Journal of Neuroscience Methods 178 (2009) 327–333

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

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

Statistical precision and sensitivity of measures of dynamic gait stability

Sjoerd M. Bruijn[a], Jaap H. van Dieën[a,*], Onno G. Meijer[a,b], Peter J. Beek[a]

[a] Research Institute MOVE, Faculty of Human Movement Sciences, VU University Amsterdam, Van der Boechorststraat 9, NL-1081 BT Amsterdam, The Netherlands
[b] Second Affiliated Hospital of Fujian Medical University, Quanzhou, 362000 Fujian Province, PR China

ARTICLE INFO

Article history:
Received 6 June 2008
Received in revised form 9 December 2008
Accepted 11 December 2008

Keywords:
Gait stability
Lyapunov exponents
Floquet multipliers
Statistical precision
Sensitivity
Bootstrap

ABSTRACT

Recently, two methods for quantifying a system's dynamic stability have been applied to human locomotion: local stability (quantified by finite time maximum Lyapunov exponents, $\lambda_{\text{S-stride}}$ and $\lambda_{\text{L-stride}}$) and orbital stability (quantified as maximum Floquet multipliers, MaxFm). Thus far, however, it has remained unclear how many data points are required to obtain precise estimates of these measures during walking, and to what extent these estimates are sensitive to changes in walking behaviour.

To resolve these issues, we collected long data series of healthy subjects ($n=9$) walking on a treadmill in three conditions (normal walking at 0.83 m/s (3 km/h) and 1.38 m/s (5 km/h), and walking at 1.38 m/s (5 km/h) while performing a Stroop dual task). Data series from 0.83 and 1.38 m/s trials were submitted to a bootstrap procedure and paired *t*-tests for samples of different data series lengths were performed between 0.83 and 1.38 m/s and between 1.38 m/s with and without Stroop task.

Longer data series led to more precise estimates for $\lambda_{\text{S-stride}}$, $\lambda_{\text{L-stride}}$, and MaxFm. All variables showed an effect of data series length. Thus, when estimating and comparing these variables across conditions, data series covering an equal number of strides should be analysed. $\lambda_{\text{S-stride}}$, $\lambda_{\text{L-stride}}$, and MaxFm were sensitive to the change in walking speed while only $\lambda_{\text{S-stride}}$ and MaxFm were sensitive enough to capture the modulations of walking induced by the Stroop task. Still, these modulations could only be detected when using a substantial number of strides (>150).

1. Introduction

Recently, two methods for quantifying a system's dynamic stability have been applied to human locomotion: local stability (Dingwell and Cusumano, 2000; Dingwell et al., 2007; Kang and Dingwell, 2006a) and orbital stability (Hurmuzlu and Basdogan, 1994; Hurmuzlu et al., 1996; Dingwell et al., 2007). Local stability is defined by local divergence coefficients, called "maximum finite time Lyapunov exponents", that quantify how the system responds continuously in real time to very small (i.e. "local") perturbations (Rosenstein et al., 1993). Orbital stability, as defined by Floquet multipliers, estimates stability in a different manner. It assumes periodicity, and quantifies the system's response to small perturbations in a discrete fashion, from one cycle to the next (Hurmuzlu and Basdogan, 1994).

Both local and orbital dynamic stability are mathematically well motivated measures that may be readily applied in the study of natural dynamic systems like human movement. In the latter context, it has been suggested that these measures may help to gain insight into the neural control of balance (Kang and Dingwell, 2006a). Moreover, several authors have suggested that they can be used to quantify gait stability in a variety of (neurological) pathologies that may limit walking ability (Dingwell et al., 2000, 2007; Hurmuzlu et al., 1996; Kang and Dingwell, 2006a). However, there are several methodological issues that need to be considered when using these measures.

A first concern is the length of the data series needed to calculate local and orbital dynamic stability. Rosenstein et al. (1993) suggested that the maximum time finite Lyapunov exponents as calculated with their method were relatively insensitive to the length of data series for a set of known attractors. Still, Kang and Dingwell (2006b) reported that for estimating the long-term coefficients of the divergence curve of gait data, 5 min of continuous data were not sufficient to achieve adequate intra-session reliability, which may indicate a low statistical precision of the measure at hand. As regards the length of data series needed to obtain precise estimates of orbital stability, as far as we know no guidelines have been reported in the literature, but it seems that fewer strides are needed (Hurmuzlu and Basdogan, 1994; Hurmuzlu et al., 1996).

A second issue, related to the first, is the sensitivity of the measures of interest. Previous research has indicated that both local and orbital dynamic stability are affected by walking speed even if estimated from relatively few strides (~30 England and Granata, 2007). Nevertheless other manipulations, such as the introduction

* Corresponding author. Tel.: +31 20 5988501; fax: +31 20 5988529.
E-mail address: j.vandieen@fbw.vu.nl (J.H. van Dieën).

0165-0270/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jneumeth.2008.12.015

of a Stroop dual task, had only small effects on measures of stability, even if data series of considerable length were used (Dingwell et al., 2008). The (in)sensitivity of these measures to smaller changes in walking behaviour may well be related to the precision issues mentioned above.

In sum, both local and orbital dynamic stability appear to be useful measures, but more insight is needed to resolve how many data points are required to obtain statistically precise estimates, and to what extent estimates are sensitive to subtle changes in walking behaviour.

To examine the statistical precision of measures of local and orbital dynamic stability, we collected long data series of healthy subjects walking on a treadmill, and submitted these data to a bootstrap procedure (Briggs et al., 1997; Efron and Tibshirani, 1986). In short, bootstrapping involves drawing, with replacement, a number of random samples from the data set, for different sample lengths, and provides a procedure to study the precision of an estimate of a variable for different sample sizes. Furthermore, to examine the sensitivity of those measures to specific changes in walking behaviour, we invited subjects to walk at two different speeds, i.e. 0.83 m/s (3 km/h) and 1.38 m/s (5 km/h), and included a condition in which subjects performed a Stroop task (Stroop, 1935) while walking at 1.38 m/s. In particular, subjects were required to name the colour of words that differed from the colour expressed by the word's semantic meaning (e.g. the word "red" printed in blue). Thus, our design involved manipulations with expected effects ranging from small (Stroop task) to large (walking speed).

2. Methods

2.1. Subjects

Nine healthy male volunteers (mean age 25.5 years, SD 3.6, mean weight 77.9 kg, SD 7.7 and mean length 1.85 m, SD 0.08) participated in the study. Exclusion criteria were orthopaedic or neurological disorders that could interfere with gait. Subjects provided informed consent, and the protocol was approved by the ethical committee of the Faculty of Human Movement Sciences, VU University Amsterdam.

2.2. Procedure

Neoprene bands with a cluster of three infrared Light Emitting Diodes (LED's) were attached to the trunk at the level of T6 and at the right foot. LED movements were recorded with an active 3D movement registration system (Optotrak® Northern Digital Inc., Waterloo, Ontario), consisting of a 2 × 3 camera array. Sample rate was set at 50 samples per second.

During the experiment, subjects were asked to walk on a treadmill (Biostar Giant™, Biometrics, Almere, The Netherlands) under three different conditions: walking at 0.83 m/s (3 km/h) and 1.38 m/s (5 km/h), and walking at 1.38 m/s (5 km/h) while performing a Stroop task. The 0.83 and 1.38 m/s conditions lasted 20 min each, while the Stroop condition lasted 10 min. During all conditions, subjects were instructed to look at an approximately 1 m × 1 m wide screen, placed about 2 m in front of the subject. Subjects were allowed a short break between conditions.

During the Stroop condition, 20 (5 rows × 4 columns) images of colour names printed in another colour were projected on the screen in front of the subject. Subjects were instructed to verbally report the colour of the words rather than their semantic meaning as quickly as possible. As soon as they had finished reading a set of words, the next set of words was presented. All conditions were presented in random order.

2.3. Calculations

2.3.1. Pre-processing

Raw data were analysed taking into account the problems associated with filtering nonlinear signals (Kantz and Schreiber, 1997; Mees and Judd, 1993). To overcome non-stationarities (cf. Dingwell and Marin, 2006), the first derivatives of the anterior posterior (AP), medio-lateral (ML) and vertical (VT) position time-series of the average movements of the thorax markers were used for the estimation of the stability measures of interest (see below). To allow for the calculation of normalized stride cycles, heel strikes were determined from the minimum vertical position of the average of the three heel markers.

2.3.2. Bootstrapping

A bootstrap procedure (Briggs et al., 1997; Efron and Tibshirani, 1986) was used to assess the statistical precision of both local and orbital dynamic stability for different data series length. Random samples (100 per sample length) of the state space (see below) were selected and stability measures of interest were calculated for these samples (for the Stroop condition only 10 estimates were used, as this condition was not included in the assessment of statistical precision: selecting 100 samples from 10 min of data would lead to excessive increases in estimated precision, merely reflecting overlap of the selected samples). This was done for different sample lengths, ranging from 30 to 300 strides, with increments of 30 strides. From these state space samples, measures of local and orbital dynamic stability were calculated (see below).

To assess any possible effects of sample length on the actual values of the local and orbital dynamic stability measures of interest, means of the 100 samples were calculated at each sample length. To assess the statistical precision of the local and orbital dynamic stability measures of interest the standard deviation of the 100 estimates of those measures were calculated for each sample length. Including more data should lead to a more precise estimate of the "true" value of a variable, and thus a reduced standard deviation. Note however that choosing larger sample lengths from the same data pool will increase the amount of overlap. Thus, a decrease in the standard deviation of the estimate could very well entail an overestimation of the increase in precision.

To examine the sensitivity of the local and orbital dynamic stability measures of interest to changes in walking behaviour, we performed a bootstrap on the estimates of all variables obtained from the bootstrapping procedure. From the 9 (subjects) × 100 (samples) data set of each sample length of each condition, 100 new data sets were created by randomly picking one of the values for each subject (for the Stroop condition we only had 9 (subjects) × 10 (samples) values, but this still leads to 9^{10} possible new data sets). Paired *t*-tests were then carried out, comparing each of the 100 new data sets of the 0.83 m/s condition to the corresponding 100 of the 1.38 m/s condition, and each of the 1.38 m/s condition to each of the corresponding Stroop condition, amounting to a total of 10,000 *t*-tests per comparison. For each comparison, the percentage of *P*-values above 0.05 was then calculated, and plotted against the sample lengths. If longer data series lead to more precise results, than a systematic decrease in the percentage of *P*-values > 0.05 with increasing sample length would imply that a given manipulation had a real effect.

It should be noted that in view of the time needed to calculate local dynamic stability the number of state space samples was relatively low compared to the number of samples used in the standard bootstrapping literature (Briggs et al., 1997; Efron and Tibshirani, 1986; van Dieën et al., 2002).

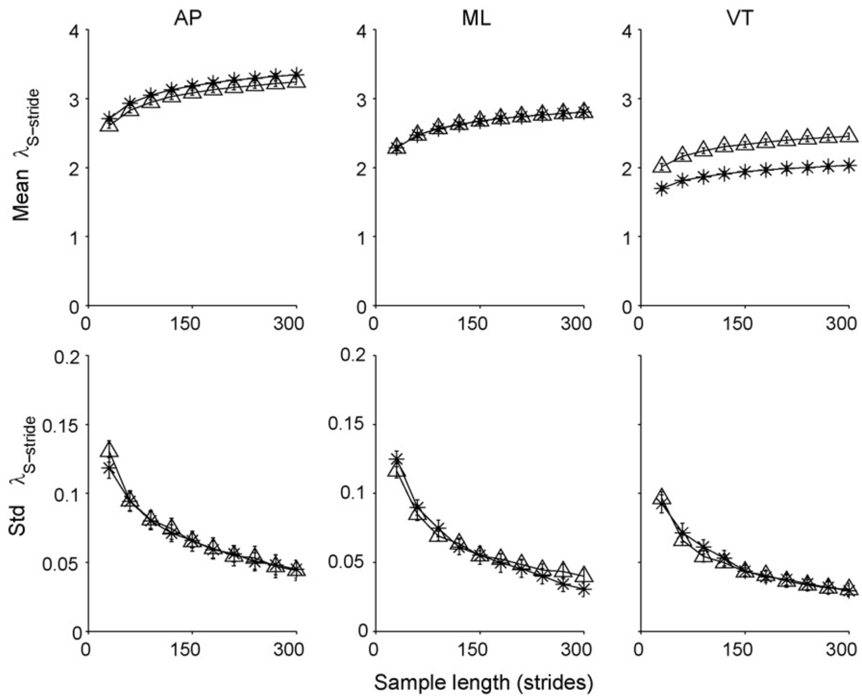

Fig. 1. Bootstrapping results for $\lambda_{S\text{-stride}}$. (*) 0.83 m/s (3 km/h) and (Δ) 1.38 m/s (5 km/h). Top panels show the mean of 100 estimates of $\lambda_{S\text{-stride}}$ for different sample lengths; bottom panels show standard deviation. Error bars represent standard errors.

2.3.3. State spaces

For each spatial dimension (i.e. AP, ML, and VT) state spaces were reconstructed for the calculation of local and orbital dynamic stability. The general form of these state spaces was:

$$(S(t) = [q(t), q(t+\tau), \ldots, q(t+(d_E-1)\tau)]) \quad (1)$$

where S(t) was the d_E-dimensional state vector, $q(t)$ was the original 1-dimensional data, τ was the selected time delay, and d_E was the embedding dimension. In the present study, an embedding dimension of $d_E = 5$ was used, because numerous studies found this to be an appropriate delay for gait data (i.e. Dingwell et al., 2000; Dingwell and Cusumano, 2000; England and Granata, 2007), and because Global False Nearest Neighbour analysis (Kennel et al., 1992) of our own data confirmed that $d_E = 5$ was appropriate.

For the calculation of local and orbital dynamic stability, the original time series were first re-sampled (cf. England and Granata, 2007), so that on average each stride was 100 samples in length. These re-sampled time series were then used to construct the state spaces. Since the time series now had the same average frequency (i.e. on average 100 samples/stride), we could choose a fixed time delay for state space reconstruction (England and Granata, 2007). We chose a delay of 10 samples.

2.3.4. Local dynamic stability

Local dynamic stability was calculated using maximum Lyapunov exponents. Maximum Lyapunov exponents express the real time response of a system to a small change in initial conditions or perturbation; a positive maximum Lyapunov exponent indicates that on average two initially neighbouring trajectories diverge, and thus that the system is unstable. Maximum Lyapunov exponents were estimated using well established techniques (e.g. Rosenstein et al., 1993; Dingwell et al., 2000). For each data point in state space, the nearest neighbour was identified, and the Euclidean distance between these points was tracked over time. Next, a divergence curve was calculated by taking the log of the mean of all these time–distance curves. Maximum Lyapunov exponents were then calculated as the slope of this divergence curve (Rosenstein et al., 1993). This slope was estimated, using strides as time basis, over 0–0.5 strides ($\lambda_{S\text{-stride}}$) and over 4–10 strides ($\lambda_{L\text{-stride}}$), respectively (Bruijn et al., submitted for publication).

2.3.5. Orbital stability

Orbital stability was estimated using maximum Floquet multipliers (MaxFm), based on previously described techniques (e.g. Hurmuzlu and Basdogan, 1994; Dingwell et al., 2007). Floquet theory assumes that a system is strictly periodic, and that the state of a system after one cycle (S_{k+1}) is a function (F) of its current state (S_k):

$$S_{k+1} = F(S_k) \quad (2)$$

From Eq. (2) it follows that limit cycle trajectories correspond to fixed points (S^*) in the Poincaré section (that is, the lower dimensional subspace perpendicular to the flow direction of the system), i.e.:

$$S^* = F(S^*) \quad (3)$$

To evaluate effects of small perturbations on S^*, we used a linearization of Eq. (2):

$$[S_{k+1} - S^*] = J(S^*)[S_k - S^*] \quad (4)$$

From Eq. (4) it can be seen that the rate by which small perturbations grow or decay is equal to the magnitudes of the eigenvalues of $J(S^*)$ (that is, the Floquet Multipliers, FM) and thus, for a limit cycle to be stable, all FM should have a magnitude <1.

In the current study, 101 Poincaré sections were made by time normalizing the state spaces into stride cycles of 101 samples (from 0 to 100%). The fixed points in these Poincaré sections were then defined as the average over all strides in the Poincaré section in question. Magnitudes of the largest Floquet Multipliers were calculated for each % of the gait cycle. For statistical analysis, the largest FM across the stride cycle was used (MaxFm), as this represents the most unstable point in the stride cycle (Dingwell and Kang, 2007).

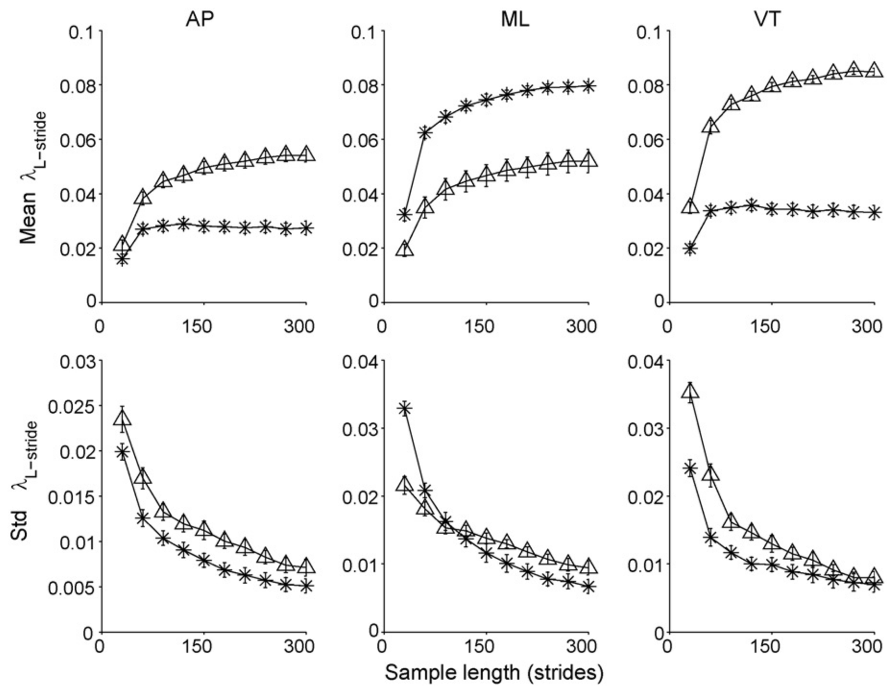

Fig. 2. Bootstrapping results for $\lambda_{L\text{-stride}}$. (*) 0.83 m/s (3 km/h) and (Δ) 1.38 m/s (5 km/h). Top panels show the mean of 100 estimates of $\lambda_{L\text{-stride}}$ for different sample lengths; bottom panels show standard deviation. Error bars represent standard errors.

3. Results

3.1. Statistical precision

Figs. 1 and 2 show the results of the bootstrapping procedure for $\lambda_{S\text{-stride}}$ and $\lambda_{L\text{-stride}}$. As can be seen, $\lambda_{S\text{-stride}}$ and $\lambda_{L\text{-stride}}$ had a clear effect of sample length for all directions; both measures started to increase (implying less stable patterns) when increasing the number of strides used in the calculation. In most cases, this increase did not appear to saturate within 300 strides; only the curves for $\lambda_{L\text{-stride}}$ for AP and VT directions started to slightly decrease between 50 and 100 strides. The standard deviation of $\lambda_{S\text{-stride}}$ and $\lambda_{L\text{-stride}}$ decreased markedly for the first few levels (up to 150 strides or so), to then decrease slightly for the higher levels.

While maximum Lyapunov exponents tended to increase when more data were used, maximum Floquet multipliers generally decreased (implying more stable patterns) with increasing sample length for all directions (Fig. 3). This decrease was seen for sample

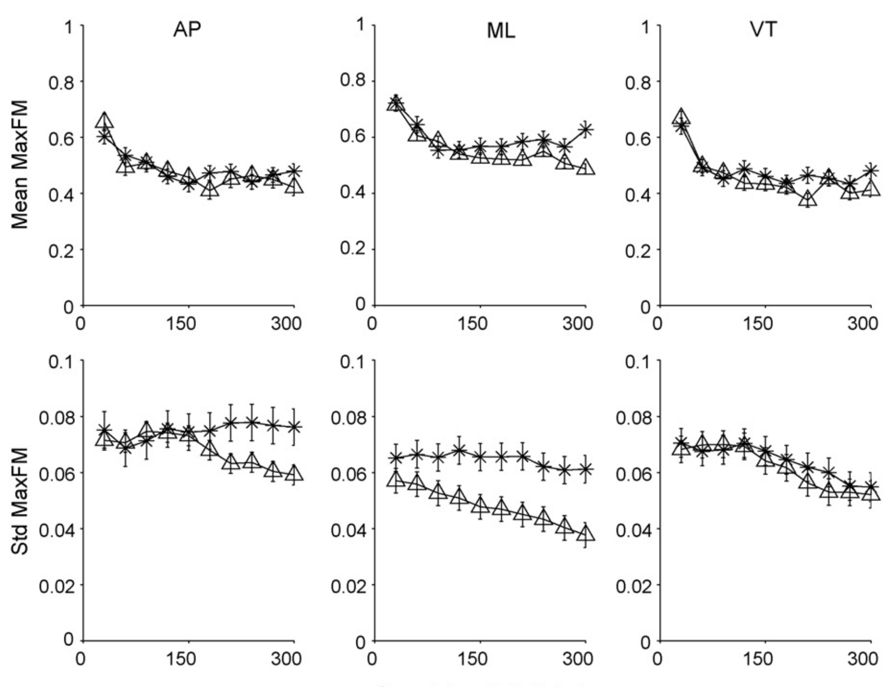

Fig. 3. Bootstrapping results for MaxFm. (*) 0.83 m/s (3 km/h) and (Δ) 1.38 m/s (5 km/h). Top panels show the mean of 100 estimates of MaxFm for different sample lengths; bottom panels show standard deviations. Error bars represent standard errors.

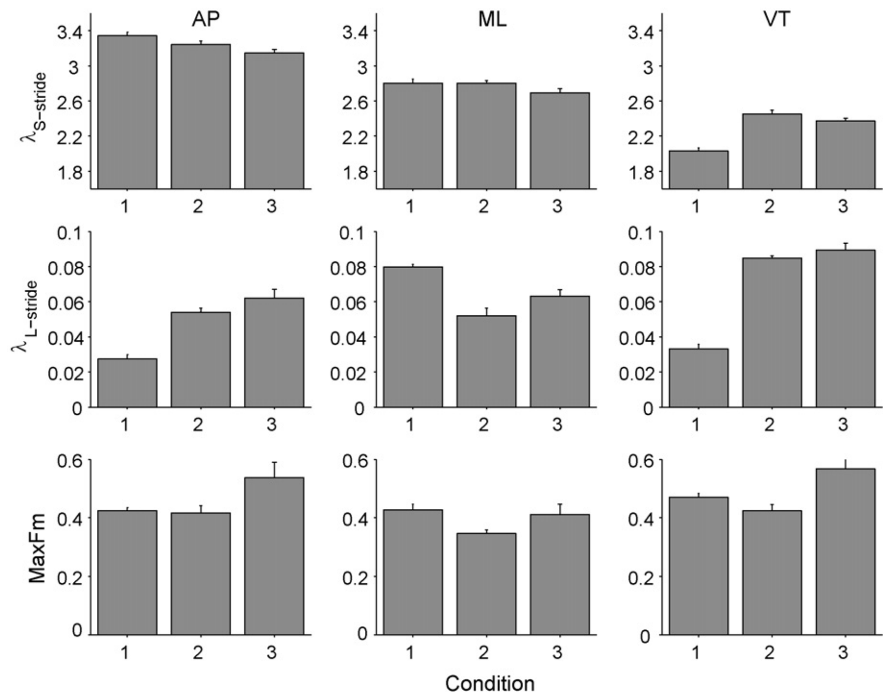

Fig. 4. $\lambda_{s\text{-stride}}$ (top panels), $\lambda_{L\text{-stride}}$ (middle panels) and MaxFM (bottom panels) for a random sample of 300 strides for the three conditions: (1) normal walking at 0.83 m/s (3 km/h), (2) normal walking at 1.38 m/s (5 km/h), (3) walking at 1.38 m/s (5 km/h) while performing a Stroop task. Error bars represent standard errors.

lengths up to about 150 strides, after which the estimate of MaxFm remained more or less constant. As for the measures of local dynamic stability, the standard deviation of the Floquet multipliers decreased with increasing sample lengths, but for MaxFm there was no clear point after which standard deviations started to decrease less.

3.2. Sensitivity

Fig. 4 gives an indication of the mean values (based on 10 estimates per subject obtained from a 300 strides data series) of the stability measures in the three experimental conditions. From this figure, it can be seen that for $\lambda_{S\text{-stride}}$ and $\lambda_{L\text{-stride}}$ the effects of

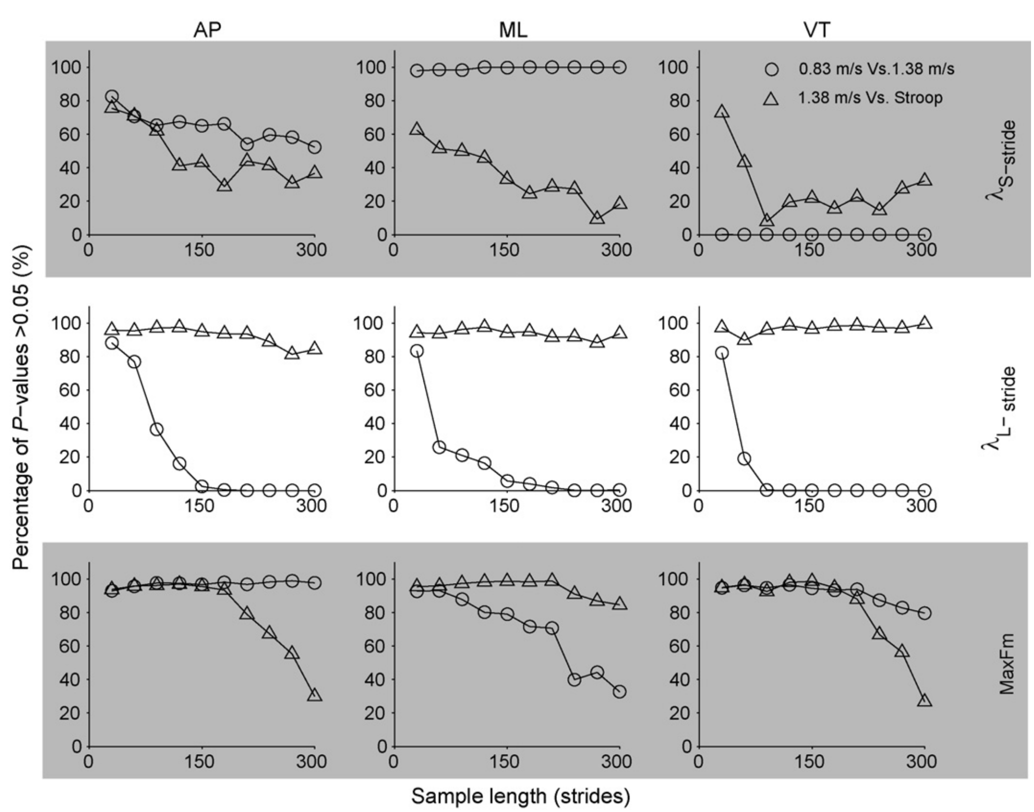

Fig. 5. The percentage of *P*-values > 0.05 (paired *t*-test) for the comparison between walking at 0.83 m/s (3 km/h) and 1.38 m/s (5 km/h) (○) and between 1.38 m/s (5 km/h) and Stroop conditions (△) when performed for different sample lengths for $\lambda_{S\text{-stride}}$ (top panels), $\lambda_{L\text{-stride}}$ (middle panels) and MaxFM (bottom panels).

walking speed and Stroop task are different per plane investigated. In addition, there were even differences in the effect of a given manipulation between $\lambda_{\text{S-stride}}$ and $\lambda_{\text{L-stride}}$. MaxFm showed more consistent effects of manipulations over the different planes.

Fig. 5 shows the percentage of comparisons between walking at 0.83 and 1.38 m/s and between walking at 1.38 m/s with and without a Stroop task yielding a non-significant result.

For the comparison between 0.83 and 1.38 m/s, $\lambda_{\text{S-stride}}$ showed a decreasing number of P-values > 0.05 with increasing sample length for the AP direction, while ML showed almost 100% of P-values > 0.05 and VT almost 0% of P-values > 0.05 for all sample lengths. For the comparison of $\lambda_{\text{S-stride}}$ between walking at 1.38 m/s with and without the Stroop task, a more or less irregularly decreasing percentage of P-values > 0.05 with increasing sample length was found for all directions.

$\lambda_{\text{L-stride}}$ showed a clear decrease in the percentage of P-values > 0.05 with increasing sample length for all directions for the comparison between walking at 0.83 and 1.38 m/s. For sample lengths longer than 150 strides there were almost no P-values > 0.05 for this comparison. When comparing walking at 1.38 m/s with and without the Stroop task, the percentage of P-values > 0.05 remained almost 100% for all sample lengths of all directions of $\lambda_{\text{L-stride.}}$

MaxFm for ML and VT directions showed a decrease in the percentage of P-values > 0.05 with increasing sample length for the comparison between 0.83 and 1.38 m/s, while AP remained almost at 100% of P-values > 0.05 for this comparison. When comparing walking at 1.38 m/s with the Stroop condition for MaxFm, a decreasing percentage of P-values > 0.05 with increasing sample length was seen for all directions. This effect was more pronounced in AP and VT directions than in the ML direction. Even when the percentage of P-values > 0.05 decreased with increasing sample length, it remained rather high in several comparisons.

4. Discussion and conclusion

We studied the effect of data series length on the statistical precision of measures of dynamic gait stability, as well as the sensitivity of these measures to changes in walking behaviour. We found that longer data series led to markedly lower standard deviations, implying more precise estimates for $\lambda_{\text{S-stride}}$ and $\lambda_{\text{L-stride}}$, but less so for MaxFm. Moreover, all variables showed an effect of data series length; when using longer data series, $\lambda_{\text{S-stride}}$ and $\lambda_{\text{L-stride}}$ increased, while MaxFm decreased. In assessing the sensitivity of these measures to changes in walking behaviour, we found that $\lambda_{\text{S-stride}}$, $\lambda_{\text{L-stride}}$, and MaxFm were sensitive to a change in walking speed in at least one direction. Only $\lambda_{\text{S-stride}}$ and MaxFm seemed sensitive enough to capture the more subtle variations in walking behaviour induced by concurrent performance of a Stroop task. Still, these variations could only be detected when using long data series (>150 strides).

Since the normal walking conditions lasted 20 min, fatigue and/or boredom may have affected the walking patterns, and thus the results of the bootstrapping procedure. However, we found no evidence for this; when comparing a random sample of data from the first 10 min with a random sample from the last 10 min for all variables, we obtained P-values ranging from 0.51 to 0.90.

4.1. Precision

We cannot exclude the possibility that the observed increase in precision reported was, at least in part, due to the increase in overlap in the samples. However, standard deviations of MaxFm decreased much slower than those of $\lambda_{\text{S-stride}}$ and $\lambda_{\text{L-stride}}$, which suggests that the increase in precision found for $\lambda_{\text{s-stride}}$ and $\lambda_{\text{L-stride}}$ was not, or at least not exclusively, caused by the overlap of the samples. All in all, it seems that a considerable number of strides are required to obtain precise estimates of the measures of interest, especially for $\lambda_{\text{S-stride}}$ and $\lambda_{\text{L-stride}}$. However, increases in precision seemed limited when using more than 150 strides.

We found that estimates of maximum Lyapunov exponents increased with the number of strides analysed. This is in agreement with findings by Kang and Dingwell (2006b), who reported that when using longer data series estimates of maximum Lyapunov exponents increased, but not with the results of Tenbroek et al. (2007), who found rather irregular changes in estimates of maximum Lyapunov exponents with increasing data series length, which may be attributed to the fact that Tenbroek et al. (2007) used only one sample per data series length. Such a change in estimates with increasing sample length may be caused by the fact that there probably exists an inverse relation between the proximity of nearest neighbours and the rate of their divergence for a given level of variability. Nearest neighbours that are initially very far apart cannot diverge all that far, as this would increase variability; however, close nearest neighbours can diverge more. Of course the probability of the existence of very close nearest neighbours also increases as the number of points in the state space increases, thus probably increasing the total rate of divergence. Alternatively, the changes with increasing sample length may be signs of the existence of processes in walking that can only be captured using more strides, which is in line with the long range correlations in stride intervals as found by several authors (e.g. Hausdorff et al., 1995), which may (e.g. Hausdorff et al., 1995) or may not (Gates et al., 2007) have a physiological basis. Note that it is unlikely that maximum Lyapunov exponents will keep increasing with further increases in data series length, as this eventually would suggest instantaneous divergence. Thus, it seems that a “true” value of the maximum Lyapunov exponent for walking may exist, although more than 300 strides may be required to estimate it.

For MaxFm, we found the opposite pattern: as more strides were analysed, MaxFm tended to become smaller. To our knowledge, this dependence of MaxFm on data series length has never been reported. However, if we view MaxFm as a measure of convergence towards an attractor (that is, the fixed point in the Poincaré section, S^*), also this result is understandable: using less data leads to less accurate estimates of the true attractor, and thus to less convergence to this incorrectly estimated attractor. Again, it is unlikely that using more data at some point leads to instantaneous return to the attractor, and thus, it is likely that a true maximum Floquet multiplier for human walking exists, and inspection of Fig. 3 appears to suggest that this value can be estimated within 300 strides.

4.2. Sensitivity

We found trends towards 0% of P-values > 0.05 for the t-tests between conditions in several comparisons, suggesting that walking behaviour was indeed altered. The effects of walking speed (see Fig. 4) found in the present study are in agreement with our earlier study (Bruijn et al., submitted for publication), including the fact that we did not find any changes in $\lambda_{\text{S-stride}}$ for the ML direction with increasing velocity, since we previously found that $\lambda_{\text{S-stride}}$ and $\lambda_{\text{L-stride}}$ exhibited a quadratic relationship with walking velocity for the ML direction (Bruijn et al., submitted for publication). Moreover, the observed effects of the Stroop task (see Fig. 4) on $\lambda_{\text{S-stride}}$ MaxFm are in agreement with the findings by Dingwell et al. (2008), but the effects of the Stroop task on $\lambda_{\text{L-stride}}$ seem different, which may indicate the presence of a type I error in the Dingwell et al. (2008) study.

We found that several comparisons showed a decrease in percentage of P-values > 0.05 with increasing sample length. If we assume that the trend toward a lower percentage of P-values > 0.05 indicates the presence of a real effect, the percentage of P-values > 0.05 may be regarded as the probability of making a type II error. Given that this percentage is rather high for some of the comparisons, it may be necessary to measure multiple (longer) data series (cf. Kang and Dingwell, 2006b) to reduce this possibility. Moreover, increasing the number of subjects will probably also lead to a decrease in type II error. Still, it is somewhat surprising to see that several comparisons (e.g. the walking speed comparison in the AP direction for $\lambda_{\text{S-stride}}$) show 50% of P-values > 0.05, and it remains an open question how this is possible, and whether these comparisons would eventually tend to 0% of P-values > 0.05 if more data would be used.

Similar to the findings of the current study regarding the statistical precision of these measures, and in relation to this, it appears that a considerable number of strides are needed to reveal the effects of any manipulation. Although the current study only tested sensitivity to two very specific manipulations, we are inclined to believe that these represent two extremes; walking speed was shown to have considerable effects, even when using very few strides (e.g. England and Granata, 2007), while the effects of a Stroop test seemed rather small, or even absent (Dingwell et al., 2008). Thus, although the results of the current study cannot be simply extrapolated to the study of other independent variables, we believe they are consequential for the design of studies using other manipulations.

4.3. Implications of the current study

The dependence of the estimates of local and orbital dynamic stability upon the number of strides included in the analysis implies that when estimating stability at different walking speeds (Dingwell and Marin, 2006; England and Granata, 2007), or in different patient groups (Dingwell and Cavanagh, 2001; Dingwell and Cusumano, 2000; Dingwell et al., 1999, 2000, 2007; Granata and Lockhart, 2008; Hurmuzlu et al., 1996), a fixed number of strides should be analysed. After all, (experimentally or pathologically induced) variations in cadence will yield different numbers of strides for the same time interval. The increase in precision with increasing data series length indicates the need to use long data series. How many strides are needed exactly for a particular study depends on its goal and the expected within and between group variances, but the present results may be of help in the design of future studies. Most notably, the present results clearly indicate that the gain in precision tends to be limited when using more than 150 strides. On the other hand, the results from the sensitivity analysis revealed that type II errors remain likely, even when using longer data series. This would suggest that it may be necessary to measure multiple trials to reduce this possibility (cf. Kang and Dingwell, 2006b).

Acknowledgements

This study was partly funded by a grant from Biomet Nederland. Joost Haeck assisted in the data collection.

References

Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. Health Econ 1997;6:327–40.
Bruijn SM, Dieën JH, Meijer OG, Beek PJ. Stability and variability in treadmill walking: is slow walking more stable?; submitted for publication.
Dingwell JB, Cavanagh PR. Increased variability of continuous overground walking in neuropathic patients is only indirectly related to sensory loss. Gait Posture 2001;14:1–10.
Dingwell JB, Cusumano JP. Nonlinear time series analysis of normal and pathological human walking. Chaos 2000;10:848–63.
Dingwell JB, Cusumano JP, Sternad D, Cavanagh PR. Slower speeds in patients with diabetic neuropathy lead to improved local dynamic stability of continuous overground walking. J Biomech 2000;33:1269–77.
Dingwell JB, Gu KH, Marin LC. The effects of sensory loss and walking speed on the orbital dynamic stability of human walking. J Biomech 2007;40:1723–30.
Dingwell JB, Kang HG. Differences between local and orbital dynamic stability during human walking. J Biomech Eng 2007;129:586–93.
Dingwell JB, Marin LC. Kinematic variability and local dynamic stability of upper body motions when walking at different speeds. J Biomech 2006;39:444–52.
Dingwell JB, Robb RT, Troy KL, Grabiner MD. Effects of an attention demanding task on dynamic stability during treadmill walking. J Neuroeng Rehabil 2008;5:12.
Dingwell JB, Ulbrecht JS, Boch J, Becker MB, O'Gorman JT, Cavanagh PR. Neuropathic gait shows only trends towards increased variability of sagittal plane kinematics during treadmill locomotion. Gait Posture 1999;10:21–9.
Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci 1986;1:54–77.
England SA, Granata KP. The influence of gait speed on local dynamic stability of walking. Gait Posture 2007;25:172–8.
Gates DH, Su JL, Dingwell JB. Possible biomechanical origins of the long-range correlations in stride intervals of walking. Physica A 2007;380:259–70.
Granata KP, Lockhart TE. Dynamic stability differences in fall-prone and healthy adults. J Electromyogr Kinesiol 2008;18:172–8.
Hausdorff JM, Peng CK, Ladin Z, Wei JY, Goldberger AL. Is walking a random walk? Evidence for long-range correlations in stride interval of human gait. J Appl Physiol 1995;78:349–58.
Hurmuzlu Y, Basdogan C. On the measurement of dynamic stability of human locomotion. J Biomech Eng 1994;116:30–6.
Hurmuzlu Y, Basdogan C, Stoianovici D. Kinematics and dynamic stability of the locomotion of post-polio patients. J Biomech Eng 1996;118:405–11.
Kang HG, Dingwell JB. A direct comparison of local dynamic stability during unperturbed standing and walking. Exp Brain Res 2006a;172:35–48.
Kang HG, Dingwell JB. Intra-session reliability of local dynamic stability of walking. Gait Posture 2006b;24:386–90.
Kantz H, Schreiber T. Nonlinear Time Series Analysis. Cambridge, New York: Cambridge University Press; 1997.
Kennel MB, Brown R, Abarbanel HDI. Determining minimum embedding dimension using geometrical construction. Phys Rev A 1992;45:3403–11.
Mees AI, Judd K. Dangers of geometric filtering. Physica D 1993;68:427–36.
Rosenstein MT, Collins JJ, Deluca CJ. A practical method for calculating largest Lyapunov exponents from small data sets. Physica D 1993;65:117–34.
Stroop JR. Studies of interference in serial verbal reactions. J Exp Psychol 1935;18:643–62.
Tenbroek TM, van Emmerik REA, Hasson JH. Lyapunov exponent estimation for human gait acceleration signals. In: ISB 2007; 2007.
van Dieën JH, Hoozemans MJ, van der Beek AJ, Mullender M. Precision of estimates of mean and peak spinal loads in lifting. J Biomech 2002;35:979–82.