Possibilities for real-time DFA based injury detection and skill level differentiation.

M. Norris*a, I.C. Kennya, R. Healya & R. Andersona

*Biomechanics Research Unit, University of Limerick, Ireland.

Abstract

Detrended fluctuation analysis (DFA) is applied to running stride time series for the detection of long range correlations. However, there is no concurrent real-time data collection and DFA feedback mechanism. We aimed to verify a running analysis system for production of real-time DFA α values. Data were collected utilising an accelerometer, attached to the tibia. The accelerometer data were transmitted to MATLAB for processing. Results demonstrated DFA α value output < 2 seconds post individual running Epochs, and < 5 seconds post run completion. Our running analysis system provides rapid advanced variability information, important in both training, and injury prevention.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Feedback; gait; locomotion; variability; stride.

1. Introduction

Due to their wireless and lightweight nature accelerometers have become increasingly popular in running gait analysis [1], and are an integral part of research and consumer-led wearable running devices such as sports bracelets, smartwatches and smart clothing [2]. Accelerometers are commonly embedded within these devices to provide feedback to consumers on parameters such as distance and speed [3], and stride frequency [4] during running.

While these are basic running parameters recently there has been increased interest in measuring running variability through investigating long range correlations in stride time series using the advanced statistical analysis, Detrended Fluctuation Analysis (DFA) [5]. DFA is quantified through output of an α value. When applied to stride time series α values of 0.5 indicate a random walk, where one stride is uncorrelated to previous strides, whilst α values of between 0.5 and 1 indicate the presence of long range correlations, where stride time patterning continues over a number of strides resulting in predictable pattern [6]. Previously, DFA α values have been found to distinguish training status of runners [7], and identify previously injured runners from their previously non-injured counterparts [6]. It could therefore provide vital information to researchers and recreational runners alike as to changes in stride time variability due to training or perhaps onset of injury.

However, DFA requires extended data collection due to its investigation into long range correlations, and studies investigating DFA on running populations have typically collected greater than 500 strides [6],[7]. Therefore, to complete DFA increased post-processing capabilities are needed along with advanced statistical knowledge for correct implementation. This has led to DFA being underutilised within laboratory settings, and never utilised in consumer wearable running devices or real-time feedback environments.

Thus, the aim of this research is to develop an advanced running analysis system, utilising accelerometry and the PhysioNet DFA algorithm [8], to provide real-time DFA α values. Whilst real-time feedback in sporting technology has been described as “a method that allows participants to observe their movements for the purpose of making immediate biomechanical adjustments” [9], DFA does not investigate discrete values which are susceptible to immediate adjustments. Therefore, longer periods of data collection prior to result output is more appropriate within this system when referring to “real-time” DFA α value output.

If verified this system could enhance researcher and coach data collection alike, in terms of rapid data output and statistically
advanced running variability information.

2. System Description

The advanced running analysis system is comprised of a Shimmer 2r accelerometer (range ± 6 g, sensitivity = 200 mV/g) (SHIMMER™, Dublin, Ireland) and a laptop equipped with Bluetooth capability, MATLAB (Mathworks, Cambridge, UK) and the PhysioNet C+ DFA programme [8] (Fig 1).

2.1 Pre-data collection process

The system is designed to be adapted to research interests, along with the individual running styles of the participants and therefore there are a number of user inputs which can be varied prior to data collection to enhance the accuracy of data output (Table 1). Prior to data collection the Shimmer 2r accelerometer should be programmed with Bluetooth streaming capability and paired with the laptops appropriate comport, to ensure real-time data streaming via the Shimmer MATLAB Instrument Driver.

Table 1. Pre-data collection user input parameters.

<table>
<thead>
<tr>
<th>Parameter Title</th>
<th>Parameter Explanation</th>
<th>Input Example or Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name_</td>
<td>Enter name of subject or experiment bordered by apostrophes.</td>
<td>Any input eg, ‘John’</td>
</tr>
<tr>
<td>comPort</td>
<td>Designated comPort due to Shimmer Bluetooth pairing, enter number bordered by apostrophes.</td>
<td>Number designated by pairing eg, ‘9’</td>
</tr>
<tr>
<td>Epoch_1</td>
<td>Enter length of 1st running Epoch in seconds. *Must be a long enough period to collect &gt; 400 running strides for valid DFA result for Epoch 1.</td>
<td>Any number of seconds eg, 480</td>
</tr>
<tr>
<td>Trials</td>
<td>Enter number of DFA values and/or number of overlapping windows, maximum 10.</td>
<td>Any value 1-10 eg, 10</td>
</tr>
<tr>
<td>TrialPeriod</td>
<td>Enter length of each overlapping window in seconds.</td>
<td>Any value eg, 60</td>
</tr>
<tr>
<td>fs</td>
<td>Enter sampling frequency of Shimmer 2r.</td>
<td>Must be a multiple of 51.2 and ≤ 1024 eg, 51.2 OR 102.4 ... 1024</td>
</tr>
<tr>
<td>fc</td>
<td>Filtering cutoff for butterworth lowpass filter</td>
<td>Any value, eg 2</td>
</tr>
<tr>
<td>Order</td>
<td>Enter butterworth lowpass filter order</td>
<td>2 OR 4</td>
</tr>
<tr>
<td>SignalAnalysis</td>
<td>Axis which strides will be derived from (bordered by apostrophes)</td>
<td>‘Filtered_X’ OR ‘Filtered_Y’ OR ‘Filtered_Z’</td>
</tr>
<tr>
<td>FolderName</td>
<td>Enter folder pathway (ending in \ and bordered by apostrophes) to save pre-processed files in.</td>
<td>Any folder pathway eg, ‘C:\Users\John\Documents\MATLAB\data’</td>
</tr>
<tr>
<td>SaveFolderName</td>
<td>Enter folder pathway (ending in \ and bordered by apostrophes) to save post-processed files in.</td>
<td>Any folder pathway eg, ‘C:\Users\John\Documents\MATLAB\data’</td>
</tr>
</tbody>
</table>

2.2 Data collection initiation

Firstly, a Shimmer 2r accelerometer is turned on and attached to the participants’ right distal anterio-medial tibia (Fig. 1, A). Participants are then given a treadmill warm-up period in which they are allowed to reach the required running speed. On having
reached their running speed the system user/researcher then runs the custom developed ‘DFA_running_analysis.m’ script. This script contains the Shimmer MATLAB Instrument Driver streaming, plotting and text writing capabilities, along with a data analysis function to perform data processing to create a stride time series and run the PhysioNet DFA C+ Programme to calculate a DFA α value (Fig. 1, B), at user specific time intervals (defined in pre-data collection user input). The Shimmer accelerometer then starts logging tri-axial data which it transmits, via Bluetooth, to a laptop situated within 5m of the treadmill, for a designated capture period, (Fig. 1, A and B). When accelerometry data begins streaming to MATLAB a timer is automatically generated incrementing elapsed time.

![Fig 1. System image](image)

The participant runs continuously for the designated capture period,

\[
\text{capture period } = (\text{Trials} \times \text{TrialPeriod}) + \text{Epoch}_1 \tag{1}
\]

### 2.3 Accelerometry text file generation.

During this capture period accelerometry data is written to 11 text files, 'Epoch_1' + 10 trials named 'Epoch_2', 'Epoch_3' etc. The end of each Epoch period is defined as,

\[
\text{Epoch}_2 = \text{Epoch}_1 + \text{TrialPeriod}, \\
\text{Epoch}_3 = \text{Epoch}_2 + \text{TrialPeriod etc.} \tag{2}
\]

As previously stated, when the user runs the ‘DFA_running_analysis.m’ script there is an inbuilt data analysis function to perform data processing and DFA. This data analysis function requires 2 - 3 seconds to perform data analysis, calculate and display the DFA α value on a message dialog box. To allow for this, variables \(A1 - A11\) are created automatically to identify the point of at which each Epoch ceases writing to text file.

\[
\begin{align*}
A1 &= \text{Epoch}_1 - 2, \\
A2 &= \text{Epoch}_2 - 2 \text{ etc.} \tag{3}
\end{align*}
\]

Therefore, the first data file containing accelerometry data, 'Epoch_1', is generated and continuously appended to whilst,

\['\text{Epoch}_1' = 0 \leq \text{elapsed time} < A1 \tag{4}\]

The period in which the remaining text files are written to, 'Epoch_2', 'Epoch_3' etc, are defined by,

\[
\begin{align*}
\text{Epoch}_2 &= \text{Epoch}_1 \leq \text{elapsed time} < A2, \\
\text{Epoch}_3 &= \text{Epoch}_2 \leq \text{elapsed time} < A3 \text{ etc.} \tag{5}
\end{align*}
\]

Therefore, there is a 2 second period within each Epoch in which accelerometry data are streamed, but not written to file and therefore will not be included in subsequent analyses. The effect of this 2 second loss of recorded data is minimal, since using the example of the minimal suggested Epoch length of \(\text{Epoch}_1 = 8 \text{ minutes}\) (Fig 3.), it results in a 0.4 – 3.3% loss of data within an 8 minute Epoch and a 2% loss of data in an overall capture period of 18 minutes.

When \(\text{elapsed time} \geq \text{capture period}\) acceleration data collection ceases.
2.4 Data processing and DFA calculation.

The data analysis and DFA initiates 1 second post variables A1 – A11 allowing for sufficient accelerometry collection time and text files to cease writing before processing. Therefore data analysis occurs at,

\[ A_{-1} \leq \text{elapsed time} < (A_{-1} + 1), \]
\[ A_{-2} \leq \text{elapsed time} < (A_{-2} + 1) \text{ etc.} \] (6)

At each data analysis initiation, firstly all accelerometry text files are concatenated to previous accelerometry files. This merged file is then cut from the beginning down to mirror ‘Epoch_1’, using Fraction = TrialPeriod / Epoch_1, multiplied by the length of Epoch_1. At each data analysis initiation this creates overlapping accelerometry data files, which all contain an acceleration collection period equal to Epoch_1 (Fig.2).

![Fig. 2. Process of data merging and cutting.](image)

This merged file undergoes filtering, using a Butterworth Filter, with pre-data collection user input defined frequency cut off and filter order. Then, a stride time series is produced using peak identification in the axis chosen through pre-data collection user input. Due to the process of merging multiple acceleration files which contain 2 seconds of missing data and therefore not continuous data there is the possible creation of stride times not representative of actual strides. Therefore within the data analysis there are stride time threshold boundaries, which can be altered by the user, to eliminate stride time outliers. Lastly, the PhysioNet DFA C+ programme generates a DFA α value from the calculated stride time series, which is displayed automatically in a message dialog window (Fig. 1, C).

3. Programme verification experimental protocol

To verify the advanced running analysis programme generates real-time DFA α values in an efficient manner and in line with expected α results, a healthy active subject (female, age: 26.6 years, height: 1.80 m, mass: 70.1 kg) performed a running protocol whilst completing the advanced analysis programme. Firstly, an 8 minute period for Epoch_1 was selected, as it suggested by the current researchers as the minimum length of time required to allow enough time within you are guaranteed to collect over 500 strides, regardless of running speed. Further pre-data collection user input parameters were as follows, Trials = 11, TrialPeriod = 1 min, fs = 102.4, fc = 2, Order = 4, SignalAnalysis = ‘Filtered_Z’ (Fig. 3). Utilizing a 2 Hz filter cut off to derive stride time has previously been validated [11]. Along with this a stride time upper threshold was identified as 0.8 seconds with a lower threshold of 0.6 seconds, based on previous literature investigating both runners and non-runners [7].

The running protocol used aimed to verify the analysis programme over a range of running speeds. For this, the participant ran at their preferred running speed (PRS), 80% of their PRS and 120% of their PRS. To establish the participant’s PRS the same protocol as that used by Nakayama et al. [7] and Jordan et al. [5] was employed,. In short, the participant ran at a range of speeds which they indicated were “comfortable” or “uncomfortable”. The average speed of the “comfortable” speeds was estimated as the participant’s PRS. From the participant’s PRS, 120% and 80% of their PRS were calculated. The participant was then required to run for 18 minutes at 80% PRS, 100% PRS and 120% PRS in randomized order. The participant was allowed as long as necessary to rest between runs to mitigate the effect of fatigue. This resting period was further supported by a return to resting heart rate as confirmed by a heart rate monitor.
To verify the system met the specified requirement, real-time output of repeated DFA $\alpha$ values, the time difference in seconds ($\Delta t$) between the end of each Epoch and the related $\alpha$ value display time were calculated, via a record of the MATLAB Command Window displaying elapsed time and each data analysis. To verify the system produced reliable $\alpha$ values over a range of running speeds $\alpha$ values were also recorded and compared to previous literature.

4. Results & Discussion

The participant’s PRS was 2.8 m/s, with 80% of PRS 2.2 m/s and 120% of PRS 3.3 m/s, similar to that previously seen in trained runners (PRS of 3.0 m/s) [7]. The running analysis programme resulted in collection of 653 – 711 stride time values within a data analysis epoch (A1- A11) across all running speeds (average 663 ± 6 for 80% of PRS, average of 677 ± 3 for 100% of PRS and average of 704 ± 3 for 120% PRS). This is similar to the amount of stride time values collected by both Nakayama et al. [7] (512 strides time values) and Meardon et al. [6] (661 stride time values) when investigating running DFA. The number of overall stride time values (A12) collected within the capture period (1,490 for 80% of PRS, 1,526 for 100% of PRS and 1,591 for 120% of PRS) are greater amounts than any represented in the current literature of DFA in running.

In relation to the real-time output of DFA $\alpha$ values, all $\alpha$ values were displayed to the researcher within 0.83 – 2.19 seconds of Epoch completion time (average of 1.49 ± 0.41 seconds for 80% of PRS, average of 1.55 ± 0.34 seconds for 100% of PRS, and average of 1.28 ± 0.32 seconds for 120% PRS). For 80% of PRS and 100% of PRS average stride intervals of 0.71 seconds and 0.70 seconds were identified, which indicates DFA $\alpha$ value output within 3 running strides of the next successive Epoch, at these running speeds. For 120% of PRS an average stride interval of 0.67 was identified, which indicates DFA $\alpha$ value output also within 2 running strides of the next successive Epoch. Post overall run, or at cessation of the capture period, DFA $\alpha$ value output occurred within 5 seconds, across all running speeds (average 3.61 ± 1.03 seconds). All previous literature around running DFA [5] [6] [7] appears to indicate that DFA occurred within a post processing environment and therefore the current author believes this may be the first running analysis system to produce real-time DFA output.

Lastly, DFA $\alpha$ values ranged 0.70 – 0.86, within the data analyses epochs (A1- A11), across all running speeds. Whilst our results are similar to those found in previously injured runners (average $\alpha$ values of 0.68 – 0.92) [6], Meardon et al. [6] were unable to determine that previous injury and the $\alpha$ values they reported were explicity linked. They suggest further investigation into the clinical interpretation of DFA $\alpha$ values, which this system may provide. However, these results are similar to those found by Jordan et al. [5], who identified $\alpha$ values of 0.70 – 0.90 whilst running at various percentages of PRS. This may verify our advanced running analysis system generates valid DFA $\alpha$ values over a range of running speeds. Interestingly, we also found that the participants overall run $\alpha$ was lowest at 100% of PRS (0.80, compared to 0.85 at 80% of PRS and 0.92 at 120% of PRS). This was previously identified by Jordan et al. [5], and is explained as a runner being most adaptable, and therefore less predictible in their stride time, at their PRS. This further supports our system within a training and skill level identification setting, as it appears the system is able to detect $\alpha$ value differences previously identified within the literature.
Table 2. Number of strides (n), difference in elapsed time and $\alpha$ value output time (secs) $\Delta t$, and DFA $\alpha$ values over three running conditions at 80% PRS, PRS and 120% PRS.

<table>
<thead>
<tr>
<th>Analysis No.</th>
<th>Strides (n)</th>
<th>$\Delta t$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>80% PRS</td>
<td>PRS</td>
<td>120% PRS</td>
</tr>
<tr>
<td>A1</td>
<td>669</td>
<td>679</td>
<td>711</td>
</tr>
<tr>
<td>A2</td>
<td>671</td>
<td>679</td>
<td>709</td>
</tr>
<tr>
<td>A3</td>
<td>669</td>
<td>680</td>
<td>706</td>
</tr>
<tr>
<td>A4</td>
<td>668</td>
<td>680</td>
<td>704</td>
</tr>
<tr>
<td>A5</td>
<td>665</td>
<td>679</td>
<td>703</td>
</tr>
<tr>
<td>A6</td>
<td>663</td>
<td>679</td>
<td>702</td>
</tr>
<tr>
<td>A7</td>
<td>661</td>
<td>678</td>
<td>702</td>
</tr>
<tr>
<td>A8</td>
<td>658</td>
<td>677</td>
<td>701</td>
</tr>
<tr>
<td>A9</td>
<td>656</td>
<td>676</td>
<td>702</td>
</tr>
<tr>
<td>A10</td>
<td>656</td>
<td>673</td>
<td>702</td>
</tr>
<tr>
<td>Average</td>
<td>663</td>
<td>677</td>
<td>704</td>
</tr>
<tr>
<td>(± stdev)</td>
<td>(± 6)</td>
<td>(± 3)</td>
<td>(± 3)</td>
</tr>
</tbody>
</table>

Overall Run

A12          | 1490       | 1526       | 1591     | 2.68      | 4.72      | 3.44      | 0.85      | 0.80      | 0.92      |

5. Conclusion

The aim of this research was to verify a running analysis system allowing the concurrent collection of stride time series and real-time DFA $\alpha$ value output. This aim was achieved as the running analysis system resulted in DFA $\alpha$ value output < 2 secs post Epoch and < 5 seconds post overall run. This provides access to further feedback information important in both a training and injury prevention context, for coaches and researchers alike. The programme also supports the future use of inertial sensors in a sports engineering context, as advances in Bluetooth technology will lead to further development of the system, with an aim of performing advanced stride variability analysis in an outdoor, ecological environment.

Acknowledgements

This work was supported by the Education and Health Sciences Faculty, University of Limerick.

References