Human Movement Science 44 (2015) 91–101

Contents lists available at ScienceDirect

Human Movement Science

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

Do intermediate- and higher-order principal components contain useful information to detect subtle changes in lower extremity biomechanics during running?

Angkoon Phinyomark [a], Blayne A. Hettinga [a,c], Sean Osis [a,c], Reed Ferber [a,b,c,*]

[a] Faculty of Kinesiology, University of Calgary, Calgary, Alberta, Canada
[b] Faculty of Nursing, University of Calgary, Calgary, Alberta, Canada
[c] Running Injury Clinic, Calgary, Alberta, Canada

ARTICLE INFO

Article history:
Received 30 May 2014
Revised 25 August 2015
Accepted 26 August 2015

PsycINFO Classification:
2240
4120

Keywords:
Gait analysis
Kinematics
Linear discriminant analysis
Patellofemoral pain
Principal component analysis
Support vector machine

ABSTRACT

Recently, a principal component analysis (PCA) approach has been used to provide insight into running pathomechanics. However, researchers often account for nearly all of the variance from the original data using only the first few, or lower-order principal components (PCs), which are often associated with the most dominant movement patterns. In contrast, intermediate- and higher-order PCs are generally associated with subtle movement patterns and may contain valuable information about between-group variation and specific test conditions. Few investigations have evaluated the utility of intermediate- and higher-order PCs based on observational cross-sectional analyses of different cohorts, and no prior studies have evaluated longitudinal changes in an intervention study. This study was designed to test the utility of intermediate- and higher-order PCs in identifying differences in running patterns between different groups based on three-dimensional bilateral lower-limb kinematics. The results reveal that differences between sex- and age-groups of 128 runners were observed in the lower- and intermediate-order PCs scores ($p < 0.05$) while differences between baseline and following a 6-week muscle strengthening program for 24 runners with patellofemoral pain were observed in the higher-order PCs scores ($p < 0.05$), which exhibited a moderate correlation with self-reported pain scores ($r = -0.43$; $p < 0.05$).

1. Introduction

Biomechanical gait investigations have typically analyzed discrete 3-dimensional (3D) kinematic variables using traditional univariate and multivariate statistical techniques in an attempt to understand gait patterns for runners at an increased risk of injury. For example, studies have investigated differences between male and female runners (Ferber, Davis, & Williams, 2003), differences between young and elderly runners (Lilley, Stiles, & Dixon, 2013), and differences in gait patterns between healthy and injured runners, such as those experiencing iliotibial band syndrome (ITBS) (Ferber, Noehren, Hamill, & Davis, 2010; Noehren, Davis, & Hamill, 2007; Phinyomark, Osis, Hettinga, Leigh, & Ferber, 2014b) or patellofemoral

* Corresponding author at: Running Injury Clinic, University of Calgary, 2500 University Drive NW, Calgary, Alberta T2N 1N4, Canada.
E-mail address: rferber@ucalgary.ca (R. Ferber).

http://dx.doi.org/10.1016/j.humov.2015.08.018

pain (PFP) (Earl & Hoch, 2011; Ferber, Kendall, & Farr, 2011; Willy, Manal, Witvrouw, & Davis, 2012). However, these studies have reported that following a muscle strengthening protocol, significant increases in muscle strength and decreases in self-reported pain were not accompanied by concomitant changes in discrete kinematic angles, as compared with their baseline values. Therefore, different statistical methods, such as a principal component analysis (PCA) approach, have been recently used to provide insight into changes in gait pathomechanics (Eskofier, Federolf, Kugler, & Nigg, 2013; Foch & Milner, 2014; Soares, de Castro, Mendes, & Machado, 2014).

In brief, in gait analysis a PCA converts a set of original variables, such as discrete biomechanical variables (Olney, Griffin, & McBride, 1998), gait waveforms (Deluzio & Astephen, 2007), or marker trajectories (Federolf, Boyer, & Andriacchi, 2013), into a set of linearly uncorrelated variables called the principal components (PCs). This approach has demonstrated classification rates between 80% and 99% in identifying differences in running gait patterns between different cohorts (Deluzio & Astephen, 2007; Federolf et al., 2013; Reid, Graham, & Costigan, 2010). Often, however, researchers account for nearly all of the between-group variance from the original data using the first few, or lower-order, PCs (Bennett, Russell, Sheth, & Abel, 2010; Lamoth, Daffertshofer, Huys, & Beek, 2009; Liebl, Willwacher, Hamill, & Brüggemann, 2014; Sadeghi, Allard, & Duhaime, 1997; Witte, Schobesberger, & Peham, 2009). In these analyses, only a reduced set of the first few PCs are generally retained and interpreted, whereas the remaining set of intermediate- and higher-order PCs are ignored. However, when using PCs in discriminant analysis, there is no guarantee that the difference between the groups of interest will be in the direction of the first few, or high-variance, PCs. For example, Jolliffe (2002, chap. 2, 3, 6, 8–10, pp. 201–204) showed that the first few PCs are useful for identifying differences between groups only in the case where the between-group variation is much larger than within-group variation, and within- and between-group variation have the same dominant directions. If this situation does not occur, excluding the higher-order or low-variance PCs may actually remove most of the information in the original data concerning between-group variation. These authors also suggested that when classifying between two distinct groups, the best subset of PCs do not necessarily consist of those with the largest variances (Jolliffe, 2002, chap. 2, 3, 6, 8–10, p. 206) and one should consider intermediate- and higher-order PCs.

Intermediate- and higher-order PCs have been generally associated with minor or subtle movement patterns of running but may contain valuable information about between-group variation and specific test conditions (Daffertshofer, Lamoth, Meijer, & Beek, 2004; Lamoth, Daffertshofer, Meijer, & Beek, 2006; Maurer, Federolf, von Tscharner, Stirling, & Nigg, 2012). Jolliffe (2002, chap. 2, 3, 6, 8–10, pp. 11–13) demonstrated the mathematical and statistical property of the higher-order PCs (i.e., Property A2) based on the assumption that the higher-order PCs are not simply unstructured left-overs after removing the large-variance PCs and involve linear functions of the original variables whose variances are as small as possible yet are uncorrelated with previous linear functions. These higher-order PCs, because of their small variances, may therefore be useful to detect unsuspecting near-constant linear relationships between variables (Jolliffe, 2002, chap. 2, 3, 6, 8–10, pp. 27, 43, 56) and can still be highly correlated with the dependent variable (Jolliffe, 2002, chap. 2, 3, 6, 8–10, p. 202).

The utility of higher-order PCs for classification has been reported to be useful for separating two different populations (Chang, 1983) as well as thematic 3D land mapping (Townshend, 1984) wherein it has been shown that significant discriminatory power may be lost if all available PC data are not used. However, specific to running gait analysis, Nigg, Baltich, Maurer, and Federolf (2012) used only lower-order PCs and reported that three, nonconsecutive PCs (8, 9, and 19), from the first 20 PCs, exhibited large effect sizes when attempting to separate male and female runners while PCs 1 and 2 exhibited large effect sizes in separating age groups. Similar results were also found when separating data across shoe midsole hardness conditions using PCs 3, 5, 6, and 19. These authors reported that these lower-order PCs were related to sagittal and frontal plane kinematics and were the "dominant movement components of running." In contrast, Maurer et al. (2012) analyzed a complete set of lower-, intermediate-, and high-order PCs and reported significant differences between male and female runners within intermediate-order PCs (10, 12–14, 20, 21, 29, 30, 34, 36, and 41) but significant differences between shoe conditions were only found in higher-order PCs. Finally, in order to achieve the maximum classification accuracy between young and elderly gait patterns, Eskofier et al. (2013) reported that 75–81.25% of the total number of PCs were required for a support vector machine (SVM) approach. However, these aforementioned running studies were observational cross-sectional analyses of different cohorts and to our knowledge, no study has experimentally tested the utility of higher-order PCs.

Focusing only on lower-order PCs, which represent the dominant movement patterns of running gait, may exclude important information necessary for class separability between groups of interest. Therefore, the purpose of this study was to determine whether intermediate- and higher-order PCs can identify the differences in gait patterns between groups of runners based on discrete 3D lower extremity bilateral joint kinematic angles. This study was designed to test three hypotheses: (1) differences in running biomechanics before and after a 6-week exercise program, for runners with PFP, can be observed in the higher-order, or low-variance, PCs scores (i.e., linear combinations of subtle changes in discrete kinematic angles between groups). In order to provide interpretation that is clinically relevant to PCs, we based our analysis on the construct validity that has been established in previous studies by comparing the PC results to clinical pain outcome measures (Deluzio, Wyss, Costigan, Sorbie, & Zee, 1999; Deluzio, Wyss, Zee, Costigan, & Sorbie, 1997). The second hypothesis was that (2) clinical reductions in self-reported PFP pain, based on Visual Analog Scale (VAS) scores, would be correlated to changes in the higher-order PC scores, as compared to baseline values.

Further, considering the well-established and significant differences in lower-extremity kinematics between male and female runners, as well as between young and elderly runners, we also hypothesized that (3) gender- and age-related dif-