Interpreting principal components in biomechanics: Representative extremes and single component reconstruction

Scott C.E. Brandon a,b,⁎, Ryan B. Graham c, Sivan Almosnino b,d, Erin M. Sadler d, Joan M. Stevenson b,d, Kevin J. Deluzio a,b

a Department of Mechanical and Materials Engineering, Queen’s University, McLaughlin Hall, 130 Stuart Street, Kingston, Ontario, K7L 3N6, Canada
b School of Physical and Health Education, Nipissing University, North Bay, Ontario, P1B 8L7, Canada
c School of Kinesiology and Health Studies, Queen’s University, 28 Division Street, Kingston, Ontario, K7L 3N6, Canada
d Human Mobility Research Centre, Syl & Molly Apps Medical Research Centre, Kingston General Hospital & Queen’s University, Kingston, Ontario, K7L 2V7, Canada

⁎ Corresponding author at: Department of Mechanical and Materials Engineering, Queen’s University, McLaughlin Hall, 130 Stuart Street, Kingston, Ontario, K7L 3N6, Canada. Tel.: +1 613 533 6737; fax: +1 613 533 6489.
E-mail address: Brandon@me.queensu.ca (S.C.E. Brandon).

ARTICLE INFO

Article history:
Received 21 January 2013
Received in revised form 13 May 2013
Accepted 30 September 2013
Available online xxxx

Keywords:
Principal component analysis
Interpretation
Waveforms
Biomechanics
Multivariate
Gait
Lifting

ABSTRACT

Principal component analysis is a powerful tool in biomechanics for reducing complex multivariate datasets to a subset of important parameters. However, interpreting the biomechanical meaning of these parameters can be a subjective process. Biomechanical interpretations that are based on visual inspection of extreme 5th and 95th percentile waveforms may be confounded when extreme waveforms express more than one biomechanical feature. This study compares interpretation of principal components using representative extremes with a recently developed method, called single component reconstruction, which provides an uncontaminated visualization of each individual biomechanical feature. Example datasets from knee joint moments, lateral gastrocnemius EMG, and lumbar spine kinematics are used to demonstrate that the representative extremes method and single component reconstruction can yield equivalent interpretations of principal components. However, single component reconstruction interpretation cannot be contaminated by other components, which may enhance the use and understanding of principal component analysis within the biomechanics community.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Principal component analysis (PCA) has emerged as a useful method for analyzing human motion data (Chau, 2001; Deluzio et al., 1999, 1997). The appeal of this multivariate statistical tool is that it can be applied to a set of temporal waveforms without a priori selection of important features (e.g. peaks), thereby providing an objective characterization of waveform features that differ between subjects in the dataset. However, while important features are identified objectively, it may be difficult to interpret the biomechanical meaning of these features (Ryan et al., 2006).

Principal component analysis of the pooled biomechanical waveforms from multiple subjects yields an orthogonal set of waveform features, called principal components (PCs), and a corresponding set of dimensionless PC-scores for each subject (Jackson, 1991). Many biomechanical studies have interpreted each PC by visually comparing the raw waveforms of representative extreme subjects who exhibit high (95th percentile) or low (5th percentile) PC-scores (Astephen et al., 2008b; Deluzio and Astephen, 2007; Jones and Rice, 1992; Kirkwood et al., 2011; Reid et al., 2010). Interpretation based on extremes is complicated and subjective because there may be multiple features that differ between the extreme raw waveforms, while only one of these features is actually associated with the PC of interest.

Principal component analysis reduces the dataset to a linear combination of orthogonal components of variation; therefore, each PC can be analyzed independently. Ramsay and Silverman (1997) showed that adding and subtracting a scalar multiple of the PC in question to the overall group mean provides a clear visual representation of the a single PC feature. This method of interpretation, which we call single component reconstruction, has been used in recent biomechanical studies that employed PCA on both functional data (Donoghue et al., 2008; Ryan et al., 2006) and discrete data (Brandon, 2009; O’Connor and Bottum, 2009). In this paper, PCA is applied to discrete data, meaning that the data have not been represented with basis functions or temporally registered to create functional data. While single component reconstruction still relies upon visual inspection of waveforms, the removal of confounding features captured by other PCs may facilitate a more accessible interpretation of PCA.

The purpose of this paper is to contrast the representative extremes method with single component reconstruction for interpretation of principal components in biomechanical applications.
2. Methods

2.1. Principal component analysis

Principal component analysis is a statistical technique that can be used to decompose a complex dataset into a series of orthogonal patterns of variance, called principal components (PCs) (Jackson, 1991). Principal components are computed to describe a maximal amount of variance using the fewest possible parameters, and are ranked according to the percent of the total variance that they explain (Ramsay and Silverman, 1997). For a single PC, each subject’s waveform receives a PC-score, representing the degree to which the subject expresses the specific pattern of variance, or PC loading vector. Because the PCs are orthogonal, the PCA estimate of the dataset can be expressed as a linear combination of PCs (Jackson, 1991):

\[
\hat{x}_i = \bar{x} + u_1 z_{1i} + u_2 z_{2i} + \cdots + u_k z_{ki}
\]  

(1)

where \(\hat{x}_i\) is the \((1 \times n)\) estimated waveform for subject \(i\); \(n\) the number of waveform observations; \(\bar{x}\) the \((1 \times n)\) mean temporal waveform for all subjects; \(u_k\) the \((1 \times n)\) pattern of variance, or loading vector, for PC \(j\); \(z_{ki}\) is the PC-score for subject \(i\), PC \(j\); \(k\) is the number of PCs retained in model \((k < n)\).

A common procedure is to retain only the first \(k\) PCs required to explain \(90\%\) of the total variance (Chau, 2001; McKean et al., 2007). Thus, instead of comparing raw waveforms described by \(n\) correlated variables, it is possible to consider the PCA estimated waveforms from Eq. (1), which are described by \(k\) orthogonal PCs (Jackson, 1991; Jones and Rice, 1992).

2.2. Representative extremes interpretation

Principal components are often interpreted through simultaneous inspection of the shape of the PC loading vector and the differences between extreme subject waveforms chosen from some quantiles of the data (Deluzio et al., 1997; Jones and Rice, 1992). For the following examples, extreme waveforms were taken as the raw waveforms corresponding with subjects who exhibited 5th (low) and 95th (high) percentile PC-scores for the PC of interest (Landry et al., 2007). These representative extreme subject waveforms should differ from the mean waveform by positively (95th percentile, high PC score) or negatively (5th percentile, low PC score) expressing the PC feature of interest. Thus, if at a given instant the 95th percentile waveform has a large magnitude, and the 5th percentile waveform has a small magnitude, one can infer that the PC of interest captured a magnitude change at this instant.

Representative extremes interpretation relies primarily on a visual comparison of the differences between two extreme waveforms. When the primary difference between 5th and 95th percentile waveforms is a vertical shift throughout the majority of the waveform, the PC of interest is likely a “magnitude” feature. A change in the timing of the extreme waveforms indicates a “phase shift” feature, while a “difference” feature is evident when there is a change in the magnitude of one or more local peaks (Wrigley et al., 2005).

However, any observed differences between these representative extreme waveforms can include both the biomechanical feature captured by the PC of interest and additional inter-subject variance captured by other PCs. In order to isolate the biomechanical feature captured by the PC of interest, from the additional inter-subject variance, it is necessary to simultaneously consider the shape of the PC loading vector. At locations along the abscissa where the magnitude of the loading vector is great, the observed difference between extreme waveforms can be associated with the chosen PC: when the loading vector approaches zero, the difference between extreme waveforms is due to other variance. For the following examples, vertical dashed lines indicate locations where the loading vector has a large positive or negative magnitude, indicating regions of interest for PC interpretation.

A magnitude feature captures a vertical shift in the overall temporal waveform, and was characterized by its monophasic (almost entirely positive or entirely negative) loading vector. Both “difference features” and phase shift features have multiphasic (alternating positive and negative phase) loading vectors. A “difference feature” captures a relative change in the magnitude of local peaks, and was identified when the PC loading vector had a positive peak aligned with one local peak in the mean data waveform, and a negative peak aligned with a second peak in the mean data waveform. Conversely, a phase shift feature indicates a change in the timing of local peaks in the data. Phase shift features were identified when the PC loading vector had a positive peak before, and negative peak after, the timing of the local peak in the mean data waveform.

2.3. Single component reconstruction

In order to interpret the biomechanical meaning of a single PC, the goal is to isolate the variance captured by the PC of interest. This can be visualized on a single figure by plotting upper and lower bands about the mean that differ only in the feature captured by a single PC (Ramsay and Silverman, 1997). According to Eq. (1), the PCA model is a linear combination of the mean waveform and the PCs; thus, upper and lower bands that include only the contribution of the \(k\)th PC, denoted \(PC_k\), can be computed by simply discarding all other confounding PCs from the linear model to yield:

\[
\hat{x}_R = \hat{x} + u_R z_{95}
\]

(2)

\[
\hat{x}_L = \hat{x} + u_R z_{5}
\]

(3)

where \(\hat{x}_R\) is the \((1 \times n)\) reconstructed upper waveform for \(PC_R\); \(\hat{x}_L\) the \((1 \times n)\) reconstructed lower waveform for \(PC_R\); \(n\) the number of waveform observations; \(\hat{x}\) the \((1 \times n)\) mean temporal waveform for all subjects; \(u_R\) the \((1 \times n)\) pattern of variance, or loading vector, for \(PC_R\); \(z_{95}\) is the 95th percentile (high) scalar PC-score for \(PC_R\); \(z_5\) is the 5th percentile (low) scalar PC-score for \(PC_R\).

The scalar weight factors, \(z_5\) and \(z_{95}\), can be any values that adequately scale the loading vector, \(u_R\), to make the reconstructed waveform, \(\hat{x}_R\), visually differ from the mean temporal waveform, \(x\) (Ramsay and Silverman, 1997). In this paper, the biomechanical effect of a given PC is shown using the 5th (low) and 95th (high) percentile PC-scores as scalar weight factors, \(z_5\) and \(z_{95}\), respectively, for the lower and upper bands. These percentiles were selected so that the lower and upper bands correspond with the same two subjects that were previously selected for each PC of interest using the method of representative extremes interpretation (Section 2.2). As shown in Fig. 1, principal component scores for biomechanical waveform data tend to be normally distributed; therefore, the 5th and 95th percentile bands in this paper are roughly equivalent to \(\pm 2\) standard deviations. For the single component reconstruction interpretation, magnitude, “difference”, and phase shift features were identified by visual comparison of upper and lower bands with respect to the mean waveform.

2.4. Data

In this paper, PCA was applied to previously collected data from a sampling of motion analysis applications: knee joint adduction moments during gait (Brandon and Deluzio, 2011), lateral gastrocnemius EMG during gait (Hubley-Kozey et al., 2006), and lumbar spine kinematics during lifting (Sadler et al., 2013, 2011). The collection of each dataset was approved by the Institution’s Research Ethics Board.
Unilateral lower-limb gait data were collected at self-selected walking speed for 44 subjects with moderate unilateral knee osteoarthritis and 44 asymptomatic control subjects (Brandon and Deluzio, 2011). Knee adduction moments were calculated for each subject using a standard inverse dynamics approach. Moments were amplitude-normalized to body weight, time-normalized to the stance phase of the gait cycle and re-sampled at each percent of the stance phase, then ensemble averaged within each subject across at least five walking trials to yield an 88 x 101 matrix representing 101 stance phase cycle observations for each of the 88 subjects.

Surface EMG data were simultaneously collected from the lateral gastrocnemius during the entire gait cycle for the same 88 subjects. Data collection and analysis has been previously described (Hubley-Kozey et al., 2006). Briefly, EMG data were full-wave rectified and low-pass filtered at 6 Hz, amplitude-normalized to maximum voluntary isometric contraction, time-normalized to 100% of the gait cycle, and ensemble averaged within each subject across at least five walking trials to obtain an 88 x 101 matrix representing 101 gait cycle observations for each of the 88 subjects.

Lifting data consisted of lumbar spine flexion angles for 30 asymptomatic healthy subjects performing 30 consecutive lift cycles under four different lifting conditions (Sadler et al., 2013, 2011). The load was lifted at a pace of 10 cycles per minute from a target on the floor to target on a table that was set at 50% of the subject’s height. The 30 lifting cycles of lumbar spine flexion were time-normalized to lifting cycle and re-sampled at each percent of the lifting cycle, then ensemble averaged within each subject to yield a 120 x 101 matrix representing 101 lifting cycle observations for each of the 30 subjects in each of the 4 conditions.

2.5. Data analysis

For each of the three datasets, principal component analysis was performed using an m x n matrix, with m subjects and n = 101 temporal observations throughout the cycle (Deluzio et al., 1997). A minimal set of PCs was retained to describe at least 90% of the variance within each dataset (Jackson, 1991; McKean et al., 2007).

One PC was selected from each of the three datasets to compare the two methods of interpretation (Fig. 2). Additionally, each of the three examples was chosen to illustrate a different type of biomechanical waveform feature that can be captured using PCA: magnitude, difference, and phase shift features (Wrigley et al., 2005).

Finally, to provide a complete example of the application of PCA to a dataset, the first k PCs explaining greater than 90% of the total variance from the knee adduction moment data were interpreted using both representative extremes interpretation and single component reconstruction (Fig. 3).

3. Results

3.1. Magnitude feature: lumbar spine flexion PC1 (Fig. 2A, D, G)

Subjects in the lifting study began from a neutral lumbar spine flexion angle of approximately 15°, stooped to about 45° to pick up the load, and returned to neutral posture (Fig. 2A and G). There was a large amount of inter-subject variation in lumbar spine flexion angle throughout the lifting cycle (Fig. 2A); however, PC1 captured 86% of the total variance in the temporal waveforms.

The extreme raw waveforms (Fig. 2A) for PC1 suggested that a high (95th percentile) PC-score corresponded with greater flexion angle magnitude throughout the lifting cycle, while a low (5th percentile) PC-score was associated with reduced flexion angle magnitude throughout the lifting cycle. Additionally, the low (5th percentile) waveform returned to neutral lumbar spine flexion much earlier in the lifting cycle than the high (95th percentile) waveform, indicating the presence of a phase shift. To complete the representative extremes interpretation, the loading vector was examined to confirm the magnitude and phase shift features described above. The PC1 loading vector (Fig. 2D) exhibited a relatively constant magnitude between 20% and 80% of the cycle. This monophasic loading vector confirmed that PC1 captured inter-subject variance in magnitude during this period. However, as described in greater detail in Section 3.3, a phase shift feature cannot occur unless the loading vector crosses zero at the same instant as the peak in the mean waveform. In this example (Fig. 2D), the PC1 loading vector did not cross zero at any point in the cycle, which means that the phase-shift exhibited by extreme 5th and 95th percentile subjects in Fig. 2A was not actually captured by PC1. Instead, the phase difference between extreme subjects was present because the extreme subject waveforms are contaminated by variance that was captured by other PCs. Thus, using representative extremes interpretation PC1 was correctly described as a magnitude feature because it captured the large inter-subject variation in magnitude throughout the waveform, but could have mislead users regarding the phase shift.

Lumbar spine flexion angle PC1 was also interpreted using single component reconstruction (Fig. 2G). In this method, the mean flexion angle was bounded by upper (95th percentile) and lower (5th percentile) waveforms that differed from the mean only in the pattern captured by PC1, as calculated using Eqs. (2) and (3). These upper and lower bounds were from the same 5th and 95th percentile subjects as the extremes in Fig. 2A; however, all variance from the raw waveforms that was not associated with PC1 was removed by single component reconstruction. It was evident from Fig. 2G that PC1 described variation in the overall lumbar spine flexion magnitude throughout the lifting cycle, primarily between 20% and 80% of the cycle. Subjects with high scores on PC1 tended to initiate the lifting cycle with greater flexion angles, achieve a greater range of motion, and sustain peak flexion longer than subjects with low scores on PC1. Using single component reconstruction, there was no evidence of a phase shift between 5th and 95th percentile reconstructed waveforms, therefore PC1 was correctly classified as a magnitude feature.

In summary, both representative extremes and single component reconstruction methods were used to identify PC1 as a magnitude feature. Biomechanically, this means that subjects with high PC1 scores exhibited greater lumbar spine flexion angle magnitudes between 20% and 70% of the lifting cycle, while subjects with...
low PC1 scores had reduced spine flexion angle magnitudes during this period. This magnitude feature, which is consistent with observation of raw waveforms, can be statistically tested using the set of PC1 scores from all subjects.

3.2. Difference feature: knee adduction moment PC2 (Fig. 2B, E, H)

Subjects in the gait study walked with a positive knee adduction moment throughout most of the stance phase (Fig. 2B). The second PC (PC2) for the knee adduction moment during gait explained 16% of the total variance in the temporal waveforms, and was selected to illustrate a “difference feature”.

Extreme subject waveforms (Fig. 2B) suggest that subjects with high (95th percentile) PC2-scores had greater first peak adduction moments around 20% of the cycle than subjects with low (5th percentile) PC2-scores. The PC2 loading vector (Fig. 2E) showed large absolute magnitudes at 20% and 70% of the cycle, which indicated that the PC2 feature was primarily present in these regions. However, it was difficult to fully interpret PC2 because there was only a small difference between extreme subjects (Fig. 2B) around 70% of the cycle. Therefore, interpretation using representative extremes indicated that PC2 captured a large change in the first peak knee adduction moment amplitude during gait, and a lesser change around 70% of the cycle.

The second PC for the knee adduction moment was also interpreted using single component reconstruction (Fig. 2H). When contaminating variance from other PCs was removed (Fig. 2H), it was evident that PC2 described a change in the relative amplitude of the first and second peak adduction moments around 20% and 70% of the cycle, respectively. Subjects with high (95th percentile) PC2-scores exhibited a large first peak and a reduced second peak, while subjects with low (5th percentile) PC2-scores exhibited a reduced first peak and greater second peak for an overall flatter waveform throughout the cycle. Thus, single component reconstruction identified PC2 as a difference feature, because it captures a relative difference between two locations within the waveform.

In summary, both representative extremes and single component reconstruction methods were used to identify PC2 as a “difference feature”. Biomechanically, this means that subjects with high PC2 scores exhibited greater knee adduction moment peak amplitudes around 20% of the cycle, and reduced peak knee adduction moment amplitudes around 70% of the cycle. Conversely, subjects with low PC2 scores had reduced first peak (20% cycle) and increased second peak (70% cycle) knee adduction magnitudes for an overall flatter waveform. This “difference feature”, which captures the relative magnitude of the two peaks in the knee adduction moment waveform, can be statistically tested using the set of PC2 scores from all subjects.

3.3. Phase shift feature: lateral gastrocnemius EMG PC2 (Fig. 2C, F, I)

Lateral gastrocnemius EMG during gait showed a large (50% maximum voluntary contraction) peak during propulsion around 50% of the gait cycle (Fig. 2C). The second PC (PC2) captured 19% of the total variance between raw waveforms, and was selected to illustrate how PCA reveals a phase shift feature.

Extreme raw waveforms (Fig. 2C) showed that PC2 captured a difference in the onset of excitation and the timing of the peak; subjects with high PC2 scores exhibited earlier peaks and excitation onset compared to subjects with low scores for PC2. The PC2 loading vector (Fig. 2F) had large positive amplitude before and large negative amplitude immediately after the raw waveform peaks (Fig. 2C), which confirmed that PC2 captured a phase shift feature.

Similarly, using single component reconstruction for PC2 it was demonstrated that subjects with high PC2 scores had earlier onset and peak lateral gastrocnemius excitation compared to subjects with low PC2 scores (Fig. 2I). Additionally, the reconstructed upper
and lower bands in (Fig. 2I) revealed that delayed excitation (a low PC2-score) was also associated with higher peak excitation, indicating that PC2 captured a combination of a phase shift and a change in peak amplitude. In contrast, the extreme waveforms in Fig. 2C had nearly identical magnitudes (although much greater than most other raw waveforms) and appeared to differ only in their phase. Therefore, the method of extreme subject interpretation did not show the presence of an amplitude effect in the PC2 feature.

In summary, both representative extremes and single component reconstruction methods were used to identify PC2 as a phase shift feature, but the representative extremes method was misleading regarding the associated amplitude effect. Biomechanically, this means that subjects with high PC2 scores exhibited earlier onset and peak lateral gastrocnemius excitation around 40% of the cycle, while subjects with low PC2 scores exhibited delayed onset and peak muscle excitation. Additionally, it was evident using single component reconstruction that the phase shift feature includes an amplitude effect, such that subjects with high PC2 scores tended to have reduced peak amplitude in conjunction with earlier onset, while subjects with low PC2 scores tended to have higher peak amplitudes as well as delayed onset. This phase shift feature represents a meaningful biomechanical change in muscle excitation which can be statistically tested using the set of PC2 scores from all subjects.

3.4. Detailed example: knee adduction moment PC1–PC4 (Fig. 3)

Four PCs were required to explain greater than 90% of the total variation in the knee adduction moment dataset. This reduced the dimensionality of the dataset from a matrix of 88 subjects × 101 observations to a matrix of 88 subjects × 4 PC scores with minimal loss of information. Representative extremes, PC loading vectors, and single PC reconstructions for each PC are shown in Fig. 3.

Representative 5th and 95th percentile extreme raw waveforms were similar in shape to the single component reconstruction waveforms for PC1 (Fig. 3A versus Fig. 3I) and PC2 (Fig. 3B versus Fig. 3J). Using either method of interpretation, it was possible to interpret the first PC (PC1) as a magnitude feature, active primarily between 20% and 70% of the cycle. Both methods of interpretation also showed that PC2 captured a difference feature between early and late-cycle peaks.

There was less agreement between representative extremes and single component reconstruction waveforms for PC3 and PC4 (Fig. 3C and D versus Fig. 3K and L). The third and fourth PCs only captured 8% and 4% of the total waveform variance, respectively. Therefore, their representative extreme waveforms differed not only according to the PC of interest, but also for other PCs (e.g. PC1 and PC2). Using representative extremes interpretation as explained previously (Sections 3.1–3.3), it was determined that PC3 captured a “difference feature” expressed as a change in mid-cycle concavity between 5th and 95th percentile waveforms. This interpretation was supported by the single component reconstruction (Fig. 3K).

The representative extreme waveforms for PC4 were so different that it was difficult to extract a biomechanical interpretation. Examining the PC4 loading vector (Fig. 3H), it is clear that PC4 captured variance around 10% and 20% of the gait cycle. However, from the loading vector alone it is difficult to determine whether PC4 is a “difference feature” or a phase shift feature. In contrast, the upper and lower bands constructed using single component reconstruction clearly show that high PC4 scores are associated with an earlier peak (around 15% cycle), while low PC4 scores are associated with a delayed peak (close to 20% cycle) adduction moment. Therefore, only single component reconstruction was able to identify PC4 as a phase shift in the early-cycle loading rate (Fig. 3L).

Using single component reconstruction, the difference between representative waveforms (Fig. 3I–L) is proportional to the amount of variance captured by each PC. There was a large separation between 5th and 95th reconstructed waveforms for PC1 (Fig. 3I) because it explained a large proportion of the total variance (65%). The fourth PC (PC4, Fig. 3L) explained only 4% of the total inter-subject variance; therefore the difference between 5th and 95th reconstructed waveforms was so small that they were barely distinct from the mean waveform, except around 25% of the gait cycle. This aspect of single component reconstruction could help a user to assess the clinical importance of the PC feature. In contrast, the separation between 5th and 95th percentile extreme raw waveforms (Fig. 3A–D) is independent of the percentage of variance explained. For example, the vertical difference between extreme waveforms was very large for PC4 (Fig. 3D), when in fact PC4 only captured a minor (4% variance) change in the loading rate. Therefore, representative extremes interpretation cannot be used to infer the size of the effect captured by each PC.

4. Discussion

Both representative extremes and single component reconstruction methods rely upon visual inspection of waveforms to interpret the biomechanical differences captured using PCA. In accordance with previous studies that used PCA (Astapheon et al., 2008a; Deluzio et al., 1997; O’Connor and Bottum, 2009) our results demonstrate that either method can be used to correctly interpret a biomechanical feature.

Single component reconstruction offers one primary advantage over the representative extremes method: the reconstructed waveforms for a single PC of interest are not contaminated by variance from other PCs. This fact was important in the interpretation of the first principal component of the lumbar flexion angle during lifting (Fig. 2A, D, G). An untrained observer may not have recognized that the phase shift exhibited by the 5th percentile extreme subject waveform in Fig. 2A does not correspond with the shape of the loading vector in Fig. 2D, and is therefore not part of the feature captured by PC1. Selecting an individual raw waveform that represents only the PC of interest is difficult, particularly when the biomechanical meaning of the PC of interest has not yet been established. The representative extremes method requires a synthesis of waveform data, which may be perceived as subjective interpretation. Single component reconstruction provides a more robust procedure for visualizing individual PC waveform features.

In the knee adduction moment data, the difference feature captured by PC2 is particularly interesting because it may not have been detected using alternative data analysis techniques (Fig. 2B, E, H). The second PC has been shown to differentiate between osteoarthritis and control subjects (Newell et al., 2008) and could be important for the development of gait interventions. Some gait-alteration strategies, such as toe-out gait (Fregly et al., 2007), act to reduce the second peak of the adduction moment. However, this might not be an appropriate clinical intervention for subjects with a high score on PC2 who already exhibit a reduced second peak moment.

The three types of features (magnitude, difference, and phase) presented in this study are consistent with previous studies that used PCA (Deluzio et al., 1997; O’Connor and Bottum, 2009; Wrigley et al., 2005). When analyzing a dataset of temporal biomechanical waveforms, the first PC almost always captures a magnitude feature. This is partly because the method of PCA describes the pattern of greatest variance within the data, where variance is defined as the vertical difference between observations and the
mean value. Thus, a vertical magnitude feature will introduce much more variance across the entire temporal waveform than localized difference or phase shift features. However, this does not necessarily diminish the clinical importance of smaller PCs or their ability to discriminate between subjects and treatments. Although the difference and phase shift features explain less of the overall variance than the magnitude feature (PC1), even small components can be discriminatory between subjects (Ast Stephen and Deluzio, 2004). Therefore, it is important to obtain a biomechanical interpretation even for these smaller PCs.

The general biphasic shape of the phase shift loading vector in Fig. 2F was quite similar to the loading vector from the previous “difference feature” example (Fig. 2E). However the difference between these two features is the timing of the loading vector peaks with respect to the raw data. For a difference feature, the loading vector peaks indicated by vertical event lines are directly aligned with peaks in the raw data (Fig. 2B and E). However, for a phase shift feature, the loading vector will have a positive peak before and a negative peak after (or vice versa) the corresponding peak in the raw data (Fig. 2C and F). For the example phase shift feature (Fig. 2C), the vertical event lines indicating large loading vector peaks clearly fall before and after the peaks in the raw data. A phase shift loading vector crosses zero at the same instant as the peak in the mean data (Fig. 2F and I).

The phase-shift feature captured by PC2 in the EMG data could have been interpreted by examining extreme subjects in conjunction with the relative timing of the peaks in the loading vector (Fig. 2F) and in the raw data (Fig. 2C). However, only single component reconstruction identified that the PC2 phase shift feature also includes a change in the peak amplitude. In some studies, it may be desirable to extract a parameter that does not combine phase and amplitude effects. When a phase-shift feature emerges as one of the primary modes of variance within the dataset, more valuable information might be gleaned by isolating the temporal variance through registration of the functional data (Ramsay and Silverman, 1997; Sadeghi et al., 2000), or using an alternative data analysis technique such as cross-correlation analysis (Wren et al., 2006).

Principal component analysis is a powerful tool for objectively exploring biomechanically relevant waveform differences between subjects (Deluzio et al., 1999; Gaudreault et al., 2011; Kipp et al., 2011; Kirkwood et al., 2011). The biomechanical features captured by principal component analysis can be visually interpreted using either the representative extremes method or single component reconstruction. In fact, the most robust and clinically relevant interpretation might be enabled by employing both methods. However, single component reconstruction provides a simplified, uncontaminated visualization of the biomechanical difference captured by each PC, which may make conclusions from PCA more accessible to the broader biomechanics community.

**Acknowledgements**

Financial support for S.C.E. Brandon, R.B. Graham, E.M. Sadler, and S. Almosnino was provided through awards and stipends granted by the Natural Sciences and Engineering Research Council of Canada (NSERC).

**References**


Scott C.E. Brandon is a PhD candidate in Mechanical and Materials Engineering at Queen’s University, under the supervision of Dr. Kevin Deluzio. His research focus is biomechanical causes and treatments for knee osteoarthritis. He is currently using computer simulations to investigate the effectiveness of knee braces that are designed to unload the medial tibiofemoral joint. Scott received his B.Sc.Eng from The University of Western Ontario, and M.Sc.Eng from Queen’s University, and is now a trainee in the HMRC-CREATE program at Queen’s.

Ryan B. Graham is an Assistant Professor in the School of Physical and Health Education at Nipissing University. He obtained his doctorate in biomechanics and ergonomics in 2012 from the School of Kinesiology and Health Studies at Queen’s University, where he currently holds an Adjunct Assistant Professor appointment. Dr. Graham’s current research focuses on the quantitative assessment of low back pain mechanisms and risk, with specific reference to spine instability and impaired neuromuscular control.

Sivan Almosnino is a doctoral candidate in Biomechanics and Ergonomics at the School of Kinesiology and Health Studies, Queen’s University. He is also a member of Queen’s University Ergonomics Consulting Program. His current research includes development of tools and methods for clinical decision making regarding readiness to return to activity following knee injuries; assessment of load carriage gait; and, development and implementation of preventive ergonomic aids in a variety of work settings. In addition, he serves as an editorial board member for the journal Isokinetics & Exercise Science.

Erin M. Sadler completed her Master of Science in the School of Kinesiology and Health Studies at Queen’s University in Kingston, Ontario. Her thesis involved principal component analysis of three dimensional kinematics of the spine during lifting tasks. Erin is now completing her MD at the University of Toronto.

Joan M. Stevenson (PhD) is a full professor in occupational biomechanics and ergonomics at Queen’s University in the School of Kinesiology and Health Studies with a cross-appointment to the School of Rehabilitation Therapy. She has led various research teams on applied projects for the military and many private sector companies. Her current research initiatives include: development of ergonomic assistive devices for industry, biomechanical evaluation of spinal mechanics, and creation of objective measurement tools.

Kevin J. Deluzio, PhD, a Professor at the department of Mechanical and Materials Engineering, Queen’s University at Kingston, Canada, studies human locomotion to investigate the biomechanical factors of musculoskeletal diseases such as knee osteoarthritis. He is also interested in the design and evaluation of non-invasive therapies as well as surgical treatments as total joint replacement.