Computers in Biology and Medicine 50 (2014) 9–13

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Gait variability and stability measures: Minimum number of strides and within-session reliability

F. Riva [a,*], M.C. Bisi [a], R. Stagni [a,b]

[a] DEI – Department of Electrical, Electronic, and Information Engineering 'Guglielmo Marconi', University of Bologna, via Venezia 52, 47521 Cesena (FC), Italy
[b] Health Sciences and Technologies – Interdepartmental Center for Industrial Research (HST – ICIR), University of Bologna, Italy

ARTICLE INFO

Article history:
Received 13 January 2014
Accepted 5 April 2014

Keywords:
Within-session reliability
Stride number
Gait stability
Gait variability
Inertial sensors

ABSTRACT

Background: Several methods are proposed in the literature for the quantification of gait variability/stability from trunk accelerations. Since outputs can be influenced by implementation differences, reliability assessment and standardization of implementation parameters are still an issue. The aim of this study is to assess the minimum number of required strides and the within-session reliability of 11 variability/stability measures.
Method: Ten healthy participants walked in a straight line at self-selected speed wearing two synchronized tri-axial Inertial Measurement Units. Five variability measures were calculated based on stride times namely Standard deviation, Coefficient of variation, Inconsistency of variance, Nonstationary index and Poincaré plot. Six stability measures were calculated based on trunk accelerations namely Maximum Floquet multipliers, Short term/long term Lyapunov exponents, Recurrence quantification analysis, Multiscale entropy, Harmonic ratio and Index of harmonicity. The required minimum number of strides and the within-session reliability for each measure were obtained based on the interquartile range/mean ratio. Measures were classified in five categories (namely excellent, good, average, poor, and very poor) based on their reliability.
Results: The number of strides required to obtain a reliable measure was generally larger than those conventionally used. Variability measures showed average to poor reliability, while stability measures ranged from excellent to very poor reliability.
Conclusion: Recurrence quantification analysis and multiscale entropy of trunk accelerations showed excellent reliability and a reasonable number of required strides. Based on these results, these measures should be taken into consideration in the assessment of fall risk.

1. Introduction

Ageing and pathology can worsen gait performance at multiple levels and in selective ways [1], and quantitative assessment of gait pattern has been proven to be useful in the early identification and prediction of pathology or cognitive decline [2–4]. In particular, trunk acceleration-based measures of gait variability and stability are proposed in the literature aiming at quantifying subject specific gait characteristics such as gait impairment, degree of neuro-motor control and balance disorders, in pathologic and healthy subjects, and are often related to fall at risk [5–9]. However, no standard implementation procedure for these measures is defined, potentially explaining the incoherent conclusions [10], as implementation differences can affect outputs. Thus, a standardization of the implementation parameters is necessary to perform a consistent evaluation. Moreover, these measures must reproduce the same results in the same experimental conditions.

Many strides can be required to obtain reliable measures, but treadmill walking differs significantly from over-ground walking [11]; hence, long walking trials have to be analyzed. The use of Inertial Measurement Units (IMU) allows to obtain both stride time variability and stability measures from trunk acceleration signals during long over-ground outdoor walking trials.

In order to further define implementation features for future effective exploitation of measures in research or clinical practice, an assessment of the repeatability of variability/stability measures is hence needed, together with an assessment of the number of necessary strides. The aim of the present study was to assess the minimum number of required strides and the within-session reliability of 11 temporal variability/stability measures proposed in the literature and applied to stride time and trunk accelerations during over-ground walking.

* Corresponding author. Tel.: +39 0547 33 89 53.
E-mail address: f.riva@unibo.it (F. Riva).

http://dx.doi.org/10.1016/j.compbiomed.2014.04.001
0010-4825/

2. Methods

Ten healthy participants [28 ± 3 years, 174 ± 11 cm, 67 ± 13 kg] walked in a straight line at self-selected natural speed on a 250 m long dead-end road (about 180 strides), wearing two synchronized tri-axial IMU (Opal, APDM, USA) located on the trunk (at the level of the fifth lumbar vertebra) and on the right shank. Sample size was chosen in agreement with previous literature [12]. Range of accelerometers was ±6 G and sampling rate 128 Hz. Right heel strike instants were obtained from the angular velocity measured by the sensor on the shank with a wavelet-based method [13]. Gait initiation and termination phases were excluded. The average walking speed was 1.43 ± 0.15 m/s.

Since our aim was the characterization of variability/stability measures from a strictly methodological point of view, adequately segmented data coming from the same experimental trial (long overground walk) were analyzed. This approach was adopted in order to maintain the experimental conditions unvaried, being thus able to ascribe every variation found in the results to the intrinsic variability of each measure only rather than to actual differences in gait patterns depending on specific trial analyzed.

The following variability measures were applied to stride time:

i. Standard deviation (SD) [14];
ii. Coefficient of variation (CV) [14];
iii. Inconsistency of variance (IV) [15];
iv. Nonstationary index (NI) [15];
v. Poincaré plots (PSD1/PSD2) [16].
SD represents the standard deviation of stride time. CV is the variability of stride time normalized to the mean stride time value (CV = 100 × SD/mean) [14]. IV and NI quantify the temporal "structure" of the time series (independent of the overall variance); each time series was first normalized with respect to its mean and SD and then divided into blocks of five strides each. In each segment, the local average and the local SD were computed. NI is then defined as the SD of the local averages, while IV is defined as the SD of the local SDs [15]. Stride time data plots between successive gait cycles, known as Poincaré plots, show variability of stride time data. Statistically, the plots display the correlation between consecutive stride times data in a graphical manner. PSD1 and PSD2 represent, respectively, width and length of the long and short axis describing the elliptical nature of the plots, and hence the short-term and long-term variability of stride time [16].
The following stability measures were calculated on trunk accelerations in vertical (V), medio-lateral (ML) and anterior–posterior (AP) directions:
vi. Maximum Floquet multipliers (maxFM) [5,10];
vii. Short term/long term Lyapunov exponents (sLE/lLE) [5];
viii. Recurrence quantification analysis (RQA) [17];
ix. Multiscale entropy (MSE) [18];
x. Harmonic ratio (HR) [6];
xi. Index of harmonicity (IH) [19].

maxFM quantify orbital stability of a periodic or pseudo-periodic dynamic system, that is the tendency of the system state to return to the periodic limit cycle orbit after small perturbations [5,10]. On the other hand, sLE and lLE quantify local dynamic stability of a system and are used for systems that do not necessarily exhibit a discernable periodic structure [5]. RQA provides a characterization of a variety of features of a given time series, including a quantification of deterministic structure and non-stationarity [17], based on the construction of recurrence plots [20]. All of these measures imply the reconstruction of the state space of the system; in this study, four different state spaces were constructed: one 3-dimensional state space composed by acceleration signals in the V, ML and AP direction and three (one per direction) 5-dimensional state spaces composed by delay-embedding of each acceleration component (delay = 10 samples). Such parameters were chosen based on previous literature, stating that an embedding dimension of 5 and a 10 samples delay are appropriate for gait data [21–23], and on a false nearest neighbors analysis performed on our data.

Several measures were extracted from RQA, namely recurrence rate (rr), determinism (det), averaged diagonal line length (avg), maximum diagonal line length (max) and divergence (diverg). In the calculation of RQA measures, a radius of 40% was chosen to make sure that recurrence rate (rr) responded smoothly and was not too high, and that determinism (det) did not saturate at the floor of 0 or the ceiling of 100, as approaching these limits would tend to suppress variance in the measure [20]. Time series derived from complex systems, like biological systems, are likely to present structures on multiple spatio-temporal scales; MSE has been introduced to quantify the complexity or irregularity of a time series [18]. MSE has been obtained calculating sample entropy (consecutive data points $m = 2$, distance $r = 0.2$ [24]) on six consecutively more coarse-grained (scale factor $\tau = 1, \ldots, 6$) time series. HR quantify the smoothness of acceleration patterns of the trunk based on amplitudes in the frequency spectra. It provides information on how smoothly subjects control their trunk during walking and it is directly related to whole body balance and coordination [6,25]. In this study, HR was not calculated stride by stride, but decomposing the whole signal into its harmonics [7]. Similarly to HR, IH assesses the contribution of the oscillating components to the observed coordination patterns by means of spectral analysis [19], quantifying the contribution of the stride frequency to the signal power relative to higher harmonics.

For the quantification of the **minimum number of strides**, measures were calculated on windows of decreasing length (from 150 to 10 strides, 1 stride resolution). Interquartile range and median value of measures were calculated for all the windows. Percent interquartile range/median ratio (*imr*) was then calculated, starting from the 150 strides window (which gave the lowest ratio) and proceeding backwards.

Adding an increasing number of strides to the calculation would cause the measure to reach a steady outcome, which represents a compromise between reliability of the measure and experimental limitations. Percent *imr* is then an indication of the variations of the measures around the median value. A low *imr* indicates small variations of the measure around its median value with the increase of the number of strides; this means that the measure reached a steady value, and it is not likely to change with the inclusion of further strides. On the contrary, a high *imr* indicates that the measure undergoes significant variation with the increase of the number of strides, and hence is still not fully reliable.

Thresholds for the *imr* were fixed at 10%, 20%, 30%, 40% and 50%. The required number of strides was defined as the smallest one at which the ratio remained below the lowest possible threshold. The minimum number of strides was calculated per index and per subject at first, and then the largest number of strides over subjects was selected for each index.

The assessment of **within-session reliability** was performed calculating variability/stability measures on a window sliding along the trial with 1 stride steps. The sliding window size was set to 85 strides, since most measures (51 out of 57) required less than 85 strides. lLE (tot, V, ML, AP) and RQA V (max, diverg) did not satisfy this criterion. Interquartile range and median values of the measures over the windows, together with the percent *imr* for each measure, were calculated. Measures were grouped in five reliability categories, ranging from very poor (*imr* > 40%) to excellent (*imr* < 10%), based on the maximum inter-subject *imr*.