has comparable reliability to force platforms used for computerized posturographic analysis in standing positions. Thus, WBB could substitute conventional force platforms for analysing slow movements (0.1 – 10 Hz).

The conclusions of Clark *et al.* (2018) are in agreement with Ruff *el al*. (2015) study that after a systematic review evaluating the validity and reliability of WBB and Kinect concluded in favour of the utility of these devices in clinical diagnosis and functional assessments in orthopaedics clinics.

3.2 Microsoft Kinect and balance applications

Several studies also proposed the use of Kinect for assessing postures related to human balance (Yang *et al.*, 2014; Yeung *et al.*, 2014; Clark *et al*., 2015; Lim *et al.*, 2015; Diaz-Centeno, 2016; Eltoukhy *et al.*, 2018). The reported advantages would be the low cost, the facility of use, as well as the quality of information provided by Kinect. (Yeung *et al*., 2014). Kinect obtains joint position without fixing markers on body joints or corporal segments. The device can estimate whole body´s center of mass using the method of segmentation analysis and it would have an advantage on capturing more accentuated, quick and frequent

---

[6] Clark *et al.* (2018).

oscillations when compared to commonly used methods for calculating CoP in force platforms (Yeung *et al.*, 2014).

Yang and colleagues (2014) studied the reliability and validity of center of mass (CoM) properties estimated by Kinect v1 during three semi-static standing positions in nine healthy subjects. The study compared the results to a laboratory-grade 3D motion capture system and estimated CoM by a kinematic method using a 15-segment body model and anthropometric data of Chinese subjects. Individuals performed two successful trails with 1-minute interval in double stance, feet shoulder width apart, feet together (30s), and unipodal standing (15s). Kinect and Optotrak Systems captured data simultaneously. The horizontal displacement of CoM was quantified, the root mean square (RMS), and the average velocity were estimated. ICC scores between the trials was >.75 for all positions, and for all variables. Moreover, ICC scores between devices were > .88 for all properties as well. The study concluded that Kinect v1 had excellent test-retest reliability and good concurrent validity for assessing standing balance.

Yeung and colleagues (2014) investigated the reliability and accuracy of Kinect v1 in relation to a force platform and to eight 3D cameras Vicon system. CoP and CoM comparison was implemented with a zero-point-to-zero-point integration technique that estimates the gravity line projection from CoP data. CoM was estimated using the kinematic method using 15-body segments model. Ten healthy individuals were recorded during 60s in each of the four test conditions: double stance with eyes open, with eyes closed, on foam with eyes open, and with eyes closed. Only the central 40s were considered for analysis and each participant performed two trials in each condition. Sway range and standard deviation of CoM trajectory in anterior-posterior (A-P) and medial-lateral (M-L) directions were computed. ICC scores demonstrated a large range of variation as low as 0.16 in M-L range during eyes closed without foam, and as high as 0.81 in A-P range in the same test condition. Coefficient of variation also presented a high range from 15.5% to 59.3% depending the test condition. Compared to Vicon system, Kinect overestimated CoM in both directions by a mean of 2-4mm and it was larger in A-P axis than M-L direction. Paired t-tests comparing $ICC_{2,1}$ scores and coefficients of variation among methods demonstrated no significant differences.

Kinect v2 was analysed by Clark *et al.* (2015) during a test-retest condition and concurrent validity with a marker-based 3D-motion analysis system with nine

VICON cameras. Kinect and VICON cameras recorded data simultaneously during the tests. Thirty healthy individuals performed two sessions with an interval of one week of semi-static balance tests in single and double limb support with eyes open and closed. In addition to that, participants also executed dynamic balance tests consisted of forward and lateral reach. Single limb test had the duration of 15s while double limb support took 30s. In double stance, feet were 10cm apart. Anterior-posterior (A-P) and medial-lateral (M-L) displacements were estimated using default joint coordinates provided by Kinect, such as shoulder center and spine base. Double leg tests had modest to excellent reliability on both devices ($ICC_{2, 3}$ >= .70) in A-P axis and path length values were more consistent than range values. Single leg presented lower reliability than double leg trials in A-P direction for both devices. In M-L plane the unique test that were considered valid for Kinect was the single leg with eyes closed.

That study also estimated standard error of measurement, which presented a high range of results from 8.5% of trunk lateral flexion angle during lateral reach trial up to 60.9% for M-L pelvis range during double stance with eyes closed. Authors suggested that Kinect A-P measures could be used confidently during double stance, but M-L values should be taken with caution, except for single leg with eyes closed task. Finally, the study concluded that Kinect v2 has potential to be used in tests to evaluate some aspects of body balance.

Otte and colleagues (2016) also explored Kinect v2 characteristics and verified accuracy of 21 joint information during six different balance and lower limb tasks using a 16-camera VICON system as a gold-standard reference. The study also assessed several clinical parameters derived from those tasks and compared their precision. Nineteen healthy participants attended one test session. Dynamic tasks like stand up and sit down, short comfortable walk, short maximum speed walk, short line walk, and walking on the spot were performed. A semi-static task composed by stance with feet together and keeping eyes open and closed for 20s each. All tasks were repeated 5 times, except for walking on spot that was executed only 3 times. Accuracy in A-P and M-L planes was good to excellent with only moderate correlations. Most of the 45 clinical parameters demonstrated good to excellent agreement ($ICC_{2, 1}$ > .70, for 30 parameters) and consistent (r > .70 for 38 parameters). Trunk movements had high agreement in sway speed during quiet standing. $ICC_{1, 1}$ was used to measure repeatability and Kinect showed values >.60

for 33 parameters while VICON for 30 parameters. Standard error of measurement as a percentage of mean was lower than 20% for 31 Kinect parameters. $ICC_{1,1}$ for quiet balance test had a range of results of .41 (.18, .69) up to .82 (.67, .93) for deflection roll and resultant mean sway velocity, respectively. Then, authors concluded that Kinect v2 had potential to be used as a clinical measurement tool.

Grooten *et al.* (2018) verified the reliability and validity of a system named Quinematic[tm] that used Kinect v2 to capture information and derive posture, balance and side-bending properties. Using this system, the majority of analyzed variables demonstrated $ICC_{3,1} < .40$ in a test-retest reliability design with 37 participants performing a complete set of tasks in two different occasions, 6-8 days apart from each other. The complete test included 7 different tasks like stand still with arms at sides, side-bending, two-leg squat, one-leg balance on right leg, on left leg, one leg squat right, and left sides. The conclusion of this study was that this system should not be used for monitoring balance posture over time nor in research.

Eltoukhy *et al.* (2018) investigated Kinect v2 in a group of 10 young and another group of 10 old heathy participants. Twenty individuals were recorded by Kinect v2 and eight infrared cameras processed by Vicon Nexus software. The study aimed at examining the validity and reliability between instruments. They performed three tests consisting in single leg stance with open eyes, closed eyes, and leg sways. Each task was executed twice in a one-day session. During the dynamic trial, the subject was asked to sway the free leg as long as possible. Segmental method using 15 body segments model was applied to calculate total body center of mass (CoM). Consistency and agreement between instruments were considered excellent for all analyzed variables for the whole group (ICC >.75). Despite concordance between systems varied from poor to almost perfect when analyzing through Lin´s Concordance Correlation Coefficients depending on group, task and property. Nonetheless, when groups were combined, CCC scores were >.90. Properties analyzed were CoM range and speed in anterior-posterior and medial-lateral directions, as well as, mean distance. Authors suggested that Kinect might be used for assessing balance parameters in clinical and research environments.

3.3 Simultaneous use of Kinect and Wii Balance Board

Few studies investigated the concurrent use of Kinect and WBB during assessments of static or dynamic human balance. Reliability of these devices was evaluated separately and we did not find studies of reliability during body balance tasks using both devices simultaneously.

Galeano and colleagues (2014) proposed the use of both devices with an electric muscle stimulator as a non-expensive tool for using in neurorehabilitation treatments. The solution aimed at performing computerized posturographic analysis together with balance training in the medial-lateral and anterior-posterior axis through stimulation of synergist muscles using electric impulses. WBB was responsible for registration and follow-up of CoP and Kinect provided body position information while the electric stimulator triggered the target muscles. However, the major challenge of this solution was the method for synchronizing all devices in order to generate the electric impulses for the correct synergist muscles at the right time.

Dutta *et al.* (2014) also proposed a solution using Kinect and WBB using a technique of non-invasive cerebral stimulation for improving balance. Kinect calculated and kept track of center of mass (CoM) trajectory while WBB followed the center of pressure (CoP). Using both data, it was possible to execute a posturographic analysis using CoM and CoP trajectory. This information was then compared to the scores got using the scale of BERG balance test, a popular qualitative balance test. The authors concluded that both devices were appropriate for giving posture feedbacks during transcranial direct-current stimulation.

In a case study (Levinger *et al.*, 2016), both devices provided visual feedback during rehabilitation program sessions in patients who had undergone total knee replacement surgeries. The authors stressed the importance of the solution in the expressive improvement of knee function after only six weeks of treatment.

Finally, Chakravarty and colleagues (2017) utilized Kinect and WBB in conjunction with the software OpenSim for creating a musculoskeletal model from the data acquired by Kinect, the ground reaction force and CoP derived information provided by WBB. The objective was to analyze the function of lower limb muscles for maintaining body balance in a single stance position and compare the

differences of muscle activation between obese and non-obese groups. The authors did not relate device limitations for performing the analysis and successfully identified distinct patterns of muscle activation between groups.

Reliability of Kinect v2 and WBB had not been verified simultaneously during balance tests of different leg support tasks using the kinematic method for calculating whole body CoG. Moreover, using static manikins and replicating the tests with participants in different locations following the same procedures. Beyond of providing a vast set of reference values for these tests using the devices, our study also intended to compare the reliability of both devices during the same test and compare them between test and retest conditions and between obtained data from each instrument. Additionally, it will be able to examine the differences between values of the balance test properties with the unavoidable variability when testing humans and without it. Different methods of checking reliability were also investigated in order to amplify the comparison with other studies and to observe their strengths and weaknesses. Methods for evaluating relative and absolute reliability given a whole set of parameters regarding the repeatability consistency of these devices. These tests may provide valuable references for expanding the use of WBB and Kinect during balance test due to their reliability, low-cost, easiness of use and portability, what may increase accessibility to health services.

4
Implemented solution

This study performed a set of standardized procedures in order to obtain signal information from WBB and Kinect during the experiment. A customized software was built-up in order to record participants' profiles, setup trials parameters, organize test sequences, capture raw data from devices, calculate center of mass and log coordinate data points for future transformation and analysis.

Additionally, a set of script routines were developed in order to prepare, analyze collected information, perform statistical analysis, and pre-generate graphs and tables.

4.1 Application requirements and functionalities

Although, the system was not intended to be distributed or to be used outside the scope of this research, some requirements were outlined.

The functional requirements were:

- Establish and configure the communication and access to internal parameters and properties of WBB and Kinect.
- Register participants' profiles.
- Register balance tests administrative information, such as date, time, trial identification, participant or manikin identification etc.
- Manage balance test application procedures, standardizing data collection and controlling test durations.
- Provide real-time feedback of estimated center of pressure and center of gravity displacements during the tests.
- Provide flexibility for performing as many repetitions as necessary.
- Prevent overwritten stored information during test repetitions.
- Allow self-application of balance tests.
- Store structured data obtained from devices during the test
- Allow verifying consistency of obtained data.

The non-functional requirements were:

- Be friendly, providing a graphical user interface.
- Obtain data from Kinect and WBB simultaneously at 30Hz.
- Utilize only open-source or free of charge development frameworks, libraries or any software resource.
- The application should not require special hardware and should be able to run on an affordable notebook.

Since Kinect is native of Windows operational system, this was chosen in order to avoid incompatibilities considering that WBB libraries were also available for Windows. The chosen programming language was C# and the .NET framework using Visual Studio as the integrated development environment (IDE). This development environment beyond of being native for Windows, it is relatively easy to develop graphical user interfaces and a windows-based application.

Some requirements such as connecting to Kinect were reached during the configuration of the project within Visual Studio while others, such as the connection with WBB, beyond of configuring the environment it was necessary to implement or adapt small routines aimed at having a better control over the Bluetooth connection between the system and WBB device.

The C# application followed roughly the Model-View-ViewModel architectural pattern using the Windows Presentation Foundation graphical system. The application has four main modules that are accessed through tab items (Configuration, Participants, Tests, and Review) on the main screen as shown in Figure 9**Error! Reference source not found.**.

In the configuration's tab, the Bluetooth connection with WBB is established and some general parameters, such as delay time for starting test, test duration, calibration time and other parameters for real-time evaluation of the balance tests.

Within the Participants tab, it is possible to list, add, edit, delete and select a participant of the executing test. Inside the Tests tab, the first screen presents instructions for the trials and the experimental group identification is set. The images of the task positions can be also shown to participants if they have any doubts about them. The following screen is aimed at test execution. Double stance, single stance and tandem tests are organized in sequence. During the tests a stick body image with the body segments constructed from Kinect 3D joint information and a white point representing the movement of the estimated center of mass is

showed on screen, as well as the estimated center of pressure location related to WBB dimensions is shown in another image frame (Figure 10). Captured values of load cells and estimated body´ total weight are also shown on screen.

Dynamic time warping algorithm was used to identify falls or loss of stability and to count errors during the test. This was only used for providing real-time feedback, not used in our reliability analysis.
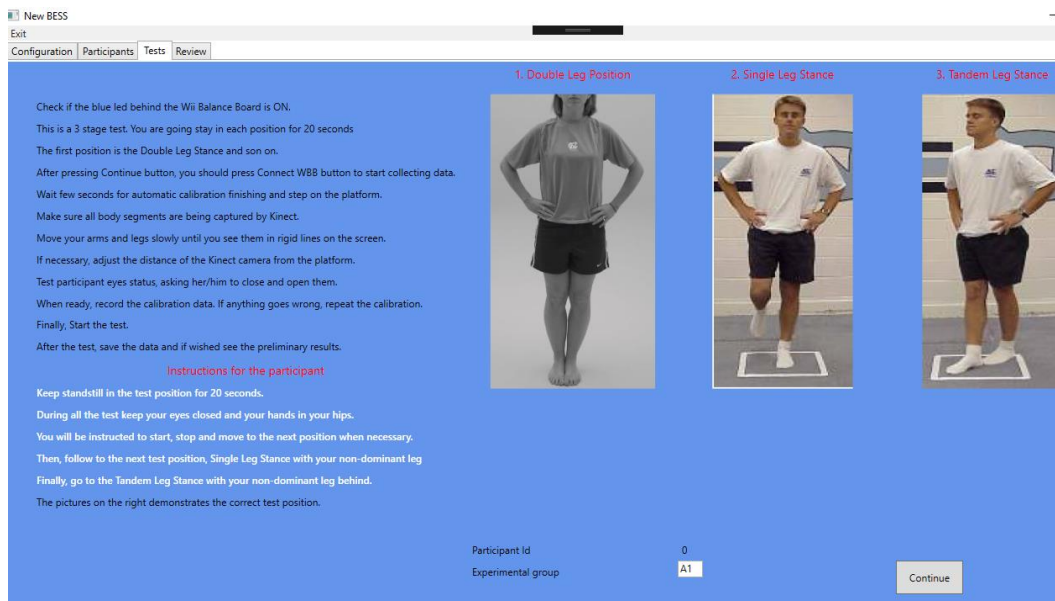
Figure 9 Customized C# application. Test tab.

Kinect joint positions were obtained at 30Hz and data of the four load cells from WBB was acquired using the same rate. Center of mass was estimated using a kinematic method based on body segments. All raw information was logged into text files. Test can be reviewed in the Review tab that showed the stick figure and some variables values from CoP and CoG position. It was also possible to perform some basic additional analysis using generated statocinesigram and stabilogram charts of the selected trial.

Therefore, the C# application was implemented with the following functionalities:

1. Activate the Bluetooth connection with WBB.
2. Allow parametrization of test procedures such as delays in recording data, task duration, and amount of logged information.
3. Register and store participants' profile.
4. Organize and standardize application of the tests.
5. Register test condition (test or retest).

6. Control sampling rates from devices.

7. Capture load cells signals from WBB and joint positions from Kinect.

8. Estimate Center of Gravity using a kinematic method from body segments frame by frame.

9. Record acquired information from WBB and Kinect into log files.

10. Allow verifying success or fail of data collection using logged information.
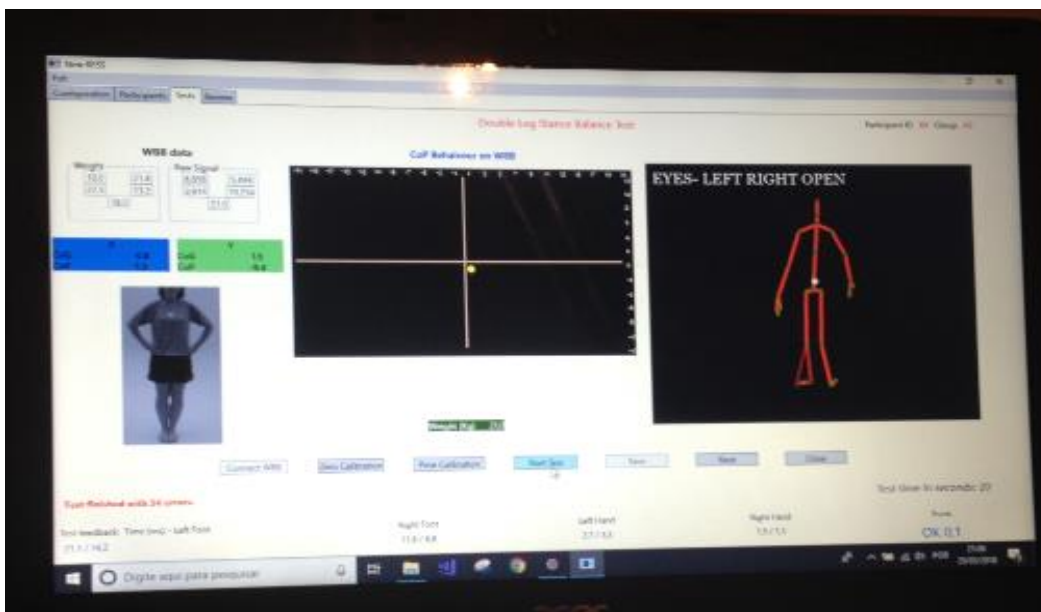
Figure 10 Screen shot of the Test tab during a test. In the middle, the position of the center of pressure represented by a yellow point. On the right, the stick avatar and the center of gravity represented by the white point.

For applying a balance test, the standard procedure was:

1. Once the application was started, double clicking on its executable file, it was necessary to establish the Bluetooth connection with WBB at Configuration tab.

2. Register the participant into the system in the Participants tab if he/she was not already registered.

3. Select the participant in the Participants tab who is going to be evaluated.

4. Input trial identification into the first screen of the Test tab.

5. Start balance test tasks clicking continue buttom after inputing trial identification what will open the second screen of Test tab (Figure 10).

   Participant and trial identifications are shown in top right of the screen. Besides that, raw signal information and estimated weights in kilograms on each WBB load cell are also shown on the top left of the screen.

Additionally, estimated coordinates of center of pressure and center of gravity are shown just below load cells information. On the bottom of the screen, it is shown the status of the position of some body segments captured by Kinect which were compared to correspondent joint positions captured during preparation of the task (step C described below). The status of Ok/NotOk was reached using an experimental dynamic time warping algorithm that calculated differences between current position and "calibrated position". This information was not used in this study.

a. Before the participant stepping on WBB and staying just behind the platform, unloaded WBB signal was captured for about 5s. Meanwhile, Kinect should have already identified body joints and a stick avatar appeared into the screen. Tracked body segments were shown in solid lines and inferred ones in thinner lines.

b. The participant steps on WBB and he/she may rehearsal the task position if necessary and when ready, preparation started, clicking on a buttom in screen.

c. During the preparation the participant stays in test position with eyes open as static as possible for three seconds.

d. After finish preparation, it was said to participant to close their eyes and then, clicking on start test buttom, test starts and a stopwatch is triggered showing the test time on screen. When finished, if everything was ok, it is possible to follow to the next task, doing that, a log file is created on filesystem with all collected information of this test task. Otherwise, it is possible to repeat the task again. Only double stance task (first task) has to read unloaded WBB signals.

e. When finish the last task (tandem stance). It is possible to go back to first test screen and input another trial identification and perform all tasks again, or go to Participants tab and select another participant.

Internally, it was used collection data structures provided by .NET framework to store log information such as List<T> class. There was one log list for each phase of test task and device. For example, one list for WBB zero loading, WBB preparation, Kinect preparation and so on. They were kept in

memory until saving this information into log files. Standard System.IO.File methods were used for that while System.Xml and System.Xml.Linq classes were used to manipulate XML structures and files. For controlling stopwatches, it was used System.Timers.Timer class and customized flags to set start and end times of controlled routines.

These routines were combined with WBB and Kinect libraries that basically performed the following tasks:

- Instantiate a sensor object (WBB or Kinect).
- Open a connection to receive sensor information, configuring parameters of filters, and ruling this link behavior using the sensor object.
- Register the callback method which will receive updates from sensor link.
- Close or reset sensor connection when finish.

Application logic was implemented mostly inside the callback methods that received frame by frame information from devices.

4.2 Application environment

It was generated an instance of the application for testing individuals and another instance for testing manikins. Each instance utilizes its own directory structure composed with a main directory where the application runs and a data directory to store log files.

The application stores two types of data, participants or manikins profiles and log information. Profiles were stored into XML format, one for each application. Basically, each participant has an XML entry with his/hers identification, name and birth date, height and weight were obtained from logged information considering joint distances of head and foot, or load cell values, respectively from Kinect and WBB. Regarding log files, it was generated one log file for each leg task test. The log's filename stores the identification of the leg position (DS, SS, or TS), the participant name, the trial identification (test or retest with a symbol), and the full date-time up to seconds (YYYY-MM-DD-HH-MM-SS).

Within each log file, there are five subsections. One with the records of the unloaded platform, one for preparation time with load cells data, another one with body joints 3D positions, and two other sections with load cells data and body joints for test time.

This system was developed in C#, using a 64-bit .NET framework v. 4.0.30319, within the Visual Studio 2017 version 4.7.02.046 development IDE. For accessing Kinect device information, the Microsoft Kinect SDK 2.0 14.10.19000 was installed. In addition, WiiMoteLib version 1.7.0.0 was used for interacting with WBB.**Error! Reference source not found.**

The software run under Windows 10 Pro 64-bit installed into an Aspire F5-573G notebook, with an Intel i5-7200U CPU @ 2.50-2.70 GHz, 8 GB of RAM, and a GeForce 940MX, NVIDIA 2GB GDDR5 video board. Kinect was connect through a standard USB 3.0 and WBB via a Bluetooth link connection.

4.3 Scripts for data preparation and analysis

The C# application was responsible for collecting the balance test information in a standardized way generating log files with signals raw information. It was necessary to extract from the log files the necessary information for data analysis and equalize units, normalize data, reduce signal noise and so on. This section gives an overview of the scripts objectives.

The first step was extract data from log files and transform them into 2D or 3D time series with center of pressure and center of gravity coordinates, respectively. Data from WBB was collected using millimeters and from Kinect meters, all data were converted to millimeters and time in milliseconds. Additionally, mean values were calculated and subtracted from data series, a process known as "detrend", which normalize the series capturing only the oscillation of values over the mean value.

Remembering that procedures were applied on manikins and individuals log files separately. Some routines were duplicated to avoid generating mixed information.

Secondly, it was performed the filter analysis over normalized data series. Besides that, a new set of files with filtered data series were generated in different directories, one directory per chosen cut-off filter frequency.

Thirdly, using filtered data series of CoP and CoG coordinates with chosen filter cut-off frequencies, temporary Bland-Altman charts were generate to observe first and second trials differences and estimate an ideal threshold for outliers. Besides that, descriptive statistical tests routines over filtered coordinates of CoP and CoG generated tables for comparing effects of using different outliers' thresholds on mean values and distributions. These routines also checked normality of data and equality of variances. Most of statistical functions were native to scientific Python packages, such as Numpy and Scipy.

In parallel, another set of routines were implemented to derive properties from filtered data series generating a new file consolidating all properties for CoP and another one for CoG. It was generated two consolidated files per each cut-off filter frequency.

Next, a new set of routines were implemented to perform inferential statistical tests dealing with transformations on non-parametric data series only for individuals' data series. For manikins' data series, transformations were not implemented.

For Lin´s Concordance Correlation Coefficient and Intraclass Correlation Coefficient tests data series were imported into R environment and tests were performed using R scripts. It was necessary to implement a routine to read properties consolidated file and create an exported file per each property, which had the results of each pair test-retest condition per participant, one for CoP property and the correspondent one for CoG. Moreover, based on coordinates data series or consolidated properties files, a series of scripts were prepared to generate graphs, some requiring special data handling, such as color map histograms and others special treatment for designing outputs..

Scripts were developed in Python version 3.6.1, 64-bit platform. As mentioned before, Intraclass Correlation tests and Lin´s Correlation Coefficient tests were performed in R version 3.5. Spyder development environment version 3.1.4 was used for writing, debugging and running Python scripts, while R Studio was used for writing and running R scripts. A more detailed specification of the implemented solution can be found in the Appendix I on page 146.

5
Methodology

This study was composed of two main parts, one aimed at testing intrinsic reliability of MS Kinect and WBB, and the other one, aimed at verifying reliability of those devices under different site locations with apparently healthy participants, simulating a real application of balance tests that takes advantage of devices portability and setup easiness. For the former objective, it was used an improvised static manikin and for the latter a cohort of 70 healthy individuals.

5.1 Participants

Seventy-two subjects gave written consent for participating in the experiment and allowed the use of collected data. Among the subjects, there were Physical Education and Computer Science students, doctors, and hospital employees, liberal professionals, rowing professional athletes, schoolteachers, and young high school students. Subjects were asked if they had any balance impairment and if they were able to stay standstill for 20s in one leg. Data from one subject with multiple sclerosis and one with Down syndrome were removed from the analysis, remaining 70 apparently healthy subjects (Males, 41 and Females, 29). Mean age (SD) 28.8 (8.7) y-o, median age 28 y-o, minimum/maximum age 16/52 y-o. Mean weight (SD) 67.6 (14.7) Kg and mean height 166 (0.11) cm with no reported balance impairment.

5.2 Manikins

A static structure simulating a person was put on WBB and in front of Kinect in order to capture concurrently CoP and CoG coordinates as shown in Figure 11. It was important that Kinect recognized the structure as a human body for calculating total body center of mass using the same algorithm used with people.

The idea of using a manikin was aimed at isolating the implicit variance of the devices and the associated software solution from that one of intra and inter-subjects' variances.

Figure 11 Improvised manikin.

An arbitrary choice of manikin weights was chosen (Table 3), which was limited by the dimensions of the platform and the improvised manikin structure. The weight load was not distributed in any specific place along the platform. However, it was aimed at setting up the center of pressure around the center of the board. Manikin weight mean (SD): 44.0 (21.5) Kg

Table 3. Manikin sample description and histogram.

Manikin Weight range		# Samples	
>= 10	< 20 Kg	30	
>= 20	< 30 Kg	54	
>= 30	< 40 Kg	9	
>= 40	< 50 Kg	24	
>= 50	< 60 Kg	30	
>= 60	< 70 Kg	33	
>= 70	< 80 Kg	30	
Total		210	
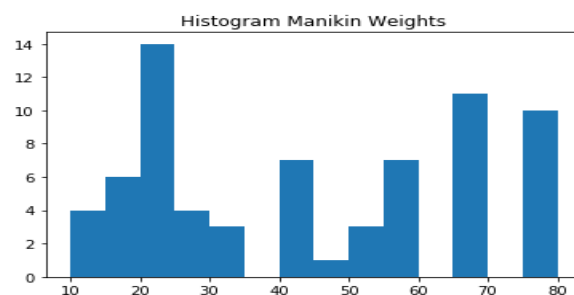
5.3   Test Procedures

Tests were performed in different sites: university classrooms, hospital rooms, schools and at home with the objective of simulating the portability and the simple setup characteristics of the solution.

The WBB was positioned in front of the Kinect in a distance approximately of 2-3m from each other. Kinect was about 1m above the ground and it was connected via an USB 3.0 cable to a notebook next to it as shown in Figure 12.

Figure 12 Position of the devices to perform the balance tests.

Volunteers performed two tests in the same session. The test was adapted from Balance Error Scoring System test that was composed of a sequence of three different feet positions: double (DS), single (SS) and tandem (TS) as shown in Figure 13. Before starting the test in each position, the subject was recorded in the required posture for three seconds with open eyes. Then, it was asked to close the eyes and stay as standstill as possible, so the test started.

In double stance, the feet were apart in a comfortable position, it was not asked for participants to put feet together. For single leg task, it was asked to place the non-dominant foot in the middle of the platform and raise up the other foot from

the ground, with no special instruction about the position of the free foot in the air. For tandem position, subjects were asked to put their feet on the diagonal of the platform and to turn their trunk to face Kinect and avoid the occlusion of the arm, what was not always possible.

The individual should stay in each position for 20 seconds with closed eyes and keep their hands on their hips around iliac crest. In our routine, during instructions presentation figures of the test position were showed to the subjects.

Figure 13 Test tasks. Double, single and tandem positions.

For single and tandem leg positions, before performing the test, it was required to the participant to try closing the eyes and get used to it. For double stance, it was optional. Most of the subjects in few tries were ready to continue the test. If the subject loses his /her stability, it was instructed to return to test position as soon as the balance was recovered.

After the finish of the first trial, it was asked to the participant to unload the platform and relax for a few seconds and invited to perform a second time when ready. Meanwhile, the signal from the unloaded platform was captured for about 5 seconds and a new trial identification was entered manually into the system. Then, when the participant was ready, the test was performed a second time. It generally occurred in less than 1 minute.

The participant was instructed to try to execute all positions in the same way as he/she had done during the first trial. Table 4 shows the general procedures applied during tests with individuals.

Table 4. General description of test procedures with individuals.

#	Procedure	Time	Eyes condition	Who
0	General instructions			Observer
1	Set trial condition (test or retest)			Observer

2	Instructions for Double Stance task			Observer
3	Rehearsal (only for test condition)	Free	Closed	Participant
4	Rest (only for test condition)	Free	Open	Participant
5	Step down WBB	5s or more		Participant
6	Step on and prepare for double leg position	3s	Open	Participant
7	Start command			Observer
8	Test	20s	Closed	Participant
9	Instructions for Single Stance task			Observer
10	Rehearsal – mandatory	Free	Closed	Participant
11	Rest	Free	Open	Participant
12	Preparation on single leg position	3s	Open	Participant
13	Start command			Observer
14	Test	20s	Closed	Participant
15	Instructions for Tandem stance task			Observer
16	Rehearsal – mandatory	Free	Closed	Participant
17	Rest	Free	Open	Participant
18	Preparation on tandem position	3s	Open	Participant
19	Start command			Observer
20	Test	20s	Closed	Participant
21	Step down WBB	Free	Open	Participant
22	Save test			Observer
23	Repeat from #1 for retest condition	When ready		-

5.3.1. Manikin test procedures

For each sample, two set of tests were performed following the same routine as the balance test done with individuals, except from having no pause between the preparation (3s) and the start of the test (20s). Test started automatically after preparation, there were no steps 7, 13, and 19 showed in Table 4. Additionally, the second trial was performed just after the first trial ended without unloading the board or repositioning the manikin, and, of course, there was no instructions, rehearsal or rest time for manikins. However, it was necessary to insert the new trial identification manually into the system before starting the retest condition.

There was no different stance positions with the manikins, because they were supposed to be static and our objective was to verify instruments' reliability during a base case. Therefore, for each trial, it was collected three samples of the same manikin position. Figure 14 illustrates one test with a manikin.
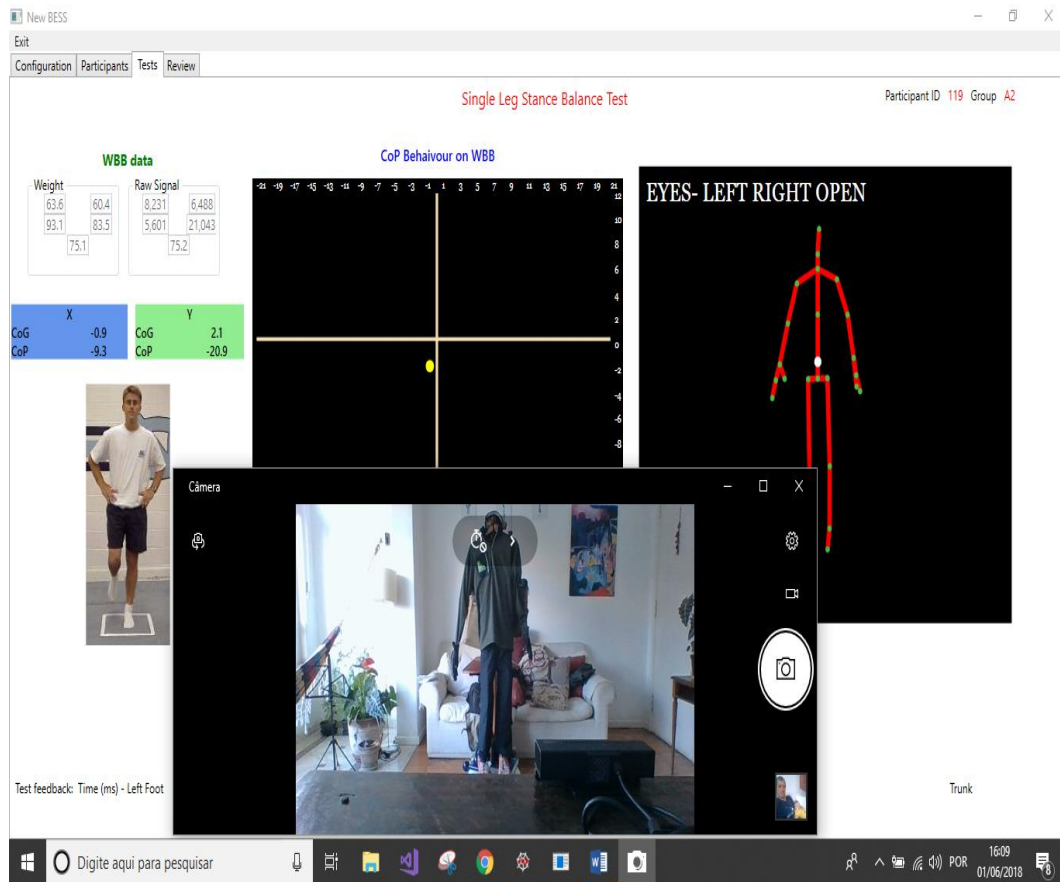
Figure 14 Image of the Test tab screen during a real test.

5.3.2. Center of pressure calculation

The system was configured to acquire 30 frames per seconds from WBB. WBB transmits in each frame the information about the load on each of its load cells. Top Right (TR), Bottom Right (BR), Top Left (TL) and Bottom Left (BL) variables received the estimated weight in kilograms sent by WBB (Figure 15). Some empirical tests were done using raw signal data and internal calibration parameters, but it seemed more inconsistent when compared to manufactory provided information in kilograms. Then, it was used the manufactory estimated weight in kilograms of each sensor. For calculating CoP position, it was used the following formulas (Leach *at al.*, 2014):

$$CoPx = \frac{X}{2} * \frac{(TR + BR) - (TL + BL)}{TR + BR + TL + BL}$$

$$CoPy = \frac{Y}{2} * \frac{(TR + TL) - (BR + BL)}{TR + BR + TL + BL}$$

CoPx represents the M-L displacement of CoP, while CoPy represents the A-P displacement. X and Y are the M-L and A-P distances (433, 238, respectively) in mm of load cells on the board, respectively.

No custom calibration was implemented and information provided directly from WBB was used. Bartlett et al. (2014) found that internal calibration was comparable to customized calibration data. Custom calibration described in Clark et al. (2010), had minimal effect on the noise inherent in the WBB. Furthermore, following Leach *et al.* (2014), custom calibration would be expensive, time intensive, and neither feasible nor affordable for most of the users.
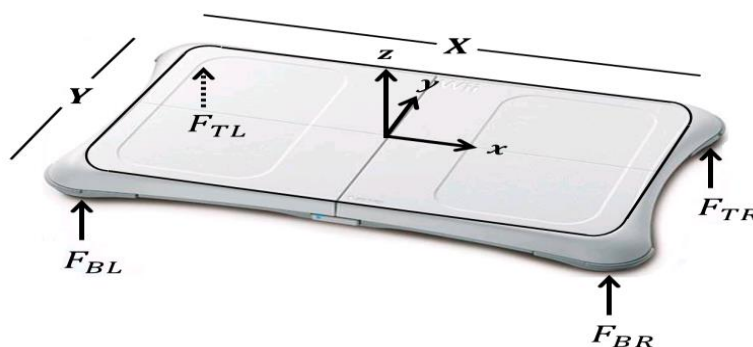
Figure 15 Wii Balance Board axes.[7]

During all tests, it was used the same WBB device. Bartlett *et al.* found that using the same device reduces the mean CoP error during static sway conditions. Additionally, Leach *et al.* (2014) found no error differences among 12 different WBBs when testing controlled dynamic sway conditions.

5.3.3. Center of gravity calculation

The system received Kinect data at a rate of 30 frames per second. The device provided twenty-five 3D joint positions of one participant in each frame. Using this information, body segments lengths were calculated using the Euclidian distance of neighbor joints. The body model was segmented in: head, trunk, two upper arms, two forearms, two hands, two thighs, two shanks, and two feet. Distance from rotation axis of each segment was found using anthropometrical data from University of Exeter Kinesiology and Biomechanics Laboratory Manual (2013-2014) based on the tables provided by Clauser *et al.* (1969), which estimated

---

[7] Image source: Leach *et al.*, 2014.

body segments center of gravity. The weight proportions were based on the study of Clauser *et al*. (1969) for calculating segments weight. Gravity acceleration was considered constant at 9.81 m/s². Total body mass was estimated by WBB.

For calculating Center of Gravity it was used the following formula:

$$M_{tb} = \sum_{i=1}^{n} Ms_i$$

M_{tb} = Total body moment; n = number of body segments;

Ms_i = Moment of each body segment.

$$Ms_i = F * d$$

Ms_i = moment, F = force (mass * gravity acceleration), d = distance from the rotation axis.

This formula is based on the principle that the resultant moment of a body is equal to the sum of its segments moments (Hamil, Knutzen, & Derrick, 2015).

5.3.4. Formulas for generating derived variables

For analyzing body sway, important properties were calculated from the time series data points of the displacement on A-P and M-L directions of CoP and CoG. Time-domain properties represent distances, speed and area of resultant, M-L, and A-P directions. Table 5 shows the formulas or the pseudo-code based on Matlab or Python syntax used to calculate these properties as specified in Prieto *et al.* (1996), Duarte and Freitas (2010), and Schubert *et al*. (2012).

Table 5. Formulas used for calculating derived properties of CoP and CoG coordinates data series. Medial-lateral plane data points is represented by x, and y the anterior-posterior plane data points. Resultant is the bivariate location of CoP or CoG.

Property	Mathematical description or pseudo-code
Total Path Length	$\sum_{i=1}^{N} \sqrt{(x_i - x_{i-1})^2} + (y_i - y_{i-1})^2$
Range resultant	Maximum (SQRT(y^2 + x^2)) – Minimum (SQRT(y^2 + x^2))
Range A-P / M-L	Maximum(y) – Minimum(y); Maximum(x) – Minimum(x)
Mean speed A-P	sum(abs(diff(y))) * frequency / length(y)
Mean speed M-L	sum(abs(diff(x))) * frequency / length(x)

Mean total speed	TPL / time
Distance root mean square (RMS)	$$\sqrt{\dfrac{\sum_{i=1}^{N}[(x_i - \bar{x})^2 + (y_i - \bar{y})^2]}{N}}$$
Distance RMS M-L	$$\sqrt{\dfrac{\sum_{i=1}^{N}[(x_i)^2]}{N}}$$
Distance RMS A-P	$$\sqrt{\dfrac{\sum_{i=1}^{N}[(y_i)^2]}{N}}$$
95% of estimated circumference area	R = mean(SQRT(y^2 + x^2)); pi*(mean(R) + 1.96*std(R))^2;
95% of estimated ellipse area	[vec, val] = eig(cov(y, x)); Pi * prod(2.4478 * sqrt(svd(val)));

5.4 Data Analysis

It was simulated the use of 70 manikins. Each one performing two trials (test and retest) of one test composed of three leg tasks, generating a total of 420 unique data series, 210 (70x3) for the first trial and 210 for the second trial. Remembering that with manikins, actually, it was not changed the base of support (Figure 11). Therefore, the three leg tasks were equal tasks.

With individuals, the 70 participants performed two trials (test and retest) as well. One test composed of three different leg positions (double, single and tandem stance), totalizing 420 unique data series.

All data obtained from WBB and Kinect were logged into text files. One log file per test with the identification of the leg position, participant and trial (test or retest). Log files contained one record per line, and each line with fields delimited by ':' (colon character). Different records were created with raw and processed information, such as read values of each load cell or CoG 3D and CoP 2D coordinates time series, for example.

Center of gravity calculation was performed during data acquisition by the C# application using Kinect joint references provided by Kinect. Each leg position test collected more than 600 instant joint information and logged each of them into records containing time in milliseconds, and x, y, z coordinates in meters, for

horizontal, vertical and distance from the center of camera displacements, respectively. A Python script converted coordinates to millimeters and separated interested data series from other log information, generating one new file per test.

Raw load information in kilograms from each WBB sensor was obtained by the C# application at 30 Hz rate. For each leg-position more than 600 frames were collected and logged into a record containing time in milliseconds, top-right, top-left, bottom-right, and bottom-left load information in kilograms. A Python script imported logged information and calculated CoP medial-lateral and anterior-posterior position for each frame. As recommended by Duarte and Freitas (2010), mean values of M-L and A-P locations were subtracted from data series before any analysis.

A routine for finding the best filter cut-off frequency based on David Winter (2009) analysis of residuals between filtered and unfiltered signals was performed. This method was applied to every test with manikins and individuals. The chosen cut-off frequency was in the range of the median of best found cut-off frequency value up to the maximum value. Then, a low-pass 2nd order Butterworth filter was applied into raw files generating filtered data series.

Additionally, tests were performed with a range of different cut-off frequencies from 4 Hz up to 12 Hz and agreement tests over Manikin data were conducted in order to observe the behavior of different cut-off frequencies on concordance scores.

Bland-Altman graphs were also generated in order to find outliers. It was defined that properties with values above or below mean +/- three standard deviations for individuals and +/- five standard deviations for manikins would be excluded from analysis.

With the chosen cut-off frequency, 2Hz for CoG and 4Hz for CoP, calculations over CoG and CoP on M-L and A-P data series were executed in order to derive posturographic properties based on the formulas presented in Table 5.

The analysis of data was performed firstly checking color-map histograms of all data points of CoP and CoG in A-P and M-L coordinates for all tests with manikins, and for each leg position for individuals. Color map histograms show the distribution of all data points on A-L and M-L directions. The histograms were generated after subtracting the mean value from data series and applying the low-pass filter. The color scale represents the occurrence of determinate value at that

position. The objective was to observe the general similarities and differences between test-retest trials and between CoP and CoG measures.

Secondly, basic descriptive information of CoP and CoG coordinates with mean and 95% confidence interval, standard deviation and minimum, maximum, 25th, 50th, 75th percentile values for test and retest trials were provided. Shapiro-Wilk test checked normality and Levene´s test check equality of variances between trials. When necessary, Boxplot and Histograms were provided for checking visually normality and series distribution.

If normality was violated in individuals' derived properties, data was transformed using two strategies, $\frac{1}{\sqrt{Variable}}$ and if it still was violated, logarithmic transformation was applied. If, normality was not met, even after transformations, it is highlighted in the text. Descriptive statistics shows back-transformed data in original scale. Transformations were not applied in Manikins data series, consequently, non-parametric statistical functions were used.

In this case, a customized Chi-squared test was implemented as follows. The distribution of expected values used in Chi-squared test was based on the amount of test cases in each 20th percentile of the first trial, generating five ranges. Consequently, N was divided by 5 and each range had 1/5 N. Then, the value found at each percentile limit was used to seek in retest trial the correspondent percentile equal or immediately greater than that value. After, the number of test cases within the limits found on the retest trial was set to that bin. Finally, the bins series were submitted to Chi-squared evaluation. Therefore, it was possible to measure the agreement between the distributions of the two sets. Considering that in Manikin N=210 assures that each range will have a reasonable number of cases for the comparisons. Furthermore, ICC and CCC tests were performed.

Additionally, similarity tests, paired T-Test or Wilcoxon signed rank tests were executed to verify bias between trials. Pearson´s correlation or Spearman-rank correlation were also executed on mean of each data point coordinate in order to check associations between trials. For all statistical tests, a significant value of $p < .05$ was used.

Only for individuals tests, a graph analysis of mean values of each data point were presented for each leg position showing the differences between test and retest trials. Descriptive analysis with basic assumptions and inferential tests were applied

over all derived properties. Beyond that, Shapiro-Wilk test was applied on test-retest differences. An analysis of difference sizes and amount of cases that were within a specified difference was performed. For example, if the difference between the test and retest was 20% of the test value, within this limit, it was checked the percent of pair tests that differences were below this threshold.

For individuals, a correlation graph with the best-fit linear model and the perfect concordance line demonstrated the bias between trials for each derived property and visually demonstrated most and least reliable properties.

It was also presented Lin´s Concordance Correlation Coefficients (CCC) for repeated measures and Intraclass Correlation Coefficients with 95% confidence intervals. Following McGraw and Wong (1996) classification the ICC model used in our work was based on two mixed effects, absolute agreement, single measurement (ICC $_{2, 1}$) with the formula explained previously on page 34.

The references for interpreting ICC results' classification were: excellent >= .75; fair .40 =< ICC <= .74; and poor < .40, similar to other reliability studies using WBB (Clark *et al.*, 2010; Bower *et* al., 2014; Severini *et al.* 2017). For, CCC, scores greater than .69 was interpreted in the satisfactory range as used in Larsen *et al.* (2014).

A graph analysis with Bland-Altman plots, limits of agreement (LOA) and other analysis were presented for individuals. LOA was originally designed for evaluating the agreement of two different methods on the same variable. However, it can be used for repeated measures adapting the limits to consider within-subject variability. As explained in Bland and Altman (1999), the repeatability coefficient represents that values of 95% of the subjects will be within $1.96 * \sqrt{2} * S_w$, where S_w is the square root of residuals mean square. Then, the plotted Bland-Altman graphs in this study present mean differences and 95% confidence intervals, within-subject variability, and 95% limits of agreement based on repeatability coefficient, and the correlation coefficients between mean values and differences. Differences were calculated subtracting the values of the retest from test (difference = test − retest). Therefore, a positive difference means a decreased value in the second trial.

Confidence intervals were calculated using standard error of mean (SE) formula:

$$SE = \frac{SD}{\sqrt{N}}$$

SD = Standard deviation of differences, N = sample size.

For mean confidence interval, mean +/- (t * SE), where t, is the value found in a t-distribution's table for degrees of freedom (N-1) and p = .05, approximately it is 1.6669.

Absolute reliability was checked through, mean differences, standard error of measurement (SEM) and minimal detectable change (MDC). Relative SEM (SEM%) and relative MDC (MDC%) were analyzed as a percentage of mean of the mean test and mean retest trials. Standard error of measurement was calculated following the formula:

$$SEM = SD * \sqrt{1-r}$$

SD = Within-subject standard deviation. r = ICC scores in our study on transformed data.

Minimal detectable change was calculated based on the formula (Cook *et al.*, 2014):

$$MDC = 1.96 * \sqrt{2} * SEM$$

MDC values were considered excellent when <10%, acceptable between 10% and 30% and poor when >30% as used by Llorens *et al.* (2015).

6
Results

This chapter presents all results of our experiments following the procedures described previously.

6.1 Filter Analysis

It was not possible to identify the best cut-off frequency for the 2nd order Butterworth filter with Manikins CoP data series. Figure 16 demonstrates the common behavior of the residuals root mean square error (RMSE) with different cut-off frequencies.
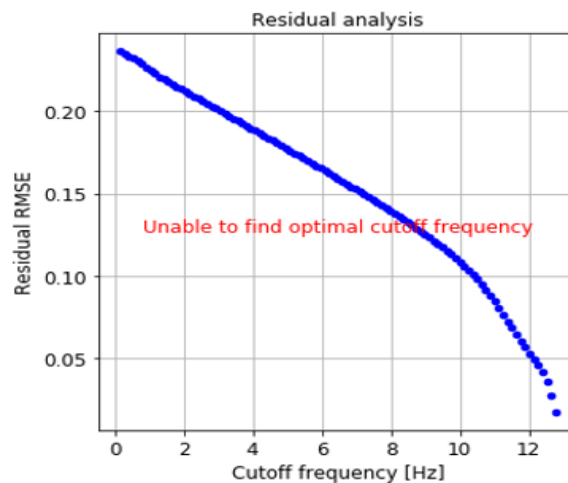
Figure 16 Manikin Center of Pressure filter analysis example. Residuals root mean square error differences were minimal across different cut-off frequencies.[8]

On the other hand, it was found in Individuals CoP data on medial-lateral (M-L) direction that the median and maximum best cut-off frequency were 3.5Hz and 4.7Hz, respectively and for anterior-posterior (A-P) direction, 3.7Hz and 4.6Hz (Table 6). Then, we decided to set the cut-off frequency at 4 Hz for all Individuals CoP data series. Figure 17 shows an example of the graphs generated to analyze the best cut-off frequency and its effect on the unfiltered signal (on top right) and in the

---

[8] This figure was generated using and adapted function from (Biomechanics and Motor Control [BMC], n.d.).

acceleration of the signal (on bottom right). The figure shows that selected frequency had almost no effect on unfiltered signal, filtered signal covers up most of the trajectory of unfiltered signal.
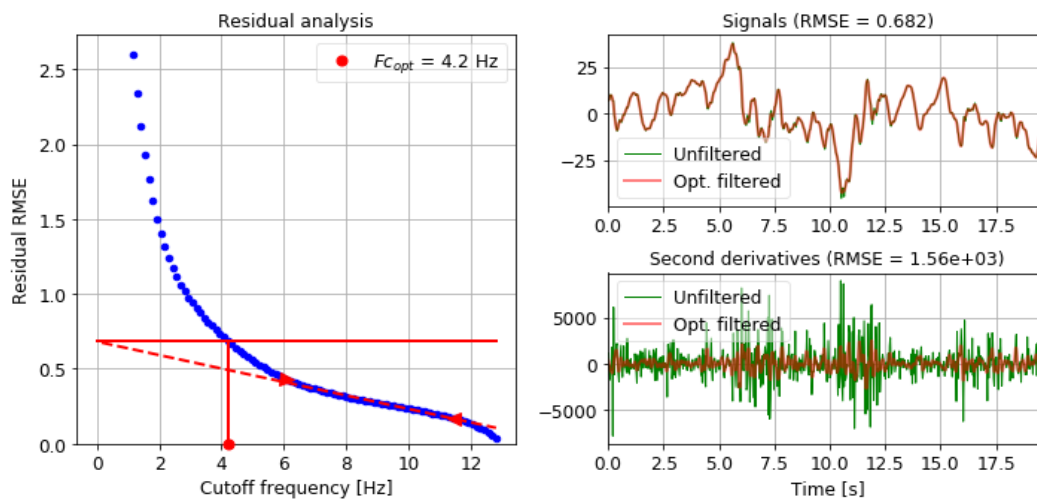
Figure 17 Filter analysis on CoP A-P data during a double-stance test. On the left, effects of applying different low-pass cut-off frequencies on root mean squared error (RMSE) between filtered and unfiltered signals. On the top right, the differences over test time with found RMSE for the found optimal cut-off frequency. There was almost no difference. On the bottom right, the effect of the optimal cut-off frequency on the frequency acceleration along test time.[9]

The same analysis with Manikin CoG data series suggested that cut-off frequencies should be set from about 1Hz up to 2.9Hz on M-L direction and up to 3.5Hz on A-P plane (Table 6).

Table 6. Analysis of best cut-off frequency on all CoG data series for Manikins and Individuals. Median and Maximum best found cut-frequencies are presented in Hz.

CoG coordinates	Manikin		Individuals	
	Anterior-Posterior	Medial-Lateral	Anterior-Posterior	Medial-Lateral
Median (Hz)	0.76	0.96	3.72	3.52
Maximum (Hz)	3.50	2.90	4.60	4.70

Based on Manikin and Individuals best cut-off frequency analysis (Table 6), we set the cut-off frequency at 2Hz for Kinect signal in both planes in order to do not lose signal information even getting more noise in some comparisons. Figure 18 and Figure 19 show examples of the applied filter on CoG signals for tests with manikins and individuals, respectively.

---

[9] This figure was generated using adapted functions from (Biomechanics and Motor Control [BMC], n.d.).
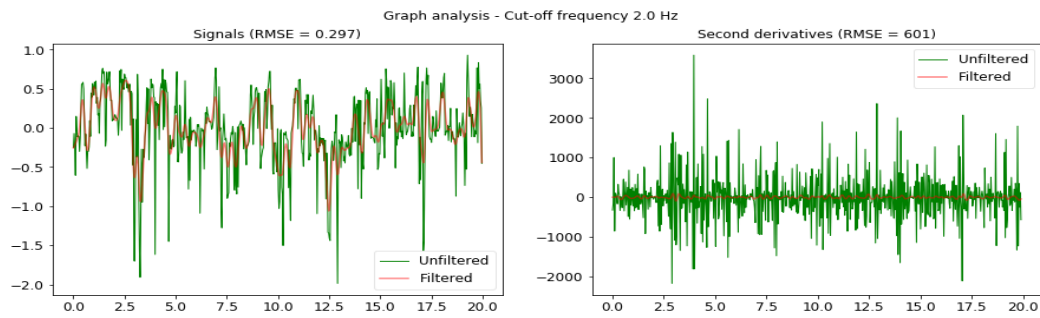
Figure 18 Manikin CoG anterior-posterior filtered and unfiltered signals during test (Left). On the right, the attenuation of filter on signal's second derivative.
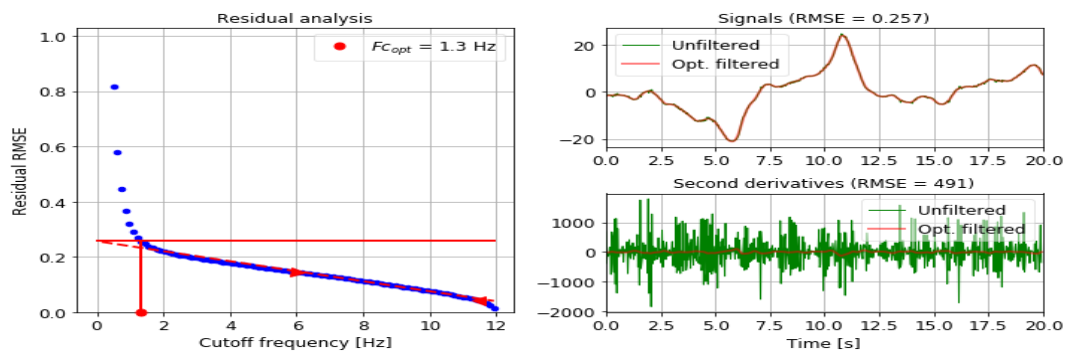
Figure 19 Residuals analysis for choosing best filter cut-off frequency on individuals CoG anterior-posterior signal.[10]

Different filter cut-off frequencies did not alter significantly the concordance indexes for all CoP variables in all tested positions with individuals and manikins. Table 7 demonstrates an example of the effects of varied cut-off frequencies on CCC scores for CoP properties of Individuals. The same lack of effect was seen in ICC scores (not shown).

Table 7. Minimal effects of different Butterworth 2nd order filter low-pass cut-off frequencies on Lin´s Concordance Correlation Coefficient over Center of Pressure analyzed variables for all test positions.

Individuals Center of Pressure Properties for each task		Lin´s Coefficients for different Filter cut-offs frequencies in Hz								
		4	5	6	7	8	9	10	11	12
Resultant										
Median Distance	DS	.57	.57	.57	.57	.57	.57	.57	.57	.57
	SS	.42	.41	.41	.41	.41	.41	.41	.41	.41
	TS	.30	.30	.30	.30	.30	.30	.30	.30	.30
Total Path Length	DS	.68	.68	.68	.68	.68	.68	.68	.68	.68
	SS	.41	.40	.39	.38	.36	.35	.34	.33	.32
	TS	.54	.55	.55	.55	.55	.55	.55	.55	.55
RMS Distance	DS	.57	.57	.57	.57	.57	.57	.57	.57	.57
	SS	.37	.37	.37	.37	.36	.36	.36	.36	.36

---

[10] This figure was generated using adapted functions from (Biomechanics and Motor Control [BMC], n.d.).

	TS	.16	.16	.16	.16	.16	.16	.16	.16	.16
	DS	.68	.68	.68	.68	.68	.68	.68	.68	.68
Average Speed	SS	.41	.40	.39	.38	.36	.35	.34	.33	.32
	TS	.54	.55	.55	.55	.55	.55	.55	.55	.55
	DS	.34	.34	.34	.34	.34	.34	.34	.34	.34
Range	SS	.19	.17	.16	.14	.12	.10	.09	.08	.07
	TS	-.08	-.08	-.08	-.09	-.09	-.09	-.09	-.09	-.09
	DS	.49	.49	.49	.49	.49	.49	.49	.49	.49
95% confidence Circle Area	SS	.32	.31	.31	.30	.30	.30	.29	.29	.28
	TS	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02	-.02
	DS	.66	.66	.66	.66	.66	.66	.66	.66	.66
95% confidence Ellipse Area	SS	.44	.43	.42	.42	.41	.40	.39	.39	.38
	TS	.15	.15	.15	.15	.15	.15	.15	.15	.15

For frequencies lower than 4Hz the choice of filter cut-off frequency may affect concordance indexes of Manikins and Individuals CoG. For example, in double stance task for Individuals CoG resultant Total Path Length, CCC score was .69 at 2 or 3Hz cut-off frequency, while at 6Hz it reached .83. Additionally, in exploratory tests in frequency domain, the median frequency CCC score for 2 or 3Hz cut-off frequency was .23, while at 6Hz frequency in DS it reached .60. Despite of that, the chosen frequency was set at 2Hz following the residuals analysis and compatibility with Manikins.

6.2 Center of mass and center of pressure coordinate analysis

Despite small inequalities between test-retest trial shown in Figure 20 and Figure 21, the similarities between trials in CoP and CoG Manikin series are clear. CoP data point distribution was more concentrated around the center than CoG data points, which were more dispersed in the A-P direction.
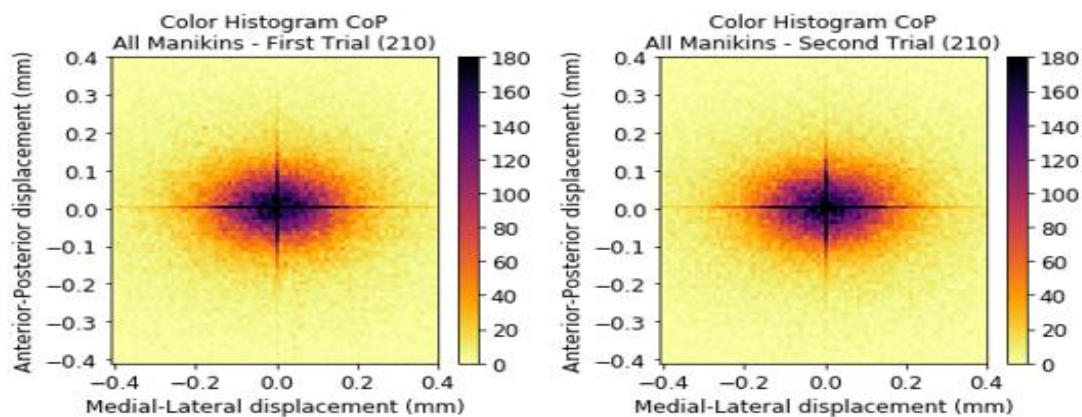
Figure 20 Manikin CoP color map histograms images.
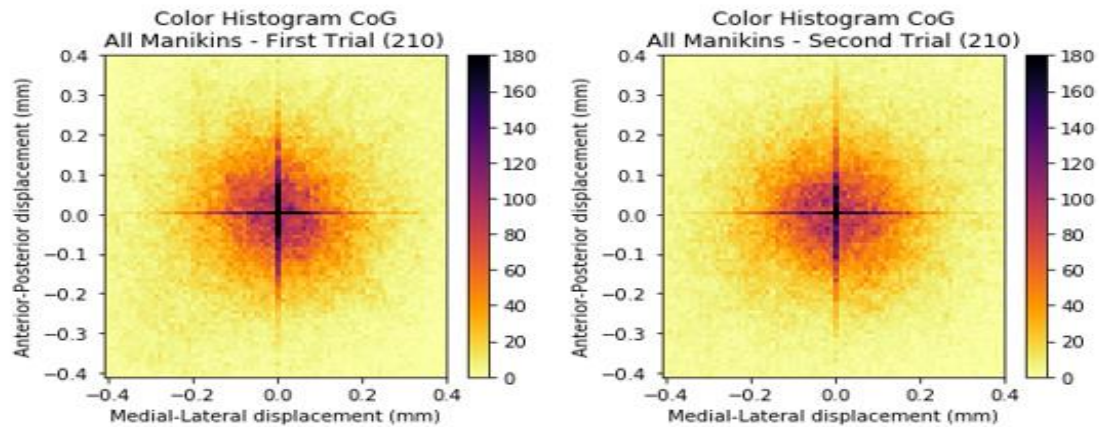
Figure 21 Manikin CoG colour map histograms images.

For Individuals, it was noticeable the different distribution patterns of the three tasks (Figure 22). In Double Stance (DS) task, the variance was larger in A-P plane compared to M-L. Additionally, the variance in Single Stance (SS) task was larger when compared to the other two leg stances. Individuals CoP data series were similar between first and second trials. Nonetheless, some small differences were apparent between test-retest trials like a more longitudinal concentration in DS data points in the test compared to the retest trial.
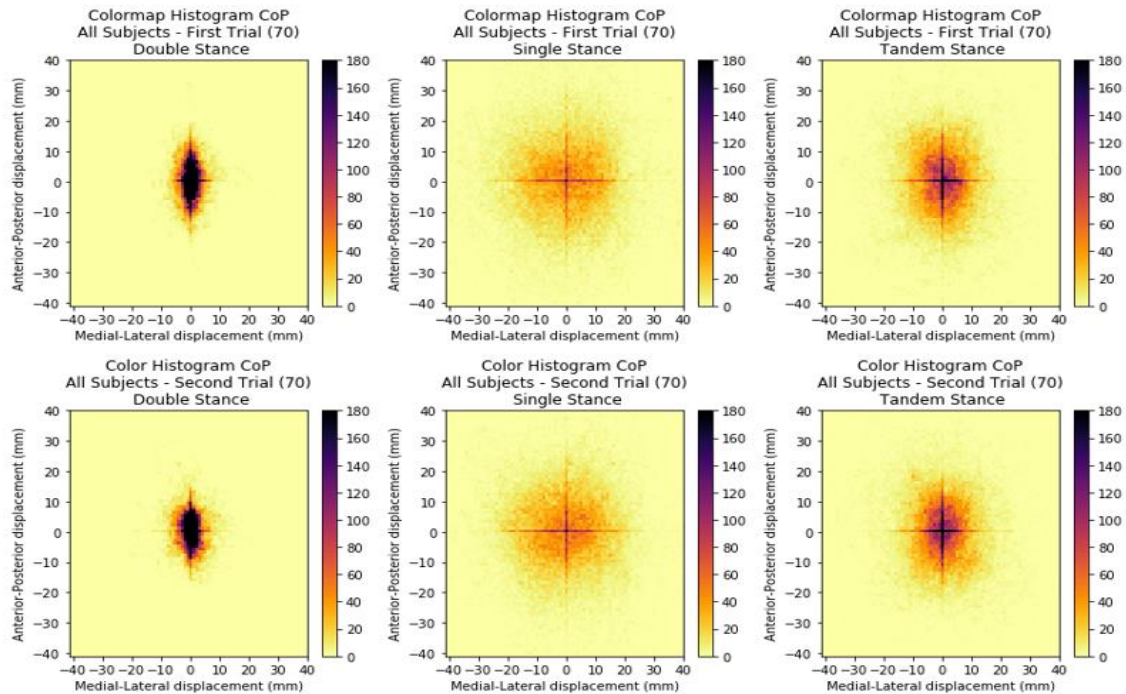
Figure 22 Individuals CoP color map histograms for all positions.

Individuals CoG color maps were very similar between test-retest trials (Figure 23). Distinction between Single Stance and Tandem (TS) was not so

apparent as compared to DS. For DS CoG image, during the retest it was noticeable a fatter shape in M-L and narrower extremities when compared to first trials. In SS, it was not clear the predominant axis of CoG oscillation. However, it was possible to note that the central black cross was longer in A-P plane during the second trials and there was a more even distribution of the points forming a circle around the center while the first series have a more spread distribution and tendency to a quadratic shape. In TS, it is important to realize that M-L and A-P planes are in relation to the displacements from the point of view of the camera, not the body.
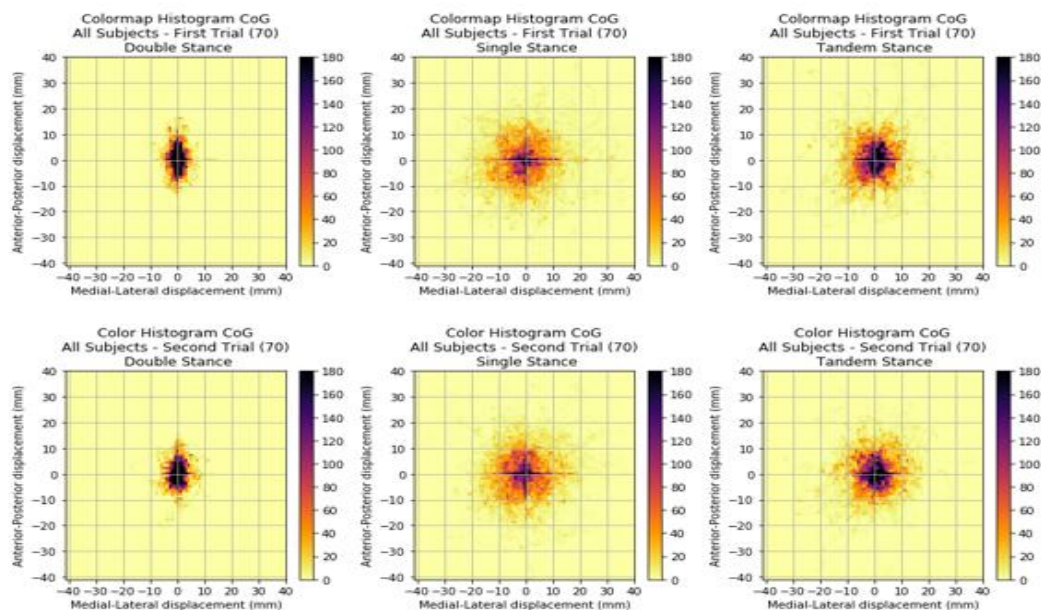
Figure 23 Color map histograms of CoG coordinates in A-P by M-L planes.

It means that A-P oscillations represent movements where the CoG got closer or further to the center of Kinect sensors and M-L the oscillations to the right or left side of the center of the camera. However, during TS task part of the body was occluded and not seen by Kinect. In Figure 2, it can be noticed that participant was in the diagonal of the platform and not parallel to the front plane causing the occlusion of the right arm. The feet were in the diagonal of the platform turned about 45 degrees to the left or right following the preference of the participant. Therefore, much of the A-P oscillation represents actually the M-L body oscillation. Considering that our objective was not to analyze the oscillation, but the reliability of the procedures and devices, we focused on the differences between test-retest. It was harder to characterize visually the two TS color map differences, but it seemed that retest area of point was more concentrated with a smaller central black area. In

relation to CoP images, it was clear the more concentrated distribution in CoG SS and TS trials. Differences were not clear between CoP and CoG DS trials.

6.3 General coordinate analysis

Descriptive information of the data demonstrated the similarity between the two Manikins CoP series of tests (Table 8). Shapiro-Wilk test were performed to check the normality of data series. Equality of variances (homoscedasticity) between trials was also verified through Bartlett´s or Levene´s test.

Table 8. Manikins CoP descriptive of (A-P) anterior-posterior and (M-L) medial-lateral coordinates. Values in (mm). (*) Barlett´s test.

CoP Series	Mean (SD)	95% CI		Percentiles					Shapiro Wilk		Levene´s test	
		Low	Up	Min	25th	50th	75th	Max	W	P	T	P
A-P trial-1	.00 (.01)	.00	.00	-.03	-.01	.00	.01	.03	.99	**.04**	.03	.86
A-P trial-2	-.00 (.01)	-.00	.00	-.03	-.01	-.00	.01	.03	1.0	.92		
M-L trial-1	-.00 (.02)	-.00	.00	-.06	-.01	.00	.01	.06	1.0	.58	.69	.41*
M-L trial-2	-.00 (.02)	-.00	.00	-.07	-.01	-.00	.01	.05	1.0	.07		

Although, Shapiro-Wilk test pointed out a significant difference from normal distribution of A-P data series of the first trial (p = .04; Table 8), analyzing the correspondent Box Plot and Histogram (Figure 24), normality was considered met for our purposes.
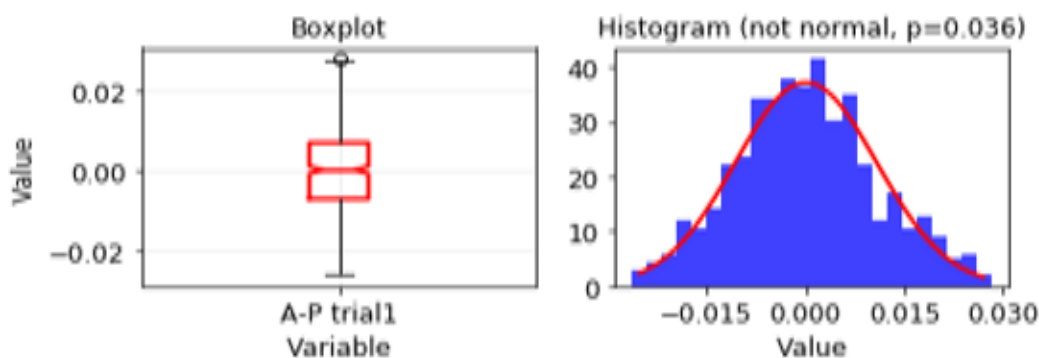
Figure 24 Normality chart analysis of anterior-posterior mean coordinates of first trail.

The general descriptive information regarding Manikin CoP (Table 8) and CoG (Table 9) series demonstrated that mean values of each registered coordinate in both planes were very close to each other. The largest registered amplitude was 1.4 mm in the M-L CoG of the second trial, and standard deviations were <=0.3 mm.

Table 9. Manikin CoG descriptive of anterior-posterior (A-P) and medial-lateral (M-L) coordinates series. Values in (mm).

CoG Series	Mean (SD)	95% C. I.		Percentiles					Shapiro Wilk		Levene´s test	
		Low	Up	Min	25th	50th	75th	Max	W	P	T	P
A-P trial-1	-.00 (.01)	-.00	.00	-.04	-.01	.00	.01	.05	.99	.00	194	.00
A-P trial-2	.00 (.03)	-.00	.00	-.07	-.02	.00	.02	.06	.98	.00		
M-L trial-1	.00 (.02)	.00	.00	-.05	-.01	-.00	.01	.07	.99	.00	64	.00
M-L trial-2	.00 (.03)	.00	.00	-.08	-.02	.00	.02	.06	.98	.00		

Table 10 shows that Individuals CoP means of mean data point values were very similar. Differences were noticed only on the second decimal digit. However, variation and range values were high compared to mean values. Moreover, absolute mean values in the first trial were greater than the ones in the second trial for DS and TS A-P, and SS M-L. They were equal only in the DS M-L, and smaller in all other leg position directions.

Variations was smaller in the retest conditions except for DS M-L direction. Normality was violated in most cases. Correspondent interquartile values (25th, 50th, and 75th) between trials were below .50 mm.

Table 10. Individuals CoP descriptive tables. Values in millimetres. CI=confidence interval.

Double Stance	Mean (SD)	95% CI		Percentiles					Shapiro Wilk		Levene´s test
		Low	Up.	Min	25th	50th	75th	Max	w	p	
A-P trial-1	.03 (.62)	.02	.03	-1.51	-.40	-.02	.48	1.86	.99	.00	w: 66
A-P trial-2	- .02 (.44)	-.01	- .01	-1.40	-.29	-.00	.26	1.36	.99	.30	p: .00
M-L trial-1	.00 (.28)	-.01	-.00	.80	- .23	.02	.21	.68	.99	.00	w: 25
M-L trial-2	.00 (.37)	-.00	.00	.89	- .26	.03	.25	.97	.99	.00	p: .00

Single Stance	Mean (SD)	95% CI		Percentiles					Shapiro Wilk		Barlett´s test
		Low	Up	Min	25th	50th	75th	Max	w	p	
A-P trial-1	.02 (1.89)	.00	.04	-5.49	-1.25	-.03	1.33	5.70	1.0	.48	w:20
A-P trial-2	- .04 (1.57)	-.06	- .03	-4.57	-1.19	-.16	.95	4.46	1.0	.07	p:.00
M-L trial-1	.09 (2.00)	.07	.11	-4.95	- 1.33	.09	1.30	6.32	1.0	.24	w: 42
M-L trial-2	.05 (1.53)	.03	.06	-4.41	- 1.04	.02	1.03	4.94	1.0	.54	p: .00

Tandem Stance	Mean (SD)	95% CI		Percentiles					Shapiro Wilk		Levene´s test
		Low	Up	Min	25th	50th	75th	Max	w	p	
A-P trial-1	- .04 (1.36)	- .06	- .03	-2.95	-1.04	- .15	.85	3.86	.98	.00	w: 5.0
A-P trial-2	- .01 (1.24)	-.02	.00	-5.02	- .84	.05	.85	3.07	.98	.00	p:.02
M-L trial-1	- .01 (1.17)	-.02	.00	-5.02	- .80	-.09	.65	5.57	.97	.00	w: 2.8
M-L trial-2	.02 (1.03)	.01	.03	-2.85	- .64	-.04	.74	2.99	1.0	.18	p: .09

None of the Individuals CoG series was normally distributed (Table 11). In general, variability was lower during the retest trials, with the exception of DS M-L tests. Largest range was seen in SS in M-L plane of trial-1 (10.35 mm), but as

expected smaller than CoP. Correspondent interquartile difference between series was very small. Variation of series was also smaller than CoP correspondent series.

Table 11. Descriptive information of Individuals CoG coordinates. CI=confidence interval.

Double Stance	Mean (SD)	95% CI		Percentiles					Shapiro Wilk		Levene´s test
		Low	Up.	Min	25th	50th	75th	Max	W	P	
A-P trial-1	-.01 (.42)	-.02	-.01	-1.59	-.27	.06	.25	.88	.97	.00	w: 30.1
A-P trial-2	.00 (.31)	-.02	- .00	-1.15	-.22	.03	.23	.78	.99	.00	p: .00
M-L trial-1	.00 (.15)	.00	.00	-.34	- .12	.03	.10	.33	.97	.00	w: 434
M-L trial-2	-.01 (.33)	-.01	.00	-.71	- .34	.02	.28	.90	.95	.00	p: .00

Single Stance	Mean (SD)	95% CI		Percentiles					Shapiro Wilk		Levene´s test
		Low	Up.	Min	25th	50th	75th	Max	W	P	
A-P trial-1	-.00 (1.33)	-.02	.01	-2.99	-1.11	-.12	1.01	3.76	.98	.00	w: 2.4
A-P trial-2	-.01 (1.29)	-.02	.00	-2.83	-.70	.10	.88	3.25	.96	.00	p: .12
M-L trial-1	.03 (1.65)	.01	.05	-4.88	- 1.04	-.07	.96	5.47	.97	.00	w: 56
M-L trial-2	.02 (1.04)	.01	.03	-3.07	- .78	-.04	.85	2.58	.98	.00	p: .00

Tandem Stance	Mean (SD)	95% CI		Percentiles					Shapiro Wilk		Levene´s test
		Low	Up.	Min	25th	50th	75th	Max	W	P	
A-P trial-1	-.00 (.92)	-.01	.01	-3.50	-.47	.14	.65	1.78	.95	.00	w: .8
A-P trial-2	-.02 (.93)	-.01	.00	-2.19	-.77	.10	.68	1.67	.97	.00	p: .40
M-L trial-1	.00 (.75)	.01	.05	-1.59	- .57	-.07	.35	2.40	.97	.00	w: 7.3
M-L trial-2	-.00 (.68)	.01	.03	-3.59	- .41	-.05	.47	1.85	.98	.00	p: .01

Table 12 shows that paired T-test did not find significant differences between Manikin CoP test-retest series. Pearson Correlation was not significant or expressive, showing no correlation between the mean values of trials data series.

Table 12. Manikin CoP similarity and correlation tests between the test-retest trails.

Series	Paired t-test		Pearson Correlation test	
	t-statistics	p-value	rc	p-value
Medial-Lateral	0	.966	.12	.00
Anterior-Posterior	0	.762	.03	.40

Table 13 shows that paired tests did not demonstrate a significant difference between Manikin CoG test-retest series. Spearman correlation found moderate relationship, r=.43 on mean M-L series and poor association r =.17 on A-P data points. CoG Manikins series showed a greater association between series when compared to CoP behavior.

Table 13. Manikin CoG similarity and correlation tests between test-retest trails.

CoG Series	Wilcoxon		Spearman Correlation tests	
	t-statistics	p-value	rc	p-value

| Medial-Lateral | -0 | .77 | .43 | .00 |
| Anterior-Posterior | 87993 | .82 | .17 | .00 |

There were no significant differences between test-retest trials in both planes for Individuals CoP coordinates (Table 14). Correlations of CoP test-retest series were very low for SS and TS, and non-existent in DS. SS data series were parametric. Then, paired T-test and Pearson´s correlation were applied, and Wilcoxon and Spearman tests to other leg tasks.

Table 14. Individuals CoP coordinates similarity and correlation tests.

Stance	Direction	Similarity tests		Correlation tests	
		t-statistics	p-value	rc	p-value
Double	Medial-Lateral	82648	.86	.06	.16
	Anterior-Posterior	80992	.55	.07	.11
Single	Medial-Lateral	.46	.65	.15	.00
	Anterior-Posterior	.73	.47	.18	.00
Tandem	Medial-Lateral	78343	.21	.14	.00
	Anterior-Posterior	77262	.13	.14	.00

Individuals CoG pairwise tests of similarity for non-parametric series were executed and no systematic error was found between M-L series in all leg positions (Table 15). In A-P plane, only in DS task there was a significant difference between trials. No significant correlations in DS between trials in both planes. In SS and TS there was a moderate monotonic association in A-P series (r=.62 and r=.46, respectively) and poor association in M-L coordinates.

Table 15. Individuals CoG coordinates similarity and correlation tests.

Stance	Direction	Wilcoxon test		Spearman Correlation test	
		t-statistics	p-value	r	p-value
Double	Medial-Lateral	86208	.51	.07	.08
	Anterior-Posterior	**79895**	**.03**	.07	.08
Single	Medial-Lateral	85992	.53	.27	.00
	Anterior-Posterior	87045	.70	.62	.00
Tandem	Medial-Lateral	82610	.15	.16	.00
	Anterior-Posterior	81175	.07	.46	.00

6.4 Individuals CoP and CoG coordinate graphical analysis

The following graphs (Figure 25 - Figure 30) show the mean values of each obtained frame for test-retest trials in all leg positions. Double stance presented best consistency between trials. Apparently, retest trials showed less variation (orange

lines). Boxplots whiskers reflect the differences in variations during trails. The largest variation was in SS, after in TS, and finally DS. Individuals CoP data series, in DS M-L plane (Figure 25, top-left) appeared to have very similar series, as well as in A-P (top-right), but with a greater variance in the second half of the test.
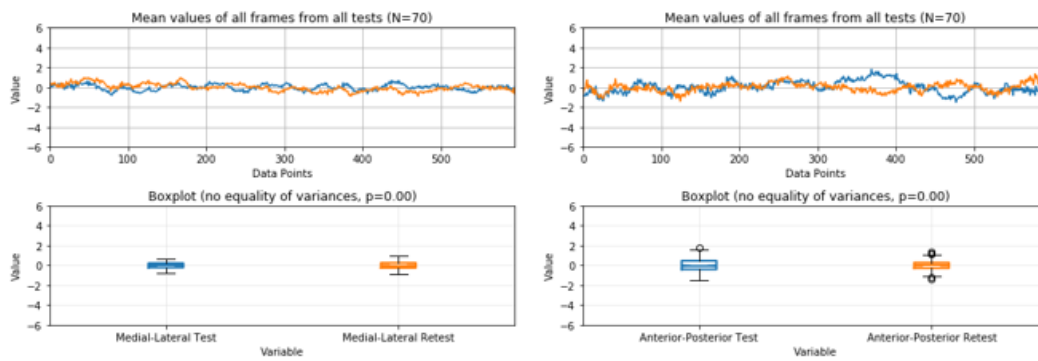
Figure 25 Individuals CoP double stance tests. On top, mean data point values for each frame. At bottom, Box Plots comparing trials variances. Blue=test, Orange= Retest.

Boxplots show interquartile and variance differences between trials. Single Stance task variances apparently were larger in first trials (blue) than those in the second trials (orange) for both planes, Figure 26.
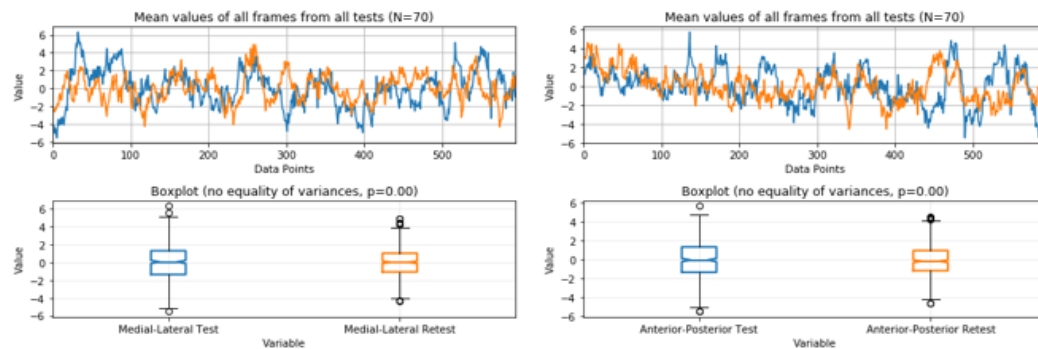
Figure 26 Individuals CoP Single Stance tests. Blue=test, Orange= Retest.
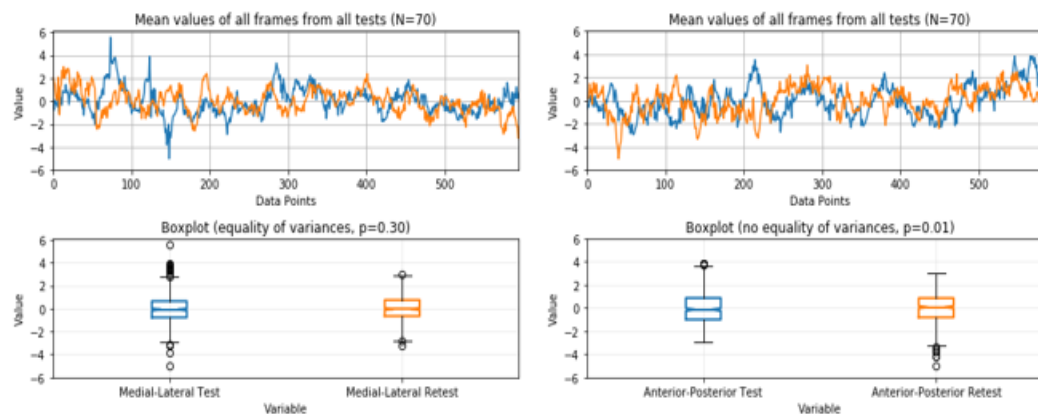
Figure 27 Individuals CoP Tandem tests. Blue=test, Orange= Retest.

In TS, M-L graph mostly represented the A-P body oscillation, and in the same way, A-P represented M-L plane in the anatomical position (Figure 27). It can be explained because during the test the subject stood up in the diagonal of WBB. Additionally, during second trials images reflected the smaller standard deviations reinforcing the suggestion that participants, in mean, presented less body sway in all positions. It was higher and larger the magnitude of signals obtained by WBB (CoP) (Figure 25 - Figure 27) when compared to that captured by Kinect (CoG) (Figure 28 - Figure 30).
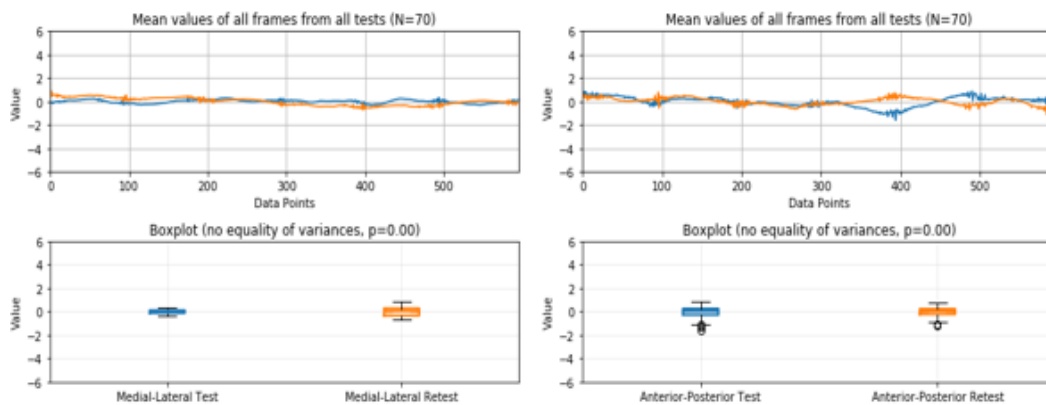
Figure 28 Individuals CoG graph analysis of double stance tests. On top, mean data point values for each obtained frame. At bottom, Box Plots comparing trials variances. Blue=test, Orange=retest.
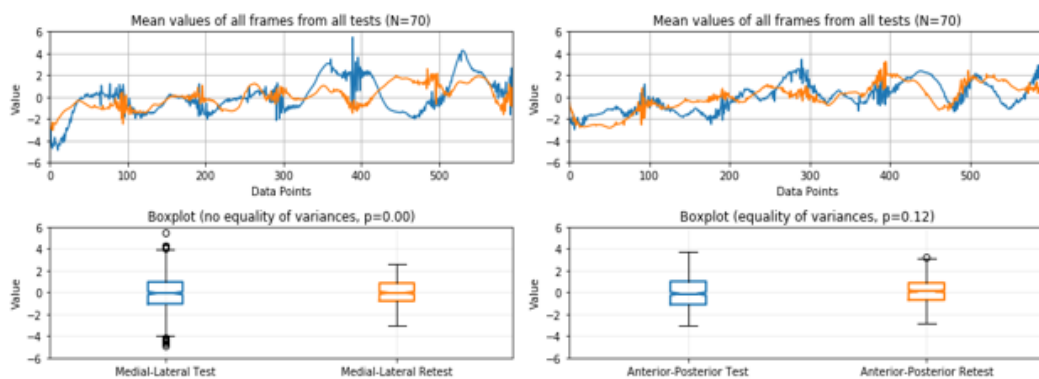
Figure 29 Individuals CoG coordinates graph analysis of single stance tests. Blue=Test, Orange=Retest.
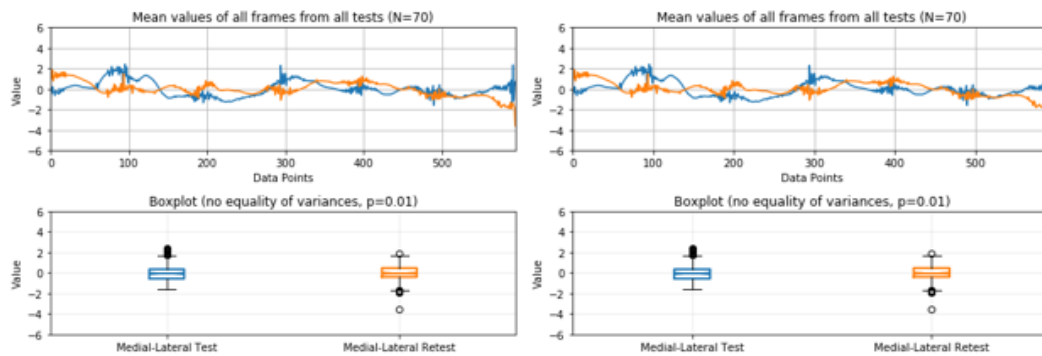
Figure 30 Individuals CoG graph analysis of tandem stance tests. Blue=Test, Orange=Retest.

6.5 Derived variables analysis

None of the Manikin CoP and CoG derived variables was normality distributed following Shapiro-Wilk test (p<.05). The differences in medians were negligible of all Manikins CoP derived variables between trials. For example, resultant TPL median difference between trials was only 0.5 millimeters (Table 16). For resultant CoP Mean Distance, all interquartile values agreed. Range property that is more sensible to extreme values, showed a maximum mean difference of only 0.3 mm between trails in M-L plane. Relative Manikins static performance could be noticed when analyzing the estimated 95% of confidence areas that were in mean less than 1mm^2 and their maximum estimated value was 5.16 mm^2. Standard error (Standard Deviation / SQRT (N), not shown in the table) in rare cases that differed between trials, and when it occurred, it was at the second decimal place of a millimeter. In the next tables all distances are in (mm), speed in (mm/s), and area (mm2), when applied.

Table 16. Descriptive of all Manikin CoP variables for test and retest trials. A refers to test and B to retest trials.

Manikin CoP Properties	Mean	(SD)	Min	Max	Percentiles			Shapiro-Wilk	
					25th	Median	75th	w	p
Mean Distance A	0.22	(0.1)	0.09	0.61	0.11	0.17	0.31	0.85	0.00
Mean Distance B	0.22	(0.1)	0.09	0.58	0.11	0.17	0.31	0.85	0.00
Total Path Length A	70.7	(42.4)	28.7	192.3	35.4	53.1	100.0	0.85	0.00
Total Path Length B	70.5	(41.7)	28.7	187.1	34.9	52.6	99.8	0.85	0.00
RMS Distance A	0.25	(0.1)	0.10	0.70	0.13	0.19	0.35	0.85	0.00
RMS Distance B	0.25	(0.1)	0.10	0.68	0.13	0.19	0.35	0.85	0.00
Average Speed A	3.56	(2.1)	1.45	9.70	1.78	2.67	5.04	0.85	0.00
Average Speed B	3.55	(2.1)	1.45	9.42	1.76	2.65	5.03	0.85	0.00
Range A	0.70	(0.4)	0.24	2.32	0.35	0.53	0.95	0.86	0.00

	Mean	(SD)	Min	Max	25th	Median	75th	W	P
Range B	0.72	(0.4)	0.24	2.15	0.38	0.53	0.99	0.86	0.00
95% Circle Area A	0.88	(1.1)	0.11	5.16	0.17	0.39	1.29	0.70	0.00
95% Circle Area B	0.88	(1.0)	0.11	4.79	0.17	0.39	1.32	0.72	0.00
95% Ellipse Area A	0.71	(0.8)	0.08	3.85	0.13	0.31	1.03	0.71	0.00
95% Ellipse Area B	0.69	(0.8)	0.08	3.74	0.13	0.31	1.04	0.73	0.00
Medial Lateral									
Mean Distance A	0.17	(0.1)	0.07	0.49	0.08	0.13	0.24	0.85	0.00
Mean Distance B	0.17	(0.1)	0.07	0.46	0.09	0.13	0.24	0.85	0.00
Total Path Length A	55.6	(33.6)	22.0	156.4	28.3	41.2	78.6	0.85	0.00
Total Path Length B	55.5	(33.0)	22.0	148.6	27.9	40.7	79.4	0.85	0.00
RMS Distance A	0.22	(0.1)	0.09	0.61	0.11	0.16	0.30	0.85	0.00
RMS Distance B	0.22	(0.1)	0.09	0.59	0.11	0.16	0.30	0.85	0.00
Average Speed A	2.80	(1.7)	1.11	7.88	1.43	2.08	3.96	0.85	0.00
Average Speed B	2.80	(1.7)	1.11	7.49	1.40	2.05	4.00	0.85	0.00
Range A	1.28	(0.8)	0.46	4.06	0.66	0.99	1.76	0.85	0.00
Range B	1.31	(0.8)	0.46	3.77	0.69	0.96	1.81	0.86	0.00
Anterior-Posterior									
Mean Distance A	0.10	(0.1)	0.04	0.29	0.05	0.08	0.14	0.86	0.00
Mean Distance B	0.10	(0.1)	0.04	0.28	0.05	0.08	0.14	0.86	0.00
Total Path Length A	33.0	(19.5)	12.6	88.2	16.5	25.1	47.0	0.86	0.00
Total Path Length B	32.8	(19.4)	12.6	89.8	16.6	25.2	46.7	0.85	0.00
RMS Distance A	0.13	(0.1)	0.05	0.37	0.06	0.10	0.18	0.86	0.00
RMS Distance B	0.13	(0.1)	0.05	0.35	0.06	0.10	0.18	0.86	0.00
Average Speed A	1.66	(1.0)	0.63	4.44	0.83	1.26	2.36	0.86	0.00
Average Speed B	1.65	(1.0)	0.63	4.52	0.83	1.27	2.35	0.85	0.00
Range A	0.78	(0.5)	0.27	2.18	0.40	0.59	1.08	0.87	0.00
Range B	0.76	(0.4)	0.27	2.11	0.40	0.59	1.06	0.87	0.00

Manikin CoG (Table 17) presented small differences in their general descriptive data when comparing trials. All properties had values similar to CoP values. However, TPL was 5 times smaller than that registered by CoP coordinates. Consequently, Average Speed was much smaller as well.

In the next tables, the letters (A) and (B) represent test and retest trials, respectively.

Table 17. Descriptive of all Manikin CoG variables for test and retest trials.

Manikin CoG Properties	Mean	(SD)	Min	Max	Percentiles 25th	Median	75th	Shapiro-Wilk W	P
Mean Distance A	0.26	(0.2)	0.07	1.28	0.15	0.21	0.33	0.77	0.00
Mean Distance B	0.27	(0.2)	0.07	1.20	0.15	0.22	0.36	0.82	0.00
Total Path Length A	14.2	(9.1)	3.59	50.0	7.02	11.76	21.0	0.86	0.00
Total Path Length B	14.5	(9.0)	3.59	47.4	7.36	11.94	21.1	0.88	0.00
RMS Distance A	0.32	(0.2)	0.08	1.37	0.17	0.24	0.41	0.78	0.00
RMS Distance B	0.33	(0.2)	0.08	1.47	0.17	0.26	0.46	0.79	0.00
Average Speed A	0.71	(0.5)	0.18	2.50	0.35	0.59	1.05	0.86	0.00

Average Speed B	0.73	(0.4)	0.18	2.38	0.37	0.60	1.06	0.88	0.00
Range A	0.79	(0.7)	0.20	4.48	0.37	0.54	0.97	0.75	0.00
Range B	0.81	(0.6)	0.20	3.32	0.38	0.57	1.02	0.77	0.00
Circle Area A	1.68	(3.0)	0.07	20.8	0.29	0.60	1.84	0.53	0.00
Circle Area B	1.73	(2.9)	0.07	18.4	0.27	0.67	2.18	0.55	0.00
Ellipse Area A	0.87	(0.8)	0.12	4.62	0.35	0.64	0.97	0.75	0.00
Ellipse Area B	0.96	(0.9)	0.12	5.07	0.35	0.71	1.12	0.79	0.00
Medial-Lateral									
Mean Distance A	0.17	(0.1)	0.04	0.98	0.09	0.14	0.20	0.70	0.00
Mean Distance B	0.17	(0.1)	0.04	0.86	0.09	0.15	0.21	0.76	0.00
Total Path Length A	9.34	(6.8)	2.24	45.8	4.78	8.06	11.8	0.78	0.00
Total Path Length B	9.62	(6.9)	2.24	44.3	5.12	7.81	11.8	0.80	0.00
RMS Distance A	0.22	(0.2)	0.06	1.07	0.12	0.18	0.25	0.74	0.00
RMS Distance B	0.23	(0.2)	0.06	1.10	0.12	0.19	0.28	0.74	0.00
Average Speed A	0.47	(0.3)	0.11	2.29	0.24	0.40	0.59	0.78	0.00
Average Speed B	0.48	(0.3)	0.11	2.22	0.26	0.39	0.59	0.80	0.00
Range A	0.99	(0.6)	0.25	3.12	0.53	0.84	1.19	0.83	0.00
Range B	1.05	(0.7)	0.25	3.50	0.57	0.90	1.27	0.84	0.00
Anterior-Posterior									
Mean Distance A	0.16	(0.1)	0.05	0.77	0.09	0.12	0.19	0.72	0.00
Mean Distance B	0.16	(0.1)	0.05	0.80	0.09	0.12	0.20	0.77	0.00
Total Path Length A	9.00	(6.6)	2.26	31.92	4.24	6.27	12.41	0.81	0.00
Total Path Length B	9.13	(6.4)	2.26	34.06	4.43	6.63	11.81	0.82	0.00
RMS Distance A	0.21	(0.2)	0.06	0.86	0.11	0.15	0.25	0.74	0.00
RMS Distance B	0.22	(0.2)	0.06	0.96	0.11	0.15	0.28	0.75	0.00
Average Speed A	0.45	(0.3)	0.11	1.60	0.21	0.31	0.62	0.81	0.00
Average Speed B	0.46	(0.3)	0.11	1.71	0.22	0.33	0.59	0.82	0.00
Range A	0.99	(0.7)	0.26	3.40	0.53	0.73	1.31	0.80	0.00
Range B	1.01	(0.7)	0.26	3.34	0.53	0.74	1.32	0.80	0.00

Normality was violated for all derived Individuals CoP and CoG variables verified by Shapiro-Wilk test ($p < .05$, not shown). Then, it was applied data transformations as described in the Methodology section, and normality was met for all CoP variables, except for Single Stance (A, B) medial-lateral Range (Table 18 - Table 20). Among CoG variables resultant Single Stance (B) Total Path Length, Average Speed, TS (A) Circle and Ellipse areas, Single Stance anterior-posterior (B) Total Path Length and Average Speed violated normality, even using transformed data ($p < .05$) (Table 21 - Table 23).

Descriptive data of Individuals CoP variables demonstrated that Double Stance properties had the smallest values, followed by Tandem Stance and Single Stance, as expected. In retest trials, properties, in general, were smaller in all positions than in test trials. Homogeneity of variance was also tested by Levene´s

test and all comparisons were satisfied by the test (p > .05), except for CoP Single Stance M-L Range (p=.03) and for CoG Single Stance Circle area (p=.04).

Table 18. Descriptive information about Individuals Resultant CoP properties on original scale based on transformed data distribution.

Position / Trial		Mean	Min	Max	Percentile			Shapiro-Wilk		Levene´s test	
					25th	Median	75th	w	P	W	p
Mean distance	DS A	4.7	1.9	10.0	3.9	5.7	4.9	.97	.08		
	DS B	4.2	1.9	12.4	3.5	5.4	4.3	1.0	.74	1.66	.20
	SS A	17.0	10.0	44.5	13.4	20.4	17.7	.98	.26		
	SS B	15.6	9.6	33.5	12.9	18.5	15.6	1.0	.72	.42	.51
	TS A	11.7	7.4	28.8	9.3	14.9	11.8	.98	.22		
	TS B	11.4	5.2	23.8	9.5	13.7	11.9	1.0	.12	.01	.93
Total Path Length	DS A	224	133	646	182	266	231	.99	.66		
	DS B	199	108	492	167	246	191	1.0	.76	.24	.63
	SS A	1584	802	3606	1286	1977	1596	.99	.73		
	SS B	1540	936	3738	1305	1903	1596	1.0	.08	.30	.59
	TS A	963	491	2278	791	1205	976	.98	.26		
	TS B	941	530	2928	732	1177	967	1.0	.72	.08	.78
RMS Distance	DS A	5.7	2.2	11.7	5.7	4.7	6.6	.99	.70		
	DS B	5.0	2.1	13.0	5.0	4.0	6.2	1.0	.94	2.0	.15
	SS A	19.9	11.3	51.7	15.5	25.1	20.0	.98	.26		
	SS B	17.9	10.9	42.6	14.7	22.8	17.6	1.0	.54	.61	.43
	TS A	13.3	8.1	24.9	10.5	16.9	13.4	.97	.16		
	TS B	13.2	6.2	27.3	11.1	16.1	13.7	1.0	.27	.00	.96
Average Speed	DS A	11.3	6.7	32.6	9.2	13.4	11.7	.99	.66		
	DS B	10.0	5.5	24.8	8.4	12.4	9.6	1.0	.77	.23	.63
	SS A	79.8	40.4	182.0	64.8	99.6	80.4	.99	.73		
	SS B	77.6	47.2	188.6	65.8	95.8	80.5	1.0	.08	.29	.59
	TS A	48.5	24.7	114.7	39.9	60.7	49.2	.98	.26		
	TS B	47.4	26.7	147.4	36.9	59.3	48.7	1.0	.72	.08	.78
Range	DS A	13.7	4.9	38.8	13.2	11.1	16.2	.97	.07		
	DS B	11.8	4.3	23.8	12.0	9.5	15.1	1.0	.73	.29	.59
	SS A	54.2	20.6	233.1	37.8	89.9	53.0	.98	.40		
	SS B	45.9	21.9	166.7	32.9	65.8	44.3	1.0	.07	.60	.44
	TS A	31.2	17.3	100.5	24.8	39.2	30.7	.98	.52		
	TS B	32.3	16.1	168.6	25.6	39.9	32.2	1.0	.26	.06	.80
95% c. circle area	DS A	348	48	1725	336	241	471	.98	.52		
	DS B	268	44	1245	264	182	410	1.0	.80	1.43	.23
	SS A	3924	1126	35435	2424	7524	3989	.98	.34		
	SS B	3184	1178	24741	2214	5624	3146	1.0	.41	.19	.66
	TS A	1918	597	7055	1815	1132	2989	.98	.26		
	TS B	1944	441	7874	1875	1182	2965	1.0	.56	.08	.78
95% c. ellipse	DS A	191	37	1214	201	125	298	.98	.40		
	DS B	168	36	1507	144	106	282	1.0	.44	.11	.74

Position / Trial		Mean	Min	Max	25th	Median	75th	W	P	w	p
SS	A	3271	1165	22078	2100	5536	3470	.98	.33		
	B	2775	1088	16753	1998	4273	2843	1.0	.54	.09	.76
TS	A	1497	496	5760	1431	946	2317	.98	.28		
	B	1457	258	7036	1647	878	2208	1.0	.73	.27	.60

Table 19. Descriptive information about Individuals Medial-Lateral CoP properties on original scale based on transformed data distribution.

Position / Trial		Mean	Min	Max	Percentile			Shapiro-Wilk		Levene´s test	
					25th	Median	75th	W	P	w	p
Mean distance											
DS	A	1.5	0.5	6.4	1.2	2.2	1.6	.97	.15		
	B	1.6	0.7	6.1	1.2	2.4	1.7	1.0	.35	.08	.77
SS	A	10.6	5.9	32.6	8.3	13.2	10.6	.98	.28		
	B	9.7	5.9	21.3	8.1	11.5	10.0	1.0	.59	1.71	.19
TS	A	6.2	3.2	13.6	5.2	7.7	6.2	.98	.36		
	B	6.1	2.6	14.8	4.8	8.1	6.5	1.0	.15	.86	.35
Total Path Length											
DS	A	98	52.5	449.5	78.3	128.9	99.6	.98	.54		
	B	95	45.4	341.9	69.7	138.0	95.5	1.0	.61	.66	.42
SS	A	1039	600	2363	843	1359	1059	.98	.46		
	B	1045	588	2049	889	1261	1115	1.0	.51	1.21	.27
TS	A	521	274	1782	403	707	505	.98	.50		
	B	509	272	2178	378	629	515	1.0	.33	.07	.78
RMS Distance											
DS	A	1.9	0.6	9.4	1.6	2.8	2.1	.97	.11		
	B	2.1	0.9	7.5	1.4	3.0	2.1	1.0	.24	.19	.66
SS	A	13.7	7.1	48.3	10.4	18.1	13.0	.97	.12		
	B	12.2	7.2	31.8	10.0	14.6	11.9	1.0	.30	2.74	.10
TS	A	7.9	3.8	18.9	6.5	9.7	7.6	.98	.32		
	B	7.9	3.3	20.1	6.1	10.6	8.5	1.0	.40	.66	.42
Average Speed											
DS	A	5.0	2.6	22.7	3.9	6.5	5.0	.98	.55		
	B	4.8	2.3	17.3	3.5	7.0	4.8	1.0	.61	.65	.42
SS	A	52.3	30.2	118.9	42.4	68.4	53.4	.98	.46		
	B	52.6	29.7	103.4	44.8	63.5	56.2	1.0	.51	1.20	.27
TS	A	26.3	13.8	89.7	20.3	35.7	25.4	.98	.48		
	B	25.6	13.7	109.7	19.0	31.7	26.0	1.0	.33	.08	.78
Range											
DS	A	10.0	3.2	65.4	7.8	14.7	10.7	.98	.18		
	B	10.5	4.7	40.3	7.3	15.9	11.2	1.0	.08	.21	.65
SS	**A**	72.8	34.7	242.5	59.5	46.4	114.3	.90	**.00**		
	B	63.2	33.7	207.6	53.5	46.0	81.9	.9	**.00**	4.73	**.03**
TS	A	42.3	16.7	273.4	33.8	61.0	40.9	.99	.71		
	B	42.7	17.3	226.1	33.2	58.9	42.7	1.0	1.00	.01	.91

Table 20. Descriptive information about Individuals Anterior-Posterior CoP properties on original scale based on transformed data distribution.

Position / Trial		Mean	Min	Max	Percentile			Shapiro-Wilk		Levene´s test	
					25th	Median	75th	w	p	w	p
Mean	DS A	4.0	1.6	8.1	4.0	3.4	4.7	.98	.30		

Category	Position	Trial	Mean	Min	Max	25th	Median	75th	w	p	w	p
		B	3.4	1.4	9.0	3.4	2.8	4.3	1.0	.95	2.84	.09
	SS	A	10.3	5.7	28.8	8.3	13.3	10.4	.99	.72		
		B	9.5	5.1	21.5	7.8	11.4	9.5	1.0	.80	.23	.63
	TS	A	8.2	4.6	14.4	6.9	10.2	8.2	.97	.14		
		B	8.0	4.0	15.1	6.9	9.8	8.1	1.0	.27	.56	.46
Total Path Length	DS	A	174	100	379	139	218	175	.98	.46		
		B	147	79	454	126	168	147	1.0	.37	.03	.85
	SS	A	940	358	2701	750	1213	969	.98	.35		
		B	881	486	2731	681	1161	922	1.0	.41	.32	.57
	TS	A	701	347	1380	710	580	841	.97	.14		
		B	687	359	1545	653	539	835	1.0	.35	.01	.91
RMS Distance	DS	A	5.0	1.9	9.8	4.9	4.2	5.9	.98	.23		
		B	4.3	1.8	10.4	4.3	3.5	5.2	1.0	.87	2.38	.12
	SS	A	13.3	7.0	37.0	10.6	17.5	13.8	.99	.84		
		B	12.3	6.6	27.7	10.0	15.1	12.2	1.0	.64	.21	.65
	TS	A	10.3	5.7	17.4	8.6	12.9	10.4	.97	.08		
		B	10.1	5.2	18.1	8.6	12.4	10.2	1.0	.41	.80	.37
Average Speed	DS	A	8.8	5.0	19.1	7.0	11.0	8.8	.98	.46		
		B	7.4	4.0	22.8	6.3	8.5	7.4	1.0	.37	.03	.86
	SS	A	47.4	18.0	136.3	37.8	61.1	48.8	.98	.36		
		B	44.4	24.5	137.8	34.4	58.5	46.4	1.0	.42	.32	.57
	TS	A	35.3	17.4	69.5	35.8	29.2	42.4	.97	.14		
		B	34.6	18.1	77.8	32.9	27.2	42.0	1.0	.35	.01	.91
Range	DS	A	24.0	9.3	46.1	24.0	20.5	29.4	.98	.57		
		B	20.3	8.3	40.5	20.2	17.3	24.1	1.0	.51	.69	.41
	SS	A	75.0	32.8	242.3	57.7	101.2	75.8	.99	.96		
		B	67.6	37.3	158.4	53.3	92.0	66.7	1.0	.18	.54	.46
	TS	A	50.7	27.4	92.7	41.7	63.0	50.8	.98	.33		
		B	50.8	27.7	104.0	42.0	62.3	48.2	1.0	.50	.84	.36

Table 21 presents the properties derived from CoG coordinates in resultant, M-L and A-P planes for all leg positions, in both trials. Normality assumption was violated in original series and data was transformed as described in section 5.4. Even after transformations, some variables were not parametric and highlighted in bold. Data was showed after applying the back-transformed functions. Mean values of CoG were smaller of the ones for CoP.

Table 21. Descriptive information about Individuals Resultant CoG properties in original scale based on transformed data distribution.

Position / Trial			Mean	Min	Max	Percentile			Shapiro-Wilk		Levene´s test	
						25th	Median	75th	w	p	w	p
Mean distance	DS	A	3.6	1.5	8.3	3.7	2.7	4.6	.99	.77		
		B	3.2	1.1	9.5	3.3	2.4	4.3	1.0	.83	2.68	.10
	SS	A	11.6	5.2	79.3	8.8	16.0	12.2	.98	.48		

			Mean	Min	Max	25th	Median	75th	w	P	w	p
		B	10.5	5.5	26.8	8.1	14.4	10.7	1.0	.33	2.98	.08
	TS	A	8.2	3.9	21.4	7.7	6.4	10.5	.97	.09		
		B	8.1	2.6	17.7	8.1	6.8	10.3	1.0	.18	.68	.41
Total Path Length	DS	A	60	30	168	47	81	61	.98	.37		
		B	51	21	151	43	72	51	1.0	.38	1.38	.24
	SS	A	322	157	818	323	229	428	.98	.27		
		B	**282**	**150**	**647**	**283**	**203**	**360**	**1.0**	**.04**	.22	.64
	TS	A	169	60	521	132	235	176	.98	.49		
		B	162	78.3	393	126	214	162	1.0	.87	1.20	.27
RMS Distance	DS	A	4.1	1.6	9.0	4.2	3.2	5.2	.99	.84		
		B	3.7	1.3	10.0	3.9	2.7	4.8	1.0	.63	2.78	.10
	SS	A	13.5	5.9	107	10	20	14	.99	.60		
		B	12.1	6.4	38	9	17	12	1.0	.49	2.60	.11
	TS	A	9.3	4.2	21.9	8.5	6.8	11.8	.97	.05		
		B	9.4	3.3	19.8	9.2	7.8	11.9	1.0	.19	.61	.43
Average Speed	DS	A	3.0	1.5	8.4	2.4	4.2	3.1	.98	.37		
		B	2.6	1.1	7.7	2.1	3.7	2.6	1.0	.40	1.39	.24
	SS	A	16.1	7.9	40.9	16.3	11.6	21.2	.98	.28		
		B	**14.4**	**7.5**	**33.9**	**14.2**	**10.2**	**18.3**	**1.0**	**.04**	.10	.75
	TS	A	8.5	3.0	26.1	6.6	11.8	8.8	.98	.46		
		B	8.2	3.9	19.6	6.5	10.7	8.1	1.0	.90	1.16	.28
Range	DS	A	8.6	2.7	21.8	8.3	6.8	11.5	.99	.65		
		B	7.2	2.4	18.2	7.1	5.8	9.5	1.0	.89	.62	.43
	SS	A	32.0	9.9	282.2	22.7	52.7	34.5	.99	.77		
		B	26.9	13.2	150.2	18.8	41.5	25.3	1.0	.14	1.07	.30
	TS	A	18.9	7.9	100.6	13.9	25.8	18.1	.99	.96		
		B	18.7	6.2	51.3	15.0	26.2	18.1	1.0	.09	.32	.57
95% c. circle area	DS	A	181	22	779	180	110	279	.98	.48		
		B	142	17	855	153	76	211	1.0	.29	1.98	.16
	SS	A	2132	325	22140	2081	1019	4457	.98	.51		
		B	1698	404	19903	1566	910	2833	1.0	.20	1.94	.16
	TS	**A**	**880**	**154**	**4919**	**755**	**485**	**1449**	**.96**	**.04**		
		B	926	128	4056	869	627	1560	1.0	.31	1.66	.20
95% c. ellipse area	DS	A	14.2	2.6	179	15	7.1	22.7	.98	.31		
		B	15.8	2.7	303	17	7.2	27.8	1.0	.29	.80	.37
	SS	A	726	94	7422	782	304	1457	.98	.36		
		B	474	83	2313	451	221	964	1.0	.10	4.21	**.04**
	TS	**A**	**146**	**26**	**4071**	**124**	**59**	**310**	**.96**	**.02**		
		B	134	20	1745	129	65	247	1.0	.76	1.87	.17

Table 22. Descriptive information about Individuals Medial-Lateral CoG properties on original scale based on transformed data.

Position / Trial			Mean	Min	Max	Percentile			Shapiro-Wilk		Levene´s test	
						25th	Median	75th	w	P	w	p
Mean distance	DS	A	1.1	0.2	4.7	1.1	0.7	1.8	.98	.40		
		B	1.2	0.3	4.6	1.3	0.7	1.9	1.0	.25	.27	.60
	SS	A	7.2	2.8	73.7	5.3	10.8	7.3	.99	.93		
		B	6.3	2.7	20.8	4.7	9.4	6.6	1.0	.19	1.37	.24

			Mean	Min	Max	25th	Median	75th	w	p	w	P
	TS	A	4.6	1.9	12.8	4.5	3.5	6.1	.98	.54		
		B	4.6	1.1	14.1	4.9	3.3	6.2	1.0	.70	.32	.57
Total Path Length	DS	A	24	8.4	127	18	36	23	.99	.70		
		B	21	8.2	106	15	34	21	1.0	.75	.39	.53
	SS	A	199	93.6	590	144	288	203	.98	.23		
		B	173	83.8	469	123	237	178	1.0	.24	.02	.88
	TS	A	98	41.4	424	97	70	133	.98	.38		
		B	87	26.8	215	89	65	110	1.0	.63	.03	.86
RMS Distance	DS	A	1.3	0.2	6.1	1.4	0.8	2.3	.98	.52		
		B	1.5	0.4	6.0	1.6	0.9	2.3	1.0	.27	.41	.52
	SS	A	9.4	3.5	100.7	6.9	15.3	9.8	.99	.94		
		B	8.1	3.2	30.6	5.9	12.0	7.9	1.0	.54	1.32	.25
	TS	A	5.6	2.3	16.6	5.5	4.2	7.3	.98	.39		
		B	5.8	1.4	16.8	6.0	4.3	7.9	1.0	.59	.09	.76
Average Speed	DS	A	1.2	0.4	6.4	0.9	1.8	1.2	.99	.68		
		B	1.1	0.4	5.3	0.8	1.7	1.0	1.0	.74	.46	.50
	SS	A	10.0	4.7	29.5	7.3	14.4	10.2	.98	.26		
		B	8.7	4.2	23.5	6.2	12.0	8.9	1.0	.29	.02	.89
	TS	A	4.9	2.1	21.2	4.9	3.5	6.7	.98	.39		
		B	4.4	1.3	10.8	4.5	3.3	5.6	1.0	.64	.03	.86
Range	DS	A	5.6	1.0	24.7	5.7	3.6	9.0	.99	.67		
		B	6.0	1.7	26.5	6.3	3.6	9.2	1.0	.27	.60	.44
	SS	A	42.7	15.1	334.6	29.6	68.0	42.9	.99	.71		
		B	37.0	12.9	154.0	28.8	53.6	36.5	1.0	.54	1.53	.22
	TS	A	23.4	8.3	63.9	23.0	17.9	32.2	.98	.25		
		B	24.2	5.8	60.7	24.0	19.2	33.7	1.0	.57	.02	.88

Table 23. Descriptive information about Individuals Anterior-Posterior CoG properties on original scale based on transformed data.

Position / Trial			Mean	Min	Max	Percentile			Shapiro-Wilk		Levene´s test	
						25th	Median	75th	w	p	w	P
Mean distance	DS	A	3.0	1.2	7.2	3.1	2.5	3.7	.98	.31		
		B	2.6	1.0	7.6	2.7	1.8	3.4	1.0	.69	2.38	.12
	SS	A	7.2	2.5	23.3	7.1	5.7	9.4	.99	.68		
		B	6.6	2.9	18.6	6.6	5.0	9.1	1.0	.54	1.20	.27
	TS	A	5.3	2.1	12.6	4.2	7.2	5.4	.98	.24		
		B	5.2	2.3	11.5	4.4	6.5	5.3	1.0	.07	2.15	.14
Total Path Length	DS	A	48	18	142	38	65	51	.97	.05		
		B	41.6	18	129	35	54	42	1.0	.63	.78	.38
	SS	A	196	91	486	199	137	260	.97	.16		
		B	**179**	**83**	**393**	**170**	**130**	**243**	**1.0**	**.03**	.03	.85
	TS	A	128	37	313	126	100	168	.98	.55		
		B	123	47	294	116	95	164	1.0	.59	.65	.42
RMS Distance	DS	A	3.7	1.4	8.2	3.8	2.9	4.5	.98	.43		
		B	3.2	1.1	8.8	3.3	2.2	4.0	1.0	.61	2.19	.14
	SS	A	8.7	3.2	31.5	7.1	12.8	9.0	.97	.05		
		B	8.1	3.5	22.4	6.5	11.2	8.1	1.0	.51	1.27	.26

	TS	A	6.5	2.5	16.0	5.3	8.8	6.6	.98	.32		
		B	6.5	2.9	15.8	5.4	8.3	6.7	1.0	.17	2.10	.15
Average Speed	DS	A	2.4	0.9	7.1	1.9	3.3	2.6	.97	.05		
		B	2.1	0.9	6.6	1.7	2.7	2.1	1.0	.65	.77	.38
	SS	A	9.9	4.6	24.3	10.0	6.9	13.0	.98	.19		
		B	**9.0**	**4.2**	**19.7**	**8.6**	**6.5**	**12.3**	**1.0**	**.02**	.03	.85
	TS	A	6.5	1.9	15.7	6.3	5.0	8.4	.98	.52		
		B	6.2	2.4	14.7	5.8	4.9	8.2	1.0	.60	.63	.43
Range	DS	A	14.7	5.1	29.1	14.9	11.3	18.8	.98	.45		
		B	12.6	4.9	33.8	12.9	9.8	16.4	1.0	.97	.51	.48
	SS	A	39.0	17.3	121.9	29.3	60.0	42.5	.98	.23		
		B	35.4	14.3	117.0	27.4	47.4	36.5	1.0	.85	.16	.69
	TS	A	27.5	11.4	76.1	20.7	37.3	28.9	.98	.45		
		B	28.3	11.0	78.6	22.7	39.2	29.4	1.0	.16	.64	.43

6.6 Derived properties inferential analysis

Table 24 shows the great similarity and agreement of trials in most of Manikin CoP variables. Wilcoxon test demonstrated that none of the variables presented significant differences between trials, except for Range. Spearman correlation showed excellent correlation levels, $r > .95$ and $p < .01$.

In the majority of the properties above 85% of the comparisons had differences sizes up to 20% of the size of the first measure. Distance and speed related variables reached above 85% of the cases within 10% of first trail measure. Goodman and Kruskal´s Gamma test demonstrated high level of association between the two measures. Shapiro-Wilk tests showed that differences were not normally distributed ($p < .05$) for all CoP Manikin properties. Finally, Chi-squared test were performed demonstrating agreement in distributions of all properties.

Table 24. Manikins CoP results of statistical tests performed for comparisons of test and retest trials. Shapiro-Wilk over series differences (S-W Diffs.).

Manikin CoP Property	Wilcoxon		Spearman		Gamma	S-W Diffs.	Differences vs % of cases				Chi-Squared	
	w	p	r	p			5%	10%	15%	20%	chi	P
Resultant												
Mean Dist.	10280	.43	.98	.00	.88	.00	59%	89%	96%	99%	.90	.92
T. Path Len	10873	.82	.98	.00	.87	.00	61%	93%	99%	100%	.76	.94
RMS Dist.	10831	.78	.97	.00	.86	.00	60%	87%	96%	98%	1.19	.88
Av. Speed	10844	.79	.98	.00	.87	.00	60%	93%	99%	100%	.76	.94
Range	9024	**.03**	.93	.00	.77	.00	23%	45%	60%	74%	3.45	.49
Circle	10787	.93	.96	.00	.86	.00	33%	59%	78%	87%	3.11	.54
Ellipse	10298	.44	.97	.00	.86	.00	33%	58%	79%	89%	1.98	.74

Medial Lateral

	w	p	r	p	Gamma	S-W Diffs	10%	30%	50%	80%	chi	P
Mean Dist.	10692	.84	.98	.00	.87	.00	51%	82%	95%	100%	4.29	.37
T. Path Len	10893	.83	.97	.00	.86	.00	47%	81%	94%	99%	1.52	.82
RMS Dist.	10733	.70	.97	.00	.86	.00	56%	84%	96%	98%	5.33	.25
Av. Speed	10872	.82	.97	.00	.86	.00	47%	81%	94%	99%	1.33	.86
Range	8642	**.01**	.95	.00	.81	.00	26%	54%	70%	84%	7.88	.10

Anterior-Posterior

	w	p	r	p	Gamma	S-W Diffs	10%	30%	50%	80%	chi	P
Mean Dist.	10418	.53	.97	.00	.84	.00	39%	76%	90%	96%	.95	.92
T. Path Len	10994	.92	.97	.00	.86	.00	49%	81%	95%	99%	1.57	.81
RMS Dist.	10222	.39	.96	.00	.84	.00	41%	78%	90%	94%	3.12	.54
Av. Speed	10975	.91	.97	.00	.86	.00	49%	81%	95%	99%	1.57	.81
Range	9812	.22	.94	.00	.78	.00	29%	55%	72%	84%	.77	.94

None of the Manikin CoG properties was parametric. For all variables, Wilcoxon and Spearman tests investigated differences and correlations between trials. Table 25 demonstrates that Mean Distance, RMS Distance, and Ellipse Area showed significant differences between trials (p < .05). Most of variables achieved r>.80 of correlation. Resultant TPL demonstrated the highest level of correlation (r=.93, p < .01). Gamma test also presented moderate level of association between trials while Chi-Squared test demonstrated agreement between distributions for most of variables. Error distributions showed that most of the test cases were within 30% of difference size. Manikin CoG variables, in general, had worse agreement when compared to CoP variables.

Table 25. Manikin CoG statistic tests executed over all tests comparing the similarities, relationships and agreement between test-retest trials.

Manikin CoG Properties	Wilcoxon		Spearman-r		Gamma	S-W Diffs	Differences vs % of cases				Chi-Square	
	w	p	r	p			10%	30%	50%	80%	chi	P
Resultant												
Mean Dist.	8601	**.04**	.86	.00	.69	.00	39%	82%	92%	95%	5.50	.24
T. Path Len	9302	.11	.93	.00	.79	.00	56%	92%	95%	98%	1.36	.85
RMS Dist.	8994	**.03**	.85	.00	.68	.00	39%	80%	89%	94%	7.05	.13
Av. Speed	9338	.12	.93	.00	.79	.00	56%	92%	95%	98%	1.36	.85
Range	9143	.10	.88	.00	.69	.00	30%	74%	88%	96%	5.02	.28
Circle	8892	.05	.86	.00	.69	.00	18%	54%	76%	86%	9.51	.05
Ellipse	8040	**.00**	.81	.00	.62	.00	21%	52%	75%	88%	9.19	.06
Medial-Lateral												
Mean Dist.	8447	**.01**	.80	.00	.61	.00	29%	73%	86%	94%	3.07	.55
T. Path Len	9802	.26	.91	.00	.76	.00	43%	88%	94%	98%	4.04	.40
RMS Dist.	8700	**.01**	.81	.00	.63	.00	34%	72%	87%	95%	2.90	.58
Av. Speed	9722	.27	.91	.00	.77	.00	42%	88%	94%	98%	4.04	.40
Range	8559	**.02**	.87	.00	.69	.00	35%	75%	91%	96%	.99	.91
Anterior-Posterior												

Mean Dist.	9984	.75	.79	.00	.62	.00	33%	78%	91%	95%	1.60	.81
T. Path Len	9837	.34	.90	.00	.74	.00	47%	88%	95%	98%	9.31	.05
RMS Dist.	9550	.28	.80	.00	.63	.00	41%	78%	89%	95%	5.89	.21
Av. Speed	9829	.33	.90	.00	.74	.00	47%	88%	95%	98%	7.95	.09
Range	-1	.35	.88	.00	.64	.00	33%	76%	91%	95%	3.49	.48

Different leg tasks reached distinct results for Individuals CoP properties. Similarity test between test-retest for resultant DS CoP properties showed that trials differed significantly ($p < .01$). Medial-lateral coordinate properties did not differ while A-P plane variables demonstrated significant differences. Despite found differences, there was still moderate-positive association between trials following Pearson´s test ($p < .01$ and $r > .50$ for most of variables, Table 26).

Residuals were normally distributed following Shapiro-Wilk test ($p>.05$, Table 26 - Table 28). For the great majority of variables (Table 26) the differences between trials up to 50% of the first measurement included more than 70% of the comparisons, with some variables such as resultant TPL or Average Velocity reaching 100% of the cases. Inferential tests were implemented on transformed data not in original scale. However, tests to calculate difference sizes were performed with original scale.

Table 26. Individuals CoP Double Stance properties inferential statistics on test-retest trials. Shapiro-Wilk test of residuals (S-W Diffs).

DS – CoP Variables	Paired T-Test		Pearson´s Test		S-W Diffs.		Difference Sizes by % of Cases			
	w	P	r	p	w	p	5%	15%	30%	50%
Mean Distance	-2.96	0.00	0.66	0.00	0.98	0.40	18%	44%	69%	96%
Total Path Length	-3.91	0.00	0.65	0.00	0.97	0.05	9%	32%	74%	99%
RMS Distance	3.31	0.00	0.61	0.00	0.99	0.76	14%	40%	70%	93%
Average Speed	-3.91	0.00	0.65	0.00	0.97	0.05	9%	32%	74%	99%
Range	3.58	0.00	0.49	0.00	0.99	0.78	6%	24%	66%	91%
Circle Area	3.52	0.00	0.56	0.00	0.99	0.94	9%	16%	40%	70%
Ellipse Area	2.08	0.04	0.76	0.00	0.99	0.72	6%	23%	52%	74%
Medial-Lateral										
Mean Distance	0.99	0.32	0.60	0.00	0.99	0.73	10%	34%	58%	78%
Total Path Length	-0.77	0.44	0.68	0.00	0.99	0.97	12%	34%	68%	90%
RMS Distance	1.05	0.30	0.61	0.00	0.98	0.50	13%	33%	61%	81%
Average Speed	-0.77	0.45	0.68	0.00	0.99	0.97	10%	34%	68%	90%
Range	0.97	0.34	0.60	0.00	0.99	0.83	13%	30%	51%	81%
Anterior-Posterior										
Mean Distance	3.65	0.00	0.50	0.00	0.98	0.36	9%	36%	61%	90%
Total Path Length	-5.71	0.00	0.64	0.00	0.96	**0.02**	10%	32%	65%	99%
RMS Distance	4.07	0.00	0.51	0.00	0.99	0.58	16%	30%	61%	91%
Average Speed	-5.71	0.00	0.64	0.00	0.96	**0.02**	10%	32%	65%	99%
Range	4.51	0.00	0.51	0.00	0.98	0.19	10%	30%	67%	93%

Most of SS Individuals CoP properties (Table 27) did not demonstrate significant differences between trails (p>.05). For most of the variables the difference size of 50% based on first measure also included most of the pair comparisons (>70%) with the exception of Circle Area.

Table 27. Individuals CoP Single Stance properties inferential statistics on test-retest trials. Shapiro-Wilk test of residuals (S-W Diffs).

SS CoP Variables	Paired T-Test		Pearson´s Test		S-W diffs.		Difference Sizes by % of Cases			
	w	p	r	p	w	p	5%	15%	30%	50%
Mean Distance	-2.33	0.02	0.40	0.00	0.98	0.29	15%	46%	69%	90%
Total Path Length	-0.99	0.33	0.65	0.00	0.98	0.42	21%	47%	81%	94%
RMS Distance	-2.31	0.02	0.35	0.00	0.98	0.18	13%	35%	68%	90%
Average Speed	-0.98	0.33	0.65	0.00	0.98	0.42	21%	47%	81%	94%
Range	-2.13	0.04	0.17	0.15	0.98	0.48	3%	24%	41%	71%
Circle Area	-2.08	0.04	0.27	0.02	0.98	0.27	3%	22%	35%	59%
Ellipse Area	-2.00	0.05	0.36	0.00	0.97	0.09	2%	16%	47%	73%
Medial-Lateral										
Mean Distance	-2.21	0.03	0.45	0.00	0.97	0.05	14%	45%	78%	97%
Total Path Length	0.18	0.86	0.57	0.00	0.98	0.18	29%	45%	77%	94%
RMS Distance	-2.67	0.01	0.50	0.00	0.96	**0.03**	15%	34%	70%	91%
Average Speed	0.19	0.85	0.57	0.00	0.98	0.19	29%	45%	77%	94%
Range	2.16	0.03	0.43	0.00	0.95	**0.01**	12%	33%	58%	75%
Anterior-Posterior										
Mean Distance	-1.57	0.12	0.30	0.01	0.99	0.67	16%	36%	59%	86%
Total Path Length	-1.76	0.08	0.59	0.00	0.97	0.14	18%	45%	73%	89%
RMS Distance	-1.53	0.13	0.24	0.04	0.98	0.56	6%	31%	61%	88%
Average Speed	-1.75	0.08	0.59	0.00	0.97	0.14	18%	45%	73%	89%
Range	-1.70	0.09	0.13	0.27	0.97	0.12	6%	29%	58%	83%

Tandem stance CoP properties (Table 28) did not present significant differences between trials for all analyzed variables. Associations were moderate for most of properties. Residuals were equally distributed and most of differences were within half of the size of the first measure.

Table 28. Individuals CoP Tandem Stance properties inferential statistics on test-retest trials. Shapiro-Wilk test of residuals (S-W Diffs).

TS CoP Variable	Paired T-Test		Pearson´s Test		S-W Diffs.		Difference Sizes by Cases			
	w	p	r	p	w	p	5%	15%	30%	50%
Mean Distance	-0.85	0.40	0.42	0.00	0.99	0.89	10%	38%	72%	91%
Total Path Length	-0.68	0.50	0.63	0.00	0.99	0.62	7%	41%	74%	93%
RMS Distance	-0.15	0.88	0.41	0.00	0.99	0.72	7%	33%	65%	91%
Average Speed	-0.68	0.50	0.63	0.00	0.99	0.61	7%	41%	74%	93%
Range	0.65	0.52	0.14	0.23	0.99	0.82	6%	31%	62%	78%
Circle Area	-0.16	0.87	0.31	0.01	0.98	0.56	10%	19%	34%	46%

Ellipse Area	0.34	0.74	0.45	0.00	0.98	0.32	1%	16%	34%	53%
Medial-Lateral										
Mean Distance	-0.35	0.73	0.49	0.00	0.99	0.86	16%	33%	67%	87%
Total Path Length	-0.57	0.57	0.60	0.00	0.99	0.92	12%	36%	64%	91%
RMS Distance	-0.03	0.98	0.47	0.00	0.99	0.92	12%	32%	61%	83%
Average Speed	-0.57	0.57	0.60	0.00	0.99	0.92	12%	36%	64%	91%
Range	0.17	0.87	0.33	0.01	0.99	0.76	11%	32%	59%	74%
Anterior-Posterior										
Mean Distance	-0.68	0.50	0.37	0.00	0.99	0.99	9%	37%	67%	89%
Total Path Length	0.63	0.53	0.65	0.00	0.99	0.65	13%	46%	74%	94%
RMS Distance	-0.49	0.63	0.33	0.01	0.99	0.83	9%	36%	64%	90%
Average Speed	0.63	0.53	0.65	0.00	0.99	0.65	13%	46%	74%	94%
Range	0.03	0.98	0.25	0.04	0.98	0.20	14%	37%	61%	81%

Similarity tests demonstrated that there was a systematic difference in most of DS and SS CoG properties, while TS CoG variables did not show systematic differences (Table 29 - Table 31). There were moderate-strong correlations between CoG DS properties (Table 29). However, SS and TS (Table 30 - Table 31) variables demonstrated none or weak correlations.

Shapiro-Wilk test checked normality of differences that was met by all CoG properties except for TS resultant Range. Analyzing the percent size of difference in relation to first trial, with 50% difference size, got different results depending on analyzed variable. It could include only 35% of the cases, as in TS Ellipse Area or 96% of the cases in DS A-P TPL and Average Speed.

Few CoG variables in some trials did not meet normality test: SS resultant and A-P TPL and Average Speed second trial, and TS resultant Mean Distance, RMS Distance, Circle and Ellipse areas, first trial condition, for example. However, we assumed that all CoG data were normal distributed for the similarities and correlation tests presented in Table 29 - Table 31.

Table 29. Individuals CoG Double Stance properties inferential statistics comparing first and second trials. Shapiro-Wilk test of residuals (S-W Diffs).

CoG DS Variable	Paired T-Test		Pearson´s Test		S-W Diffs.		Difference Sizes by Cases			
	W	P	r	P	w	p	5%	15%	30%	50%
Mean Distance	1.96	0.05	0.52	0.00	1.00	1.00	12%	25%	51%	83%
Total Path Length	-4.47	0.00	0.74	0.00	0.98	0.52	10%	37%	66%	94%
RMS Distance	2.22	0.03	0.51	0.00	0.99	0.89	7%	31%	50%	81%
Average Speed	-4.56	0.00	0.74	0.00	0.98	0.53	11%	37%	66%	94%
Range	3.78	0.00	0.50	0.00	0.99	0.85	9%	26%	54%	79%
95% C. Circle Area	2.37	0.02	0.46	0.00	0.99	0.88	4%	14%	25%	48%
95% C Ellipse Area	-0.93	0.36	0.45	0.00	0.98	0.35	11%	20%	32%	55%
Medial-Lateral										

	W	p	r	p	w	p	5%	15%	30%	50%
Mean Distance	-0.99	0.32	0.44	0.00	0.99	0.96	6%	18%	37%	59%
Total Path Length	-2.21	0.03	0.65	0.00	0.99	0.96	7%	22%	52%	78%
RMS Distance	-1.06	0.29	0.47	0.00	0.99	0.91	10%	19%	44%	56%
Average Speed	-2.28	0.03	0.66	0.00	0.99	0.90	7%	22%	52%	78%
Range	-1.11	0.27	0.52	0.00	0.99	0.73	15%	27%	43%	64%
Anterior-Posterior										
Mean Distance	2.70	0.01	0.43	0.00	0.98	0.30	6%	14%	45%	77%
Total Path Length	-4.15	0.00	0.69	0.00	0.98	0.41	6%	31%	76%	96%
RMS Distance	2.94	0.00	0.43	0.00	0.98	0.30	4%	17%	49%	73%
Average Speed	-4.22	0.00	0.70	0.00	0.98	0.42	6%	33%	75%	96%
Range	3.18	0.00	0.44	0.00	0.97	0.11	10%	25%	51%	80%

Table 30. Individuals CoG Single Stance properties inferential statistics comparing first and second trials. Shapiro-Wilk test of residuals (S-W Diffs).

CoG SS Variable	Paired T-Test		Pearson´s Test		S-W Diffs.		Difference Sizes by Cases			
	W	p	r	p	w	p	5%	15%	30%	50%
Mean Distance	-1.50	0.14	0.10	0.41	0.98	0.27	14%	30%	50%	76%
Total Path Length	2.65	0.01	0.48	0.00	0.99	0.81	11%	34%	62%	84%
RMS Distance	-1.59	0.12	0.08	0.49	0.98	0.43	15%	29%	47%	76%
Average Speed	2.17	0.03	0.44	0.00	0.99	0.60	9%	33%	61%	84%
Range	-1.98	0.05	0.06	0.61	0.98	0.29	9%	19%	39%	61%
95% C. Circle Area	1.58	0.12	0.15	0.22	0.98	0.41	8%	19%	29%	40%
95% C. Ellipse Area	3.11	0.00	0.35	0.00	0.97	0.07	7%	18%	26%	36%
Medial-Lateral										
Mean Distance	-1.72	0.09	0.11	0.35	0.97	0.13	11%	22%	52%	74%
Total Path Length	-2.34	0.02	0.34	0.00	0.98	0.56	4%	19%	49%	76%
RMS Distance	-1.89	0.06	0.15	0.21	0.98	0.20	11%	28%	45%	72%
Average Speed	-2.39	0.02	0.35	0.00	0.98	0.54	4%	21%	47%	76%
Range	-1.84	0.07	0.22	0.07	0.98	0.36	9%	24%	42%	66%
Anterior-Posterior										
Mean Distance	1.17	0.25	0.18	0.14	0.99	0.66	6%	21%	42%	73%
Total Path Length	1.68	0.10	0.37	0.00	0.97	0.18	15%	33%	64%	82%
RMS Distance	-1.01	0.32	0.13	0.27	0.99	0.62	9%	21%	45%	73%
Average Speed	1.73	0.09	0.39	0.00	0.97	0.17	16%	33%	63%	81%
Range	-1.35	0.18	0.11	0.35	0.98	0.32	15%	32%	45%	70%

Table 31. Individuals CoG Tandem Stance properties inferential statistics comparing first and second trials. Shapiro-Wilk test of residuals (S-W Diffs).

CoG TS variables	Paired T-Test		Pearson´s Test		S-W Diffs.		Difference Sizes by Cases			
	W	p	r	p	w	p	5%	15%	30%	50%
Mean Distance	0.39	0.70	0.28	0.02	0.97	0.17	13%	36%	61%	78%
Total Path Length	-0.88	0.38	0.43	0.00	0.99	0.99	13%	33%	55%	78%
RMS Distance	-0.22	0.82	0.31	0.01	0.98	0.33	13%	28%	62%	77%
Average Speed	-0.87	0.38	0.43	0.00	0.99	0.97	13%	36%	55%	78%
Range	-0.15	0.88	0.03	0.79	0.95	**0.01**	4%	16%	38%	71%
95% C. Circle Area	-0.47	0.64	0.28	0.02	0.99	0.69	4%	14%	25%	42%
95% C. Ellipse Area	0.58	0.56	0.22	0.07	0.98	0.27	2%	8%	19%	35%
Medial-Lateral										

Mean Distance	0.08	0.93	0.41	0.00	0.98	0.26	6%	21%	48%	73%
Total Path Length	1.95	0.05	0.36	0.00	0.98	0.28	12%	29%	56%	78%
RMS Distance	-0.55	0.59	0.37	0.00	0.98	0.39	9%	22%	46%	74%
Average Speed	1.95	0.06	0.36	0.00	0.98	0.25	12%	29%	56%	78%
Range	-0.52	0.61	0.33	0.01	0.98	0.48	10%	25%	51%	68%
Anterior-Posterior										
Mean Distance	-0.45	0.65	0.35	0.00	0.99	0.74	10%	31%	51%	74%
Total Path Length	0.81	0.42	0.38	0.00	0.98	0.49	9%	27%	53%	77%
RMS Distance	-0.02	0.99	0.31	0.01	0.99	0.88	9%	24%	51%	79%
Average Speed	0.80	0.42	0.38	0.00	0.98	0.45	9%	30%	53%	77%
Range	0.44	0.66	0.20	0.09	0.98	0.19	6%	24%	40%	66%

In order to check sensibility of WBB and Kinect CoP and CoG derived information, similarities were checked with paired T-tests among leg conditions separately (DS versus SS, and TS, and SS versus TS) comparing correspondent properties considering consolidated tests in one unique set of trial-1 (test) and trial-2 (retest). All properties comparisons resulted in significant $p < .0001$ differences among them. The same outcome was obtained taking individual (one by one) data sets in test and retest conditions among different leg positions.

6.7 Graphical analysis of correlations for individuals properties

The Figure 31 and Figure 32 showed differences between correlations and agreement between test and retest only for resultant properties. Graphs were constructed based on original scale and non-parametric correlation tests. Best-fit (orange) and concordance lines (green) were plotted only for Individuals resultant CoP and CoG variables. Despite many of those variables had presented moderate to good correlations ($r >= 70$, $p < .001$), they were distant from the diagonal line, demonstrating a systematic error between data series. Differences occurred on location (correlation line not over the concordance line) and scale (lines with different inclinations).

Double stance TPL and Average Speed graphs showed the best agreement among resultant CoP properties. Tandem Stance 95%-Estimated Confidence Circle Area and Range were examples of very poor association and agreement between trials (Figure 31).

Estimated 95% confidence Ellipse Area outperformed Circumference Area analyzing correlations and visual graph observations of best-fit line and the diagonal line.

Individuals resultant CoG TPL and Average Speed showed the best monotonic correlation coefficients while resultant CoG TS variables had the worse correlations and agreements (Figure 32).
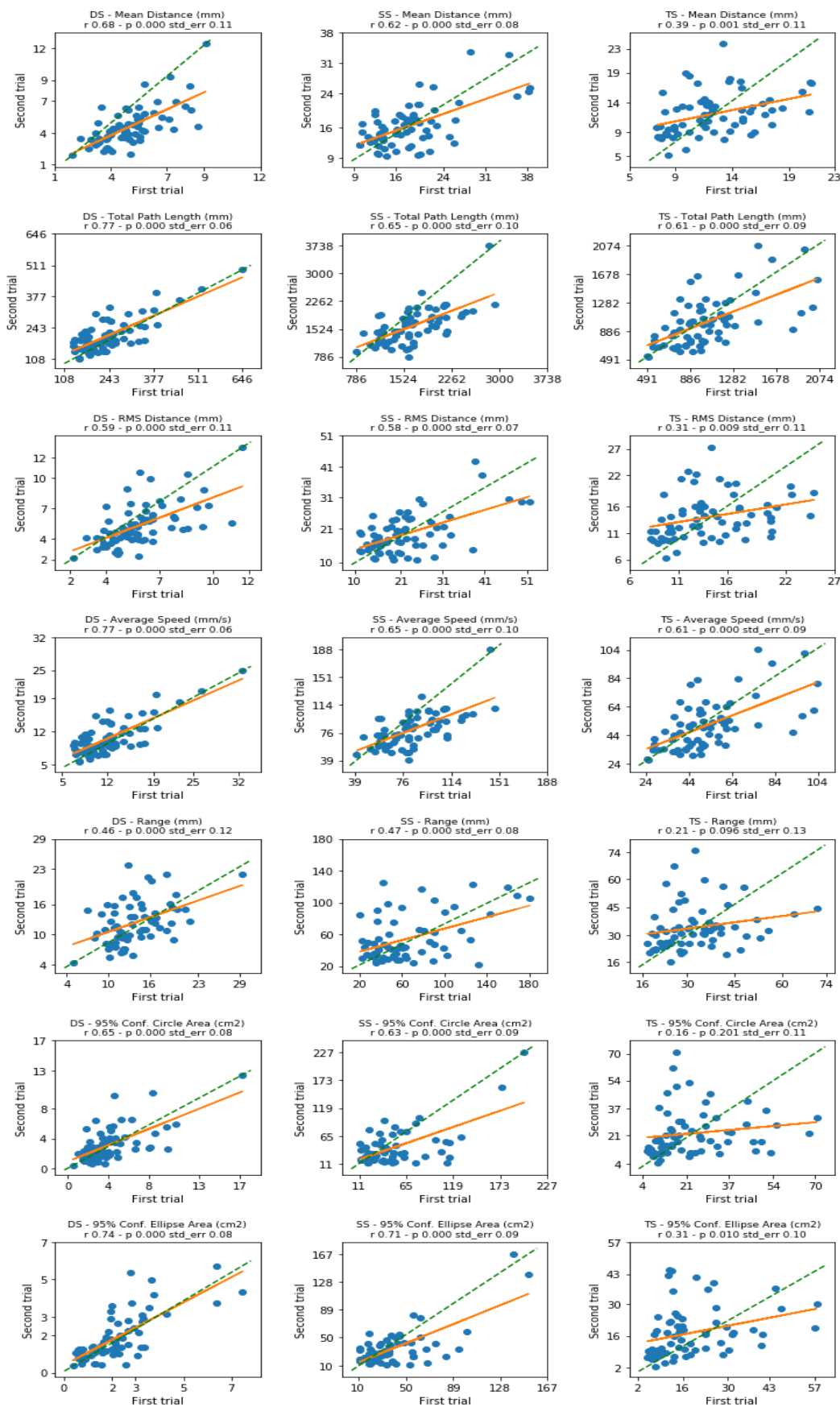
Figure 31 Individuals CoP resultant properties. First versus second trials. Best-fit line (orange) and concordance line (green). Spearman-r correlation and standard error based on non-parametric original scale.

PUC-Rio - Certificação Digital Nº 1621783/CA
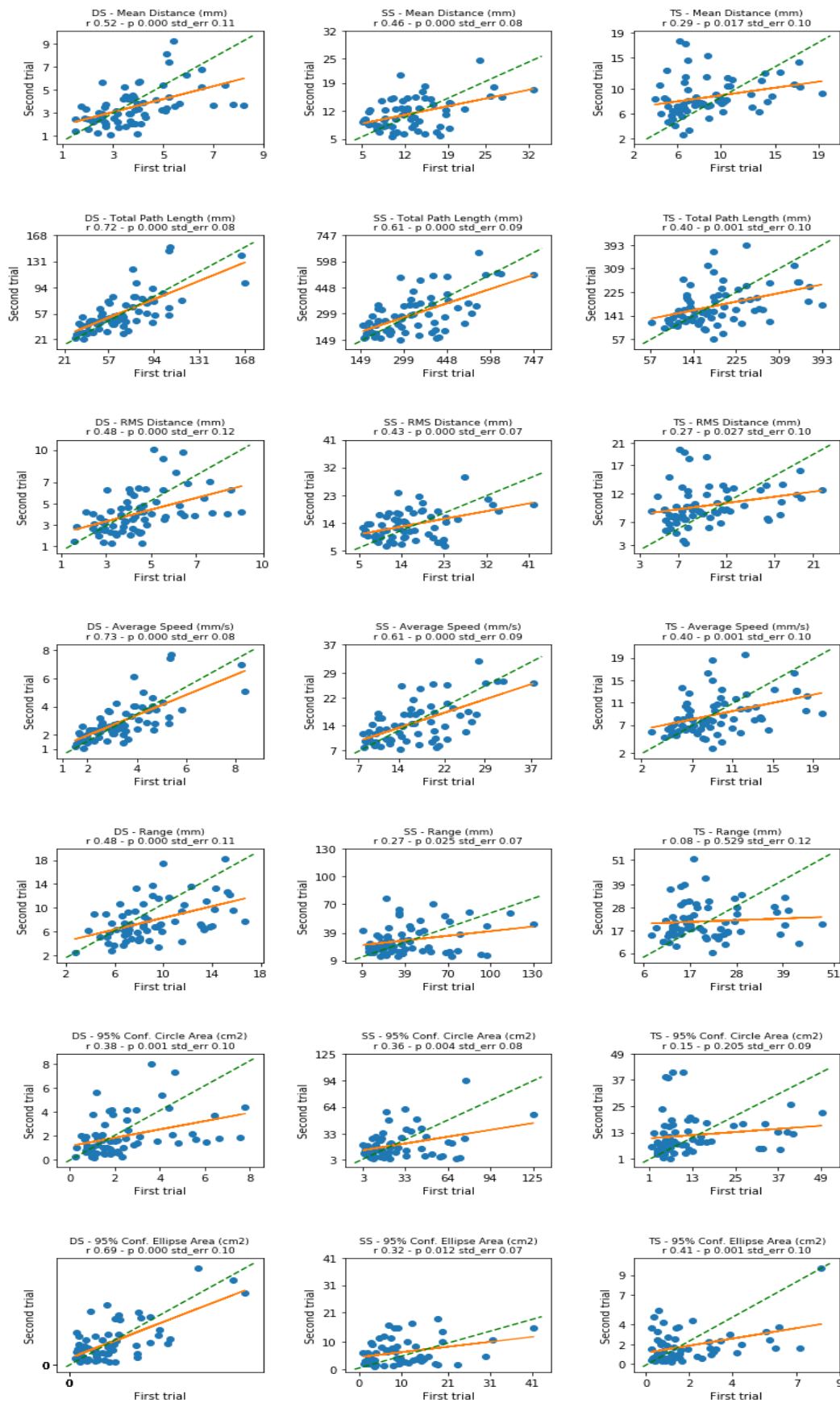
Figure 32 Individuals CoG resultant best-fit agreement chart analysis. Green line represents the perfect agreement. Orange line represents best-fit linear Spearman's correlation.

6.8 Coefficient of variance analysis

Relatively high coefficients of variation in most of Manikins properties may be interpreted with caution (Table 32). Means very close to zero are very sensible to little variations. Moreover, skewed distribution increases variance because mean will be distant of the majority of data points. Nevertheless, our objective was to verify the consistency of the distributions between the two sets of trials what seemed to be very similar with very small differences for most of the variables. Although the very small differences, it was noticeable that retest trials had lower coefficients than test trials for almost all variables in CoP related to M-L direction and CoG for both directions. Additionally, CoP variables presented lower variation in all properties compared to correspondent CoG ones, except for 95% estimated confidence Ellipse Area (Table 32).

Table 32. Coefficients of variation – Manikin variables (%)

Properties	CoP		CoG	
	CV trial1	CV trial2	CV trial1	CV trial2
Mean Distance	60	59	66	63
Total Path Length	60	59	64	62
RMS Distance	60	59	71	70
Average Speed	60	59	64	62
Range	60	61	83	79
95% Confidence Circle Area	121	117	177	165
95% Confidence Ellipse Area	120	117	93	90
Medial-Lateral				
Mean Distance	60	59	73	70
Total Path Length	60	59	72	71
RMS Distance	60	59	74	76
Average Speed	60	59	72	71
Range	60	61	63	63
Anterior-Posterior				
Mean Distance	58	58	75	71
Total Path Length	59	59	73	71
RMS Distance	58	58	74	76
Average Speed	59	59	73	71
Range	59	58	68	69

Individuals CoP coefficients of variation (Table 33) demonstrated that variation increased from test to retest trials in DS CoP resultant for some variables, except for TPL and Average Speed. However, it decreased in SS, except for Circle, and Ellipse Areas. The most stable variable for all positions was resultant TPL. Beyond that, it had the lowest coefficient of variation (27%) among properties.

With humans, variation decreased significantly compared to Manikins series. The largest change in variation was registered in DS resultant Mean Distance (33%). It increased from 0.30 to 0.40 (Table 33).

Table 33. Individuals CoP properties coefficient of variation. Based on original data. (%)

CoP Properties	Double Stance		Single Stance		Tandem Stance	
	Test	Retest	Test	Retest	Test	Retest
Resultant						
Mean Distance	30	40	35	30	29	28
Total Path Length	37	34	27	29	35	35
RMS Distance	31	38	41	33	30	29
Average Speed	38	34	27	29	35	35
Range	29	34	57	52	34	35
95% Circle Area	67	71	76	84	65	60
95% Ellipse Area	64	67	69	76	67	62
Medial – Lateral						
Mean Distance	58	54	30	29	37	38
Total Path Length	47	46	28	25	38	40
RMS Distance	60	54	47	36	38	40
Average Speed	47	46	28	25	38	40
Range	48	47	63	54	45	43
Anterior-Posterior						
Mean Distance	32	36	35	35	28	26
Total Path Length	33	29	31	31	31	35
RMS Distance	30	37	36	35	28	25
Average Speed	33	29	31	31	31	35
Range	28	32	40	38	30	29

For Individuals CoG properties (Table 34), variation increased in DS task from test to retest in all properties except for TPL and Average Speed in M-L plane. For SS task, all properties decreased variation in retest condition, except for resultant TPL that stayed in the same level in both trials, while for TS, in general, all properties had smaller variations except for TPL and Average Speed in M-L. Variation was also larger in CoG than in CoP for all variables.

Table 34. Individuals CoG coefficients of variation based on original data (%).

CoG Properties	Double Stance		Single Stance		Tandem Stance	
	Test	Retest	Test	Retest	Test	Retest
Resultant						
Mean Distance	39	46	44	34	40	35
Total Path Length	41	46	39	39	40	39
RMS Distance	37	46	47	36	41	34
Average Speed	41	46	39	39	40	39
Range	35	42	63	51	43	40

95% Circle Area	76	84	86	82	91	75
95% Ellipse Area	86	88	89	76	100	90
Medial – Lateral						
Mean Distance	66	68	49	37	48	44
Total Path Length	66	62	50	43	41	45
RMS Distance	66	67	52	40	47	41
Average Speed	66	63	50	43	41	45
Range	59	64	62	47	51	46
Anterior-Posterior						
Mean Distance	41	47	42	41	43	34
Total Path Length	31	35	41	40	40	37
RMS Distance	38	47	44	40	44	36
Average Speed	32	35	41	40	40	37
Range	36	41	46	42	45	39

6.9 Analysis of concordance indexes

Concordance indexes should be interpreted with caution for some variables, because they did not meet assumption of normality. Examining Manikin data series, violation of normality could be attributed due the negligible oscillation of manikins what pulled values close to zero as demonstrated in the following histograms (Figure 33). Commonly used data transformation did not change distributions and establishing a floor value, in our case, would cut out most of the analyzed cases, and at the risk of still presenting a strong positive skewness. Considering the number of investigated cases (N=210) and understanding the distribution of data, more important than the statistical power that a normal distribution would provide was the evidence that data series without individuals variability presented an excellent level of precision using Kinect and WBB.
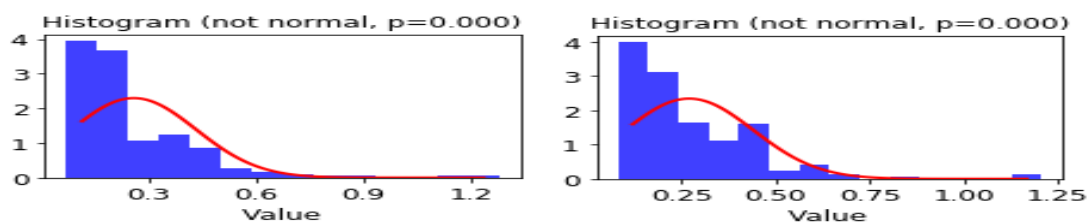
Figure 33 Manikin Mean Distance histograms for test (left) and retest (right). Manikin is a static body with negligible oscillation, then, values are positively skewed close to zero.

All properties had almost perfect relative reliability (Table 35). In order to compare with other studies and our results, it was presented the found coefficients for $ICC_{2,1}$ and CCC tests for Manikins series.

Table 35. Manikins Lin´s Concordance Correlation Coefficient (CCC) and Intraclass Correlation Coefficient (ICC) comparing test-retest trials.

Properties	Center of Pressure			Center of Gravity		
	CCC	ICC	ICC 95% CI	CCC	ICC	ICC 95% CI
Mean Distance	.99	.99	.99 – 1.0	.88	.88	.85 - .91
Total Path Length	.99	.99	.99 – 1.0	.94	.94	.92 - .95
RMS Distance	.99	.99	.99 – .99	.84	.84	.79 - .87
Average Speed	.99	.99	.99 – 1.0	.94	.94	.92 - .95
Range	.95	.95	.93 – .96	.88	.88	.84 - .91
95% Confidence Circle	.99	.99	.99 – .99	.91	.91	.88 - .93
95% Confidence Ellipse	.99	.99	.99 – .99	.74	.74	.68 - .80
Medial-Lateral						
Mean Distance	.99	.99	.99 – .99	.87	.87	.83 - .90
Total Path Length	.99	.99	.99 – .99	.92	.92	.90 - .94
RMS Distance	.99	.99	.99 – .99	.84	.84	.79 - .87
Average Speed	.99	.99	.98 – .99	.92	.92	.90 - .94
Range	.97	.97	.96 – .98	.86	.86	.82 - .90
Anterior-Posterior						
Mean Distance	.99	.98	.98 – .99	.91	.91	.88 - .93
Total Path Length	.99	.99	.99 – .99	.95	.95	.94 - .96
RMS Distance	.99	.98	.98 – .99	.91	.91	.88 - .93
Average Speed	.99	.99	.98 – .99	.95	.95	.94 - .96
Range	.96	.96	.95 – .97	.88	.88	.85 - .91

Concordance indexes for Individuals were generated based on transformed data series (Table 36 - Table 39).

Table 36 demonstrated that CoP Resultant properties demonstrated fair to good concordance scores. Although, 95% confidence intervals (CI) ranged from poor to excellent. Among CoP variables, TPL and Average Speed were more stable across leg tasks and planes. Only TPL and Average Speed had confidence intervals above poor classification in M-L and A-P planes, except in DS task in A-P plane with CI ranging from .16 to .74. Medial-lateral variables scores were higher than A-P ones. Estimated Ellipse Area achieved better relative reliability than Circle Area in all leg positions. Considering consolidated data for all leg positions, CoP variables achieved excellent scores for all variables.

Table 36. Individuals CoP Intraclass Correlation Coefficient (2, 1) for all leg positions.

CoP Variables	Double Stance			Single Stance			Tandem			All		
	r	95% CI		r	95% CI		r	95% CI		r	95% CI	
Mean Distance	.63	.44	.76	.48	.27	.64	.46	.25	.63	.89	.85	.91
Total Path Length	.61	.36	.76	.59	.42	.73	.64	.48	.76	.96	.94	.97
RMS Distance	.57	.37	.72	.42	.21	.60	.42	.20	.59	.88	.84	.91

CoP Variables	DS			SS			TS			All		
	r	95% CI		r	95% CI		r	95% CI		r	95% CI	
Average Speed	.61	.36	.76	.60	.42	.73	.64	.48	.76	.96	.94	.97
Range	.44	.21	.62	.35	.13	.54	.30	.06	.51	.83	.78	.87
95% C. Circle Area	.56	.31	.73	.28	.04	.49	.33	.10	.53	.86	.82	.90
95% C. Ellipse Area	.69	.52	.80	.37	.14	.56	.46	.24	.63	.92	.89	.94
Medial-Lateral												
Mean Distance	.65	.49	.77	.45	.23	.62	.49	.28	.65	.91	.89	.93
Total Path Length	.66	.50	.77	.60	.42	.73	.65	.48	.77	.96	.95	.97
RMS Distance	.65	.49	.77	.47	.26	.63	.47	.26	.63	.91	.89	.93
Average Speed	.66	.50	.77	.60	.42	.73	.65	.48	.77	.96	.95	.97
Range	.67	.51	.78	.44	.23	.61	.42	.20	.60	.91	.88	.93
Anterior-Posterior												
Mean Distance	.45	.20	.63	.34	.12	.53	.37	.15	.55	.81	.75	.86
Total Path Length	.54	.16	.74	.55	.36	.70	.65	.49	.77	.95	.92	.96
RMS Distance	.45	.21	.63	.34	.12	.54	.33	.10	.52	.81	.74	.85
Average Speed	.54	.16	.74	.55	.36	.70	.65	.49	.77	.95	.92	.96
Range	.47	.18	.66	.28	.05	.49	.25	.01	.46	.83	.77	.87

Individuals CoG properties also did not meet normality assumption and tests were performed over transformed data series. All CoG DS properties received fair to good scores with 95% confidence intervals varying from poor to excellent (Table 37). Double stance task received higher scores compared to other leg support tasks. With few exceptions, SS and TS variables had poor scores. Combining all tasks relative concordance was excellent.

In general, CoP scores (Table 36) were higher than CoG (Table 37) ones. However, in DS A-P properties, CoG TPL and Average Speed performed better than CoP.

Table 37. Individuals CoG Intraclass Correlation Coefficient (2, 1) for all leg positions.

CoG Variables	DS			SS			TS			All		
	r	95% CI		r	95% CI		r	95% CI		r	95% CI	
Mean Distance	.52	.32	.67	.30	.07	.51	.33	.10	.52	.79	.73	.84
Total Path Length	.68	.44	.81	.52	.31	.68	**.39**	.17	.57	.90	.86	.92
RMS Distance	.49	.29	.65	.28	.05	.49	.31	.08	.51	.79	.73	.84
Average Speed	.68	.44	.82	.51	.30	.67	**.39**	.17	.57	.90	.86	.92
Range	.46	.23	.64	.19	-.04	.40	.13	-.12	.35	.77	.71	.82
95% C. Circle Area	.44	.23	.61	.23	-.01	.45	.28	.04	.48	.79	.72	.84
95% C. Ellipse Area	.51	.30	.67	.30	.07	.51	.33	.09	.54	.83	.78	.87
Medial-Lateral												
Mean Distance	.52	.32	.67	.29	.06	.49	.41	.19	.60	.81	.76	.86
Total Path Length	.64	.47	.76	.37	.15	.56	.40	.18	.58	.89	.86	.92
RMS Distance	.54	.35	.69	.28	.05	.49	.52	.32	.68	.83	.78	.87
Average Speed	.64	.47	.76	.36	.14	.55	.40	.19	.58	.89	.86	.92
Range	.58	.39	.72	.32	.09	.51	.34	.11	.53	.85	.80	.88
Anterior-Posterior												

	DS			SS			TS			All		
Mean Distance	.41	.19	.59	.31	.07	.51	.35	.12	.54	.65	.56	.72
Total Path Length	.61	.35	.77	.50	.29	.66	.38	.16	.56	.87	.83	.91
RMS Distance	.40	.19	.58	.29	.05	.49	.31	.08	.51	.66	.58	.73
Average Speed	.62	.35	.78	.46	.25	.63	.38	.16	.57	.88	.83	.91
Range	.44	.21	.62	.27	.04	.48	.20	-.03	.42	.72	.65	.78

In general, CCC scores were higher than correspondent ICC scores. Average Speed and TPL were the most consistent properties. Some CoP and CoG variables reached the satisfactory range ($> .69$; Table 38 - Table 39) in confidence intervals. Consolidated data series outperformed scores based on separate leg tasks.

Table 38. CoP Lin´s Concordance Correlation Coefficient scores based on transformed data.

CoP Properties	DS			SS			TS			All		
	r	95% CI		r	95% CI		r	95% CI		r	95% CI	
Resultant												
Mean Distance	.63	.47	.74	.47	.28	.63	.46	.25	.62	.89	.85	.91
Total Path Length	.60	.44	.73	.59	.41	.73	.64	.47	.76	.96	.94	.97
RMS Distance	.57	.40	.70	.42	.22	.59	.41	.20	.59	.88	.84	.90
Average Speed	.60	.44	.73	.59	.41	.73	.64	.47	.76	.96	.94	.97
Range	.43	.24	.59	.35	.13	.53	.30	.06	.50	.83	.78	.87
95% C. Circle Area	.56	.39	.69	.27	.04	.48	.33	.10	.52	.86	.82	.89
95% C. Ellipse Area	.68	.53	.79	.37	.14	.56	.45	.24	.62	.92	.89	.94
Medial – Lateral												
Mean Distance	.65	.49	.77	.44	.23	.61	.48	.28	.64	.91	.89	.93
Total Path Length	.65	.50	.77	.60	.42	.73	.65	.48	.77	.96	.95	.97
RMS Distance	.65	.49	.77	.46	.27	.62	.46	.26	.63	.91	.89	.93
Average Speed	.65	.50	.77	.60	.42	.73	.65	.48	.77	.96	.95	.97
Range	.66	.50	.78	.44	.24	.60	.42	.20	.60	.91	.88	.93
Anterior-Posterior												
Mean Distance	.44	.26	.60	.34	.11	.53	.36	.14	.55	.81	.76	.85
Total Path Length	.53	.37	.66	.55	.36	.70	.65	.49	.77	.95	.93	.96
RMS Distance	.45	.26	.60	.34	.11	.53	.32	.10	.52	.81	.75	.85
Average Speed	.53	.38	.66	.55	.36	.70	.65	.49	.77	.95	.93	.96
Range	.46	.29	.61	.28	.05	.48	.24	.01	.45	.83	.78	.87

Table 39. CoG Lin´s Concordance Correlation Coefficient scores based on transformed data.

CoG Properties	DS			SS			TS			All		
	r	95% CI		r	95% CI		r	95% CI		r	95% CI	
Resultant												
Mean Distance	.51	.32	.66	.30	.08	.50	.32	.10	.52	.79	.73	.84
Total Path Length	.68	.54	.78	.51	.32	.67	.39	.17	.57	.90	.86	.92
RMS Distance	.49	.30	.64	.28	.06	.48	.31	.10	.51	.79	.73	.84
Average Speed	.68	.55	.78	.51	.31	.66	.39	.17	.57	.90	.86	.92
Range	.46	.27	.62	.18	-.04	.39	.12	-.11	.35	.77	.71	.82
95% C. Circle Area	.44	.24	.60	.23	-.00	.44	.27	.04	.47	.79	.72	.84
95% C. Ellipse Area	.50	.30	.66	.30	.07	.50	.33	.09	.53	.83	.78	.87

Medial – Lateral												
Mean Distance	.51	.31	.67	.29	.06	.48	.41	.19	.59	.81	.76	.86
Total Path Length	.63	.47	.75	.37	.16	.55	.40	.18	.58	.89	.86	.92
RMS Distance	.54	.35	.69	.28	.05	.48	.52	.31	.67	.83	.78	.87
Average Speed	.64	.48	.76	.36	.15	.54	.40	.18	.58	.89	.86	.92
Range	.57	.39	.71	.32	.10	.50	.33	.11	.53	.85	.80	.88
Anterior-Posterior												
Mean Distance	.40	.20	.57	.30	.07	.50	.34	.12	.53	.65	.56	.72
Total Path Length	.61	.45	.73	.49	.29	.65	.38	.16	.56	.87	.84	.90
RMS Distance	.40	.20	.57	.28	.05	.48	.31	.08	.50	.66	.58	.73
Average Speed	.62	.46	.74	.46	.24	.62	.38	.16	.56	.87	.84	.90
Range	.44	.24	.60	.27	.04	.47	.20	-.03	.41	.72	.65	.78

6.10 Bland and Altman chart analysis

Only graphs based on Individuals variables were presented because systematic differences were negligible with Manikins data series. All plots were generated using original data. Most of the residuals could be seen within proposed limits (2.77 * within-subject standard deviation) for all variables. Most of the properties differences were normality distributed as shown in our previous analysis, except for CoP Double Stance A-P Total Path Length and Average Speed, CoP Single Stance M-L RMS Distance and Range, as well as CoG resultant Tandem Stance Range (Table 26 - Table 31).

Resultant CoP properties demonstrated relatively small positive bias for most of the variables, representing a greater mean value in the first trial, except for resultant Range in TS that had a small negative difference of -1.58 mm (Figure 34). On M-L plane (Figure 35), the occurrence of negative differences was found in Total Path Length (SS: -7.16 mm; TS: -.09 mm), consequently in Average Speed (SS: -.38 mm/s; TS: -.01 mm/s), Root Mean Squared of Distance (TS: - 0.13 mm) and Range (DS: -0.17 mm; TS: -.18 mm). On A-P plane (Figure 36), mean differences were positive for all properties in all positions. Negative differences represent greater values on retest trial that theoretically suggests worsen performance, oppositely, a positive value suggests a performance enhancement in balance control. However, all TS variables showed non-significant differences between trials (Table 28), while all resultant and A-P DS properties demonstrated significant differences (Table 26). In addition, Single Stance resultant CoP Mean,

RMS Distance and Range presented also systematic differences. No other properties of DS and SS differed from zero significantly (Table 27).

Despite of relatively low mean difference between trials and no statistical evidence of bias for most of the properties, there was a large variance in differences. In the header of each graph was presented the mean of differences, its within-subject standard deviations, 95%-confidence interval of mean differences, and the correlation coefficients between means and differences followed by its p-value. Some variables presented weak positive correlations (CoP Resultant and M-L Single Stance RMS Distance, r=.39 and .44, p <.01, respectively, i.e., Figure 34 - Figure 35), suggesting that differences increases in companion with increased mean values, therefore, differences should be analyzed with caution in those cases.

For Individuals CoG variables, few properties showed significant moderate correlations, r>.50, p< .05 for Single Stance Ellipse Area, Single Stance M-L Range (Figure 37 - Figure 38) and some weak significant correlations, .30 < r < 50, p < .05, resultant SS Mean distance, RMS Distance, Range, Circle Area, TS Circle Area, and so on (Figure 37).
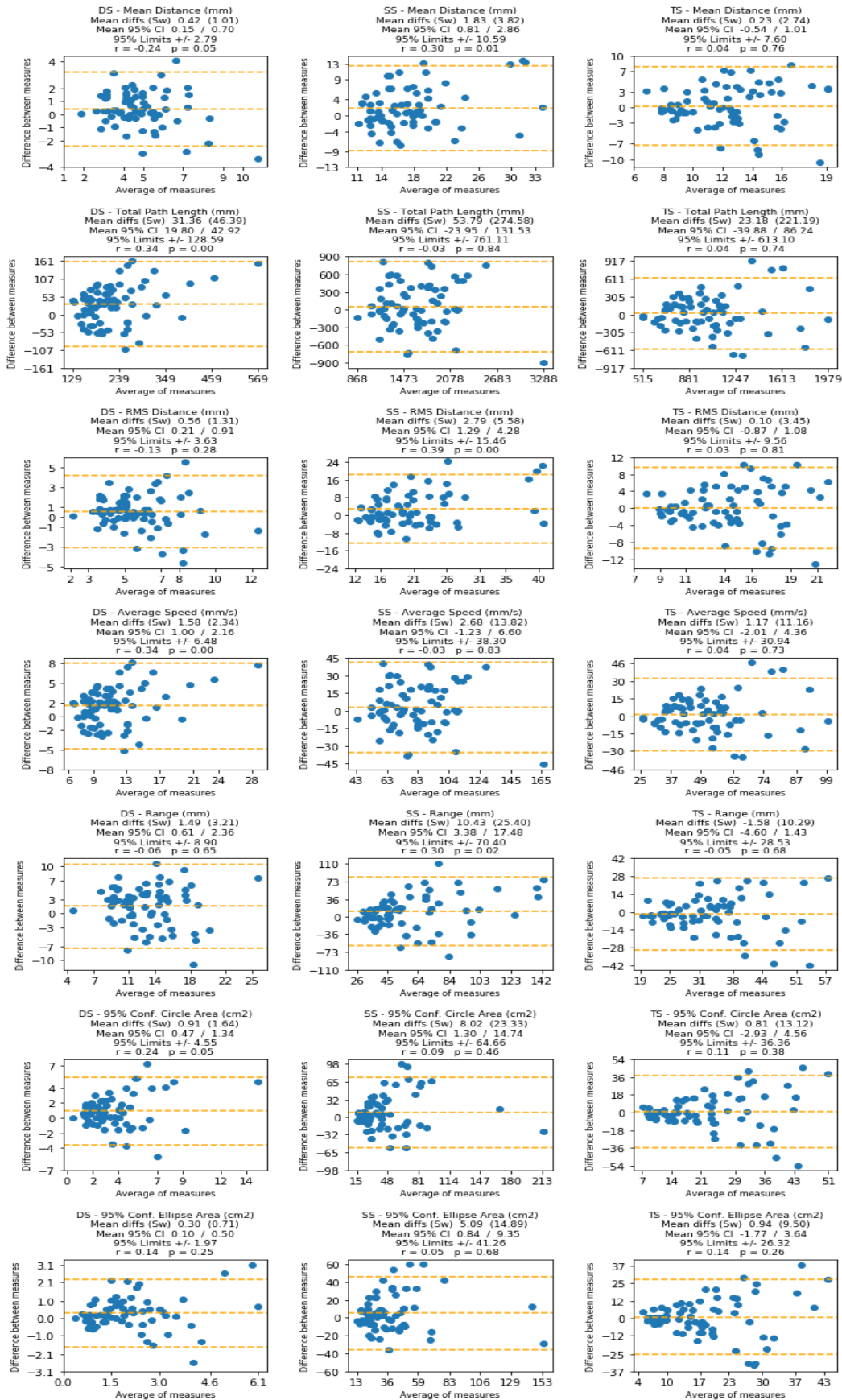
Figure 34 Bland and Altman chart analysis of resultant CoP properties. Middle dashed line is the mean difference, and top and bottom lines +/- ~2.77 within-subject standard deviation. 'r' and 'p' represent the correlation coefficient between means vs differences and its p-value.

PUC-Rio - Certificação Digital Nº 1621783/CA
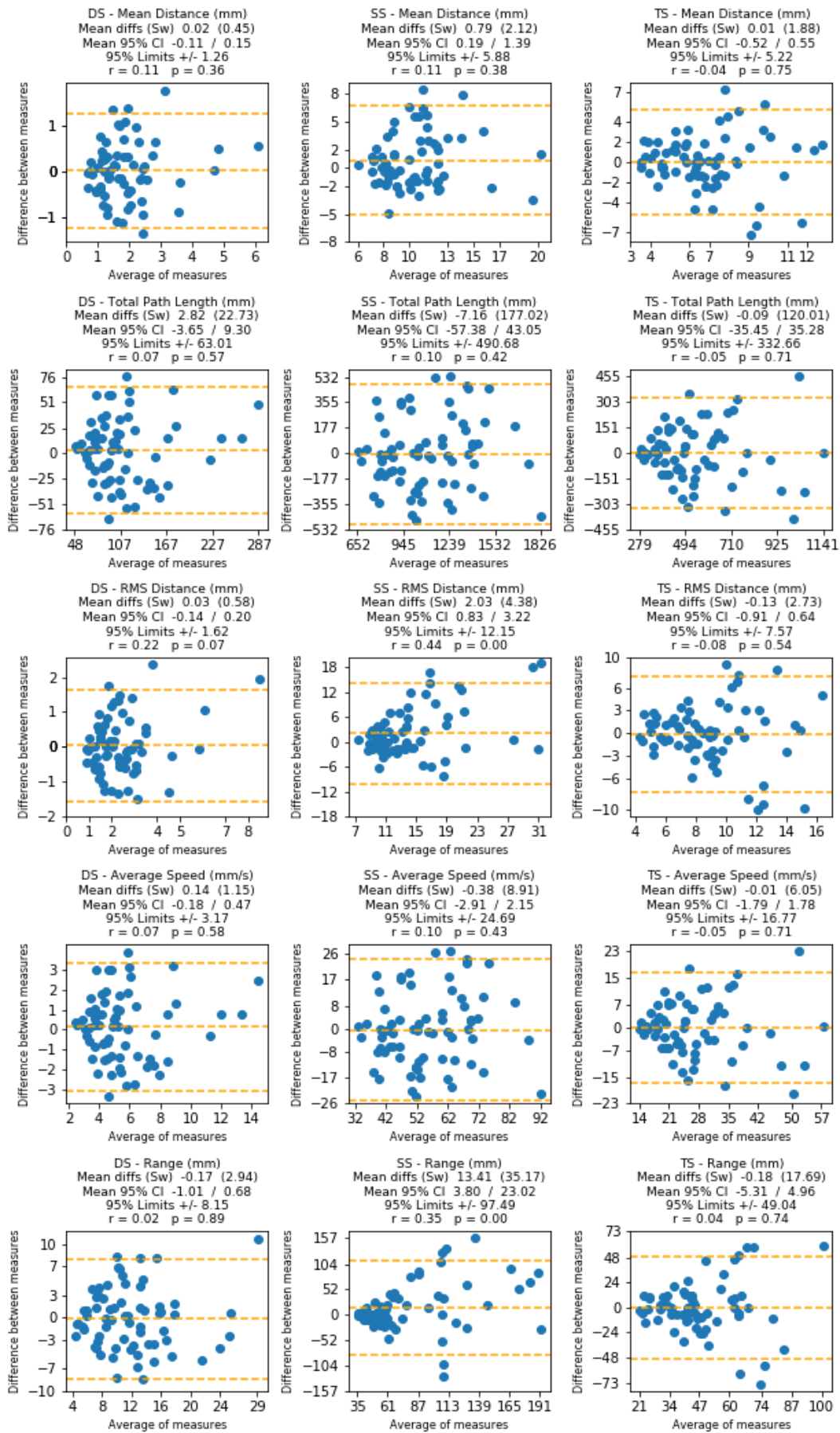
Figure 35 Individuals CoP Medial-Lateral Bland-Altman graphs for all leg positions.

PUC-Rio - Certificação Digital Nº 1621783/CA
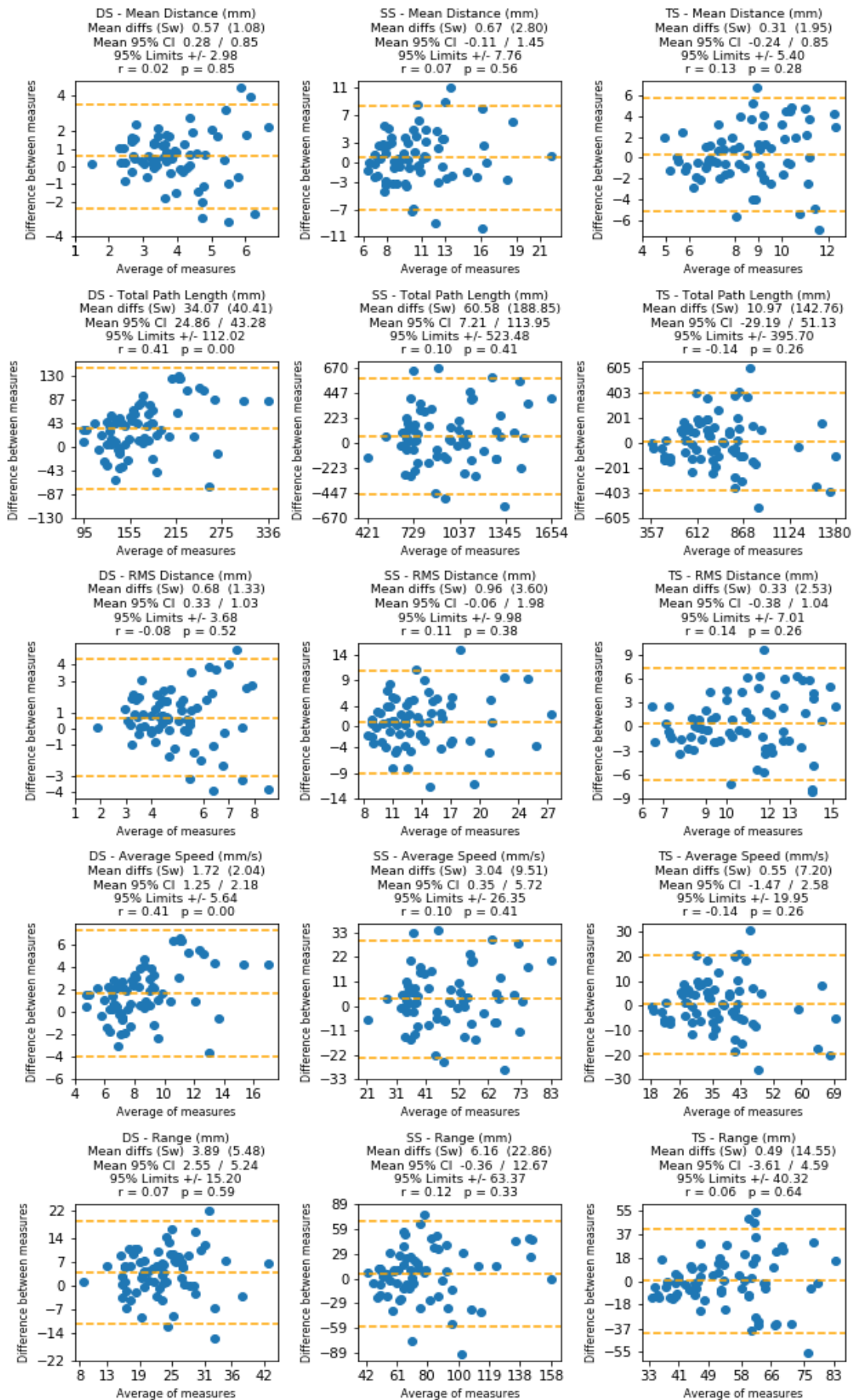
Figure 36 Individuals CoP Anterior-Posterior properties Bland and Altman analysis.
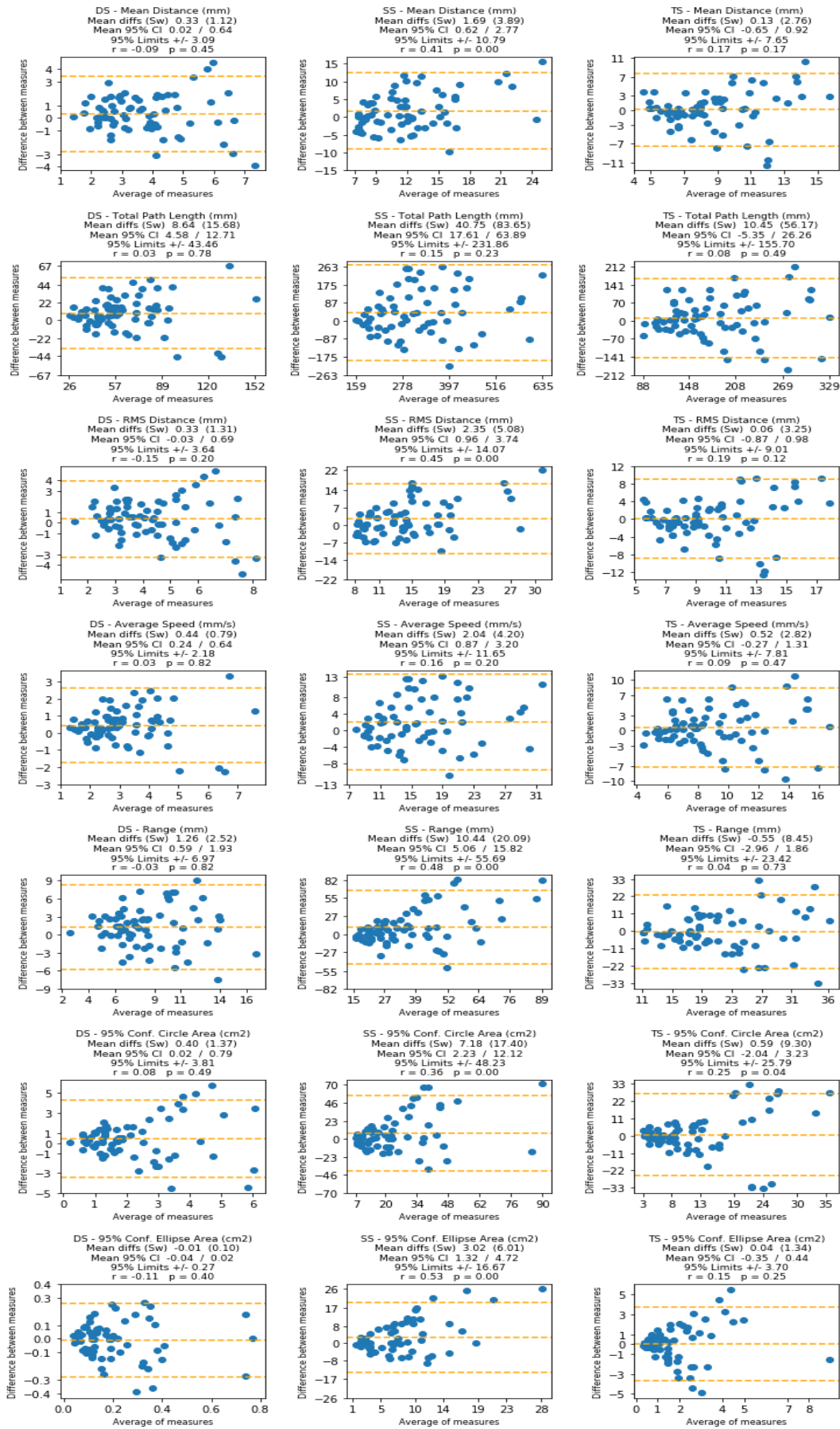
Figure 37 Individuals CoG Resultant properties. Bland-Altman chart analysis.

PUC-Rio - Certificação Digital Nº 1621783/CA
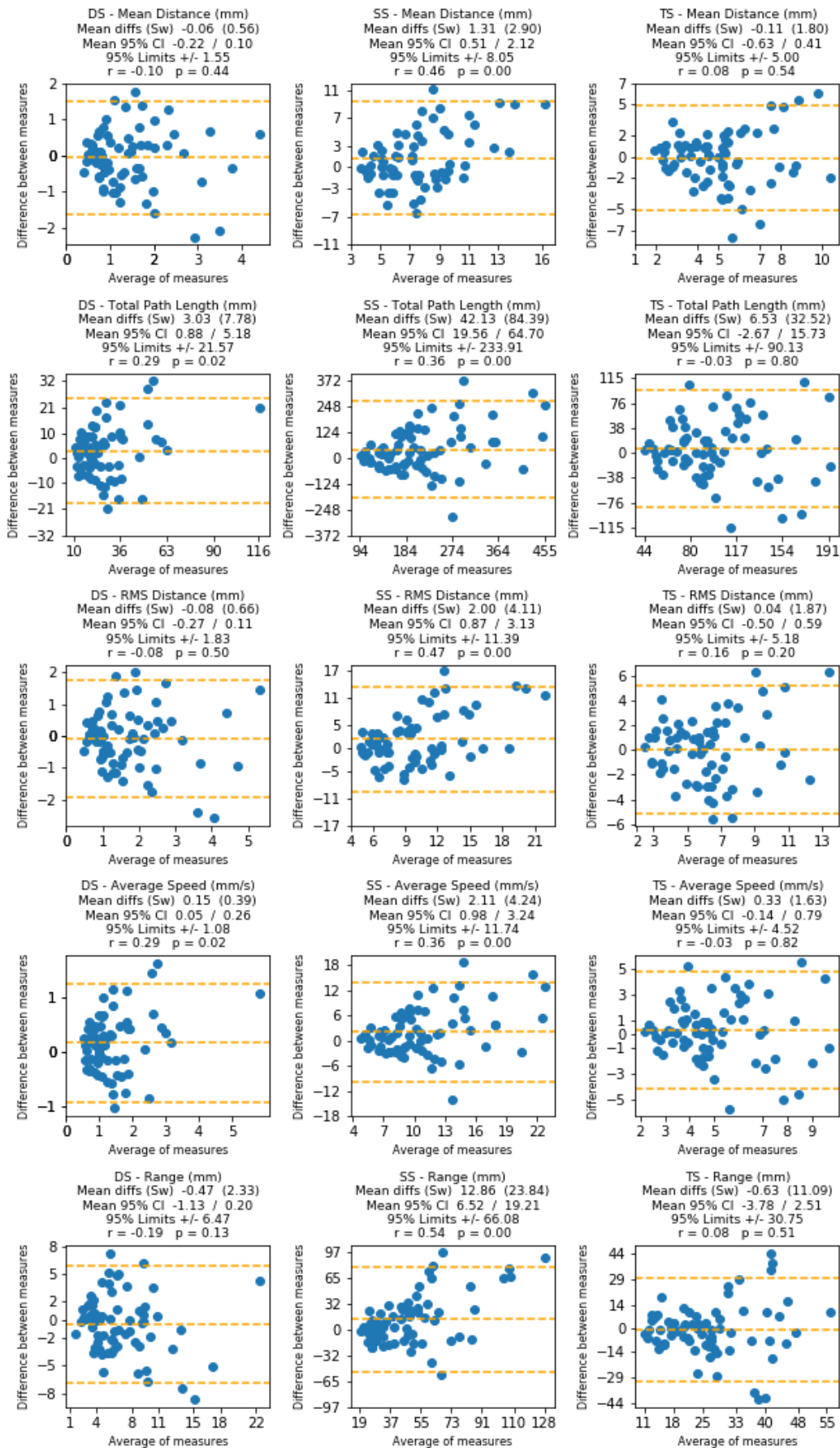
Figure 38 Individuals CoG medial-lateral properties. Bland-Altman analysis.
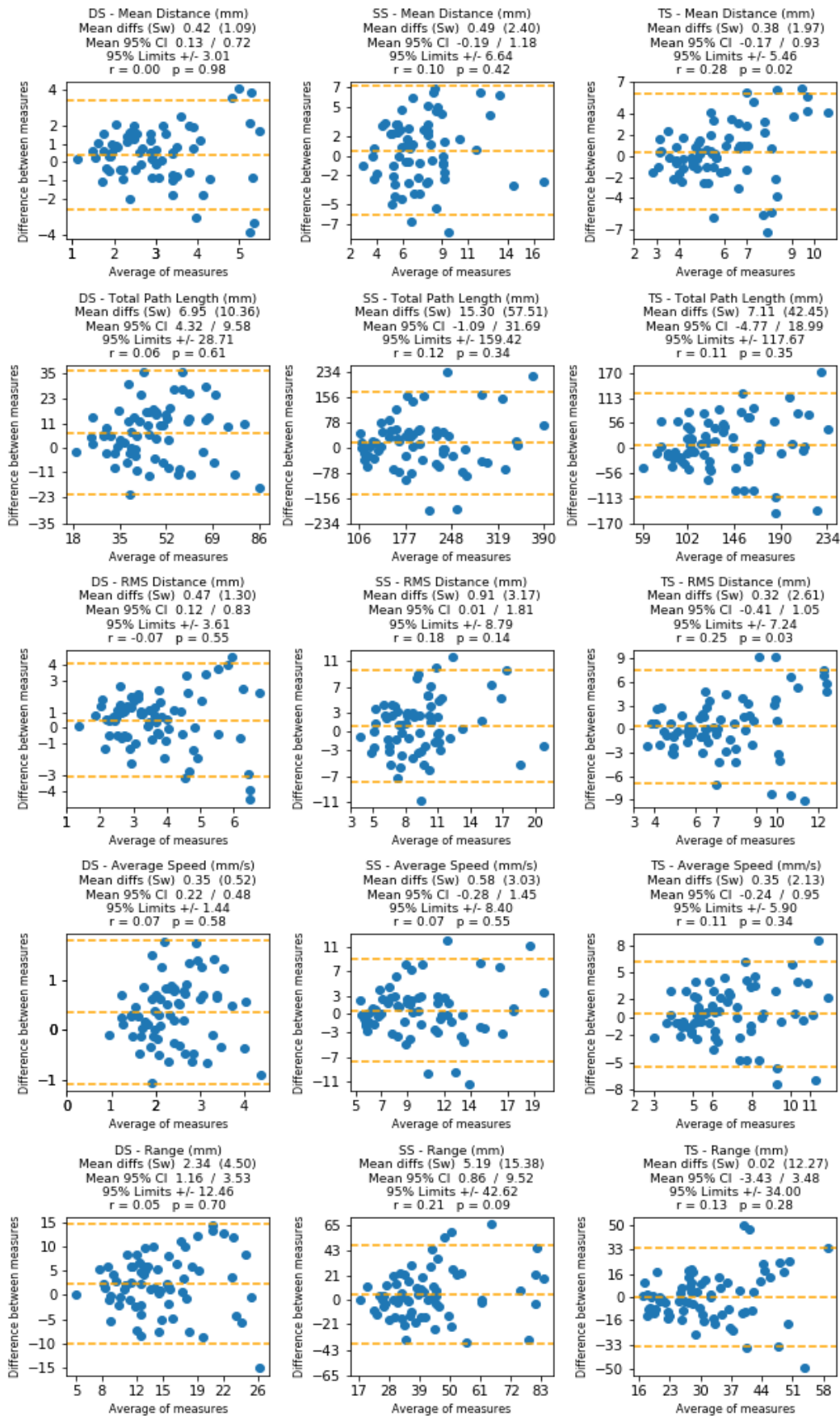
Figure 39 Individuals CoG anterior-posterior properties. Bland and Altman chart analysis.

6.11 Repeatability Coefficient

The difference between trials of TS resultant CoP Mean Distance was only 2% of great mean, while for DS reached 9% and SS 11%. However, calculating the percentage of the repeatability coefficient in relation to great mean, TS reached 125%. Yet, DS and SS got 61% and 58%, respectively (Table 40).

Each position and each plane should be analyzed individually. For resultant DS, SS and TS CoP properties (Table 40), the best relative coefficients were for TPL and Average Speed with the same scores (57%, 47%, and 60%, respectively for each leg support in both variables). The worst reliability was the estimated 95% confidence Circle Area (125%, 140%, and 160%, respectively for each leg position). CoP Ellipse Area had a better agreement than Circumference Area in all leg positions (DS – 98%; SS – 111%; TS – 149%).

CoP M-L TPL presented the better relative coefficients (DS – 58%; SS – 45%; TS – 61%) than A-P (DS – 66%; SS – 55%; TS – 54%) plane. Average Speed followed the same pattern (Table 41). Remembering that in TS, A-P values carry much of M-L body sway displacement.

CoG variables (Table 43 - Table 45) had the same behavior as CoP. However, relative repeatability coefficients (RC) were, with few exceptions, much higher than CoP, the mean relative differences between devices reached 42%, lower in resultant properties (34%) and higher in M-L plane (52%), and in A-P was about 40% in mean. Although, CoG performed better in A-P plane for DS TPL (RC = 60% versus 66%), in CoG M-L SS TPL was much higher than CoP (RC=109% versus 45%).

Table 40. Resultant CoP Properties. Analysis. Differences and repeatability coefficient analysis.

	Resultant CoP Properties	Mean Differences	Within-Subjects Standard Deviation	Great Mean	% Mean Difference of Great Mean	2x Repeatability coefficient	% Repeatability coefficient of Great Mean
DS	Mean Distance (mm)	0.42	1.01	4.79	9	5.59	58
SS		1.83	3.82	17.36	11	21.18	61
TS		0.23	2.74	12.20	2	15.19	62
DS	Total Path Length (mm)	31	46	227	14	257	57
SS		54	275	1632	3	1522	47
TS		23	221	1018	2	1226	60
DS	RMS Distance (mm)	0.56	1.31	5.65	10	7.26	64
SS		2.79	5.58	20.54	14	30.92	75

TS		0.10	3.45	14.11	1	19.12	68
DS		1.58	2.34	11.44	14	12.95	57
SS	Average Speed (mm/s)	2.68	13.82	82.20	3	76.61	47
TS		1.17	11.16	51.27	2	61.87	60
DS		1.49	3.21	13.09	11	17.80	68
SS	Range (mm)	10.43	25.40	59.67	17	140.81	118
TS		-1.58	10.29	33.40	5	57.05	85
DS	95% Conf.	0.91	1.64	3.65	25	9.10	125
SS	Circle Area (cm2)	8.02	23.33	46.23	17	129.33	140
TS		0.81	13.12	22.78	4	72.71	160
DS	95% Conf.	0.30	0.71	2.02	15	3.94	98
SS	Ellipse Area (cm2)	5.09	14.89	37.22	14	82.52	111
TS		0.94	9.50	17.63	5	52.64	149

Tandem M-L coordinates represent oscillations on A-P plane because the position of the legs on the platform. It was very difficult to isolate M-L and A-P movements because sensors were located in each corner of WBB. Therefore, movements directed in one plane were captured in some degree by the other plane.

Table 41. Individuals CoP Medial Lateral Properties. Differences and repeatability coefficient analysis.

	M-L CoP Properties	Mean Differences	Within-Subjects Standard Deviation	Great Mean	% Mean Difference of Great Mean	2x Repeatability coefficient	% Repeatability coefficient of Great Mean
DS		0.02	0.45	1.86	1	2.51	68
SS	Mean Distance (mm)	0.79	2.12	10.41	8	11.77	57
TS		0.01	1.88	6.81	0	10.45	77
DS		2.82	22.7	108.3	3	126.0	58
SS	Total Path Length (mm)	-7.16	177.0	1092.9	1	981.4	45
TS		-0.09	120.0	544.0	0	665.3	61
DS		0.03	0.58	2.37	1	3.24	68
SS	RMS Distance (mm)	2.03	4.38	14.13	14	24.29	86
TS		-0.13	2.73	8.78	2	15.13	86
DS		0.14	1.15	5.46	3	6.35	58
SS	Average Speed (mm/s)	-0.38	8.91	55.06	1	49.39	45
TS		-0.01	6.05	27.41	0	33.54	61
DS		-0.2	2.9	11.8	1	16.3	69
SS	Range (mm)	13.4	35.2	77.5	17	195.0	126
TS		-0.2	17.7	46.8	0	98.1	105

Table 42. Individuals CoP Anterior-Posterior properties. Differences and repeatability coefficient analysis.

Properties	Mean Differences	Within-Subjects Standard Deviation	Great Mean	% Mean Difference of Great Mean	2x Repeatability coefficient	% Repeatability coefficient of Great Mean

		Mean Differences	Within-Subjects Standard Deviation	Great Mean	% Mean Difference of Great Mean	2x Repeatability coefficient	% Repeatability coefficient of Great Mean
DS	Mean Distance (mm)	0.57	1.08	3.89	15	5.96	77
SS		0.67	2.80	10.69	6	15.52	73
TS		0.31	1.95	8.52	4	10.81	63
DS	Total Path Length (mm)	34.07	40.41	170.17	20	224.04	66
SS		60.58	188.85	955.82	6	1046.96	55
TS		10.97	142.76	729.73	2	791.40	54
DS	RMS Distance (mm)	0.68	1.33	4.92	14	7.36	75
SS		0.96	3.60	13.81	7	19.96	72
TS		0.33	2.53	10.74	3	14.03	65
DS	Average Speed (mm/s)	1.72	2.04	8.57	20	11.28	66
SS		3.04	9.51	48.15	6	52.70	55
TS		0.55	7.20	36.77	2	39.90	54
DS	Range (mm)	3.89	5.48	23.08	17	30.40	66
SS		6.16	22.86	78.13	8	126.73	81
TS		0.49	14.55	54.01	1	80.64	75

Table 43. CoG Resultant properties. Differences and repeatability coefficient analysis.

	CoG Resultant Properties	Mean Differences	Within-Subjects Standard Deviation	Great Mean	% Mean Difference of Great Mean	2x Repeatability coefficient	% Repeatability coefficient of Great Mean
DS	Mean Distance (mm)	0.33	1.12	3.68	9	6.18	84
SS		1.69	3.89	12.19	14	21.58	89
TS		0.13	2.76	8.68	2	15.29	88
DS	Total Path Length (mm)	8.64	15.68	63.05	14	86.93	69
SS		40.75	83.65	320.38	13	463.72	72
TS		10.45	56.17	182.36	6	311.39	85
DS	RMS Distance (mm)	0.33	1.31	4.27	8	7.28	85
SS		2.35	5.08	14.36	16	28.14	98
TS		0.06	3.25	9.98	1	18.03	90
DS	Average Speed (mm/s)	0.44	0.79	3.17	14	4.35	69
SS		2.04	4.20	16.09	13	23.31	72
TS		0.52	2.82	9.17	6	15.62	85
DS	Range (mm)	1.26	2.52	8.44	15	13.94	83
SS		10.44	20.09	36.33	29	111.37	153
TS		-0.55	8.45	21.02	3	46.83	111
DS	95% Conf. Circle Area (cm2)	0.40	1.37	2.14	19	7.62	178
SS		7.18	17.40	23.95	30	96.45	201
TS		0.59	9.30	12.02	5	51.58	215
DS	95% Conf. Ellipse Area (cm2)	-0.01	0.10	0.20	6	0.54	136
SS		3.02	6.01	7.61	40	33.33	219
TS		0.04	1.34	1.82	2	7.41	204

Table 44. Individuals CoG M-L properties. Differences and repeatability coefficient analysis.

	M-L CoG Properties	Mean Differences	Within-Subjects Standard Deviation	Great Mean	% Mean Difference of Great Mean	2x Repeatability coefficient	% Repeatability coefficient of Great Mean
DS	Mean Distance (mm)	-0.06	0.56	1.40	4	3.10	111
SS		1.31	2.90	7.55	17	16.09	107
TS		-0.11	1.80	5.01	2	10.00	100
DS	Total Path Length (mm)	3.03	7.78	27.31	11	43.13	79
SS		42.13	84.39	215.03	20	467.81	109
TS		6.53	32.52	99.40	7	180.26	91
DS	RMS Distance (mm)	-0.08	0.66	1.71	5	3.65	107
SS		2.00	4.11	9.93	20	22.78	115
TS		0.04	1.87	6.17	1	10.37	84
DS	Average Speed (mm/s)	0.15	0.39	1.37	11	2.16	79
SS		2.11	4.24	10.80	20	23.48	109
TS		0.33	1.63	5.00	7	9.04	90
DS	Range (mm)	-0.47	2.33	6.68	7	12.94	97
SS		12.86	23.84	48.41	27	132.17	137
TS		-0.63	11.09	26.66	2	61.50	115

Table 45. Individuals CoG A-P properties. Differences and repeatability coefficient analysis.

	A-P CoG Properties	Mean Differences	Within-Subjects Standard Deviation	Great Mean	% Mean Difference of Great Mean	2x Repeatability coefficient	% Repeatability coefficient of Great Mean
DS	Mean Distance (mm)	0.42	1.09	3.01	14	6.02	100
SS		0.49	2.40	7.41	7	13.29	90
TS		0.38	1.97	5.81	7	10.93	94
DS	Total Path Length (mm)	6.95	10.36	47.52	15	57.42	60
SS		15.30	57.51	200.23	8	318.84	80
TS		7.11	42.45	135.15	5	235.34	87
DS	RMS Distance (mm)	0.47	1.30	3.73	13	7.21	97
SS		0.91	3.17	9.37	10	17.57	94
TS		0.32	2.61	7.30	4	14.48	99
DS	Average Speed (mm/s)	0.35	0.52	2.39	15	2.87	60
SS		0.58	3.03	10.11	6	16.81	83
TS		0.35	2.13	6.79	5	11.79	87
DS	Range (mm)	2.34	4.50	14.60	16	24.93	85
SS		5.19	15.38	42.28	12	85.24	101
TS		0.02	12.27	31.57	0	68.00	108

6.12 Absolute reliability

Standard error of measurement (SEM) and minimal detectable change (MDC) were calculated based on original data for all properties. The relative SEM and MDC were calculated related to the mean of the two series means. First, we examined the properties for Manikins tests (Table 46).

Absolute and relative SEM values and respective MDC were small for most of CoP variables and larger for CoG variables. Except for CoP Range, all other CoP variable presented MDC% <= 4% and SEM% < 1.4%. With CoG variables MDC% varied from 15.2% (resultant TPL and Average Speed) up to 92.7% (Ellipse Area).

Table 46. Estimated Manikin tests standard error of measurement (SEM), minimal detectable change (MDC) and their relative values from mean for CoP and CoG variables.

Manikin Properties	Center of Pressure				Center of Gravity			
	SEM	SEM %	MDC	MDC%	SEM	SEM %	MDC	MDC%
Resultant								
Mean Distance (mm)	0.00	0.7%	0.00	1.9%	0.03	10.9%	0.08	30.3%
Total Path Length (mm)	0.45	0.6%	1.24	1.8%	0.79	5.5%	2.18	15.2%
RMS Distance (mm)	0.00	0.7%	0.01	2.0%	0.05	16.2%	0.15	44.9%
Average Speed (mm/s)	0.02	0.6%	0.06	1.8%	0.04	5.5%	0.11	15.2%
Range (mm)	0.03	4.5%	0.09	12.4%	0.11	13.9%	0.31	38.5%
95% Conf. Circle Area (cm2)	0.01	1.4%	0.03	3.8%	0.37	21.8%	1.03	60.6%
95% Conf. Ellipse Area (cm2)	0.01	1.4%	0.03	3.9%	0.30	33.4%	0.85	92.7%
Medial-Lateral								
Mean Distance (mm)	0.00	0.8%	0.00	2.1%	0.02	13.3%	0.06	36.9%
Total Path Length (mm)	0.47	0.9%	1.31	2.4%	0.76	8.0%	2.10	22.2%
RMS Distance (mm)	0.00	0.8%	0.00	2.2%	0.04	17.2%	0.11	47.8%
Average Speed (mm/s)	0.02	0.9%	0.07	2.4%	0.04	8.0%	0.11	22.2%
Range (mm)	0.03	2.7%	0.10	7.4%	0.13	12.4%	0.35	34.2%
Anterior-Posterior								
Mean Distance (mm)	0.00	1.4%	0.00	4.0%	0.02	9.5%	0.04	26.3%
Total Path Length (mm)	0.26	0.8%	0.72	2.2%	0.45	5.0%	1.25	13.8%
RMS Distance (mm)	0.00	1.4%	0.01	4.0%	0.02	9.7%	0.06	26.8%
Average Speed (mm/s)	0.01	0.8%	0.04	2.2%	0.02	5.0%	0.06	13.8%
Range (mm)	0.02	3.2%	0.07	8.8%	0.12	11.5%	0.32	32.0%

With Individuals, WBB measures for all leg positions showed lower relative SEM values in relation to Kinect measures (Table 47 - Table 49), consequently with lower MDC as well. The higher relative SEM values were associated with estimated confidence areas. Ellipse Area performed better than Circle Area in all leg positions. The lowest relative SEM and MDC values were associated with TPL and Average Speed.

Table 47. Individuals double stance tests estimated standard error of measurement (SEM), minimal detectable change (MDC), and their relative values from mean.

Properties Double Stance	Center of Pressure				Center of Gravity			
	SEM	SEM %	MDC	MDC%	SEM	SEM %	MDC	MDC%
Resultant								
Mean Distance (mm)	0.9	18.1%	2.4	50.2%	1.1	29.7%	3.0	82.3%
Total Path Length (mm)	41.0	18.0%	113.6	50.0%	12.5	**19.9%**	34.8	55.1%
RMS Distance (mm)	1.2	21.5%	3.4	59.6%	1.3	31.0%	3.7	86.0%
Average Speed (mm/s)	2.1	18.0%	5.7	50.0%	0.6	**19.8%**	1.7	54.9%
Range (mm)	3.4	26.0%	9.4	72.0%	2.6	31.0%	7.2	85.8%
95% Conf. Circle Area (cm2)	1.5	42.2%	4.3	116.9%	1.5	68.1%	4.0	188.7%
95% Conf. Ellipse Area (cm2)	0.6	27.7%	1.6	76.8%	0.1	48.6%	0.3	134.6%
Medial-Lateral								
Mean Distance (mm)	0.4	20.4%	1.1	56.5%	0.5	39.2%	1.5	108.7%
Total Path Length (mm)	18.7	17.3%	52.0	48.0%	6.6	24.2%	18.3	67.0%
RMS Distance (mm)	0.5	20.6%	1.4	57.1%	0.6	36.9%	1.8	102.4%
Average Speed (mm/s)	0.9	17.3%	2.6	48.0%	0.3	24.1%	0.9	66.7%
Range (mm)	2.4	20.2%	6.6	56.1%	2.1	32.0%	5.9	88.7%
Anterior-Posterior								
Mean Distance (mm)	1.1	29.0%	3.1	80.3%	1.2	39.2%	3.3	108.6%
Total Path Length (mm)	38.8	22.8%	107.4	63.1%	9.1	19.2%	25.4	53.4%
RMS Distance (mm)	1.4	28.3%	3.9	78.4%	1.4	38.2%	3.9	105.8%
Average Speed (mm/s)	2.0	22.8%	5.4	63.1%	0.5	18.9%	1.3	52.4%
Range (mm)	5.6	24.5%	15.7	67.8%	4.8	32.6%	13.2	90.4%

Surprisingly, M-L TPL and Average Speed of SS (Table 48) showed the lowest relative SEM (14.5% and 14.5%), and lowest relative MDC (40.2% and 40.1%) among all leg positions given the high instability of unipodal support. However, when measured by Kinect the lowest scores of resultant TPL and Average Speed for SS did not reach lower values than the ones in DS, 25.6% and 25.9% versus 19.9% and 19.8%, respectively.

Table 48. Individuals' single stance tests estimated standard error of measurement (SEM), minimal detectable change (MDC), and their relative values from mean.

Properties Single Stance	Center of Pressure				Center of Gravity			
	SEM	SEM %	MDC	MDC%	SEM	SEM %	MDC	MDC%
Resultant								
Mean Distance (mm)	3.9	22.4%	10.8	62.2%	4.6	37.8%	12.8	104.8%
Total Path Length (mm)	248.6	15.2%	689.2	42.2%	82.0	**25.6%**	227.2	70.9%
RMS Distance (mm)	6.0	29.3%	16.7	81.1%	6.1	42.4%	16.9	117.6%
Average Speed (mm/s)	12.4	15.0%	34.3	41.7%	4.2	**25.9%**	11.5	71.7%
Range (mm)	29.0	48.5%	80.3	134.5%	25.6	70.4%	70.9	195.1%
95% Conf. Circle Area (cm2)	28.0	60.6%	77.6	167.9%	21.6	90.1%	59.8	249.8%
95% Conf. Ellipse Area (cm2)	16.7	44.9%	46.3	124.4%	7.1	93.4%	19.7	259.0%
Medial-Lateral								

	SEM	SEM %	MDC	MDC%	SEM	SEM %	MDC	MDC%
Mean Distance (mm)	2.2	21.4%	6.2	59.3%	3.5	45.8%	9.6	127.0%
Total Path Length (mm)	158.3	**14.5%**	438.9	**40.2%**	94.7	44.1%	262.6	122.1%
RMS Distance (mm)	4.5	31.9%	12.5	88.5%	4.9	49.7%	13.7	137.6%
Average Speed (mm/s)	8.0	**14.5%**	22.1	**40.1%**	4.8	44.4%	13.3	123.0%
Range (mm)	37.2	48.0%	103.2	133.2%	27.8	57.4%	77.1	159.2%
Anterior-Posterior								
Mean Distance (mm)	3.2	30.1%	8.9	83.4%	2.8	38.0%	7.8	105.3%
Total Path Length (mm)	179.2	18.7%	496.6	52.0%	57.5	28.7%	159.4	79.6%
RMS Distance (mm)	4.1	29.9%	11.5	83.0%	3.8	40.3%	10.5	111.7%
Average Speed (mm/s)	9.0	18.7%	25.0	51.9%	3.2	31.2%	8.7	86.4%
Range (mm)	27.4	35.1%	76.0	97.3%	18.6	43.9%	51.5	121.8%

In general, TS task presented the higher resultant values than those on other leg tasks, and TPL and Average Speed were the most reliable variables (Table 49).

Table 49. Individuals' tests estimated standard error of measurement (SEM), minimal detectable change (MDC), and their relative values from mean for Tandem Stance task CoP and CoG variables.

Properties Tandem Stance	Center of Pressure				Center of Gravity			
	SEM	SEM %	MDC	MDC%	SEM	SEM %	MDC	MDC%
Resultant								
Mean Distance (mm)	2.8	23.3%	7.9	64.7%	3.2	36.8%	8.9	102.0%
Total Path Length (mm)	187.7	18.4%	520.2	51.1%	62.0	34.0%	172.0	94.3%
RMS Distance (mm)	3.7	26.3%	10.3	73.0%	3.8	38.3%	10.6	106.1%
Average Speed (mm/s)	9.5	18.5%	26.3	51.2%	3.1	33.9%	8.6	94.1%
Range (mm)	12.2	36.5%	33.8	101.1%	11.1	53.0%	30.9	147.0%
95% Conf. Circle Area (cm2)	15.2	66.7%	42.1	184.8%	11.2	92.9%	30.9	257.5%
95% Conf. Ellipse Area (cm2)	9.9	56.0%	27.4	155.1%	1.5	85.1%	4.3	236.0%
Medial-Lateral								
Mean Distance (mm)	1.9	28.0%	5.3	77.5%	2.0	39.1%	5.4	108.3%
Total Path Length (mm)	100.4	18.5%	278.3	51.2%	35.6	35.8%	98.7	99.3%
RMS Distance (mm)	2.8	32.0%	7.8	88.7%	1.8	29.7%	5.1	82.3%
Average Speed (mm/s)	5.1	18.5%	14.0	51.2%	1.8	35.8%	5.0	99.1%
Range (mm)	19.1	40.7%	52.8	112.8%	12.7	47.8%	35.3	132.5%
Anterior-Posterior								
Mean Distance (mm)	2.2	25.7%	6.1	71.2%	2.2	38.7%	6.2	107.2%
Total Path Length (mm)	119.4	16.4%	331.1	45.4%	47.3	35.0%	131.0	97.0%
RMS Distance (mm)	2.9	27.3%	8.1	75.6%	3.1	42.0%	8.5	116.4%
Average Speed (mm/s)	6.0	16.4%	16.7	45.4%	2.4	34.9%	6.6	96.6%
Range (mm)	17.8	33.0%	49.4	91.4%	15.5	49.1%	43.0	136.2%

Based on consolidated data of all leg tasks (Table 50), resultant CoP TPL and Average Speed reached relative SEM <6% and MDC of 16.3%. With CoG variables, values were still high, but also lower relatively to data analysed from each test condition.

Table 50. Standard error measurement and minimal detectable change based consolidated data series of all leg positions.

Properties All tasks	Center of Pressure				Center of Gravity			
	SEM	SEM %	MDC	MDC%	SEM	SEM %	MDC	MDC%
Resultant								
Mean Distance (mm)	1.1	9.9%	3.0	27.4%	1.6	20.0%	4.4	55.4%
Total Path Length (mm)	55.3	5.9%	153.3	16.3%	22.7	12.6%	62.8	35.0%
RMS Distance (mm)	1.4	11.1%	3.9	30.8%	1.8	19.7%	4.9	54.7%
Average Speed (mm/s)	2.8	5.9%	7.7	16.3%	1.1	12.6%	3.2	34.9%
Range (mm)	4.9	16.5%	13.5	45.6%	4.4	23.2%	12.3	64.3%
95% Conf. Circle Area (cm2)	5.3	25.2%	14.6	69.8%	3.0	32.4%	8.2	89.8%
95% Conf. Ellipse Area (cm2)	2.4	15.6%	6.8	43.2%	0.5	29.5%	1.3	81.9%
Medial-Lateral								
Mean Distance (mm)	0.6	9.5%	1.6	26.4%	1.0	22.2%	2.7	61.5%
Total Path Length (mm)	33.7	5.8%	93.4	16.2%	16.0	15.4%	44.4	42.8%
RMS Distance (mm)	0.7	9.7%	2.0	26.8%	1.1	19.7%	2.9	54.7%
Average Speed (mm/s)	1.7	5.8%	4.7	16.2%	0.8	15.7%	2.3	43.6%
Range (mm)	3.5	9.7%	9.7	26.9%	4.5	19.6%	12.6	54.2%
Anterior-Posterior								
Mean Distance (mm)	1.0	13.8%	2.9	38.3%	1.6	29.1%	4.3	80.5%
Total Path Length (mm)	35.8	6.0%	99.2	16.7%	15.1	12.6%	41.8	34.9%
RMS Distance (mm)	1.5	15.1%	4.0	41.9%	1.9	28.9%	5.4	80.2%
Average Speed (mm/s)	1.8	6.0%	5.0	16.7%	0.7	12.1%	2.0	33.4%
Range (mm)	8.0	15.9%	22.3	44.0%	6.9	24.3%	19.0	67.3%

7
Discussion

In this chapter, firstly, it is discussed the results reached in our experiment and secondly, the outcomes are compared to other studies.

7.1 Analysis of results of the experiment

The first contribution of this study was the use of a method for the selection of an optimal low pass filter cut-off frequency for WBB and Kinect signals based on residuals analysis as utilized in Salavati *et al.* (2009). Following the technique described in Winter (2009), comparing the residuals between filtered and unfiltered signals across a range of cut-off frequencies might determine the appropriate cut-off point. The optimal point would be where the difference starts changing very little. From this point, mostly noise and minimal signal would be filtered. Using this technique (Figure 19) our study found the lowest cut-off frequency among relevant reliability studies of balance tests using WBB (Clark *et al.* 2018), and probably with Kinect, as well. The chosen cut-off frequencies of our study were much lower than the ones recommended in the literature (Duarte & Freitas, 2010). However, the study demonstrated that different cut-off frequencies on a second order Butterworth filter has little or no influence in agreement indexes for CoP measures (Table 7). However, they may influence CoP and CoG properties when applying very low frequencies (< 4Hz).

The optimal low cut-off frequencies found in our study may reflect some limitations of these devices and the reported low signal to noise rate (Pagnacco *et al.*, 2011), what might cause loss of valuable information of higher frequencies in the CoP or CoM oscillations during the semi-static balance tests.

Images of color map histograms during tests with Manikins demonstrated the similarities between trials regarding CoP (Figure 20) and CoG (Figure 21) coordinates distributions. In the same way, the color maps showed marked differences between each performed task by Individuals, and not so clear, but

present, in test and retest trials for CoP (Figure 22) and CoG (Figure 23). It could be noticed in Manikins series histograms that WBB and Kinect magnitudes of variation were very small with amplitudes below 1mm (0.4mm, -0.4mm). However, regarding Individuals tests, captured variations by these devices were ranging in tenths of millimeters (40mm, -40mm).

Descriptive information of CoP and CoG coordinates for Manikin tests (Table 8 - Table 9) demonstrated very small differences between test and retest trials also seen in histograms, which was confirmed with no statistically significant differences between trials (Table 12 - Table 13). It was noticed similarities between trials when analyzing tests with Individuals if only mean or median values were considered given the small differences (Table 10 - Table 11). However, it was observed high standard deviations, and 25th, 75th percentiles, minimal and maximal higher differences between trials. Nonetheless, differences were not statistically significant neither for CoP (Table 14) nor for CoG (Table 15) series, except for CoG DS in A-P direction. As expected, there were no correlations between CoP and CoG coordinates of first and second trials given the lack of synchronism in body sway between trials.

The graphical analysis allowed observing associations and asynchronous signals behavior during test trials with Individuals (Figure 25 - Figure 30). The figures illustrated the great variability during the tests in all task condition, with lower variability in DS task, as expected. It could also be noticed an apparent reduction of variation during the second trials. Perhaps, due to learning effect.

A set of properties related to distances, speed and estimated movement area were derived from CoP and CoG coordinates. None of derived variables for the tests with Manikins (Table 16 - Table 17) and with Individuals was normally distributed. Two different transformations strategies were performed over Individuals variables in order to achieve normal distributions in the majority of variables (Table 18 - Table 23). It was not done with Manikins because the differences between trails were negligible and the reliability was clearly high.

Surprisingly, among the 12 reliability studies that used WBB cited in Clark *et al.* (2018) review, only one commented about performing a normality test. In this study (Jeter *et al.*, 2015), even though it was performed a log transformation on original data, it was not reported if with that, normality was met. Yet, following Clark *et al.* (2018) review, ICC was the most used test to report repeatability and

reproducibility. ICC is based on ANOVA assumptions that require parametric series and equality of variance between data sets (Weir, 2005). Reports of concordance indexes and confidence intervals without presenting evidences of data distributions, may cast doubts about the statistical power of found results.

Some other studies, not presented in cited review (Clark *et al.*, 2018), reported that normality was met (Martínez-Lemos *et al.,* 2017) and few others that it was not (Stadelman *et al.*, 2015). Nonetheless, violation of normality in large samples (> 30 or 40) may lead to minor problems, because distribution tends to be normal regardless the shape of data (Ghasemi and Zahediasl, 2012).

Bland-Altman´s graph analysis has been used in reliability studies (Watson & Petrie, 2010; McLaughlin, 2013), and normality of differences is also desirable to be checked (Hirakata and Camey, 2009; Giavarina, 2015). Bland-Altman (1999) made a suggestion of presenting data in reliability studies when data are not parametric, but not seen in literature.

None of the Manikin derived variables' differences between trials met normality (Table 24 - Table 25), whilst with Individuals, most of the variables showed normal distribution in residuals (Table 26 - Table 31). Probably, when considering biological variance and differences among individuals, differences tend to be normally distributed while with reliable devices, differences tend to zero generating a too skewed distribution.

CoP and CoG properties correlations between trails in Manikins tests were very high, with CoP outperforming CoG variables (Table 24 - Table 25). A customized Chi-Squared test also confirmed associations between Manikins trials series properties. The size of differences was about 20% of first measure value for most of cases for CoP variables covering more than 95% of the test cases. For CoG properties, about 80% of measured difference was necessary to cover 95% of the cases.

With Individuals tests, correlations of variables between trials were much lower than those with Manikins. CoP and CoG DS task showed the higher coefficients among other leg tasks. CoG variables correlations in other tasks were very weak or not existent (Table 26 - Table 31). However, correlations are not a good index for establishing concordance between variables (Chen & Barnhart, 2008), what was demonstrated in the graph analysis comparing linear best fit line to concordance line (Figure 31 - Figure 32). Some variables such as CoP Single

Stance Average Speed presented moderate correlations, but the lack of agreement could be seen by the differences from the best-fit line to the perfect concordance association line.

The intrinsic variability of WBB and Kinect could be observed analyzing the high coefficient of variance (CV) among all analyzed variables, achieving above 58% when testing static manikins (Table 32), and above 30% with Individuals tests in most of the cases for CoP (Table 33) and CoG (Table 34) variables. Moreover, coefficients of variation rarely met in mean between test-retest trials. Higher Manikins CV values may be explained because mean differences were very close to zero and little variations in standard deviations cause great impact in the coefficient. CoP variables had, in general, lower coefficients of variation when compared to CoG ones.

Nevertheless, even with high CV, similarity tests performed to check equalities of CoP and CoG variables between the three different leg support tasks reported very low probability by chance (p<.001) that any property be similar to correspondent property in a different leg position (page 94). Roughly, demonstrating a degree of sensibility in both devices to distinguish different tasks. Additionally, properties' values differences between participants and even between two tests of the same participant performing the same task were evidences of devices sensibility, since it was expected that an individual would not be able to reproduce the same test twice without any difference.

Another contribution of this study was that properties related to CoP outperformed, in terms of reliability, CoG properties. Tests with manikins, which isolated device error because within-subject variation was negligible, demonstrated almost perfect relative reliability measured by CCC and ICC tests for all CoP properties and lower scores for CoG variables (Table 35). However, someone may argue that data series with Manikins were not parametric. Nonetheless, looking at the ICC scores for Individuals tests that were performed on parametric data series, it was noticed that CoP variables received higher scores than correspondent CoG ones, except for resultant DS. Still, in this case, with similar results (Table 36 - Table 37). CCC scores confirmed CoP superiority over CoG (Table 38 - Table 39). The poorer performance of CoG variables, reflected in larger differences between trials, could be explained because of the calculation of whole body Center of Mass requires the position of several body joints, and the variability of Kinect in

providing each joint position, frame by frame, using a deep decision forest algorithm might have increased the total variability of the measure. On the other hand, WBB, for calculating CoP position only needs to provide four load sensors values and uses standard distances of platform dimensions.

Absolute reliability measured by relative Minimal Detectable Change (MDC) in Manikins demonstrated very low scores, in the excellent range (<10%) for all CoP variables except for resultant Range (12.4%) that still was in a reasonable level (<30%) (Table 46). For CoG variables (Table 47), only TPL and Average Speed were in the reasonable range in all directions. A-P Mean Distance and A-P RMS Distance were also within the reasonable range, but all other variables showed poor absolute reliability (MDC >30%).

Our study reinforced that CoP and CoG Average Speed and TPL present the best relative and absolute reliability in almost all tests and all directions among variables, achieving higher $ICC_{2,1}$ scores (Table 36 - Table 37), and lower relative MDC values (Table 47 - Table 50). In agreement with previous studies (Ruhe *et al.,* 2010), CoP Average Speed was the most reliable property.

Additionally, between resultant Circle and Ellipse estimated areas, the latter performed better with CoP and CoG except for Manikin and Individuals CoG Tandem Stance. Highlighting that only CoP DS Ellipse Area had a more consistent relative reliability, $ICC_{2,1}$ ranging from fair to excellent score (ICC=.69, 95% CI=.52, .0.80).

Furthermore, among CoP and CoG Mean Distance, RMS Distance and Range, the former demonstrated better relative reliability in most of the cases. Surprisingly, Range outperformed the other two variables in CoP Double Stance Medial-Lateral and Anterior-Posterior. The International Society of Posturography recommended the use of RMS Distance as one of the two measures to standardize evaluations of postural steadiness (Prieto *et al.,* 1996). The other one was Mean Velocity[11].

Finally, resultant CoP Total Path Length $ICC_{2,1}$ in our study showed scores ranging from .36 up to .76 in agreement with previous studies .27 - .997 (Clark *et al*., 2018). It reached better agreement in tandem position task with ICC = .64 (95%

---

[11] The calculation of mean velocity is based on center of pressure or center of mass distance, not displacement, and then, being a scalar not a vector unit. Then, our study named this property as Average Speed.

CI= .48 - .76). All positions demonstrated very similar ICC scores: double stance, ICC = .61 (95% CI=.36 - .76), whilst in single stance, it achieved ICC = .59 (.42, .73). Nonetheless, when verifying consolidated information from all test conditions, ICC scores for CoP TPL achieved .95 (95% CI=.92, .97), within the excellent range. It suggested that when different tasks conditions were performed, analyzing consolidated data series might provide more reliable information.

Tests with static manikins allowed measuring intrinsic Standard Error of Measurement (SEM) of devices. Almost all CoP variables demonstrated low relative SEM to mean values <=4%, except for Range (Table 46). Since Range refers to only two data points, its value for representing the whole balance test is questionable (Palmieri *et al.,* 2002). For Manikin CoG properties, relative SEM were higher with A-P TPL and Average Speed having the lowest scores (5%). Most of the CoG variables scores were between 5% and 22%. Much higher relative SEM found during tests with Individuals in our study can be attributed to test procedures and participants' variation (Table 47 - Table 50).

Analyzing Individuals SEM and MDC based on consolidated data of all leg support tasks, these indices decreased significantly (Table 50). For example, the relative lowest CoP SEM and MDC with different leg position were 14.5% and 40.1%, respectively obtained during single leg task for M-L Average Speed. However, considering all leg conditions the lowest values decreased to 5.8% and 16.2%, respectively, for M-L Average Speed. These values were comparable to previous studies that used mean or median of repeated trials.

In conclusion, our experiment could capture interesting quantitative information about body sway during tests with well-defined relative and absolute reliability indexes. Devices demonstrated sensibility to capture coarse differences during Manikins tests compared to Individuals tests. Besides that, it was possible to distinguish Individuals tasks and provide quantitative parameters in order to monitor balance sway along time. Results demonstrated that when using consolidated data from all tasks and the most reliable variables, it was possible to achieve excellent levels of relative reliability with high ICC and CCC scores. However, absolute reliability was satisfactory for the most reliable variables only for CoP data.

7.2 Analysis of results in comparison to other studies

Yang *et al.* (2014) investigated the reliability and validity of Kinect v1, as described previously on page 42. The study did not find differences in mean values or 95% confidence intervals (CI) between test-retest trials, differently from our study that only had similar result with RMS property in SS task. Their study found that estimated 95% CI of ICC $_{2,1}$ for double stance with feet apart task was .76, .99 for RMS, and .84, .99 for Average Speed, while for feet together RMS = .23, .94 and Average Speed = .68, .98, for single leg stance, RMS= .25, .94, and Average Speed = .53, .97. In our study, RMS for double stance test, 95% CI of $ICC_{2,1}$ was .29, .65 and for Average Speed = .44, .82, while for SS task, RMS = .05, .49 and Average Speed = .30, .67. All measures in our study had remarkable lower reliability scores. There were differences in test durations and conditions, our tests were performed with eyes closed which may had increased within-subject variability what partially may explain differences between trials and consequently leading to lower reliability scores.

Yeung *et al.* (2014) also assessed reliability and validity of Kinect v1 (see page 42). Results in their study for double stance with eyes closed task on M-L and A-P Range were ICC $_{2,1}$ = .27, .37, respectively. In our study, these values for the same variable were ICC =.58, .44, respectively. Differences may be attributed to characteristics of Range property, as mentioned before. Absolute reliability was also verified in their study, presenting coefficient of variation scores, which was based on typical error (standard deviation of differences over square root of 2, and the coefficient of variation equals to typical error over mean, times 100) and equaled to 23.9%, 21.9%, for Range in M-L and A-P directions, respectively. Although, calculated differently, in our study, the correspondent coefficient would be the relative SEM property and our results were higher than those presented in their study (M-L and A-P Range, 39.9%, 28.6%).

Although some studies examined reliability of Kinect Xbox 360 during balance assessment (Clark *et al.,* 2015; Otte *et al.*, 2016), we could not find any study that checked reliability of Kinect Xbox One (v2) with repeated measures that estimated CoM using the kinematic segmental model. Eltoukhy *et al.* (2018) investigated Kinect v2 using a similar kinematic method for estimating CoM as in our study, but analyzed the reliability between Kinect and VICON system. The

difficulty to compare Kinect results of our study with others, revealed another contribution of our work providing a rich set of absolute and relative reliability information for the most used posturographic variables.

On the other hand, a reasonable number of studies investigated the reliability of WBB during balance tests. Chang *et al.* (2014) performed similar set of test tasks as in our study with 30 young healthy adults. They found that tandem stance task had the worst reliability index ($ICC_{2,1}$ = .27) among the three tested leg positions. Opposite to our study. For the other leg stances tasks, single leg had the same score as in our study (ICC= .59) and a little higher score for the double stance task (.68). Moreover, using consolidate data in their study, ICC achieved .88 very similar to our result of .90 (CI =.86, .92). Although, 95% confidence intervals were not reported in their study, in general, the classifications of relative reliability agreed between the studies. The differences in tandem position may be due feet positions on the WBB. In our study, subjects stood in the diagonal of the board, whilst in Chang *et al*. (2014) work there was a longer wood board over WBB extending the support area, and participants stood aligned to anterior-posterior direction.

Scaglioni-Solano *et al.* (2014) investigated test-retest reliability with $ICC_{2,1}$ in 20 healthy participants in a double stance with eyes closed task, on a WBB, among other test conditions. Results demonstrated poor to excellent reliability .64 (95% CI = .39, .78) for that condition, what, in great extent, agreed with our results (.61, 95% CI=.36, .76). However, it was very difficult to compare reliability results because study designs are very distinct. Scaglioni-Solano study performed three tests in five different conditions with arms by the side of the body for 30s except for tandem position that lasted 10s. Moreover, tests were performed in two different days and ours, in the same session. On the other hand, consistent reliability results presented by different study designs, may suggest the actual device reliability.

Clark *et al.* (2010) tested WBB reliability during double and single leg stance with eyes open and closed tasks, with hands positioned on the hips, similar to our test, in 30 young participants, in two different days. Tests had different durations when compared to our design. Double stance tests were recorded for 30s and single stance for 10s. The study used the median of three tests in each day, reaching $ICC_{2,1}$ = .91 (95% CI=.80, .96) for DS and .81 (95% CI=.59, .91) for SS when analyzing TPL property, a better coefficient than ours. Perhaps, the use of

median of three trials decreased within-subject variability and increased ICC scores. Participants were also younger than in our study (23.7 +/- 5.6 versus 28.8 +/- 8.7 years).

Liang et al. (2014) assessed reliability of WBB in 48 healthy individuals, utilized $ICC_{2,2}$, during 2 successful trials of 60s in double stance and feet hip-apart, with results varying from .58 to .88 on time-domain variables. On M-L plane 95% confidence intervals for Range (.29, .82), Mean Velocity (.68, .95), and on A-P plane, Range, .58, .93, and Mean Velocity .55, .94. It was curious that comparing with our results where M-L double stance task values ranged from fair to excellent (95% CI=.49, .78) (Table 36), although still lower than Liang´s study in the higher end, but much higher in the lower end of the confidence interval (.29 versus .49). However, in the A-P plane, our scores ranged from poor (.16) to fair (.74), while their study varied from fair to excellent.

Jeter *et al.* (2015) assessed WBB reliability with corrected-to-normal visual (N=21) and visual impaired subjects (N=14), tests lasted 30s each, and arms were crossed around the chest in a double leg position. The study utilized the mean of three trials and $ICC_{2,k}$ with outcomes varying from .78 up to .91 depending on analyzed variable (A-P and M-L Standard Deviation of amplitude, velocity, and mean total velocity), and participant condition. All results were in the excellent range differently from our study that only achieved excellent range using consolidated data.

Park *et al.* (2014) verified intra-rater reliability in 20 healthy adults in single and double leg stances with eyes open and closed. Mean of three repetitions was used for calculating reliability indexes on CoP TPL and CoP Average Velocity. Single leg data was obtained for 10s and double leg for 30s. With eyes closed double leg ICC = .89, .89 for TPL and Average Speed, and for single leg .86, .87 for the same variables, respectively. Although in our study TPL and Average Speed got the higher scores they were not in the same level as in Park *et al*. (2014).

After examining all these studies results and methods, probably, taking the mean or median of three repetitions really reduces intra-subjects variation in great extent, what increases agreement in test-retest trials.

Larsen *et al.* (2014) investigated repeatability of WBB during a set of tests in 54 children (10-14 y-o). They perform three rounds of four tests, each one lasting 30s. The examined variable was CoP TPL and one of the tests was with double leg

support and eyes closed. The relative reliability measure was CCC with scores of .76 (95% CI=.65, .88), and relative MDC of 26.5%. Our CCC score for a similar task (Table 38) with shorter duration (20s) test with adults was .60 (95% CI=.44, .73), quite lower than theirs. Our relative MDC was 50%, almost the double achieved in their study. We did not find other similar studies that had used CCC to compare with our results.

Bower *et al.* (2014) assessed balance in 30 post-stroke survivors using WBB and got relative Minimal Detectable Change values for mean speed during double leg stance test with eyes closed for 30s of about 35.1%, 35.2%, and 44.4% in resultant, A-P, M-L planes, respectively. Our study got 50%, 63% and 48% in the same variable, much higher scores for the same test (Table 47). However, in our study, absolute speed values were approximately half of the values obtained in Bower´s study.

Martínez-Lemos *et al.* (2017) also verified relative MDC for CoP Total Path Length and estimated Ellipse Area assessing subjects with different levels of intellectual disabilities using WBB. For the whole group (n=48), MDC was 29.3% and 33.25% for those variables. The single trial consisted of staying in a double-leg stance with eyes open for 30s, feet position was registered for consistency in a repeated trial, and a minimal sampling rate was set at 50Hz. These results were also lower if compared to those in our study for correspondent variables and leg position, 50% and 76.8%, respectively. Still, that study, utilized a formula based on the standard deviation of differences for calculating MDC, while our study used the within-subject standard deviation. Post-hoc calculations using the same parameter showed a lower relative MDC for CoP TPL and Ellipse Area (43.9%, 73.3%, respectively) in our study, but still higher than theirs. Formulas may still differ when using different ICC sources. In our study, we utilized the found ICC score of our tests, using transformed data, these scores differed from other studies as well.

Clark *et al.* (2010) verified relative MDC for resultant TPL and found values 24.5% for DS, 28.3% for SS, lower in comparison to those found in our study. It was not clear which standard deviation and agreement index was used for calculating SEM, which is the basis for MDC. Beyond that, participants performed three successful trials in each of the two visits, whilst in our experiment subjects performed only one trial (successful or unsuccessful) for test and another trial for retest in the same day. As mentioned before, utilizing only one trial should have

increased variability in our experiment. One evidence of that can be seen in Scaglioni-Solano *et al.* (2014) study, which found significant differences among three repeated tests with eyes closed in a test-retest comparison. However, when taking the mean value of the three repetitions of each trial, the differences disappeared.

Therefore, increasing the number of repetitions and taking the mean value of repeated trials, reduces the variability between test-retest as stated by (Ruhe *et al.,* 2010) after performing a systematic review on test-retest reliability during bipedal task conditions. Literature recommends two to four trials for minimizing concordance variance as well as for not suffering a test learning effect or reaching participant fatigue (Duarte and Freitas, 2010; Shoukri, Asyali, and Donner, 2004). Furthermore, Hopkins (2000) suggested that for achieving reasonable reliability results it was necessary about 50 participants and at least 3 trials.

However, in our study we wished to replicate a real balance test situation in high demanding sets where time is an issue. Considering that, replicating two or three times the same procedures does not seem to be an efficient way to evaluate balance in large groups. Besides that, in our procedures we tested three different leg support conditions, and if considering the derived properties based on all conditions, ICC scores achieved much higher values. All CoP and the majority of CoG properties reached the excellent range, above .75, especially resultant CoP TPL (.96). The effect of higher between-subjects variation inflates ICC values. Three trials, even in different positions, in mean, increases within-subject variability, reducing random error and then, achieving higher ICC scores.

It is argued that high variability in participants, like in elderly, might lead to high ICC values, and it may mask poor test-retest consistency (Ruhe et al, 2010). Between-subject variability is not part of the error term. Then, it favors increased ICC values (Weir, 2005). Monteiro-Junior and colleagues (2015) assessed WBB reliability in a group of 21 old women (64 +/- 7 y-o) who practiced aqua-aerobics two to three times a week. Participants attended two separate sessions when they performed three times a double leg test with eyes open and closed. Mean values were considered for calculations of four variables: CoP total displacement, mean velocity and standard deviation in M-L and A-P directions. For all properties during eyes closed tasks, ICC was greater than .98 and p < .01, showing almost perfect relative reliability. Although not published, in an e-mail contact with authors, it was

clarified that normality was checked, series were parametric, $ICC_{3,1}$ model was applied, and each test lasted 1 minute, but only the middle 30s were considered. Taking the published means and standard deviations of trials with eyes closed for calculating the respective coefficients of variation (CV), it was noticed a high between-subjects variability for analyzed variables: 50%, 51%, 63%, and 52%, respectively. Comparing with our correspondent CV results, 36%, 36%, 57%, and 34%, and our ICC values (.61, .61, .65, and .45), it may justify in part, the inflated ICC values achieved by that study. The choice of $ICC_{3,1}$ also does not account for systematic error, which also mask test and retest constant error and bias (Weir, 2005).

Severini *et al.* (2017) used WBB to perform posturographic analysis in healthy subjects and with multiple sclerosis (MS) patients during tasks with eyes open and closed. In this study, tests with MS subjects and eyes closed got the higher agreement scores, given support to the concept that the higher the variability of participants' condition and task, the higher the obtained ICC value, which does not reflect necessarily, repeatability quality.

Additionally, there is no consensus in literature about the ideal time duration for capturing body sway signal in postural control tests. Kapteyn *et al.* (1983) recommended start recording 10 seconds before the time of analysis, which should be, preferably, 50s, in order to eliminate transient phenomena. It was suggested by the authors, that enough time is required for discarding initial records and remaining a usable period for valid analysis. Despite this recommendation, the authors emphasized that there was no evidence for an optimal recording duration. Time duration would depends on task parameters (Duarte and Freitas, 2010; Ruhe *et al.*, 2010). Our test procedure initiated recording 3 seconds with open eyes in each position as standstill as possible, before starting data collection that was used for analysis. This may have reduced the transient phenomena, even though there was a change in visual stimuli. Beyond that, before more challenging leg support bases (single and tandem leg stances), it was required that participants got used to closed eyes condition, what also might have reduced the variation when starting the tests.

As reported by Martínez-Lemos *et al.* (2017) few studies with WBB informed relative SEM scores, only reporting absolute values. As mentioned by authors, relative values allow comparisons with other studies because they are dimensionless and easier for interpretation. Although, as we have evidenced, there

may be a significant difference in using distinct sources for feeding the formulas for calculating SEM. Compared to Martínez-Lemos' study SEM% scores 28% and 33% for CoP TPL and Ellipse Area, our study registered for the same variables, 18% and 28%, respectively. However, when the formulas were equalized, our adjusted values were 16% and 26%, respectively. Still, comparing to Park *et al.* (2014) study that reported 4.9% and 8.5%, for CoP TPL in DS and SS tasks, respectively, these scores were much lower than in our study (18.0% and 15.2%). There are different methods to calculate SEM (Cook *et al.*, 2014), what may mislead comparisons when formulas are not clearly reported.

From the analysis of differences with Bland-Altman charts, it was possible to assess another absolute reliability index based on the repeatability coefficient and the limits of agreement. Relative SEM and MDC could be compared to relative mean differences and limits of agreement. Relative SEM scores ranged from 14.5% in Single Stance M-L TPL up to 66.7% got in Tandem Stance Circumference Area, and they were much higher than relative mean differences from 0.02% in TS M-L TPL up to 25% in DS Circumference Area between trials over mean measured values. One possible explanation is that SEM is based on squared differences, while simple differences cancel each other in large samples.

On the other hand, relative MDC values were lower than repeatability coefficients for TPL or Average Speed in all leg positions (MDC = 50%, 42%, 41% and RC = 57%, 47%, 60% for DS, SS, TS respectively for both properties). TPL and Average Speed had the same scores because Average Speed represents TPL over test duration. Minimal detectable change and limits of agreement are supposed to indicate the thresholds to consider a significant change in performance in repeated trials. For example, using limits of agreement for Double Stance Total Path Length, a next trial in our study would result in -31 +/- 129 mm of previous value and using MDC +/- 114 mm based on last individual measure. The mean value for this variable in our study was 227 mm. A change in base of support area, such as from DS to SS, has a large effect in Total Path Length property. The found mean value for SS resultant TPL was 1632 mm and for TS, 1018 mm, easily detectable using any of those parameters (MDC or LOA). Regarding Manikins tests, MDC for this variable was only 1.24 mm in agreement with Bartlett *et al.* (2014) that found a sensibility of 1.5mm in WBB and suggested that the device would be able to detect differences of 10mm, what would differ a healthy from an unhealthy status.

Someone may argue that TPL has an absolute value that depends on duration of the test. However, if it was selected the resultant RMS Distance, MDC was 3.4, 16.7, and 10.3 mm for DS, SS, TS tasks, and RC, 7.3, 30.9, 19.1 mm, respectively. The unique task theoretically able to detect such differences would be double stance (values < 10mm) following our procedures. It seemed that Kinect and WBB having large limits of agreement would not detect subtle changes in performance.

Furthermore, a recent study (Sempere-Rubio *et al.,* 2018) has demonstrated that WBB was able to distinguish healthy subjects and patients with fibromyalgia investigating estimated Ellipse Area and RMS Distance. Differences between groups were statistically significant, especially Ellipse Area that was 2.03 times larger in average in the patient group. Nevertheless, distinguishing averages in groups may be useless as a clinical parameter when it is not accompanied by sensibility and specificity rates. For example, in our group we would have detected about 15% of our participants with fibromyalgia using this parameter (2.03 times the Ellipse Area of body sways from group mean value). Howsoever, it does not put down the merit of this valid use of WBB replacing a high-cost force platform.

Few CoP properties presented significant moderate correlations of residuals and observed values. Single stance M-L RMS Distance (r = .44), double stance A-P TPL and Average Speed (r = .41) were the variable with highest coefficients. Weak correlations (r < .40) were found in resultant DS TPL and Average Speed and resultant SS RMS Distance and Range, while all other CoP variables presented no significant correlations (p < .05) (Figure 34 - Figure 36). For CoG properties (Figure 37 - Figure 39), considerable correlations were found in single stance task in resultant and M-L planes, such as Ellipse Area (r > .53), and weak correlations in some TS variables. It means that there was a tendency that errors may grow with growing measures and the presented limits of agreement would be not respected in higher observed values.

In addition, systematic error pointed out in some variables and observable in Bland-Altman charts (Figure 34 - Figure 39) could be seen in simple similarity tests between trials (Table 26 - Table 31). For example, all CoP resultant, A-P DS variables and SS Mean Distance, RMS Distance and Range showed significant differences between test-retest conditions (p < .01). Exploring mean differences, it was noticeable that those properties showed higher relative mean differences compared to other leg position tasks. These systematic errors may be due learning

effect or the participant´s desire to perform better, considering that mean 95% confidence intervals of most variables achieved lower values in retest condition when compared to first trials.

Mean differences and limits of agreement may not represent ideally the amount of agreement between test-retest trials. For example, analyzing Individuals CoP variables, the differences related to great mean (mean of mean measures) for all CoP resultant properties, TS task had apparently the best agreement, with differences between trials <= 5% of great mean. However, TS had the highest relative within-subject standard deviations (within-subject standard deviation over mean's differences) evidenced in the highest relative repeatability coefficients related to great mean (Table 40), except when compared to resultant SS Range. Therefore, with huge variation in differences it is advisable to investigate the differences in relation to the relative repeatability coefficient.

Following the percentage of repeatability coefficient (RC) related to great mean WBB also outperformed Kinect. CoP variables scores, in mean, were 42% lower than CoG correspondent variables. The lowest relative RC value was 45% for CoP M-L SS Total Path Length (Table 41) while the lowest CoG relative RC score was 60% for DS A-P TPL and Average Speed (Table 45), characterizing the differences between CoP and CoG variables.

This study presented some limitations regarding the adopted procedures such as performing only 1 test repetition per task. Following Ruhe *et al.* (2010) revision, test-retest reliability would suffer more influence by the number of trials and duration than the selection of properties to be analyzed. Additionally, in our experiment, unsuccessful trials were accepted with few exceptions when the participant spent most of the task time recovering from a loss of stability. Besides that, in most of the locations, it was allowed that waiting participants watched another participant execution what might have influenced performance during the tests. During the first test condition it was not put any marker on feet positions in order to remember and reproduce it during retest condition.

Another limitation was the chosen sampling rate of WBB that was equalized at 30Hz, the maximum sampling rate allowed by Kinect and maybe below the ideal sampling rate to explore the whole potential of WBB.

In future work, we may replicate procedures using stricter rules for getting better general agreement between trials. For example, it was allowed participants

wear shoes or not since they kept their choice in the retest trial. It was not marked feet positions between test-retest trials. Many tests were performed in open spaces with external noise and other participants watching test procedures of colleagues. It was considered also some unsuccessful trials with participants stepping out of the platform and returning to position. All of these may have influenced reliability outcomes. Furthermore, regarding experimental design, we may exclude tandem stance position, in general, participants' feet did not fit in platform space, and it caused occlusion of body parts using Kinect. Additionally, WBB sampling rate could be adjusted to 50 or 60 Hz. Finally, it should be asked to participants to repeat the test two or three times, and consider the mean out of those repetitions for reliability comparison. Regarding software implementation, algorithms predicting future test results may be applied, as well as, applying different algorithms to stablish differences between data series, such dynamic time warping. Other algorithms could be used to estimate CoM such as statically equivalent serial chain as proposed by Gonzàlez *et al.* (2015). Finally, high order of Butterworth filter analysis may be applied for reducing noise influence and to compare the effect of different cut-off frequencies on reliability indexes.

8
Conclusions

The use of affordable, portable and reliable devices for evaluating balance skills is extremely relevant considering that world population is getting older and the number of diseases related to impairments in balance control. The experiments validating the use of WBB and Kinect in clinical applications aimed at assessing body balance have been indicating the benefits and restrictions of those devices. Relative and absolute reliability of WBB and Kinect seemed to be specific depending on the studied population, task, evaluated variables, and applied procedures. Therefore, this study assessed the reliability of WBB and MS Kinect simultaneously during a balance test composed of three different leg support tasks in healthy individuals. The obtained information from those devices derived Center of Pressure and Center of Gravity locations, respectively. Then, the most used time-domain properties related to posturography were calculated and a deep examination of tests to check repeatability were applied on those properties. Test procedures were replicated using static manikins and persons. Furthermore, in order to enhance ecological validity and explore the claimed portability advantage of the devices, test and retest trials with subjects took place in different locations.

In additional to the novelty of our investigating methods, it was utilized a procedure based on residuals analysis to determine the low pass filter cut-off frequency and the found frequencies values were lower than most of posturographic studies using Kinect and WBB. The found optimal low cutoff frequencies might indicate that those devices were not appropriate to capture significant oscillations in body sway frequencies that were estimated to be up to 10Hz. However, our tests also indicated that filter selection ranging from 4Hz to 12Hz had little or no influence in CCC and ICC scores.

Quiet balance tests applied in static manikins and individuals with different leg supports demonstrated that information provided by WBB is more reliable than

that provided by Kinect using estimated center of mass calculated using a kinematic method based on body segments.

As far as we know, this study was the first to examine reliability with repeated measures of the CoM calculated through the kinematic method during three different balance tasks using Kinect v2. Our tests demonstrated that depending on task, direction and analyzed variable, relative reliability varies from poor to excellent. However, when considering the consolidated information from all tasks, the great majority of variables showed excellent ICC scores (>.75).

In agreement with literature, Average Speed, also called Total Mean Velocity, showed the best relative and absolute reliability scores in all conditions and planes when compared to other CoP and CoG analyzed variables.

The study also contributed providing a vast set of data using different methods to test relative and absolute reliability with repeated measures, mainly with dimensionless variables, allowing the comparison across different studies.

Published literature about test-retest reliability using WBB and Kinect is not clear about normality checking and consequently the statistical power of presented outcomes. Even for absolute reliability, the proposed methods for calculating SEM are based on standard deviations and ICC scores, or on the square root of the mean square obtained from 2-way ANOVA, in all cases, requiring normal distribution of data to infer confidence intervals casting doubts about statistical power of their results.

Partial description of methodological procedures and formulas of previous studies hinders the comparison and analysis of results. Information about statistical assumptions, use of confidence intervals, use of the appropriate models for assessing reliability indexes following the study design, and clarification of sources of data used in variables for calculating absolute reliability properties were issues faced reviewing specialized literature.

In addition, this study presented absolute reliability variables such as SEM and MDC. These variables demonstrated higher values when compared to previous studies and, theoretically, unacceptable for clinical use. Furthermore, a comparison with repeatability coefficient, another variable that may contribute to describe reliability characteristics of CoP and CoG analyzed by WBB and Kinect, was also presented in this study. The advantage of this variable is that it does not utilize any reliability coefficient in its formula.

The lower reliability scores in this experiment may be due the differences in obtaining data from only one trial when most of the superior results appeared in studies that used the mean or median of three trials in each test-retest condition. Additionally, test duration of 30s for double stance and 10s-15s for single stance with eyes closed instead of 20s may also have influenced results. Moreover, the use of higher sampling rate with WBB, above or equal 50Hz, instead of 30Hz as used in our study for compliance with Kinect maximum sampling rate may had enhanced CoP variables reliability in other studies. However, based on consolidated data of all leg positions relative and absolute reliability scores presented better outcomes, which were comparable to best reliability scores published in the literature. It indicated that when using different tasks with these devices, the analysis of consolidated information may result in assessments that are more reliable. Additionally, adopting a stricter control of environmental conditions during test application may also reduce variability of results.

In conclusion, this study demonstrated that some WBB and Kinect derived variables had satisfactory reliability for providing additional quantitative information of balance tests in situations where do not require great precision. Moreover, both devices may be used to provide quantitative data during qualitative tests such as BESS. They may capture body sway parameters not observed by naked eyes and register body oscillation behavior over test time. Additionally, Kinect could be used to log whole body joint positions or even record test images and reproduce it later, if necessary. Finally, considering their low cost, setup easiness, and portability our findings corroborate to demonstrate the utility of Wii Balance Board and Microsoft Kinect during semi-static balance tests in different site locations what may increase accessibility of this assessments to more people.

9
Bibliography

BARTLETT, H. L.; TING, L. H.; BINGHAM, J. T. (2014). **Accuracy of force center of pressure of the Wii balance board**. Gait & Posture.39(1).

BASTOS, A. G. D.; LIMA, M. A. M. T.de; OLIVEIRA, L. F.de. (2005). **Evaluation of patients with dizziness and normal electronystagmography using stabilometry**. Revista Brasileira de Otorrinolaringologia, 71(3), 305-310.

BELL, D. R. et al. (2011). **Systematic review of the balance error scoring system**. Sports Health, 3(3), 287–295.

Biomechanics and Motor Control (n.d.). DUARTE, M. **Notes on Scientific Computing for Biomechanics and Motor Control**. https://github.com/demotu/BMC [Accessed: April-August, 2018]

BLAND, J. M; ALTMAN, D. G. (1999) Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8, pp. 135-160.

BOWER, K. J. et al. (2014). **Instrumented static and dynamic balance assessment after stroke using Wii Balance Boards: reliability and association with clinical tests**. PLoS ONE, 9(12), e115282. http://doi.org/10.1371/journal.pone.0115282

BROWN, J. D. (1999). **Standard error versus standard error of measurement.** Shiken Research Bulletin: JALT testing & evaluating SIG newsletter, 3 (1), 20-25.

BURNS, E.; KAKARA, R. (2018). **Deaths from falls among persons aged ≥65 years – United States, 2007-2016**. Centers for Disease Control and Prevention. Morbidity and Mortality Weekly Report, vol. 67(18).

CASTELLI, L. et al. (2015). **We-Measure: toward a low-cost portable posturography for patients with multiple sclerosis using commercial Wii balance board**. Journal of Neurological Sciences, http://dx.doi.org/10.1016/j.jns.2015.10.016/.

CHAKRAVARTY, K. D. et al. (2017) **Analysis of muscle activation in lower extremity for static balance**. 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Seogwipo, pp. 4118-4122. doi: 10.1109/EMBC.2017.8037762

CHANG, J. O. et al. (2014). **An alternative to the balance error scoring system: using a low-cost balance board to improve the validity/reliability of**

sports-related concussion balance testing. Clinical Journal of Sport Medicine, 24(3), pp. 256-262.

CHEN, C.; BARNHART, H. X. (2008). **Comparison of ICC and CCC for assessing agreement for data without and with replications**. Computational Statistics and Data Analysis, 53, pp. 554-564.

CLARK, R. A. et al. (2010). **Validity and reliability of Nintendo Wii balance board for assessment of standing balance**. Gait & Posture, 31, pp. 307-310.

CLARK, R. A. et al. (2015). **Reliability and concurrent validity of the Microsoft Xbox One Kinect for assessment of standing balance and postural control**. Gait & Posture, 42, pp. 210-213.

CLARK, R. A. et al. (2018). **Reliability and validity of the Wii balance board for assessment of standing balance: a systematic review**. Gait & Posture, 61, pp. 40-54.

CLAUSER, C. E. et al. (1969). **Weight, volume, and center of mass of segments of human body**. AMRL Technical Report. Wright-Patterson Air Force Base, 69-70.

COOK, J. A. et al. (2014). **Assessing methods to specify the target difference for a randomised controlled trial: DELTA (Difference ELicitation in TriAls) review**. Health Technology Assessment, 18(28).

DÍAZ CENTENO, A. (2016). **Evaluation of physical-motor status of people with reduced mobility using motion capture with Microsoft Kinect**. Masters Dissertation. Pontifical Catholic University of Rio de Janeiro.

DUARTE M.; FREITAS, S.M.S.F. (2010). **Revision of posturography based on force plate for balance evaluation**. Revista Brasileira de Fisioterapia, v14, n. 3, pp. 183-192.

DUTTA, A. et al. (2014). **Point-of-care-testing of standing posture with Wii balance board and Microsoft Kinect during transcranial direct current stimulation: a feasibility study**. Neuro Rehabilitation, 34(4), pp. 789-98.

ELTOUKHY, M. A. et al. (2018). **Validation of static and dynamic balance assessment using Microsoft Kinect for young and elderly populations**. IEEE Journal of Biomedical and Health Informatics, vol. 22, n. 1, pp. 147-153.

FLEISS, J. L. (1986). **Reliability of measurement. The design and analysis of clinical experiments**. New York: Wiley. p. 1-32.

GALE, C. R. et al. (2018). **Risk factors for incidental falls in older men and women: English longitudinal study of ageing**. BMC Geriatrics, 18:117

GALEANO, D. et al. (2014). **A tool for balance control training using muscle synergies and multimodal interfaces**. Biomed Research International, article id, 565370, http://dx.doi.org/10.1155/2014/565370.

GHASEMI, A.; ZAHEDIASL, S. (2012). **Normality tests for statistical analysis: a guide for non-statisticians**. International Journal of Endocrinology and Metabolism, 10(2), 486–489. http://doi.org/10.5812/ijem.3505

GIAVARINA, D. (2015). **Lessons in biostatistics. Understanding Bland Altman analysis**. Biochemia Medica, 25(2), 141-51. http://dx.doi.org/10.11613/BM.2015.015

GOBLE, D. J.; CONE, B. L.; FLING, B. W. (2014). **Using the Wii Fit as a tool for balance assessment and neurorehabilitation: the first half decade of "Wii-search"**. Journal of NeuroEngineering and Rehabilitation, 11:12.

GONZÀLEZ, A; HAYASHIBE, M.; FRAISSE, P. (2012). **Estimation of the center of mass with Kinect and Wii Balance Board**. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems.

GONZÁLEZ, A. M.; PAGANELLI, A. I.; RAPOSO, A. (2018). **Analysing balance loss in VR interaction with HMDs**. SBC Journal on Interactive Systems, 9(2), pp. 68-81.

GROOTEN, W. J. A., et al. (2018). **Reliability and validity of a novel Kinect-based software program for measuring posture, balance and side-bending**. BMC Musculoskeletal Disorders BMC series, 19:6.

HALL, S. J. (2009) **Biomecânica básica**, 5th Edition. Editora Manole. Barueri, São Paulo.

HAMILL, J.; KNUTZEN, K. M.; DERRICK, T. R. (2015). **Biomechanical basis of human movement**. Fourth edition. Lippincott Willians & Wilkins, Wolters Kluwer. Philadelphia, PA.

HIRATA, V. N.; CAMEY, S. A. (2009). **Bland-Altman analysis of agreement between methods**. Revista Hospital de Clínicas de Porto Alegre, 29(3).

HONDORI, H. M.; KHADEMI, M. A. (2014). **A review on technical and clinical impact of Microsoft Kinect on physical therapy and rehabilitation**. Journal of Medical Engineering, article ID 846514.

HOPKINS, W.G. (2000). **Measures of reliability in sports medicine and science**. Sports Medicine, 30(1), pp. 1-15.

HORAK, F. B.; WRISLEY, D. M.; FRANK, J. (2009). **The balance evaluation system test (BESTest) to differentiate balance deficits**. Physical Therapy, 89(5):484-498. doi:10.2522/ptj.20080071.

JETER, P. E., et al. (2015). **Intra-session test-retest reliability of magnitude and structure of center of pressure from the Nintendo Wii Balance BoardTM for a visually impaired and normally sighted population**. Gait & Posture, 41(2), 482–487. http://doi.org/10.1016/j.gaitpost.2014.11.012

JURAS, G., et al. (2008). **Evaluation of the limits of stability (LOS) balance test**. Journal of Human Kinects, 19, pp. 39-52.

KAEWKAEN, K., et al. (2018). **Reliability of Wii balance board with MFU static balance test software® for assessing balance in chronic stroke patients**. [https://researchgate.net accessed in July 29th, 2018]

KAPTEYN, T. S., et al. (1983). **Standardization in platform stabilometry being a part of posturography**. International Society of Posturography. Agressologie, 24, 7 pp. 321-326.

KING, T. S.; CHINCHILLI, V. M.; CARRASCO, J. L. (2007). **A repeated measures concordance correlation coefficient**. Statistics in Medicine, 26, pp. 3095-3113.

KNIPPENBERG, E., et al. (2017). **Markerless motion capture systems as training device in neurological rehabilitation: a systematic review of their use, application, target population and efficacy**. Journal of Neuro Engineering and Rehabilitation, 14:61

KOO, T. K.; LI, M. Y. (2016). **A guideline for selecting and reporting intraclass correlation coefficients for reliability research**. Journal of Chiropractic Medicine, 15, pp. 155-163.

LARSEN, L. R., et al. (2014). **Field assessment of balance in 10 to 14 year old children, reproducibility and validity of the Nintendo Wii board**. BMC Pediatrics, 14, 144. http://doi.org/10.1186/1471-2431-14-144

LEACH, J. M., et al. (2014). **Validating and calibrating the Nintendo Wii Balance Board to derive reliable center of pressure measures**. Sensors, 14(10), 18244–18267. http://doi.org/10.3390/s141018244

LEVINGER, P., et al. (2016). **A real time biofeedback using Kinect and Wii to improve gait for post-total knee replacement rehabilitation: a case study report**. Disability and Rehabilitation: Assistive Technology, 11(3).

LIANG, Z., et al. (2014). **Neck musculature fatigue affects specific frequency bands of postural dynamics during quiet standing**. Gait & Posture, 39(1), pp. 397-403.

LIM, D., et al. (2015). **Use of the Microsoft Kinect system to characterize balance ability during balance training**. Clinical Interventions in Aging, 10, pp. 1077-1083.

LIN L. I. (1989). **A concordance correlation coefficient to evaluate reproducibility**. Biometrics, 45:255–268.

MARTÍNEZ-LEMOS, R. I.; AYÁN-PÉREZ, C.; BOUZAS-RICO, S. (2017). **Test-retest reliability of the Wii Balance Board for assessing standing balance in young people with intellectual disability**. International Journal of Developmental Disabilities. doi: 10.1080/20473869.2017.1403065

MCGRAW, K. O.; WONG, S. P. (1996). **Forming inferences about some intraclass correlation coefficients**. Psychological Methods, v. 1, n. 1, pp. 30-46.

MCLAUGHLIN, P. (2013). **Testing agreement between a new method and the gold standard. How do we test?** Journal of Biomechanics. http://dx.doi.org/10.1016/j.jbiomech.2013.08.015i

OTTE, K., et al. (2016). **Accuracy and Reliability of the Kinect Version 2 for Clinical Measurement of Motor Function**. PLoS ONE, 11(11), e0166532. http://doi.org/10.1371/journal.pone.0166532

PAGNACCO, G.; OGGERO E.; WRIGHT, C. H. (2011). **Biomedical instruments versus toys: A preliminary comparison of force platforms and the Nintendo Wii Balance Board**. Biomedical. Science Instruments, 47, 12–17.

PALACIOS-NAVARRO, G.; GARCÍA-MARAGIÑO I.; RAMOS-LORENTE, P. (2015). **A Kinect-base system for lower limb rehabilitation in Parkinson´s disease patients: a pilot study**. Journal Medical Systems, 39:103.

PALMIERI, R. M., et al. (2002). **Center of pressure parameters used in the assessment of postural control**. Journal of Sport Rehabilitation, 11:1, 51-66

PARK, D.; LEE, G. (2014). **Validity and reliability of balance assessment software using the Nintendo Wii balance board: usability and validation**. Journal of Neuro Engineering and Rehabilitation, 11:99.

PRIETO, T. E., et al. (1996). **Measures of postural steadiness: differences between healthy young and elderly adults**. IEEE Transactions Biomed Eng. , 43(9), pp. 956-966.

PROCHÁZKAA, A., et al. (2015). **Bayesian classification and analysis of gait disorders using image anddepth sensors of Microsoft Kinect**. Digital Signal Processing, 47, pp. 169-177.

RIES, J. D., et al. (2009). **Test-retest reliability and minimal detectable change for the timed "Up & Go" test, the six-minute walk test, and gait speed in people with Alzheimer disease**. Journal of American Physical Therapy Association, 89, pp. 569-579.

RUFF, J., et al. (2015). **Commercially available gaming systems as clinical assessment tools to improve value in the orthopaedic setting: A systematic review. Injury**, International Journal Care Injured, 46, pp. 178–183.

RUHE, A.; FEJER, R.; WALKER B. (2010). **The test-retest reliability of center of pressure measures in bipedal static task conditions - A systematic review of the literature**. Gait & Posture, 32: 436–445.

SALAVATI, M., et al. (2009). **Test–retest reliability of center of pressure measures of postural stability during quiet standing in a group with musculoskeletal disorders consisting of low back pain, anterior cruciate ligament injury and functional ankle instability**. Gait & Posture, 29, pp. 460-464.

SCAGLIONI-SOLANO, P.; ARAGÓN-VARGAS, L. F. (2014). **Validity and reliability of the Nintendo Wii Balance Board to assess standing balance and**

sensory integration in highly functional older adults. International Journal of Rehabilitation Research, 37:138–143

SCHUBERT, P., et al. (2012). **About the structure of posturography: Sampling duration, parametrization, focus of attention (part I)**. Journal of Biomedical Science and Engineering, 5, 496-507. doi: 10.4236/jbise.2012.59062.

SEMPERE-RUBIO, N., et al. (2018). **Characterization of postural control impairment in women with fibromyalgia**. PLoS ONE, 13(5): e0196575. https://doi.org/10.1371/journal.pone.0196575

SEVERINI, G., et al. (2017). **Use of Nintendo Wii balance board for posturographic analysis of multiple sclerosis patients with minimal balance impairment**. Journal of NeuroEngineering and Rehabilitation, 14:19.

SGRÒ, F., et al. (2013). **Exergames for physical education: an overview about interaction design perspective**. World Journal on Educational Technology, 5(2), pp. 248-256.

SGRÒ, F., et al. (2014). **Validity analysis of Wii Balance Board versus baropodometer platform using an open custom integrated application**. AASRI Conference on Sports Engineering and Computer Science. (Eds) Deng W. SECS 2014, vol. 8, pp. 22-29.

SHOTTON, J., et al. (2011). **Real-time human pose recognition in parts from single depth images**, in 'Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on', IEEE, pp. 1297–1304.

SHOUKRI, M. M; ASYALI, M. H; DONNER A. (2014). **Sample size requirements for design of reliability study: review and new results**. Statistical Methods in Medical Research, 13 pp. 1-21.

SHROUT, P. E.; FLEISS, J. L. (1979). **Intraclass correlations: uses in assessing rater reliability**. Psychological Bulletin, v. 86, n. 2, pp. 420-428.

SILVERTHORN, D.U. (2003). **Fisiologia humana**. 2nd ed. São Paulo: Manole, 2003.

STADELMANN, K., et al. (2015). **Impaired postural control in healthy men at moderate altitude (1630 M and 2590 M): Data from a randomized trial**. PLoS ONE, 10(2), e0116695. http://doi.org/10.1371/journal.pone.0116695

TYSON, S. F., et al. (2006). **Balance disability after stroke**. Physical Therapy, 86(1), pp. 30-38.

VALGMA, L. (2016). **3D reconstruction using Kinect v2 camera**. Bachelor´s thesis. University of Tartu. Estonia.

WALTER, S. D.; ELIASZIW, M.; DONNER A. (1998). **Sample size and optimal designs for reliability studies**. Statistics in Medicine, 17:101–11.

WANG, Q., et al. (2015). **Evaluation of pose tracking accuracy in the first and second generations of Microsoft Kinect**, 2015 International Conference on Healthcare Informatics, Dallas, TX, 2015, pp. 380-389.

WASENMULLER, O.; STRICKER, D. (2017). **Comparison of Kinect v1 and v2 depth images in terms of accuracy and precision**. In: Chen CS., Lu J., Ma K. K. (Eds) Computer Vision – ACCV 2016 Workshops. ACCV 2016. Lecture Notes in Computer Science, vol. 10117. Springer, Cham.

WATSON, P. F.; PETRIE, A. (2010). **Method agreement analysis: A review of correct methodology**. Invited Review. Theriogenology, 73, pp. 1167-1179.

WEIR, J. P. (2005). **Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM**. Journal of Strength and Conditioning Research, 19(1), pp. 231-240.

WII BREW Website. http://wiibrew.org/wiki/Wii_Balance_Board. Accessed in May 2018.

WIKIPEDIA KINECT. http://Wikipedia.org/wiki/Kinect, accessed in May 2018.

WHITING, W. C.; RUGG, S. (2006). **Dynatomy: dynamic human anatomy**. Human Kinects. Champaign, Illinois.

WINTER, D. A. (1995). **Human balance and posture control during standing and walking**. Gait & Posture, Vol. 3: pp. 193-214.

WINTER, D. A. (2009). **Biomechanics and motor control of human movement**. 4th edition. Hoboken: Wiley.

WHO, WORLD HEALTH ORGANIZATION (2007) **Global report on falls prevention in older age**.

YANG, Y.; LI, S.; LI, D. (2014). **Reliability and validity of Kinect RGB-D sensor for assessing standing balance**. IEEE Sensor Journal, v. 14 (5), pp. 1633-1638.

YANG, S. T., et al. (2016). **Standing balance assessment of elderly using Kinect sensor with Wii Balance Board**. Korean Journal of Sport Biomechanics, v. 26 (4), pp. 391-396.

YEUNG, L. F., et al. (2014). **Evaluation of the Microsoft Kinect as a clinical assessment tool of body sway**. Gait & Posture, 40, pp. 532-538.

Appendix I

This appendix gives details of the customized applications developed for supporting this experiment.

Figure 40 shows the class diagram of all implemented classes. The ApplicationView class controls application navigation using a list of IPageViewModel objects and changes pages according to events triggered in the screen. According to the selected option, a UserControl is loaded into the main Window and linked to the associated ViewModel object. All ViewModel objects are specializations of the ObservableObject class, which implements the interface INotifyPropertyChanged used to update content on the screen.

KinectBodyViewModel implements the events to capture Kinect frame information and CalibrationTest class implements most of the code associated with WBB events. Auxiliary classes are used to implement signal filters, center of mass estimation, data logs handling, general vector calculations, and dynamic time warping execution, most of them used only for giving feedback on the screen. Log files store raw information except for center of mass coordinates that were calculated in the customized application.
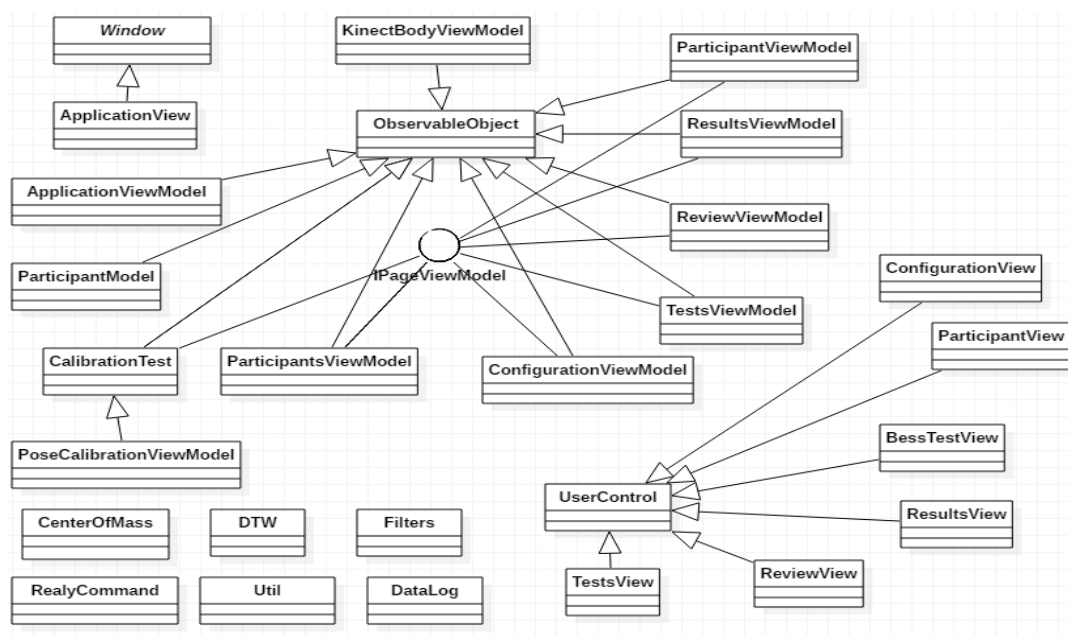

Figure 40 Class diagram of the application that manages tests.

R Package 'irr' version 0.8[12] was used for calculating Intraclass Correlation Coefficient and package 'epiR' version 0.9-96 for calculating Lin´s Concordance Correlation Coefficient. (https://cran.r-project.org/web/packages/epiR/epiR.pdf).

Scripts were organized by name within a Python project. Filenames starting with the prefix Manikin_ and Ind_ differed procedures aimed at manikin and individuals tests processing. Scripts without these prefixes were used for all other scripts. Scripts for calculating estimated circumference area, analyzing filter cut-off frequencies and some other basic descriptive and graphical generation were downloaded from https://github.com/demotu/BMC project, version 1.0.2, developed by Professor Marcos Duarte under MIT license.

Scripts were developed to organize and clean up log files, which were generated by the C# application. Log records were converted into CoP and CoG filtered 2D or 3D coordinates data series. Statistical descriptive, inferential tests, and graphic generation of raw data series were performed by functions into these scripts. Additionally, from those files, a property database file consolidates all derived variables. Formulas used to calculate time-domain variables are specified in the Methodology chapter of this document. From these database files, a new set of statistical tests and graphs were generated. The following table summarizes each Python script functionality.

Table 51. List of implemented scripts and their objectives.

Script Name	Objective
Manikins.py and Individuals.py	Read manikin/participant XML that stores profile information and calculates weight, stature, age etc.
	Generate descriptive information about manikins and participants.
[Ind|Manikin]_MountDataSeries.py	Read log files and generate raw A-P and M-L axes coordinates for CoP and CoG applying unit transformations, mean de-trend, and optionally a low-pass 2nd order Butterworth filter with configured cut-off frequency.
[Ind|Manikin]_FilterAnalysis.py	Get unfiltered data series and analyse different filter cut-off frequencies in order to find the best one based on residuals between filtered and unfiltered signals.

---

[12] https://cran.r-project.org/web/packages/irr/irr.pdf - accessed in April/2018.

[Ind|Manikin]_CoPG_Descriptive.py	Based on chosen filtered cut-off frequency generate graphs, descriptive and basic inferential information about data series.
[Ind|Manikin]_MountDB.py	From CoP and CoG data series files, calculate time domain properties (mean distance, range, root mean square, total path length, average speed, 95% estimated circle and ellipse areas) and frequency domain properties generating properties database files.
[Ind|Manikin]_Cutoff_Comparison.py	Using the properties databases with different filter cut-off frequencies calculate CCC and $ICC_{2,\,1}$ and compare results. It also export properties data tables to be read into R scripts for analysing ICC and CCC.
[Ind|Manikin]_BasicComparisonProperties.py	Using CoP and CoG properties database with chosen filter cut-off frequency perform graph generation and statistical analysis on properties.
[Ind|Manikin]_SEMMDC.py	Calculates standard error of measurement and minimal detectable change of CoP and CoG properties.
[Ind|Manikin]_Coef_Variation.py	Calculates coefficient of variation of CoP and CoG properties.
Manikin_Chi2_Percentiles.py	Calculates Chi2 percentiles analysis in individuals and manikins.
Properties_Computation.py	Functions to calculate properties from coordinate data series.
Filter_Analysis.py	Adapted residuals analysis function got from Duarte. (https://www.github.com/demotu/BMC)
ReadFiles.py, ReadGrous.py, Classes.py, DataSeries.py, RawData.py, Util.py, hyperellipsoid.py, psd.py, StatDesc.py, Feedback.py, DescriptiveAnalysis.py	Auxiliary functions for handing files, internal structures, calculate areas and generate some statistical graphs and analysis.