# Characterizing and Modeling the Joint-level Variability in Human Walking

Anne E. Martin[a,*], Dario J. Villarreal[b], Robert D. Gregg[b,c]

[a] *Department of Mechanical and Nuclear Engineering, The Pennsylvania State University, University Park, PA 16802, USA*
[b] *Department of Bioengineering, University of Texas at Dallas, Richardson, TX 75080, USA*
[c] *Department of Mechanical Engineering, University of Texas at Dallas, Richardson, TX 75080, USA*

**Abstract**

Although human gait is often assumed to be periodic, significant variability exists. This variability appears to provide different information than the underlying periodic signal, particularly about fall risk. Most studies on variability have either used step-to-step metrics such as stride duration or point-wise standard deviations, neither of which explicitly capture the joint-level variability as a function of time. This work demonstrates that a second-order Fourier series for stance joints and a first-order Fourier series for swing joints can accurately capture the variability in joint angles as a function of time on a per-step basis for overground walking at the self-selected speed. It further demonstrates that a total of seven normal distributions, four linear relationships, and twelve continuity constraints can be used to describe how the Fourier series vary between steps. The ability of the proposed method to create curves that match human joint-level variability was evaluated both qualitatively and quantitatively using randomly generated curves.

*Keywords:* Variability, Human Gait, Modeling

## 1. Introduction

A major motivation for investigating gait variability is quantifying fall risk. Mean temporal measures have consistently failed to differentiate between fallers and non-fallers (Hamacher et al., 2011; Hausdorff, 2007). In contrast, measures of gait variability can successfully identify individuals at risk for falls (Hamacher et al., 2011; Hausdorff, 2007) and with fear of falling (Sawa et al., 2014). In general, increased variability indicates increased fall risk and is often assumed to be correlated with decreased stability (Hamacher et al., 2011). For some physiological systems, a moderate amount of variability is indicative of a healthy, flexible system, and this may be true for gait (Hausdorff, 2007; Stergiou and Decker, 2011). Further, variability and stability may not be equivalent (Dingwell et al., 2001). Regardless of the exact connection between variability and stability, measures of variability likely provide different information than mean values.

Variability has traditionally been viewed as purely random fluctuations about a desired mean (e.g., (Pailhous and Bonnard, 1992)), and thus, the structure of the variation is assumed to be uninteresting. Variability has typically been quantified using standard deviations (or equivalent, e.g., (Danion et al., 2003; McGinley et al., 2009; Winter, 1984)), which provides no information about how variability at one instant affects variability at a later instant, either for the current or subsequent steps. Similarly, stochastic statistical methods can provide subject-specific variability magnitudes but have not been used to describe the structure of the variability (Yun et al., 2014).

For step-to-step measures, the variability has structure (Hausdorff, 2007). Regardless of speed, human walking exhibits long-range correlation in stride duration (Hausdorff, 2007; Hausdorff et al., 1995). Multiple models have been proposed, ranging from correlated neural networks (e.g., (Hausdorff et al., 1995)), to goal-directed control (Dingwell and Cusumano, 2015), to natural dynamics (e.g., (Ahn and Hogan, 2013)). An important feature of all of the models is that the next step explicitly depends on the previous step. Work on the structure of variability typically investigates a single step-to-step measure such as step duration and only

*Corresponding author
*Email addresses:* aem34@psu.edu (Anne E. Martin), dario.villarreal@utdallas.edu (Dario J. Villarreal), rgregg@utdallas.edu (Robert D. Gregg)

implicitly captures joint angle variation. Direct measures of joint angle variability can be obtained through nonlinear dimensionality reduction techniques (Dingwell and Cusumano, 2000; Zhang et al., 2010), but these methods do not provide an intuitive representation of variability as a function of time.

Modeling joint-level variability as a function of time is a necessary first step for incorporating variability into predictive dynamic models of human walking (Martin and Schmiedeler, 2014; Martin and Gregg, 2016). Such dynamic models may aid in the design of assistive devices because interventions can be tested using realistic simulations (Martin and Gregg, 2015) prior to experimental testing with human subjects. Predictive simulations can also be used to investigate research questions that are impractical or impossible to answer via human subject testing, such as quantifying fall risk (Byl and Tedrake, 2009). If clinicians can accurately diagnose patients at risk for falls, then it may be possible to provide targeted physical therapy to improve walking ability and reduce fall-related injuries. Such simulations could also be used to answer the theoretical questions of if and why variability is advantageous in human walking. Although a dynamic model is needed so that the forces that generate variable movement are accounted for, a functional representation of kinematic variability is needed before a variable dynamic model can be developed.

This work develops a functional representation of joint-level variability over time during healthy human walking (Fig. 1). This variability can be categorized on three levels. The first level explains how the joint kinematics differ from the mean as a function of time for a single step. The second level explains how the step-specific functions are distributed in function space, where this space was characterized by the distributions of and relationships between coefficients that define step-specific functions. The third level, not fully considered here, explains how the step-specific functions vary over many steps. To validate this approach, randomly generated curves that mimic joint level variability were created using experimentally observed distributions and relationships.

## 2. Methods

### *2.1. Experimental Data*

The proposed method was evaluated using experimental data for ten young adults (6 male, height: 1.75 $\pm$ 0.06 m, weight: 67 $\pm$ 7 kg). Relative joint angles for the right hip, knee, and ankle were collected at 100 Hz using ten motion capture cameras (Model: T20S, Vicon, Oxford, UK) as subjects walked along an 8 m walkway with a single force plate (Model: 9260AA6, Kistler, Winterthur, Switzerland) in the center at their self-selected speed. The data for the center stride were processed using the Vicon software Nexus with the Plug-in-Gait module. Prior to the experiment, subjects provided informed consent and acclimated to the walkway.

Each stride was divided into stance and swing periods, and the periods were analyzed separately. Attempts to analyze the entire stride together resulted in inferior results. For each period, the joint angle data were interpolated over 400 equally-spaced, normalized time points (Fig. 1(a)). Steps that had one or more points with an angle outside of three standard deviations were discarded. Since the amplitude of the variability was much less than the range of motion, the angles were centered about the mean trajectory so that only the variability remained (Fig. 1(b)-(c)).

The centered angles for all joints and periods had a large amplitude, low frequency dominant sinusoidal component with high frequency, low-amplitude oscillations (Fig. 2). Thus, they were characterized with a second-order Fourier series for the stance period (Fig. 1(c)):

$$q_{StJ}(t) = a_{0StJ} + \sum_{i=1}^{2} \{a_{iStJ} \cos(i \cdot \omega_{StJ} \cdot t) + b_{iStJ} \sin(i \cdot \omega_{StJ} \cdot t)\} \tag{1}$$

and a first-order Fourier series for the swing period:

$$q_{SwJ}(t) = a_{0SwJ} + a_{1SwJ} \cos(\omega_{SwJ} \cdot t) + b_{1StJ} \sin(\omega_{SwJ} \cdot t) \tag{2}$$

where $J$ indicates the joint (**H**ip, **K**nee, or **A**nkle), $q_{pj}$ is the fitted centered joint angle, $p$ indicates the period (**St**ance or **Sw**ing), $\omega_{pj}$ is the dominant frequency of the centered joint angle, and $a_{ipj}$ and $b_{ipj}$ are magnitude coefficients. Time $t$ is normalized between 0% and 100%. For each step at each joint, the coefficients were determined using the `fit()` function in MATLAB (The MathWorks, Inc., Natick, Massachusetts, United

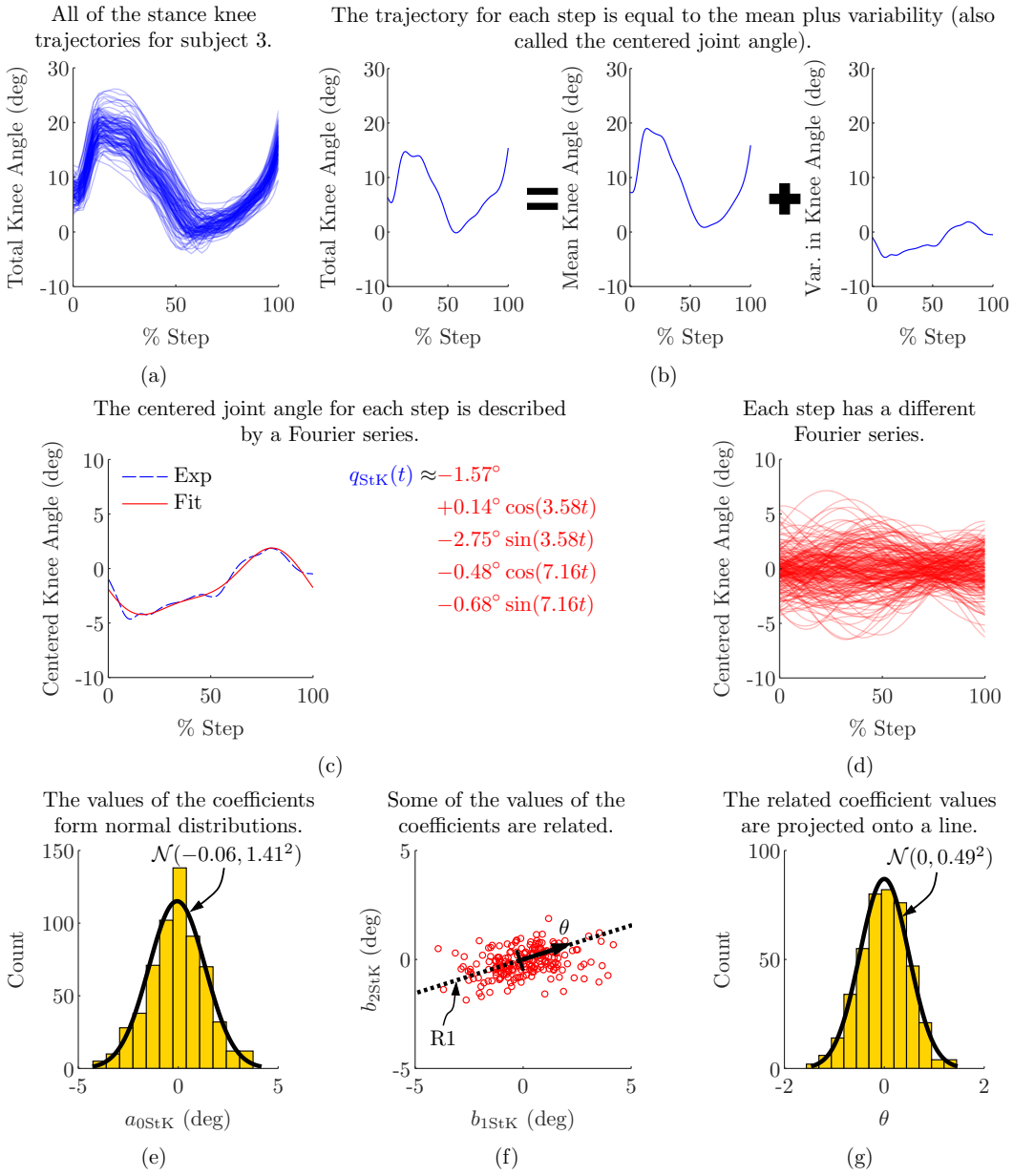



Figure 1: This figure illustrates the steps required to characterize the variability. Data for the stance knee is shown, but the procedure is identical for all joints. (a) The joint angle trajectory for each step is unique although all trajectories for a particular joint are similar. Time has been normalized to lie between 0% and 100% step. (b) For each step, the actual joint trajectory can be decomposed into a per-subject mean plus the variability. The variability is defined as the difference between the mean and the actual trajectory on a point-wise basis. (c) For each step, a Fourier series is fit to the variability. (The sin and cos functions take arguments in degrees.) This gives a unique set of coefficients that describe the within-step (first level) variability. (d) A Fourier series is found for each step. (e) Considering all of the steps, each coefficient in the Fourier series takes a range of values and these values form normal distributions. The distribution for $a_{0\mathrm{StK}}$ is shown here; the other distributions are similar. (f) There are also correlations between coefficients. For the stance knee, there is a positive correlation between $b_{1\mathrm{StK}}$ and $b_{2\mathrm{StK}}$. The relationship R1 is shown with the black dotted line. The data points (coefficient values) can be projected onto the line R1. Each data point's position along line R1 is given by $\theta$. By construction, $\theta = 0$ is at the centroid of the data points. (g) Similar to the coefficients, the $\theta$ values also form a normal distribution when all of the steps are considered. The across-step (second level) variability is captured by the combination of the normal distributions describing the range of coefficient values (e) and the relationships with their associated distributions (f-g). (In (d) and (f), every third step is shown for clarity.)

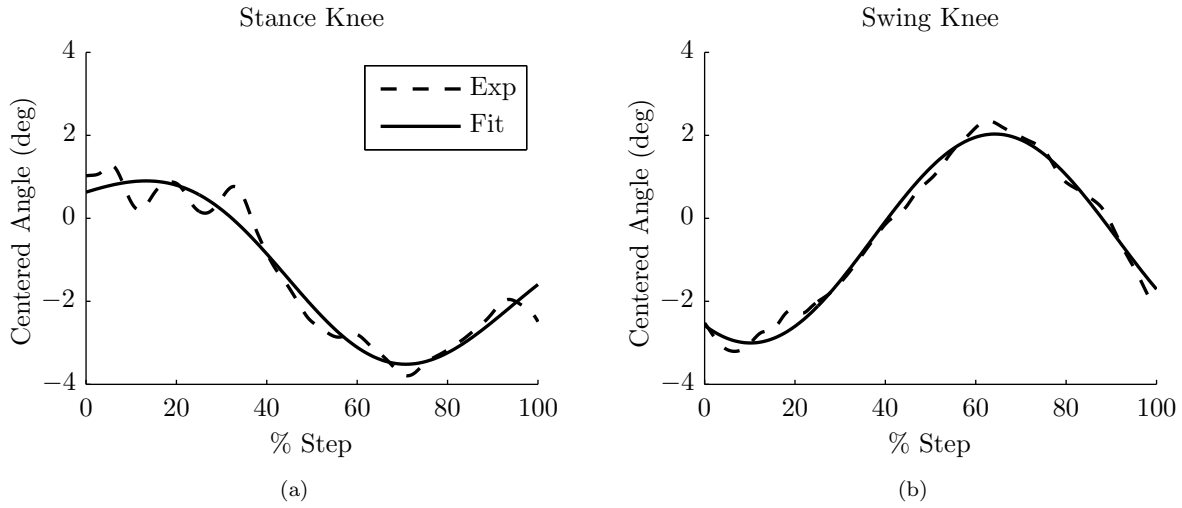



Figure 2: The centered experimental joint angles and fitted Fourier series for the (a) stance and (b) swing knee for a single representative stride. Time has been normalized to lie between 0% and 100% step. The experimental data has a dominant low frequency, large amplitude sinusoidal signal that the Fourier series captures well. The high frequency, small amplitude oscillations are not captured by the low-order Fourier series.

States). In general, the magnitude coefficients were small ($< 10°$). To remove very large magnitude terms ($> 1,000°$), any function with one or more coefficients greater than three standard deviations from the mean were discarded. Thus, if the coefficients were excessively large, we removed that joint for that step. These fitted Fourier series capture the first level of within-step variability.

Using the fitted Fourier series, the second level of variability was evaluated by looking across steps (Fig. 1(d)). Because a separate function was found for each step, the coefficients describing the variability of each step belong to 24 magnitude distributions and 6 frequency distributions (Fig. 1(e)). The data for all subjects were pooled to obtain typical distributions. All distributions were approximately Gaussian.

There were also relationships between coefficients, both across and within joints and periods (Fig. 1(f)-(g)). Except for the swing knee, all relationships were linear. The relationship strength was quantified using pairwise correlation coefficients $\rho$. Since relationships usually involved more than two coefficients, the first principle component was used to define the best-fit line (Jolliffe, 2002). For the swing knee, there was a relationship between frequency and phase shift $\phi_{\mathrm{SwK}}(a_{1\mathrm{SwK}}, b_{1\mathrm{SwK}})$:

$$\tan \phi_{\mathrm{SwK}} = \frac{a_{1\mathrm{SwK}}}{b_{1\mathrm{SwK}}}. \tag{3}$$

The phase shift vs. frequency plot has clear linear bands of data with approximately equal slope (Fig. 3). The bands were separated using the density-based clustering method DBSCAN (Daszykowski et al., 2001). The fraction of steps in each band was calculated, a linear regression line was fit to each band, the median slope of the lines was found, and the best intercept for each band using the median slope was determined. The swing knee frequency-phase shift relationship was defined using the median slope and the intercepts from the two largest bands.

### *2.2. Generating Curves*

If the distributions and relationships adequately characterize human variability, then randomly generated curves will replicate it. When creating the curves, coefficients were randomly generated and/or calculated using relationships (Table 1, Fig. 4, Supplement 1). Since human gait is continuous, the angular position and velocity at the beginning of a period must match the end of the previous period. These continuity conditions enforced relationships across periods. Rather than have a unique distribution for each coefficient, similar distributions were combined to reduce the number of distributions while maintaining sufficient accuracy.

Because each joint and period had different relationships between coefficients, the exact method of generating coefficients varied within a general framework. Fig. 4 provides a flowchart for generating stance knee

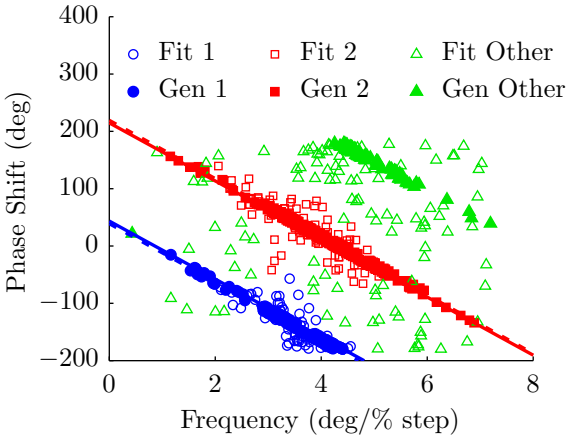



Figure 3: The swing knee phase shift vs. frequency. The coefficients from the fitted Fourier series are indicated with open markers while the coefficients from the generated Fourier series are indicated with filled markers. The marker color and shape indicates the cluster. For clarity, every third step is plotted. In both cases, the majority of points belong to one of two clusters. The best fit lines for the two largest clusters are also shown; the lines from the fitted and generated series overlap.

Table 1: General description of how coefficients were determined for the generated curves, and the distributions to which the coefficients belong. The MATLAB code used to generate the curves can be found in Supplement 1, and the flowchart for the stance knee is shown in Fig. 4. RD stands for random, PC stands for the position continuity constraint, VC stands for the velocity continuity constraint, and R1, R2, R3, and R4 stand for the four relationships. R1 is a linear relationship between $b_{1\mathrm{StH}}$, $b_{2\mathrm{StH}}$, $b_{1\mathrm{StK}}$, and $b_{2\mathrm{StK}}$. R2 is a linear relationship between $a_{0\mathrm{SwH}}$, $b_{1\mathrm{SwH}}$, and $b_{1\mathrm{SwK}}$. R3 is the phase shift-frequency relationship between $a_{1\mathrm{SwK}}$, $b_{1\mathrm{SwK}}$, and $\omega_{\mathrm{SwK}}$. R4 is a linear relationship between $a_{0\mathrm{StA}}$, $b_{1\mathrm{StA}}$, and $a_{2\mathrm{StA}}$. $\mathcal{N}(\mu, \sigma^2)$ indicates a normal distribution with mean $\mu$ and standard deviation $\sigma$. Because the swing phase uses a first-order Fourier series, those equations do not have $a_2$ and $b_2$ coefficients.

| ***Period*** | ***Joint*** | $\mathbf{a_0}$ | $\mathbf{a_1}$ | $\mathbf{b_1}$ | $\mathbf{a_2}$ | $\mathbf{b_2}$ | $\boldsymbol{\omega}$ |
|---|---|---|---|---|---|---|---|
| *Stance* | *Hip* | PC $\mathcal{N}(0, \sigma_{\mathrm{C}}^2)$ | RD & PC $\mathcal{N}(0, \sigma_{\mathrm{C}}^2)$ | RD & R1 & VC $\mathcal{N}(0, \sigma_{\mathrm{C}}^2)$ | RD $\mathcal{N}(0, \sigma_2^2)$ | RD & R1 & VC $\mathcal{N}(0, \sigma_2^2)$ | RD $\mathcal{N}(\mu_{\mathrm{St}\omega}, \sigma_{\mathrm{St}\omega}^2)$ |
| *Swing* | *Hip* | R2 & PC $\mathcal{N}(0, \sigma_{\mathrm{C}}^2)$ | RD & PC $\mathcal{N}(0, \sigma_{\mathrm{C}}^2)$ | VC $\mathcal{N}(0, \sigma_{\mathrm{C}}^2)$ | – | – | RD $\mathcal{N}(\mu_{\mathrm{SwH}\omega}, \sigma_{\mathrm{SwH}\omega}^2)$ |
| *Stance* | *Knee* | PC $\mathcal{N}(0, \sigma_{\mathrm{C}}^2)$ | RD & PC $\mathcal{N}(0, \sigma_{\mathrm{C}}^2)$ | RD & R1 & VC $\mathcal{N}(0, \sigma_{\mathrm{C}}^2)$ | RD $\mathcal{N}(0, \sigma_2^2)$ | RD & R1 & VC $\mathcal{N}(0, \sigma_2^2)$ | RD $\mathcal{N}(\mu_{\mathrm{St}\omega}, \sigma_{\mathrm{St}\omega}^2)$ |
| *Swing* | *Knee* | PC $\mathcal{N}(0, \sigma_{\mathrm{C}}^2)$ | R3 $\mathcal{N}(0, \sigma_{\mathrm{SwK}}^2)$ | VC $\mathcal{N}(0, \sigma_{\mathrm{SwK}}^2)$ | – | – | RD $\mathcal{N}(\mu_{\mathrm{SwK}\omega}, \sigma_{\mathrm{SwK}\omega}^2)$ |
| *Stance* | *Ankle* | R4 & PC $\mathcal{N}(0, \sigma_{\mathrm{C}}^2)$ | RD & PC $\mathcal{N}(0, \sigma_{\mathrm{C}}^2)$ | R4 & VC $\mathcal{N}(0, \sigma_{\mathrm{C}}^2)$ | R4 & PC $\mathcal{N}(0, \sigma_2^2)$ | RD & VC $\mathcal{N}(0, \sigma_2^2)$ | RD $\mathcal{N}(\mu_{\mathrm{St}\omega}, \sigma_{\mathrm{St}\omega}^2)$ |
| *Swing* | *Ankle* | PC $\mathcal{N}(0, \sigma_{\mathrm{C}}^2)$ | RD & PC $\mathcal{N}(0, \sigma_{\mathrm{C}}^2)$ | VC $\mathcal{N}(0, \sigma_{\mathrm{C}}^2)$ | – | – | RD $\mathcal{N}(\mu_{\mathrm{SwA}\omega}, \sigma_{\mathrm{SwA}\omega}^2)$ |

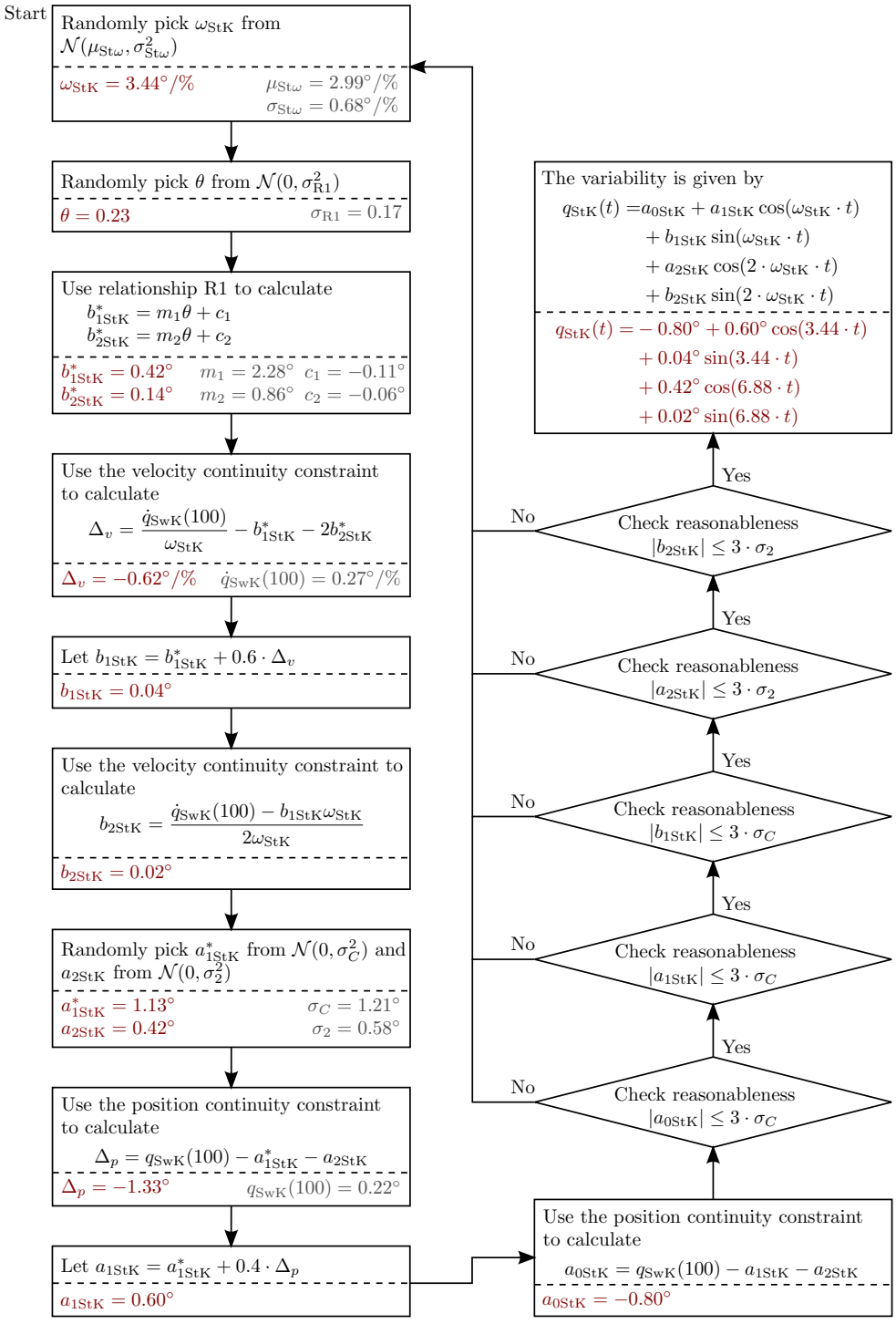



Figure 4: This flowchart shows how the stance knee coefficients are generated. Not shown for clarity is the procedure that prevents an infinite loop if all of the conditions cannot be satisfied. The top portion of each box describes how each quantity is found. The lower portion of each box presents an example case for one specific step. Values that are calculated are shown in red on the left. Constant values are shown in gray on the right. These values include the position ($q_{\mathrm{SwK}}(100)$) and velocity ($\dot{q}_{\mathrm{SwK}}(100)$) at the end of the previous swing period, which are constants for this particular step although they vary from one step to the next. The sin and cos functions take arguments in degrees. The process for the other joints is similar. In general, the frequency is randomly chosen first, then the $b_i$ terms are calculated followed by the $a_i$ terms. Finally, it is checked that each of the calculated coefficients is reasonable.

coefficients. In all cases, the frequency was chosen randomly from the appropriate distribution. In most cases, the $b_i$ coefficients were chosen next and used to enforce velocity continuity. For the swing periods, because there was only a $b_1$ term, the value was uniquely determined by the start-of-step velocity and the chosen frequency. For the stance periods, both $b_1$ and $b_2$ were used to satisfy the velocity continuity constraint. The weighting between randomly choosing $b_i$, satisfying a between-coefficient relationship, and ensuring velocity continuity was tuned for each joint to ensure a close match between the generated and experimental curves. The $a_i$ coefficients were then chosen. Because there were always at least two $a_i$ coefficients that could be used to satisfy the position continuity constraint, the weighting between randomly choosing $a_i$, satisfying a between-coefficient relationship, and ensuring position continuity was tuned to ensure a close match between the generated and experimental curves.

To avoid large ($> 20^\circ$) random walks in joint trajectories, the first set of coefficients found cannot always be used. Instead, after generating a set of coefficients for a particular joint during a particular step, it was checked that all of the calculated coefficients were within three standard deviations of the mean. If not, another set of coefficients was found. The process of finding and checking coefficients was repeated until an acceptable set was found or thirty attempts were made. If no acceptable set of coefficients was found, the set of coefficients with an angle closest to zero at the end of the step was used. More than 98% of the steps had acceptable coefficients.

This modeling method was evaluated both qualitatively and quantitatively. Because the generated curves were created randomly, only the second level of variability (behavior over many steps) was evaluated. The qualitative assessment primarily consisted of visually comparing plots with many steps. The quantitative assessment compared the distributions and relationships using standard statistical methods (t-tests for means and Levene's test for standard deviations). Similar to the pooling of the experimental data, 10 sets of 100 randomly generated strides were evaluated together.

## 3. Results

### *3.1. Experimental Data*

The original dataset had 1440 steps for each period; the fully processed data had between 477 and 786 steps depending on the joint and period. The fitted Fourier series well capture the dominate low-frequency variations (Fig. 2). Median $R^2$ values are at least 0.85 except for the swing ankle (0.68). Root mean square errors (RMSE) are low, particularly for the stance period (median $< 0.4^\circ$). The RMSE is strictly less than $2.3^\circ$ for all joints and steps, and for most steps at each joint, the RMSE $< 1.0^\circ$. Thus, a low-order Fourier series accurately captures the first level of variability, i.e., how the joint kinematics differ from the mean as a function of time for a single step.

The second level of variability (i.e., the distribution of variability in function space) can be partially evaluated visually (Fig. 5). The hip variability is approximately constant, although a few steps have a large deviation from zero for the first 50%. The stance knee also has approximately constant variance over the step except for a reduction at approximately 80%. The swing knee variability has a "bow-tie" shape, with peaks in variability at approximately 10% and 65%. The stance ankle has approximately constant variance for the first 60% before increasing for the remainder of the step. The swing ankle variability is approximately constant, although several steps have a large deviation from zero for the first 50%. It also has the largest oscillations of all joints.

The mean and standard deviation of each distribution provide a quantitative measure of the second level of variability. Because the experimental joint angles were centered, the magnitude terms were centered at $0^\circ$ (absolute value of means $< 0.10^\circ$ with large p-values, Fig. 6(a)). The standard deviations for the stance $a_2$ and $b_2$ coefficients were consistent across all three joints and relatively small ($\sigma_2 = 0.58^\circ$). The standard deviations for the swing knee $a_1$ and $b_1$ coefficients were large ($\sigma_{\mathrm{SwK}} = 2.03^\circ$). The standard deviations for all other coefficients were similar with a moderate value ($\sigma_{\mathrm{C}} = 1.21^\circ$). Thus, the magnitude coefficients can be described using three normal distributions.

The frequencies were strictly greater than zero as expected (Fig. 6(b)). The swing distributions were skewed slightly to the left such that the median was 0.20-0.44°/% smaller than the mean. The generated curves better matched the fitted curves if the median was used as the center of the distributions. Both the medians and standard deviations were greater for the swing period than for the stance period at all joints (p-values $\lll 0.001$). The stance distributions were similar, so they were combined into a single distribution

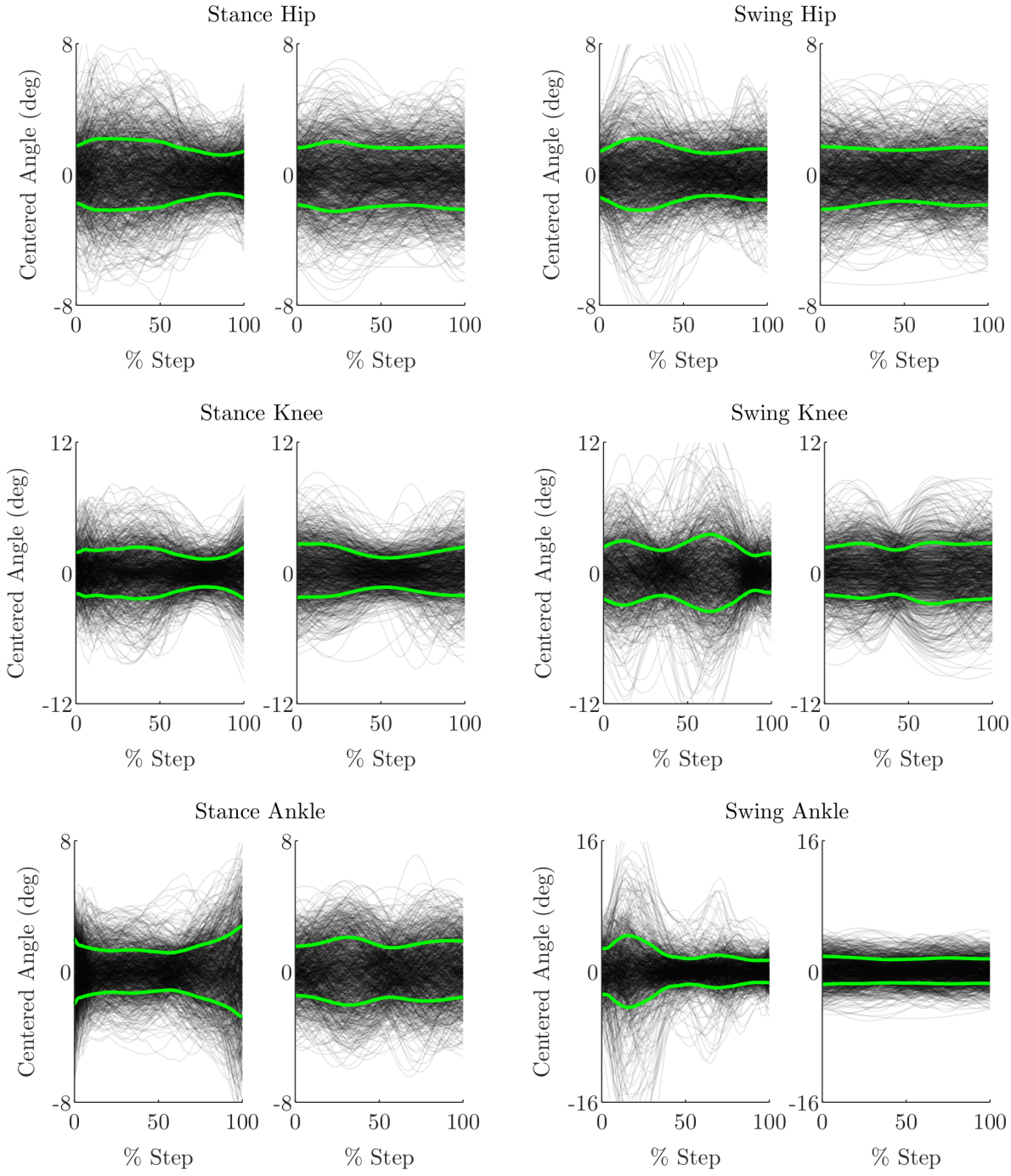
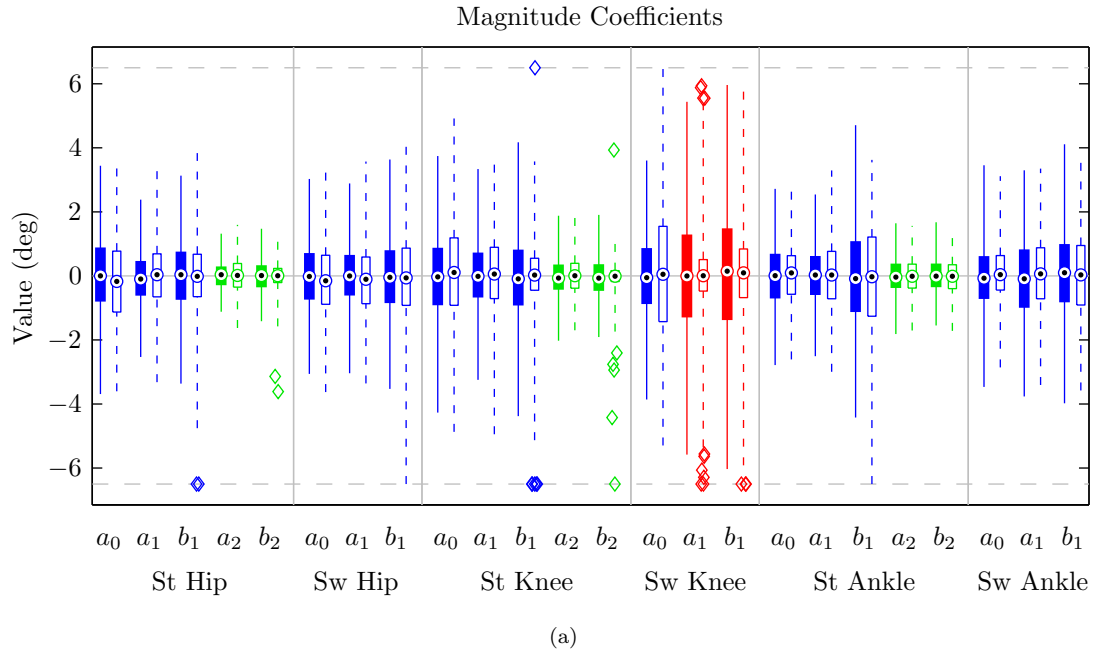
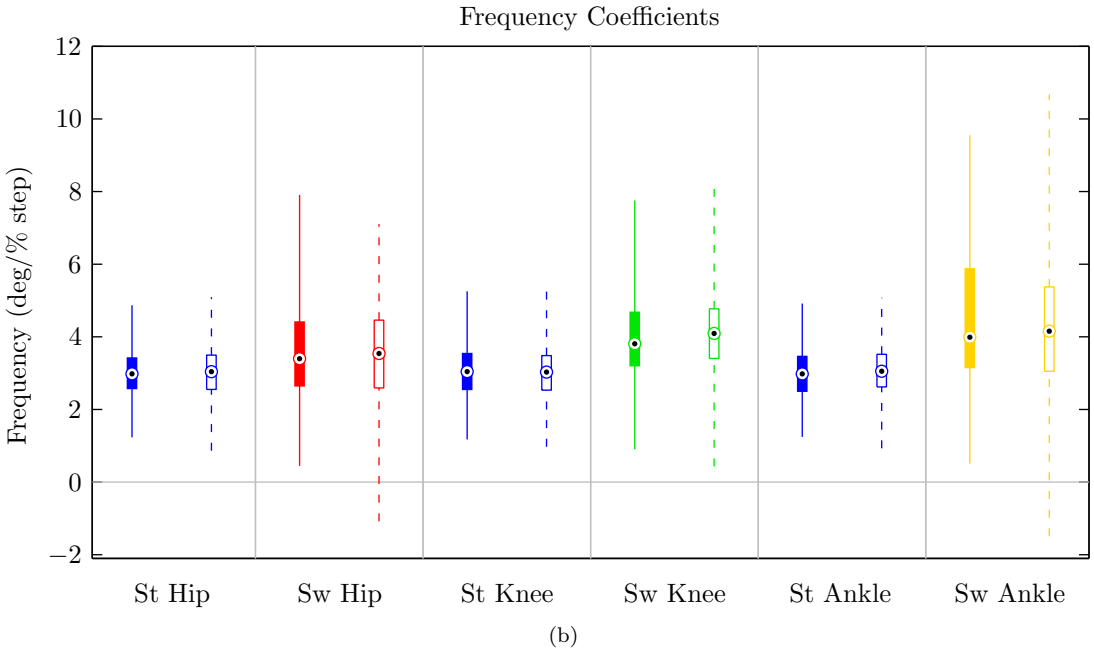



Figure 5: Each pair of plots shows the centered experimental joint angles on the left and the generated variability curves on the right. The thick green lines indicate plus or minus one point-wise standard deviation. Time has been normalized to lie between 0% and 100% step. The generated data only includes the steps that had acceptable coefficients. For the experimental swing ankle, all of the curves with the largest variability come from the same subject. For most joints, the centered experimental joint angles and the generated curves have similar structures.




(a)




(b)

Figure 6: The distributions of the (a) magnitude and (b) frequency coefficients for the fitted (filled boxes with solid whiskers) and generated (open boxes with dashed whiskers) Fourier series. 'St' stands for the stance period and 'Sw' stands for the swing period. Within each plot, each color represents a different combined distribution. The diamonds indicate outliers; extreme outliers are clipped to the broken gray line. The mean value of the magnitude coefficients were all approximately zero. In most cases, the variance in the $a_0$, $a_1$, and $b_1$ coefficients were similar, both across coefficients and between the fitted and generated curves. The variability for the $a_2$ and $b_2$ coefficients was smaller. The stance frequency distributions were all similar while each swing frequency distribution was different.

with mean/median $\mu_{\mathrm{St}\omega} = 2.99°/\%$ and standard deviation $\sigma_{\mathrm{St}\omega} = 0.68°/\%$. For the swing period, the median frequency increased as the joint became more distal (p-values $< 0.001$). Thus, each joint has its own distribution with means/medians $\mu_{\mathrm{SwH}\omega} = 3.40°/\%$, $\mu_{\mathrm{SwK}\omega} = 3.81°/\%$, and $\mu_{\mathrm{SwA}\omega} = 3.99°/\%$ and standard deviations $\sigma_{\mathrm{SwH}\omega} = 1.55°/\%$, $\sigma_{\mathrm{SwK}\omega} = 1.27°/\%$, and $\sigma_{\mathrm{SwA}\omega} = 1.77°/\%$ for the hip, knee, and ankle, respectively.

The hip and knee coefficients were correlated to each other while the ankle coefficients were only correlated with themselves. For the stance hip and knee, the coefficients $b_{1\mathrm{StH}}$, $b_{2\mathrm{StH}}$, $b_{1\mathrm{StK}}$, and $b_{2\mathrm{StK}}$ were correlated ($0.15 < |\rho| < 0.50$). For the swing hip and knee, the coefficients $a_{0\mathrm{SwH}}$, $b_{1\mathrm{SwH}}$, and $b_{1\mathrm{SwK}}$ were correlated ($0.14 < |\rho| < 0.51$). For the stance ankle, the coefficients $a_{0\mathrm{StA}}$, $b_{1\mathrm{StA}}$, and $a_{2\mathrm{StA}}$ were correlated ($0.46 < |\rho| < 0.69$). The swing knee phase shift (defined by $a_{1\mathrm{SwK}}$ and $b_{1\mathrm{SwK}}$) and frequency were negatively correlated. 40% of the swing knee steps fell along two lines with similar slopes and intercepts 171° apart (Fig. 3).

### *3.2. Generated Curves*

For the most part, the generated curves captured the key features of the variability (Fig. 5). The hip variability was approximately constant for both the experimental and generated curves. The generated stance knee variability was similar to the experimental variability, although the minimum occurred earlier (at approximately 50%). The swing knee variability for the generated curves peaked at approximately 20% and 65%, which was similar to the experimental data. At the ankle, the generated curves had more constant variance for both periods than was observed experimentally. The range of variance for the swing ankle was also smaller for the generated curves.

The distributions between the generated and experimental magnitude terms were similar (Fig. 6(a)). Just as for the experimental distributions, the mean values were small (absolute values $< 0.20°$, usually $< 0.10°$). The standard deviations for the generated curves were mostly similar to the experimental values, although the generated $a_{0\mathrm{SwK}}$ distribution had a much larger and the generated $a_{1\mathrm{SwK}}$ and $b_{1\mathrm{SwA}}$ distributions had smaller standard deviations than the distributions for the fitted curves. The generated and experimental frequency distributions had very similar medians and standard deviations except for the swing hip, which had a somewhat higher median (difference of $0.28°/\%$) and smaller standard deviation (difference of $0.13°/\%$) than was observed experimentally (Fig. 6(b)).

Just as for the fitted coefficients, many of the generated coefficients were correlated. For the stance hip and knee, $b_1$ and $b_2$ were strongly correlated ($\rho = 1.00$) at each joint. Despite including an across-joint relationship for both the stance and swing periods, the generated hip and knee coefficients were not correlated with each other. Thus, the generated curves did not capture the between-joint coordination. For the stance ankle, the coefficients $a_{0\mathrm{StA}}$ and $a_{2\mathrm{StA}}$ were correlated ($\rho = 0.51$). Unlike the fitted coefficients, the generated $b_{1\mathrm{StA}}$ coefficient was not correlated with the stance ankle $a_{i\mathrm{StA}}$ terms. When considering a single joint, the signs of the correlations between the stance and swing coefficients were usually the same for the fitted and generated coefficients, although the correlations for the generated coefficients tended to be stronger. Except for the stance hip and ankle, when the $a_0$ and $a_1$ terms shared the position constraint, there was often a moderate positive correlation between them ($\rho > 0.4$). The lack of correlation for the stance ankle may be caused by using all three $a$ terms to satisfy the position continuity constraint as opposed to just $a_0$ and $a_1$. Similarly, the velocity continuity constraint usually introduces strong correlations between the $b_1$ and $b_2$ terms ($\rho > 0.9$ for hip and knee). The generated swing knee phase shift-frequency relationship was very similar to the relationship for the fitted curves (Fig. 3).

## 4. Discussion

Low-order Fourier series are well able to capture the variability in overground walking at the self-selected speed for young, healthy adults. This agrees with motion capture validation studies, which have consistently reported similar point-wise standard deviations between subjects (McGinley et al., 2009). For swing joints, the variability is well approximated as simple pendular motion, possibly because the swing motion is largely passive, so the swing leg acts much like a multi-joint pendulum. The stance joint variability is somewhat more complex, as indicated by the second-order Fourier series. This may be due to increased active control of the stance joints. While the variation is well represented by a Fourier series, it is not periodic because the coefficients change every step and the period is typically not exactly one step. Consistent with previous work

(McGinley et al., 2009), the similar standard deviations of most magnitude coefficients indicate a similar magnitude of variability for all joints and periods.

The quality of all stance fits are similar, indicating similar amounts of high-frequency oscillation. The swing joints have more high-frequency oscillation than the stance joints. Further, more distal segments have higher magnitude oscillations, resulting in a deterioration of fit quality. The dominate frequency also increases for distal joints. More distal segments are lighter, which increases the natural frequency and decreases the effect of vibrations on whole-body motion. This may explain the increasing magnitude and frequency. The stance/swing difference may arise from the higher effective stiffness of the stance joints (Latash and Zatsiorsky, 1993) or from more precise control since it may be less critical to control the variation of distal swing segments compared to the stance joints for successful walking.

The correlation between the hip and knee is consistent with previous results on the variability of kinetics (Winter, 1984), but is somewhat in contrast to previous results which found strong synchronization between total knee and ankle movement (Zhang et al., 2010). Since it appears that only the hip and knee variability are correlated, this may suggest active control of variability. If variability simply resulted from neural control noise (Harris and Wolpert, 1998), then all of the variability should be correlated due to dynamic coupling. Alternatively, the coupling may arise from the many bi-articular muscles that cross the hip and knee joints but the relatively fewer multi-joint muscles at the ankle (Winter, 1984). These results support the proximo-distal gradient hypothesis (Daley et al., 2007) which proposes that the proximal (hip and knee) joint control is primarily feedforward and relatively insensitive to perturbations while the distal (ankle) control has a significant feedback loop. Under this paradigm, the hip and knee should likely have coupled variability since they share a feedfoward controller, while the ankle variability should be distinct.

The continuity constraints usually induce stronger correlations between terms than were observed with the fitted curves, which may be because the experimental data was isolated strides rather than a series of steps. Preliminary work indicated that matching variability for isolated steps was easier than for continuous walking, suggesting that continuity between steps has a significant impact. There are short-range correlations between sequential values of a coefficient. In addition, there may also be long-range correlations between time-series values of a coefficient, similar to the results for stride time (Hausdorff, 2007; Dingwell and Cusumano, 2000). To more fully investigate the effects of continuity constraints, further experiments with continuous walking are planned. In addition, non-parametric methods will be used to more fully characterize the variability across steps (Hausdorff et al., 1995; Taylor et al., 2016).

In summary, the joint-level variability for a single step for healthy individuals walking overground can be accurately described using a low-order Fourier series. The distribution of these Fourier series over many steps were characterized by three normal distributions for the magnitude terms and four normal distributions for the frequency terms. Coordination between joints and across steps was captured with relationships and continuity constraints, each relating two or more coefficients. Demonstrating that the coefficient distributions and relationships adequately described the experimental variation, randomly generated variability curves matched the experimental variation both qualitatively and quantitatively, particularly for the stance joints. To perform predictive studies, the purely position-based model developed in this work must be included in a dynamic model to ensure the movement is governed by the laws of physics and that the ground reaction forces are physically realizable. Depending on the choice of dynamic model controller, there may be additional constraints on the movement during double support to ensure both feet remain on the ground. Initial work has shown that the kinematic model developed in this paper can be incorporated in a moderately-complex dynamic model of human walking to more accurately simulate human locomotion (Martin and Gregg, 2016), which in turn can be used to systematically investigate the effect of variability on fall risk.

## 5. Conflict of Interest

None to report.

## 6. Acknowledgments

This work was supported by the National Institute of Child Health & Human Development of the NIH under Award Number DP2HD080349. The content is solely the responsibility of the authors and does not

necessarily represent the official views of the NIH. R. D. Gregg holds a Career Award at the Scientific Interface from the Burroughs Wellcome Fund.

## 7. References

### References

Ahn, J., Hogan, N., 2013. Long-range correlations in stride intervals may emerge from non-chaotic walking dynamics. Public Library of Science ONE 8, e73239.

Byl, K., Tedrake, R., 2009. Metastable walking machines. The International Journal of Robotics Research 28, 1040–64.

Daley, M.A., Felix, G., Biewener, A.A., 2007. Running stability is enhanced by a proximo-distal gradient in joint neuromechanical control. The Journal of Experimental Biology 210, 383–94.

Danion, F., Varraine, E., Bonnard, M., Pailhous, J., 2003. Stride variability in human gait: the effect of stride frequency and stride length. Gait & Posture 18, 69–77.

Daszykowski, M., Walczak, B., Massart, D.L., 2001. Looking for natural patterns in data part 1. Density-based approach. Chemometrics and Intelligent Laboratory Systems 56, 83–92.

Dingwell, J.B., Cusumano, J.P., 2000. Nonlinear time series analysis of normal and pathological human walking. Chaos 10, 848–63.

Dingwell, J.B., Cusumano, J.P., 2015. Identifying stride-to-stride control strategies in human treadmill walking. Public Library of Science ONE 10, e0124879.

Dingwell, J.B., Cusumano, J.P., Cavanagh, P.R., Sternad, D., 2001. Local dynamic stability versus kinematic variability of continuous overground and treadmill walking. Journal of Biomechanical Engineering 123, 27–32.

Hamacher, D., Singh, N.B., Van Dieën, J.H., Heller, M.O., Taylor, W.R., 2011. Kinematic measures for assessing gait stability in elderly individuals: a systematic review. Journal of The Royal Society Interface 8, 1682–98.

Harris, C.M., Wolpert, D.M., 1998. Signal-depended noise determines motor planning. Nature 394, 780–4.

Hausdorff, J.M., 2007. Gait dynamics, fractals and falls: finding meaning in the stride-to-stride fluctuations of human walking. Human Movement Science 26, 555–89.

Hausdorff, J.M., Peng, C.K., Ladin, Z., Wei, J.Y., Goldberger, A.L., 1995. Is walking a random walk? Evidence for long-range correlations in stride interval of human gait. Journal of Applied Physiology 78, 349–58.

Jolliffe, I.T., 2002. Principal Component Analysis. Second ed., Springer; , New York, USA.

Latash, M.L., Zatsiorsky, V.M., 1993. Joint stiffness: myth or reality? Human Movement Science 12, 653–92.

Martin, A.E., Gregg, R.D., 2015. Stable, Robust Hybrid Zero Dynamics Control of Powered Lower-Limb Prostheses. IEEE Transactions on Automatic Control In review.

Martin, A.E., Gregg, R.D., 2016. Incorporating human-like walking variability in an HZD-based bipedal model. IEEE Transactions on Robotics In press. doi:10.1109/TRO.2016.2572687.

Martin, A.E., Schmiedeler, J.P., 2014. Predicting human walking gaits with a simple planar model. Journal of Biomechanics 47, 1416–21.

McGinley, J.L., Baker, R., Wolfe, R., Morris, M.E., 2009. The reliability of three-dimensional kinematic gait measurements: a systematic review. Gait & Posture 29, 360–369.

Pailhous, J., Bonnard, M., 1992. Steady-state fluctuations of human walking. Behavioural Brain Research 47, 181–189.

Sawa, R., Doi, T., Misu, S., Tsutsumimoto, K., Nakakubo, S., Asai, T., Yamada, M., Ono, R., 2014. The association between fear of falling and gait variability in both leg and trunk movements. Gait & Posture 40, 123–7.

Stergiou, N., Decker, L.M., 2011. Human movement variability, nonlinear dynamics, and pathology: is there a connection? Human Movement Science 30, 869–88.

Taylor, P.G., Small, M., Lee, K.Y., Landeo, R., O'Meara, D.M., Millett, E.L., 2016. A Surrogate Technique for Investigating Deterministic Dynamics in Discrete Human Movement. Motor Control In press. doi:10.1123/mc.2015-0043.

Winter, D.A., 1984. Kinematic and kinetic patterns in human gait: variability and compensating effects. Human Movement Science 3, 51–76.

Yun, Y., Kim, H.C., Shin, S.Y., Lee, J., Deshpande, A.D., Kim, C., 2014. Statistical method for prediction of gait kinematics with Gaussian process regression. Journal of Biomechanics 47, 186–192.

Zhang, J., Zhang, K., Feng, J., Small, M., 2010. Rhythmic dynamics and synchronization via dimensionality reduction: application to human gait. Public Library of Science Computational Biology 6, e1001033.