# Separating Timing, Movement Conditions and Individual Differences in the Analysis of Human Movement

**Lars Lau Raket[1] \*, Britta Grimme[2], Gregor Schöner[2], Christian Igel[3], Bo Markussen[1]**

1 Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark, 2 Institut für Neuroinformatik, Ruhr-Universität Bochum, Bochum, Germany, 3 Department of Computer Science, University of Copenhagen, Copenhagen, Denmark

\* larslau@math.ku.dk

## Abstract

A central task in the analysis of human movement behavior is to determine systematic patterns and differences across experimental conditions, participants and repetitions. This is possible because human movement is highly regular, being constrained by invariance principles. Movement timing and movement path, in particular, are linked through scaling laws. Separating variations of movement timing from the spatial variations of movements is a well-known challenge that is addressed in current approaches only through forms of preprocessing that bias analysis. Here we propose a novel nonlinear mixed-effects model for analyzing temporally continuous signals that contain systematic effects in both timing and path. Identifiability issues of path relative to timing are overcome by using maximum likelihood estimation in which the most likely separation of space and time is chosen given the variation found in data. The model is applied to analyze experimental data of human arm movements in which participants move a hand-held object to a target location while avoiding an obstacle. The model is used to classify movement data according to participant. Comparison to alternative approaches establishes nonlinear mixed-effects models as viable alternatives to conventional analysis frameworks. The model is then combined with a novel factor-analysis model that estimates the low-dimensional subspace within which movements vary when the task demands vary. Our framework enables us to visualize different dimensions of movement variation and to test hypotheses about the effect of obstacle placement and height on the movement path. We demonstrate that the approach can be used to uncover new properties of human movement.

## Author Summary

When you move a cup to a new location on a table, the movement of lifting, transporting, and setting down the cup appears to be completely automatic. Although the hand could take continuously many different paths and move on any temporal trajectory, real movements are highly regular and reproducible. From repetition to repetition movements vary,

and the pattern of variance reflects movement conditions and movement timing. If another person performs the same task, the movement will be similar. When we look more closely, however, there are systematic individual differences. Some people will over-compensate when avoiding an obstacle and some people will systematically move slower than others. When we want to understand human movement, all these aspects are important. We want to know which parts of a movement are common across people and we want to quantify the different types of variability. Thus, the models we use to analyze movement data should contain all the mentioned effects. In this work, we developed a framework for statistical analysis of movement data that respects these structures of movements. We showed how this framework modeled the individual characteristics of participants better than other state-of-the-art modeling approaches. We combined the timing-and-path-separating model with a novel factor analysis model for analyzing the effect of obstacles on spatial movement paths. This combination allowed for an unprecedented ability to quantify and display different sources of variation in the data. We analyzed data from a designed experiment of arm movements under various obstacle avoidance conditions. Using the proposed statistical models, we documented three findings: a linearly amplified deviation in mean path related to increase in obstacle height; a consistent asymmetric pattern of variation along the movement path related to obstacle placement; and the existence of obstacle-distance invariant focal points where mean trajectories intersect in the frontal and vertical planes.

## Introduction

When humans move and manipulate objects, their hand paths are smooth, but also highly flexible. Humans do not move in a jerky, robot-like way that is sometimes humorously invoked to illustrated "unnatural" movement behavior. In fact, humans have a hard time making "arbitrary" movements. Even when they scribble freely in three dimensions, their hand moves in a regular way that is typically piecewise planar [1, 2]. Movement generation by the nervous systems, the neuro-muscular systems, and the body is constrained by implied laws of motion signatures which are found empirically through invariances of movement trajectories and movement paths. Among these, laws decoupling space and time are of particular importance. For instance, the fact that the trajectories of the hand have approximately bell-shaped velocity profiles across varying movement amplitudes [3] implies a scaling of the time dependence of velocity. The 2/3 power law [4] establishes an analogous scaling of time with the spatial path of the hand's movement. Similarly, the isochrony principle [5] captures that the same spatial segment of a movement takes up the same proportion of movement time as movement amplitude is rescaled. Several of these invariances can be linked to geometrical invariance principles [6].

These invariances imply that movements as a whole have a reproducible temporal form, which can be characterized by movement parameters. Their values are specified before a movement begins, so that one may predict the movement's time course and path based on just an initial portion of the trajectory [7]. Movement parameters are assumed to reside at the level of end-effector trajectories in space and their neural encoding begins to be known [8–10]. The set of possible movements can thus be spanned by a limited number of such parameters. Moreover, the choices of these movement parameters are constrained. For instance, in sequences of movements, earlier segments predict later segments [11].

A key source of variance of kinematic variables is, of course, the time course of the movement itself. The invariance principles suggest that this source of variance can be disentangled

from the variation induced when the movement task varies. In this paper, we will first address time as a source of variance, focussing on a fixed movement task, and then use the methods developed to address how movements vary when the task is varied.

Given a fixed movement task, movement trajectories also vary across individuals. Individual differences in movement, a personal movement style, are reproducible and stable over time, as witnessed, for instance, by the possibility to identify individuals or individual characteristics such as gender by movement information alone [12–14].

A third source of variation are fluctuations in how movements are performed from trial to trial or across movement cycles in rhythmic movements. Such fluctuations are of particular interest to movement scientists, because they reflect not only sources of random variability such as neural or muscular noise, but also the extent to which the mechanisms of movement generation stabilize movement against such noise. Instabilities in patterns of coordination have been detected by an increase of fluctuations [15] and differences in variance among different degrees of freedom have been used to establish priorities of neural control [16, 17].

A systematic method to disentangle these three sources of movement variation, time, individual differences, and fluctuations, would be a very helpful research tool. Such a method would decompose sets of observed kinematic time series into a common trajectory (that may be specific to the task), participant-specific movement traits, and random effects. Given the observed decoupling of space and time, such a decomposition would also separate the rescaling of time across these three factors from the variation of the spatial characteristics of movement.

The statistical subfield that deals with analysis of temporal trajectories is the field of functional data analysis. In the literature on functional data, the typical approach for handling continuous signals with time-warping effects is to pre-align samples under an oversimplified noise model in the hope of eliminating the effects of movement timing [18]. In contrast, we propose an analytic framework in which the decomposition of the signal is done simultaneously with the estimation of movement timing effects, so that samples are continually aligned under an estimated noise model. Furthermore, we account for both the task-dependent variation of movement and for individual differences (a brief review of warping in the modeling of biological motion is provided in the Methods section).

Decomposition of time series into a common effect (the time course of the movement given a fixed task), an individual effect, and random variation naturally leads to a mixed-effects formulation [19]. The addition of nonlinear timing effects gives the model the structure of a hierarchical nonlinear mixed-effects model [20]. We present a framework for maximum-likelihood estimation in the model and demonstrate that the method leads to high-quality templates that foster subsequent analysis (e.g., classification). We then show that the results of this analysis can be combined with other models to test hypotheses about the invariance of movement patterns across participants and task conditions. We demonstrate this by using the individual warping functions in a novel factor analysis model that captures variation of movement trajectories with task conditions.

We use as of yet unpublished data from a study of naturalistic movement that extends published work [21]. In the study, human participants transport a wooden cylinder from a starting to a target location while avoiding obstacles at different spatial positions along the path. Earlier work has shown that movement paths and trajectories in this relatively unconstrained, naturalistic movement task clearly reflect typical invariances of movement generation, including the planar nature of movement paths, spatiotemporal invariance of velocity profiles, and a local isochrony principle that reflects the decoupling of space and time [21]. By varying the obstacle configuration, the data include significant and non-trivial task-level variation. We begin by modeling a one-dimensional projection of the time courses of acceleration of the hand in space, which we decompose into a common pattern and the deviations from it that characterize

each participant. The timing of the acceleration profiles is determined by individual time warping functions which are of higher quality than conventional estimates, since timing and movement noise are modeled simultaneously. The high quality of the estimates is demonstrated by classifying movements according to participant. Finally, the results of the nonlinear mixed model are analyzed using a novel factor analysis model that estimates a low-dimensional subspace within which movement paths change when the task conditions are varied. This combination of statistical models makes it possible to separate and visualize the variation caused by experimental conditions, participants and repetition. Furthermore, we can formulate and test hypotheses about the effects of experimental conditions on movement paths. Using the proposed statistical models, we document three findings: a linearly amplified deviation in mean path related to increase in obstacle height; a consistent asymmetric pattern of variation along the movement path related to obstacle placement; and the existence of obstacle-distance invariant focal points where mean trajectories intersect in the frontal and vertical plane.

Software for performing the described types of simultaneous analyses of timing and movement effects are publicly available through the `pavpop` R package [22]. A short guide on model building and fitting in the proposed framework is available as Supporting Information, along with an application to handwritten signature data.

## Methods

### Experimental data set

Ten participants performed a series of simple, naturalistic motor acts in which they moved a wooden cylinder from a starting to a target position while avoiding a cylindric obstacle. The obstacle's height and positition along the movement path were varied across experiments (Fig 1).

The movements were recorded with the Visualeyez (Phoenix Technologies Inc.) motion capture system VZ 4000. Two trackers, each equipped with three cameras, were mounted on the wall 1.5 m above the working surface, so that both systems had an excellent view of the table. A wireless infrared light-emitting diode (IRED) was attached to the wooden cylinder. The trajectories of markers were recorded in three Cartesian dimensions at a sampling rate of 110 Hz based on a reference frame anchored on the table. The starting position projected to the table was taken as the origin of each trajectory in three-dimensional Cartesian space. Recorded movement paths for two experimental conditions are shown in Fig 2. The acceleration profiles considered in the following sections were obtained by using finite difference approximations of the raw velocity magnitude data, see Fig 3.

Obstacle avoidance was performed in 15 different conditions that combined three obstacle heights $S$, $M$, or $T$ with five distances of the obstacle from the starting position $d \in \{15, 22.5, 30, 37.5, 45\}$. A control condition had no obstacle. The participants performed each condition 10 times. Each *experimental condition* provided $n = 100$ functional samples for a total of $n_f = 1600$ functional samples in the dataset, leading to a total data size of $m = 175,535$ observed time points.

The present data set is described in detail in [23]. The experiment is a refined version of the experiment described in [21].

### Time warping of functional data

Not every movement has the exact same duration. Comparisons across movement conditions, participants, and repetitions are hampered by the resulting lack of alignment of the movement trajectories. For a single condition, this is illustrated in Fig 3(a) and 3(c), in which the duration of the movement clearly varies from participant to participant but also from repetition to repetition. Without alignment, it is difficult to compare movements. In an experiment such as the
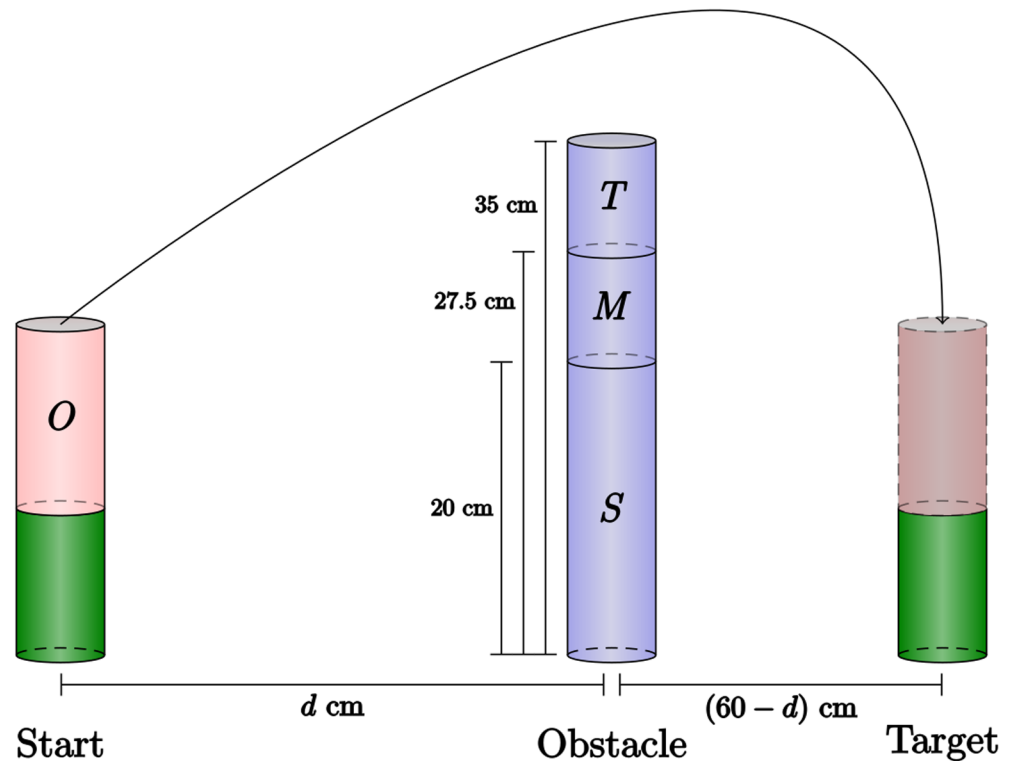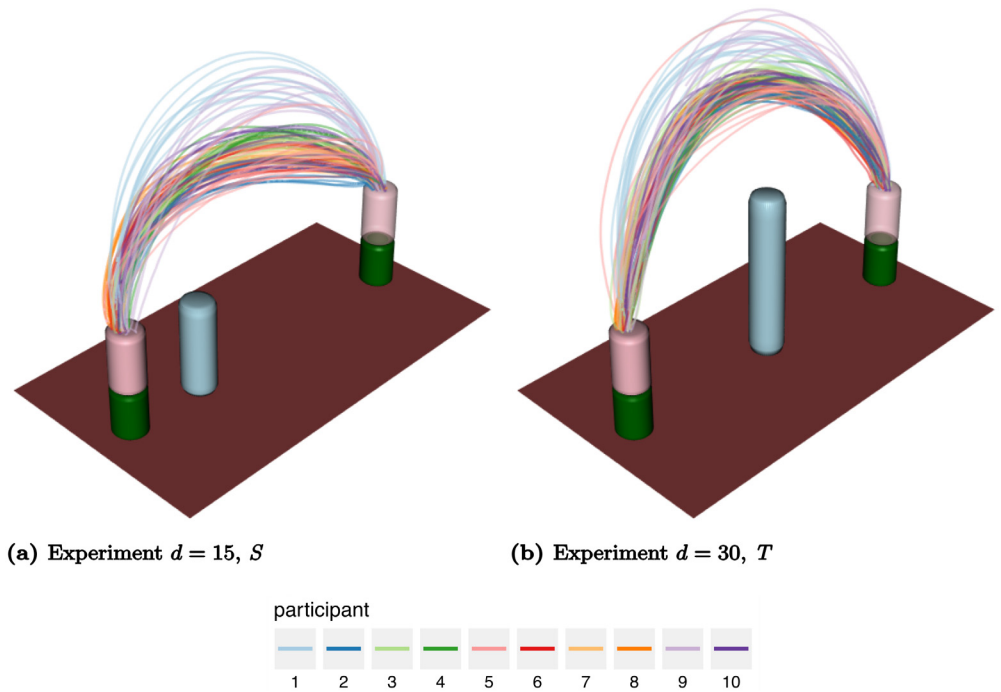
**Fig 1. Obstacle avoidance paradigm.** Participants move the cylindrical object *O* from the starting platform (green) to the target platform by lifting it over an obstacle. Obstacles of three different heights, small (*S*), medium (*M*), and tall (*T*), were used in the experiment, and the distance from starting position to obstacle *d* was varied.
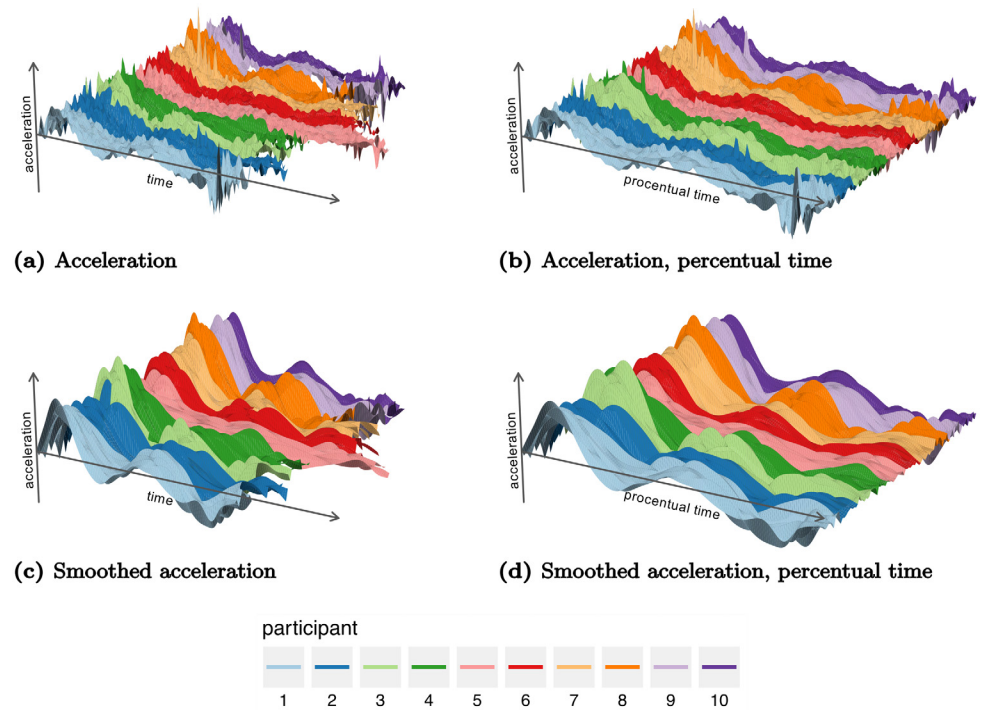
present, in which start and target positions are fixed, the standard solution for aligning samples is to use *percentual time*; the onset of the individual movement corresponds to time 0% and the end of the movement to time 100%. Such linear warping is based on detecting movement onset and offset through threshold criteria. As can be seen in Fig 3(b) and 3(c), linear warping does not align the characteristic features of the acceleration signals very well, however, as there is still considerable variation in the times at which acceleration peaks. There is, in other words, a nonlinear component to the variation of timing.

To handle nonlinear variation of timing, the signal must be time warped based on an estimated, continuous, and monotonically increasing function that maps percentual time to warped percentual time, such that the functional profiles of the signals are best aligned with each other. Such warping has traditionally been achieved by using the dynamic-time warping (DTW) algorithm [24] which offers a fast approach for globally optimal alignment under a pre-specified distance measure (for reviews of time warping in the domains of biological movement modeling see [25, 26]). DTW is both simple and elegant, but while it will often produce much better results than cross-sectional comparison of time-warped curves [27–29], it does suffer from some problems. In particular, DTW requires a pointwise distance measure such as Euclidean distance. Therefore, the algorithm cannot take serially correlated noise effects in a signal into account. As a result, basic unconstrained DTW will overfit in the sense of producing perfect fits whenever possible, and for areas that cannot be perfectly matched, either stretch them or compress them to a single point. In other words, DTW cannot model curves with systematic amplitude differences, and using DTW to naively compute time warped mean curves is

**(a)** Experiment $d = 15$, $S$          **(b)** Experiment $d = 30$, $T$

participant

1    2    3    4    5    6    7    8    9    10

**Fig 2. Recorded paths of the hand-held object in space for two experimental conditions: (a)** $d = 15$, $S$, **(b)** $d = 30$, $T$.

**(a)** Acceleration          **(b)** Acceleration, percentual time

**(c)** Smoothed acceleration          **(d)** Smoothed acceleration, percentual time

participant

1    2    3    4    5    6    7    8    9    10

**Fig 3. Surface plots of acceleration profiles ordered by repetition (y-axis) in the experiment with** $d = 30$, $T$. The figures display (a) raw acceleration in recorded time, (b) raw acceleration in percentual time, (c) smoothed acceleration in recorded time, and (d) smoothed acceleration in percentual time. The plots allow visualization of the variation across participants and repetitions.

**(a)** Recorded $z$-coordinate, observed time

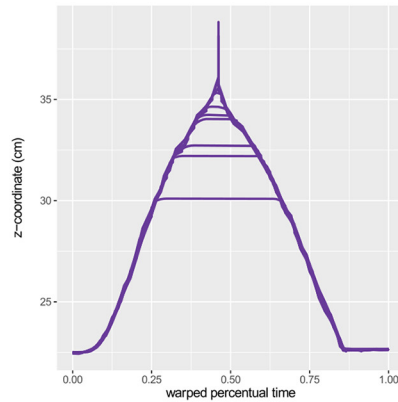**(b)** Recorded $z$-coordinate, percentual time
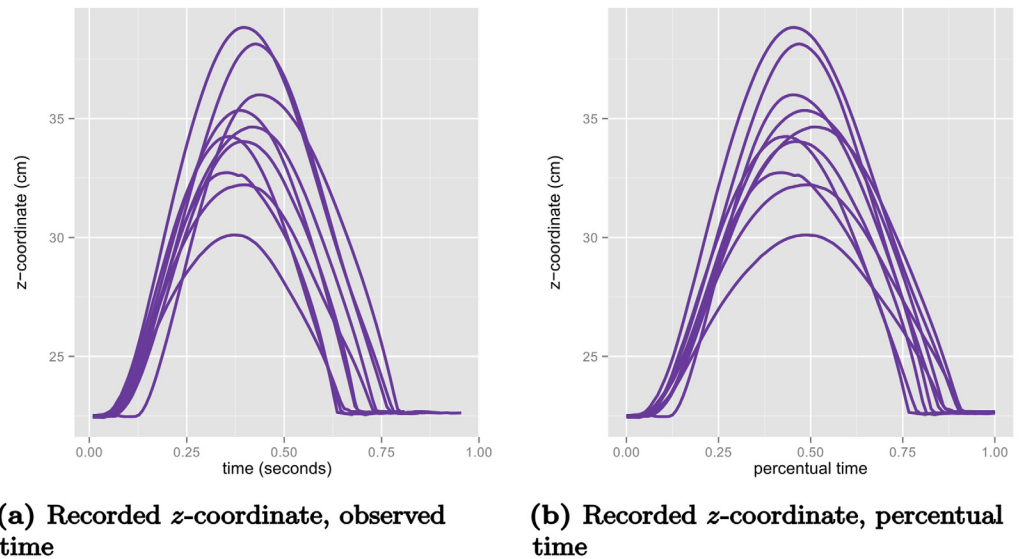
**Fig 4. Elevation ($z$-dimension) in (a) observed time and (b) percentual time for one participant's (no.10) repetitions of the control condition without obstacle plotted on different time scales.**

doi:10.1371/journal.pcbi.1005092.g004

in general problematic [30, 31]. This lacking ability to model serial correlated effects can be somewhat mitigated by restricting the DTW step pattern, in particular through a reduced search window for the warping function and constraints on the maximal step sizes. These are, however, hard model choices, they are restrictions on the set of possible warping functions, and they are a difficult to interpret since they seek to fix a problem in amplitude variation by penalizing warping variation. Instead, a much more natural approach would be to use a data term that models the amplitude variation encountered in data, and to impose warp regularization by using a cost function that puts high cost on undesired warping functions. In the following sections, we will propose a model with these properties, which, in addition, allows for estimating the data term, warp regularization and their relative weights from the data.

To illustrate the difference between DTW and the proposed method, consider the example displayed in Fig 4 where the recorded $z$-coordinates (elevation) of one participant's 10 movements in the control condition (without obstacle) are plotted in recorded time (a) and percentual time (b). These samples have been aligned using DTW by iteratively estimating a pointwise mean function and aligning the samples to the mean function (10 iterations). The three rows of Fig 5 display the results of the procedure using three different step patterns. We first note the strong overfitting of the symmetric and asymmetric step patterns, where the sequences with highest elevation are collapsed to minimize the residual. Secondly, we note the jagged warping functions that are results of the discrete nature of the DTW procedure. For comparison, we fitted a variant of the proposed model with a continuous model for the warping function controlled by 13 basis functions. We modeled both amplitude and warping effects as random Gaussian processes using simple, but versatile classes of covariance functions (see Supporting Information), and estimated the internal weighting of the effect directly from the samples. The results are displayed in Fig 6. We see that the warped elevation trajectories seem perfectly aligned, and that the corresponding warping functions are relatively simple, with the majority of variation being near the end of the movement. It is evident from the figures that both the alignment is much more reasonable, and the warping functions are much simpler than the warping functions found using DTW.
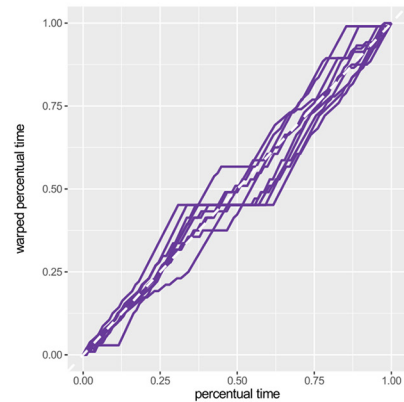
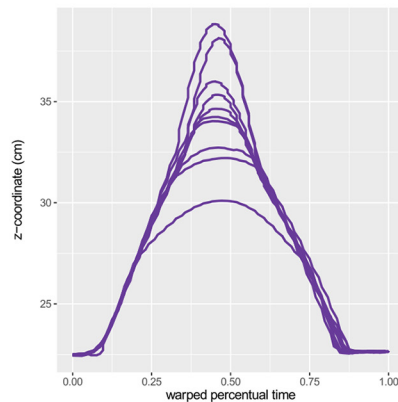**(a)** Recorded *z*-coordinate, warped percentual time, DTW symmetric step pattern
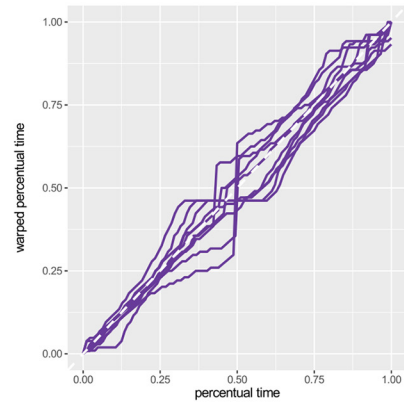
**(b)** Warping functions, white dashed line shows the identity, DTW symmetric step pattern
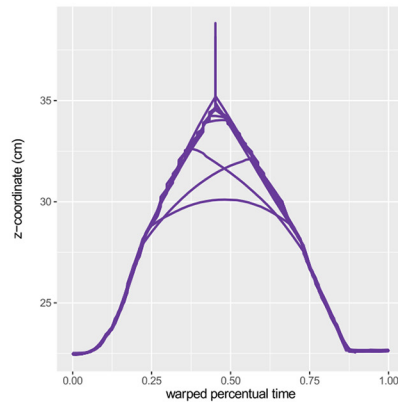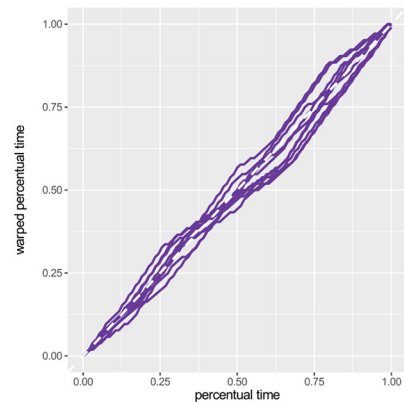
**(c)** Recorded *z*-coordinate, warped percentual time, DTW asymmetric step pattern

**(d)** Warping functions, white dashed line shows the identity, DTW asymmetric step pattern

**(e)** Recorded *z*-coordinate, warped percentual time, DTW Sakoe-Chiba step pattern

**(f)** Warping functions, white dashed line shows the identity, DTW Sakoe-Chiba step pattern

**Fig 5. The elevation trajectories from Fig 4 warped with dynamic time warping using different step patterns.** (a)-(b) *Symmetric* step pattern denotes the so-called White-Neely step pattern that has no local constraints, (c)-(d) *asymmetric* step pattern denotes a slope-constrained step pattern where local slopes are required to be between 0 and 2, and (e)-(f) *Sakoe-Chiba* step pattern denotes the asymmetric step pattern proposed in [24, Table I] with a slope constraint of 2.

doi:10.1371/journal.pcbi.1005092.g005

**(a)** Recorded *z*-coordinate, warped percentual time, proposed model

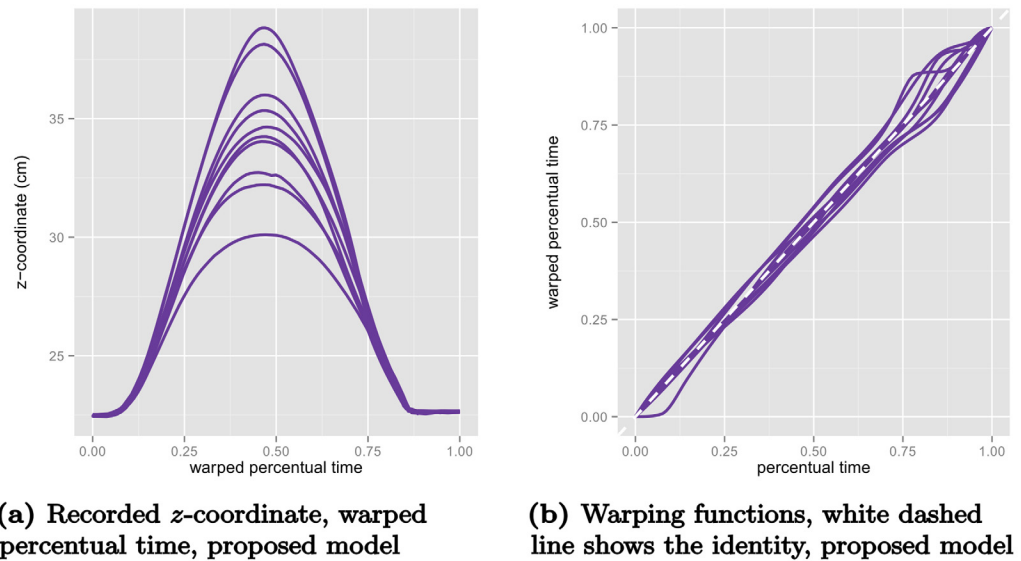**(b)** Warping functions, white dashed line shows the identity, proposed model

**Fig 6. The elevation trajectories from Fig 4 warped with the proposed model (a), and the corresponding predicted warping functions (b).**

Two questions naturally arise. Firstly, are the warping functions unique? Other warping functions could perhaps have produced similarly well-aligned data. Secondly, do we want perfectly aligned trajectories? There is still considerable variation of the amplitude in the warped *z*-coordinates of the movement trajectories in Fig 4(c), for instance. Some of the variation visible in the unwarped variant in Fig 4(b) could be due to random variations in amplitude rather than timing.

Using the time-warping functions that were determined by aligning the z-coordinates of the movements to now warp the trajectories for the *x*- and *y*-coordinates (Fig 7), we see that the
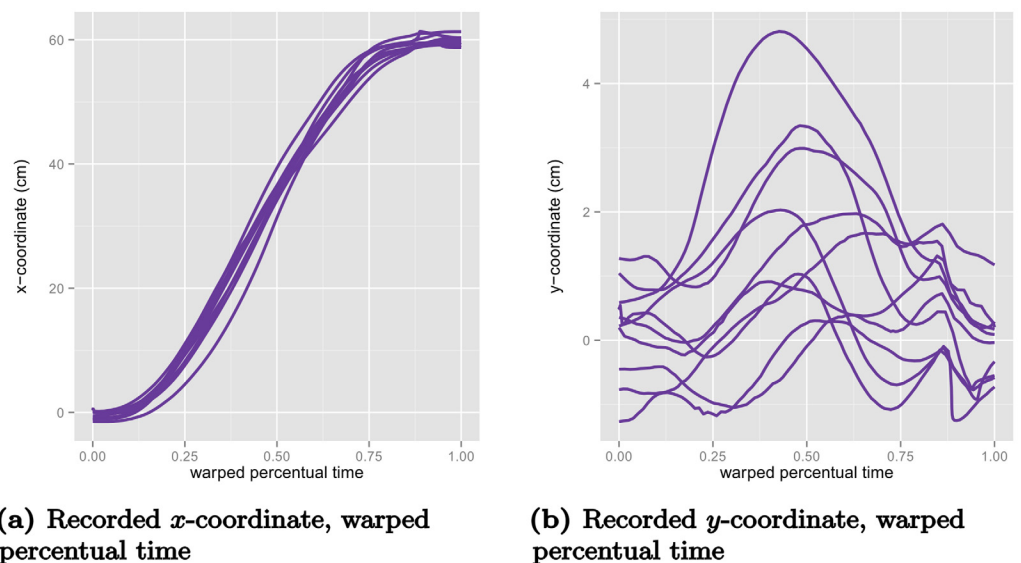


**(a)** Recorded *x*-coordinate, warped percentual time

**(b)** Recorded *y*-coordinate, warped percentual time

**Fig 7. The other two spatial coordinates of the movements from Fig 4 warped with the warping functions from Fig 6, that were estimated from the elevation component *z*.** (a) displays the warped *x*-coordinate trajectories and (b) displays the warped *y*-coordinate trajectories.

alignment obtained is not perfect. Conversely, were we to use warping functions determined for the x- or y-coordinate to warp the z-coordinate we would encounter similarly imperfect alignment. Thus, we need a method that avoids over-aligning, and represents a trade-off between the complexity of the warping functions and of the amplitude variation in data. In the next section, we introduce a statistical model that handle this trade-off in a data-driven fashion by using the patterns of variation in the data to find the most likely separation of amplitude and timing variation.

**Statistical modeling of movement data to achieve time warping.** In the following, we describe inference for a single experimental condition. For a given experimental condition—an object that needs to be moved to a target and an obstacle that needs to be avoided—we assume there is a common underlying pattern in all acceleration profiles; all $n_p = 10$ participants will lift the object and move it toward the target, lifting it over the obstacle at some point. This assumption is supported by the pattern in the data that Fig 3 visualizes. We denote the hypothesized underlying acceleration profile shared across participants and repetitions by $\theta$. In addition to this fixed acceleration profile, we assume that each participant, $i$, has a typical deviation $\varphi_i$ from $\theta$, so that the acceleration profile that is characteristic of that participant is $\theta + \varphi_i$. Such a systematic pattern characteristic of each participant is apparent in Fig 3(c) and 3(d). The individual trials (repetitions) of the movement deviate from this characteristic profile of the individual. We model these deviations as additive random effects with serial correlations so that for each repetition, $j$, of the experimental condition we have an additive random effect $x_{ij}$ that causes deviation from the ideal profile. Finally we assume that the data contains observation noise $\varepsilon_{ij}$ tied to the tracking system and data processing.

Time was implicit, up to this point, and the observed acceleration profile was decomposed into additive, linear contributions. We now assume, in addition, that each participant, $i$, has a consistent timing of the movement across repetitions, that is reflected in the temporal deformation of the acceleration profile (Fig 3(a) and 3(c)) and is captured by the time warping function $v_i$. On each repetition, $j$, of the condition, the timing of participant, $i$, contains a random variation of timing around $v_i$ captured by a random warping function $v_{ij}$ (see Fig 3).

Altogether, we have described the following statistical model of the observed acceleration profiles across participants:

$$y_{ij}(t) = (\theta + \varphi_i) \circ (v_i + v_{ij})(t) + x_{ij}(t) + \varepsilon_{ij}(t) \tag{1}$$

where $\circ$ denotes functional composition, $t$ denotes time, $\theta$, $\varphi_i$, $v_i : \mathbb{R} \to \mathbb{R}$ are fixed effects and $v_{ij}$, $x_{ij}$ and $\varepsilon_{ij}$ are random effects. The serially correlated effect $x_{ij}$ is assumed to be a zero-mean Gaussian process with a parametric covariance function $\mathcal{S} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$; the randomness of the warping function $v_{ij}$ is assumed to be completely characterized by a latent vector of $n_w$ zero-mean Gaussian random variables $\boldsymbol{w}_{ij}$ with covariance matrix $\sigma^2 C$; and $\varepsilon_{ij}$ is Gaussian white noise with variance $\sigma^2$.

Compared to conventional methods for achieving time warping, the proposed model (1) models amplitude and warping variation between repetition as random effects, which enables separation of the effects from the joint likelihood. Conventional approaches for warping model warping functions as fixed effects and do not contain amplitude effects [18]. The idea of modeling warping functions as random effects have previously been considered by [32–34] where warps were modeled as random shifts or random polynomials. None of these works however included amplitude variation. Recently, some works have considered models with random affine transformations for warps and amplitude variation in relation to growth curve analysis [35, 36]. A generalization that does not require affine transformations for warp and amplitude variation is presented in [37]. The presented model (1) is a hierarchical generalization of the

model presented in [37]. In the context of aligning image sequences in human movement analysis, morphable models [29, 38] model an observed movement pattern as a linear combination of prototypical patterns using both nonlinear warping functions (estimated using DTW) and spatial shifts. Thus morphable models are similar to the warping approaches that model both warp and amplitude effects as fixed [39].

**Maximum likelihood estimation of parameters.**    Model (1) has a considerable number of parameters, both for linear and nonlinear dependencies on the underlying state variable acceleration. The model also has effects that interact. This renders direct simultaneous likelihood estimation intractable. Instead we propose a scheme in which fixed effects and parameters are estimated and random effects are predicted iteratively on three different levels of modeling.

**Nonlinear model** At the nonlinear level, we consider the original model (1), and simultaneously perform conditional likelihood estimation of the participant-specific warping functions and predict the random warping functions from the negative log posterior. All other parameters remain fixed.

**Fixed warp model** At the fixed warp level, we fix the participant-specific warping effect $v_i$ at the conditional maximum likelihood estimate, and the random warping function $v_{ij}$, at the predicted values. The resulting model is an approximate linear mixed-effects model with Gaussian random effects $x_{ij}$ and $\varepsilon_{ij}$, that allows direct maximum-likelihood estimation of the remaining fixed effects, $\theta$ and $\varphi_i$.

**Linearized model** At the linearized level, we consider the first-order Taylor approximation of model (1) in the random warp $v_{ij}$. This linearization is done around the estimate of $v_i$ plus the given prediction of $v_{ij}$ from the nonlinear model. The result is again a linear mixed-effects model, for which one can compute the likelihood explicitly, while taking the uncertainty of all random effects—including the nonlinear effect $v_{ij}$—into account. At this level all variance parameters are estimated using maximum-likelihood estimation.

The estimation/prediction procedure is inspired by the algorithmic framework proposed in [37]. The estimation procedure is, however, adapted to the hierarchical structure of data and refined in several respects. On the linearized model level, the nonlinear Gaussian random effects are approximated by linear combinations of correlated Gaussian variables around the mode of the nonlinear density. The linearization step thus corresponds to a Laplace approximation of the likelihood, and the quality of this approximation is approximately second order [40].

Let $\boldsymbol{y}_{ij}$ be the vector of the $m_{ij}$ observations for participant $i$'s $j$th replication of the given experimental condition, and let $\boldsymbol{y}_i$ denote the concatenation of all functional observations of participant $i$ in the experimental condition, and $\boldsymbol{y}$ the concatenation all these observations across participants. We denote the lengths of these vectors by $m_i$ and $m$. Furthermore, let $\sigma^2 S_{ij}$, $\sigma^2 S_i$ and $\sigma^2 S$ denote the covariance matrices of $x_{ij} = (x_{ij}(t_k))_k$, $\boldsymbol{x}_i = (\boldsymbol{x}_{ij})_j$, and $\boldsymbol{x} = (\boldsymbol{x}_i)_i$ respectively. We note that the index set for $k$ depends on $i$ and $j$ since the covariance matrices $S_{ij}$ vary in size due to the different durations of the movements and because of possible missing values when markers are occluded.

We note that all random effects are scaled by the noise standard deviation $\sigma$. This parametrization is chosen because it simplifies the likelihood computations, as we shall see. Finally, we denote the norm induced by a full-rank covariance matrix $A$ by $\|z\|_A^2 = z^\top A^{-1} z$.

**Fixed warp level.**    We model the underlying profile, $\theta$, and the participant-specific variation around this trajectory, $\varphi_i$, in the common (warped) functional basis $\boldsymbol{\Phi} \in \mathbb{R}^{m \times K}$, with weights $\boldsymbol{c} = (c_1, \ldots, c_K)$ for $\theta$ and $\boldsymbol{d}_i = (d_{i1}, \ldots, d_{iK})$ for $\varphi_i$. We assume that the participant-

specific variations, $\varphi_i$, are centered around $\theta$ and thus $\sum_i \boldsymbol{d}_i = \boldsymbol{0} \in \mathbb{R}^K$. Furthermore, the square magnitude of the weights, $\boldsymbol{d}_i$, is penalized with a weighting factor $\eta$. This penalization helps guiding the alignment process in the direction of the highest possible level of detail in the common profile $\theta$ when the initial alignment is poor.

For fixed warping functions $v_i$ and $v_{ij}$, the negative log likelihood function in $\theta = \Phi \boldsymbol{c}$ is proportional to

$$\ell(\boldsymbol{c}) = \|\boldsymbol{y} - \Phi \boldsymbol{c}\|^2_{\mathbb{I}_n + S}$$

where $\mathbb{I}_n$ denotes the $n \times n$ identity matrix. This yields the estimate

$$\hat{\boldsymbol{c}} = (\Phi^\top (\mathbb{I}_m + S)^{-1} \Phi)^{-1} \Phi^\top (\mathbb{I}_m + S)^{-1} \boldsymbol{y}.$$

The penalized negative profile log likelihood for the weights $\boldsymbol{d}_i$ for $\varphi_i$ is proportional to

$$\ell(\boldsymbol{d}_i) = \|\boldsymbol{y}_i - \Phi_i(\hat{\boldsymbol{c}} + \boldsymbol{d}_i)\|^2_{\mathbb{I}_{m_i} + S_i} + \eta \boldsymbol{d}_i^\top \boldsymbol{d}_i,$$

which gives the maximum likelihood estimator

$$\hat{\boldsymbol{d}}_i = (\Phi_i^\top (\mathbb{I}_{m_i} + S_i)^{-1} \Phi_i + \eta \mathbb{I}_K)^{-1} \Phi_i^\top (\mathbb{I}_{m_i} + S_i)^{-1} (\boldsymbol{y}_i - \Phi_i \hat{\boldsymbol{c}}).$$

**Nonlinear level.** Similarly to the linear mixed-effects setting [41], it is natural to predict nonlinear random effects from the posterior [20], since these predictions correspond to the most likely values of the random effects given the observed data. Recall that the Gaussian variables, $\boldsymbol{w}_{ij}$, parametrize the randomness of the repetition-specific warping function $v_{ij}$. Since the conditional negative (profile) log likelihood function in $v_i$ given the random warping function $v_{ij}$ and the negative (profile) log posterior for $\boldsymbol{w}_{ij}$ coincide, we propose to simultaneously estimate the fixed warping effects $v_i$ and predict the random warping effects $v_{ij}$ from the joint conditional negative log likelihood/negative log posterior which is proportional to

$$p(v_i, \boldsymbol{w}_{ij}) = \sum_j \|\boldsymbol{y}_{ij} - (\hat{\theta} + \hat{\varphi}_i) \circ (v_i + v_{ij})(t_k)_k\|^2_{\mathbb{I}_{n_{ij}} + S_{ij}} + \sum_j \|\boldsymbol{w}_{ij}\|^2_c. \tag{2}$$

Since the variables $\boldsymbol{w}_{ij}$ can be arbitrarily transformed through the choice of warping function $v_{ij}$, the assumption that variables are Gaussian is merely one of computational convenience.

**Linearized level.** We can write the local linearization of model (1) in the random warping parameters $\boldsymbol{w}_{ij}$ around a given prediction $\boldsymbol{w}_{ij}^0$ as a vectorized linear mixed-effects model

$$\boldsymbol{y} \approx \boldsymbol{\vartheta} + Z(\boldsymbol{w} - \boldsymbol{w}^0) + \boldsymbol{x} + \boldsymbol{\varepsilon} \tag{3}$$

with effects given by

$$\boldsymbol{\vartheta} = \{(\theta + \varphi_i) \circ (v_i + v_{ij}^0)(t_k)\}_{ijk} \in \mathbb{R}^m,$$

$$Z = \mathrm{diag}(Z_{ij})_{ij}, \qquad Z_{ij} = \{\partial_t(\theta + \varphi_i) \circ (v_i + v_{ij}^0)(t_k)(\nabla_{\boldsymbol{w}} v_{ij}^0(t_k))^\top\}_k \in \mathbb{R}^{m_i \times n_w},$$

$$\boldsymbol{w} = (\boldsymbol{w}_{ij})_{ij} \sim \mathcal{N}_{n_p n_w}(0, \sigma^2 \mathbb{I}_{n_p} \otimes C), \qquad \boldsymbol{x} \sim \mathcal{N}_m(0, \sigma^2 S), \qquad \boldsymbol{\varepsilon} \sim \mathcal{N}_m(0, \sigma^2 \mathbb{I}_m),$$

where $v_{ij}^0$ indicates that the warping function is evaluated at the prediction $\boldsymbol{w}_{ij}^0$, $\mathrm{diag}(Z_{ij})_{ij}$ is the block diagonal matrix with the $Z_{ij}$ matrices along its diagonal, and $\otimes$ denotes the Kronecker product.

Altogether, twice the negative profile log likelihood function for the linearized model (3) is

$$\ell(\sigma^2, C, S) = m \log \sigma^2 + \log \det V + \sigma^{-2} \| y - \hat{\boldsymbol{\vartheta}} + Z w^0 \|_V^2 \tag{4}$$

where $V = S + Z(\mathbb{I}_n \otimes C) Z^\top + \mathbb{I}_m$.

**Modeling of effects and algorithmic approach.** So far, the model (1) has only been presented in a general sense. We now consider the specific modeling choices. The acceleration data has been rescaled using a common scaling for all experimental conditions, such that the span of data values has length 1 and the global timespan is the interval [0, 1].

To model the amplitude effects, we use a cubic B-spline basis $\Phi$ with $K$ knots [42].

We require that the fixed warping function $v_i$ is an increasing piecewise linear homeomorphism parametrized by $n_w$ equidistant anchor points in (0, 1), and assume that $v_{ij}$ is of the form

$$v_{ij}(t) = t + \mathcal{E}_{ij}(t),$$

where $\mathcal{E}_{ij}(t)$ is the linear interpolation at $t$ of the values $w_{ij}$ placed at the $n_w$ anchor points in (0, 1). In the given experimental setting, the movement path is fixed at the onset and the end of the movement. The movement starts when the cylindrical object is lifted and ends when it is placed at its target position. Thus, we would expect the biggest variation in timing to be in the middle of the movement (in percentual time). These properties can be modeled by assuming that $w_{ij}$ is a discretely observed zero-drift Brownian bridge with scale $\sigma^2 \gamma^2$ [43], which means that the covariance matrix $\sigma^2 C$ is given by point evaluation of the covariance function

$$\mathcal{C}(t, t') = \sigma^2 \gamma^2 \, t(1 - t')$$

for $t \leq t'$. When predicting the warps from the negative log posterior we restrict the search space to warps $v_i$ and $v_i + v_{ij}$ that are increasing homeomorphic maps of the domain [0, 1] onto itself. The conditional distribution of $v_{ij}$ given this restriction is slightly changed. For the used numbers of anchor points $n_w$ and the estimated variance parameters the difference is however minuscule, and we use the original Brownian model as a high-quality approximation of the true distribution.

We assume that the sample paths of the serially correlated effects $x_{ij}$ are continuous and that the process is stationary [44]. A natural choice of covariance is then the Matérn covariance with smoothness parameter $\mu$, scale $\sigma^2 \tau^2$ and range $1/\alpha$ [45], since it offers a broad class of stationarity covariance functions.

Finally, in order to consistently penalize the participant-specific spline across experimental conditions with varying variance parameters, we will use penalization weights that are normalized with the variance of the amplitude effects, $\eta = \lambda / (1 + \tau^2)$.

The algorithm for doing inference in model (1) is outlined in Algorithm 1. We have found that $i_{\max} = j_{\max} = 5$ outer and inner loops are sufficient for convergence. A wide variety of these types of models can be fitted using the `pavpop` R package [22]. A short guide on model building and fitting is available as Supporting Information.

**Algorithm 1** Maximum likelihood estimation for model (1)

```
1: procedure MLE (y, η, τ², α, γ²)
2:    Compute θ̂ and φ̂₁,…, φ̂ₘ assuming an identity warp          ▷ Initialize
3:    for i = 1, …, iₘₐₓ do                                       ▷ Outer iterations
4:      for j = 1, …, jₘₐₓ do                                     ▷ Inner iterations
5:        Estimate and predict warping functions by minimizing the posterior
              Eq (2)
6:        if Estimates and predictions do not change then break
7:      end if
```

```
 8:        Recompute $\hat{\theta}$ and $\hat{\varphi}_1, \ldots, \hat{\varphi}_m$
 9:    end for
10:    Estimate variance parameters by minimizing the linearized likeli-
          hood Eq (4)
11:  end for
12:  return $\hat{\theta}$, $\hat{\varphi}_1, \ldots, \hat{\varphi}_m$, $\hat{v}_1, \ldots, \hat{v}_m$, $\hat{\sigma}^2$, $\hat{\eta}$, $\hat{\tau}^2$, $\hat{\alpha}$, $\hat{\gamma}^2$        ▷ Maximum likelihood
          estimates
13: end procedure
```

In the following we consider two approaches, with (1) samples parametrized by recorded time (Fig 3a) and (2) samples parametrized by percentual time (Fig 3b). Parameters for the latter case will be denoted by a subscript p.

The number of basis functions $K$, the number of warping anchor points $n_w$, and the regularization parameter $\lambda$ were determined by the average 5-fold cross-validation score on each of three experimental conditions ($d = 30$ cm and obstacle heights $S$, $M$, and $T$). The models were fitted using the method described in the previous section, and the quality of the models was evaluated through the accuracy of classifying the participant from a given movement in the test set, using posterior distance between the sample and the combined estimates for the fixed effects $(\theta + \varphi_i) \circ v_i$. The cross-validation was done over a grid of the following values $\mu$, $\mu_\mathrm{p} \in \{0.5, 1, 2\}$, $K \in \{8, 13, 18, 23, 28, 33\}$, $K_\mathrm{p} \in \{8, 9, \ldots, 18\}$, $n_w$, $n_{w\mathrm{p}} \in \{0, 1, 2, 3, 5, 10\}$, $\lambda$, $\lambda_\mathrm{p} \in \{0, 1, 2, 3\}$. The best values were found to be $\mu = 2$, $K = 23$, $n_w = 2$, $\lambda = 2$ and $\mu_\mathrm{p} = 1$ $K_\mathrm{p} = 12$, $n_{w\mathrm{p}} = 1$, $\lambda_\mathrm{p} = 0$. We note that the smoothness parameter $\mu = 2$ is on the boundary of the cross-validation grid. The qualitative difference between second order smoothness $\mu = 2$ (corresponding to twice differentiable sample paths of amplitude effects) and higher order is so small, however, that we chose to ignore higher order smoothness. Furthermore, $\lambda_\mathrm{p} = 0$ indicates that we do not need to penalize participant-specific amplitude effects when working with percentual time. The reason is most likely that the samples have better initial alignment in percentual time. The estimated participant-specific acceleration profiles using percentual time can be seen in Fig 8.

A simulation study that validates the method and implementation on data simulated using the maximum likelihood estimates of the central experimental condition ($d = 30.0$ cm, medium obstacle) is available as Supporting Information.

## Results

### Identification of individual differences

A first assessment of the strength of the statistical model (1) is to examine the extent to which the model captures individual differences. Proper modeling of systematic individual differences is not only of scientific interest per se, but also provides perspective for interpreting any observed experimental effects. To validate the capability of the model to capture systematic individual differences, we use the model to identify an individual from the estimated individual templates. Such identification of individuals is becoming increasingly relevant also in a practical sense with the recent technological advances in motion tracking systems, and the growing array of digital sensors in handheld consumer electronics. Consistent with the framing of model (1), we perform identification of individual participants on the basis of the data from a single experimental condition. This is, in a sense, a conservative approach. Combining data across the different conditions of the experimental tasks would likely provide more discriminative power given that personal movement styles tend to be reproducible.

The classification of the movement data is based on the characteristic acceleration profile computed for each participant. For this to work it is important that individual movement differences are not smoothed away. The hyperparameters of the model were chosen with this
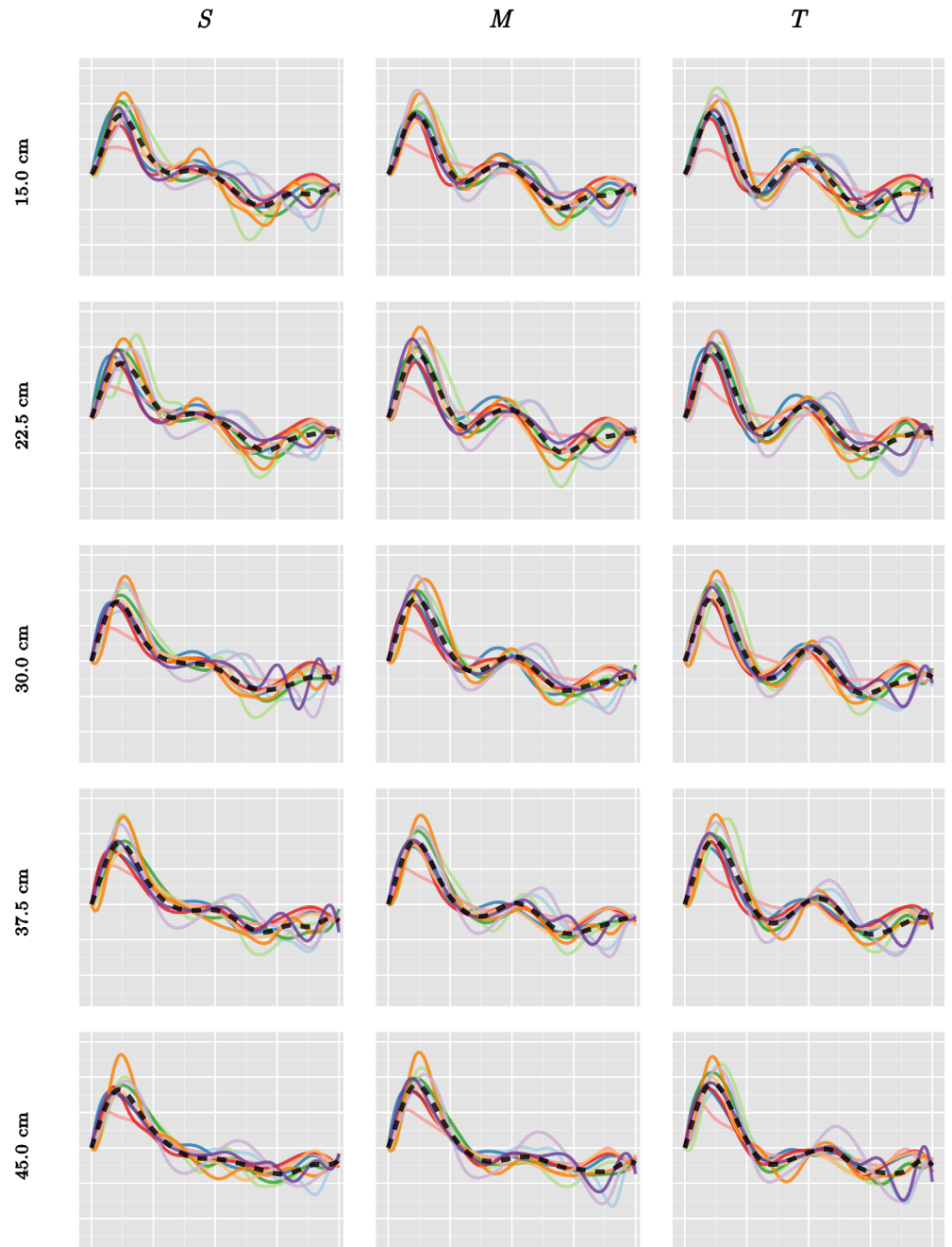
**Fig 8. Estimated fixed effects $(\theta + \varphi_i) \circ v_i$ in the 15 obstacle avoidance experiments using percentual time.** The dashed trajectory shows the estimate for $\theta$. The average percentual warped time between two white vertical bars corresponds to 0.2 seconds.

doi:10.1371/journal.pcbi.1005092.g008

requirement in mind. In the following, we describe alternative methods we considered. For all approaches, the stated parameters have been chosen by 5-fold cross-validation on the experimental conditions with obstacle distance $d = 30.0$ cm. The grids used for cross-validation are given as Supporting Information section. Recall that subscript $p$ indicates the use of percentuall time.

**Nearest Participant (NP)** NP classification classifies using the minimum combined pointwise $L^2$ distance to all samples for every individual in the training set.

**Modified Band Median (MBM)** MBM classification estimates templates using the modified band median proposed in [46], which under mild conditions is a consistent estimator of the underlying fixed amplitude effects warped according to the modified band medians of the warping functions. Classification is done using $L^2$ distance to the estimated templates. In the computations we count the number of bands defined by $J = 4$ curves [46, Section 2.2]. $MBM_p$ used $J_p = 2$.

**Robust Manifold Embedding (RME)** RME classification estimates templates using the robust manifold embedding algorithm proposed in [47], which, assuming that data lies on a low-dimensional smooth manifold, approximates the geodesic distance and computes the empirical Fréchet median function. Classification is done using $L^2$ distance to the estimated templates.

**Dynamic Time Warping (DTW)** DTW classification estimates templates by iteratively time warping samples to the current estimated personal template (5 iterations per template) using an asymmetric step pattern (slopes between 0 and 2). The template is modeled by a B-spline with 33 degrees of freedom. $DTW_p$ used 16 degrees of freedom.

**Fisher-Rao (FR)** FR classification estimates templates as Karcher means under the Fisher-Rao Riemannian metric [48] of the data represented using a single principal component [49]. Čencov's theorem states that the Fisher-Rao distance is the only distance that is preserved under warping [50], and in practice the distance is computed by using a dynamic time warping algorithm on the square-root slope functions of the data. Classification is done using $L^2$ distance to the estimated templates.

**Elastic Fisher-Rao ($FR_E$)** $FR_E$ classification estimates templates analogously to FR, but classifies using the weighted sum of elastic amplitude and phase distances [51, Definition 1 and Section 3.1]. The phase distance was weighted by a factor 1.5. $FR_{Ep}$ uses two principal components and a phase distance weight of 1.

**Timing and Motion Separation (TMS)** The proposed TMS classification estimates templates of the fixed effects $(\theta + \varphi_i) \circ v_i$ using Algorithm 1. Classification is done using least distance measured in the negative log posterior Eq (2) as a function of the test samples. The parameters were set as described in the previous section.

We evaluate classification accuracy using 5-fold cross-validation, which means that eight samples are available in the training set for every participant. The folds of the cross-validation are chosen chronologically, such that the first fold contains replications 1 and 2, the second contains 3 and 4 and so on. The results are available in Table 1. We see that TMS and $TMS_p$ achieve the highest classification rates, followed by $FR_{Ep}$, $FR_p$ and $RME_p$. Thus, the model enables identification of individual movement style. Furthermore, we note that there is little effect of using percentual time for the proposed method, which for all other methods gives a considerable boost in accuracy. This suggests that the TMS methods align data well without the initial linear warping and the endpoint constraints of percentual time.

## Factor analysis of spatial movement paths

In the previous section, the proposed modeling framework was shown to give unequalled accuracy of modeling the time series data for acceleration. In this section we use the warping functions obtained to analyze the spatial movement paths and their dependence on task conditions.

**Table 1. Classification accuracies of various methods.** **Bold** indicates best result(s), *italic* indicates that the given experiments were used for training.

| *d* | obstacle | NP | NP$_p$ | MBM | MBM$_p$ | RME | RME$_p$ | DTW | DTW$_p$ | FR | FR$_p$ | FR$_E$ | FR$_{Ep}$ | TMS | TMS$_p$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 15.0 cm | S | 0.36 | 0.48 | 0.53 | 0.43 | 0.55 | 0.57 | 0.52 | 0.52 | 0.47 | 0.54 | 0.62 | 0.51 | 0.70 | **0.76** |
| | M | 0.36 | 0.46 | 0.38 | 0.45 | 0.41 | 0.43 | 0.49 | 0.56 | 0.36 | 0.49 | 0.47 | 0.46 | **0.69** | 0.66 |
| | T | 0.41 | 0.47 | 0.41 | 0.46 | 0.49 | 0.50 | 0.43 | 0.43 | 0.32 | 0.56 | 0.49 | 0.49 | **0.64** | 0.62 |
| 22.5 cm | S | 0.36 | 0.49 | 0.34 | 0.46 | 0.37 | 0.50 | 0.45 | 0.44 | 0.44 | 0.51 | 0.50 | 0.51 | **0.70** | 0.68 |
| | M | 0.38 | 0.44 | 0.42 | 0.53 | 0.46 | 0.55 | 0.38 | 0.42 | 0.32 | 0.45 | 0.42 | 0.55 | 0.62 | **0.74** |
| | T | 0.36 | 0.49 | 0.45 | 0.54 | 0.46 | 0.57 | 0.40 | 0.53 | 0.48 | 0.57 | 0.53 | **0.64** | 0.61 | **0.64** |
| 30.0 cm | S | 0.27 | 0.29 | *0.37* | *0.47* | 0.41 | 0.45 | *0.43* | *0.44* | *0.40* | *0.43* | *0.46* | *0.55* | *0.63* | *0.65* |
| | M | 0.30 | 0.42 | *0.38* | *0.46* | 0.40 | 0.45 | *0.34* | *0.49* | *0.36* | *0.48* | *0.46* | *0.47* | *0.65* | *0.65* |
| | T | 0.37 | 0.44 | *0.42* | *0.50* | 0.50 | 0.45 | *0.44* | *0.44* | *0.37* | *0.50* | *0.39* | *0.43* | *0.74* | *0.69* |
| 37.5 cm | S | 0.28 | 0.45 | 0.41 | 0.49 | 0.42 | 0.51 | 0.45 | 0.50 | 0.36 | 0.51 | 0.39 | 0.56 | 0.69 | **0.74** |
| | M | 0.26 | 0.33 | 0.33 | 0.37 | 0.35 | 0.41 | 0.40 | 0.49 | 0.35 | 0.37 | 0.32 | 0.53 | 0.57 | **0.62** |
| | T | 0.31 | 0.43 | 0.38 | 0.43 | 0.40 | 0.46 | 0.37 | 0.29 | 0.50 | 0.48 | 0.49 | 0.55 | 0.63 | **0.65** |
| 45.0 cm | S | 0.25 | 0.38 | 0.33 | 0.45 | 0.32 | 0.42 | 0.34 | 0.51 | 0.32 | 0.45 | 0.37 | 0.41 | **0.68** | 0.65 |
| | M | 0.29 | 0.31 | 0.29 | 0.38 | 0.38 | 0.39 | 0.43 | 0.43 | 0.36 | 0.48 | 0.38 | 0.45 | 0.53 | **0.57** |
| | T | 0.29 | 0.39 | 0.45 | 0.48 | 0.48 | 0.57 | 0.38 | 0.45 | 0.39 | 0.44 | 0.44 | 0.47 | **0.66** | 0.58 |
| average | | 0.323 | 0.418 | 0.393 | 0.460 | 0.427 | 0.482 | 0.417 | 0.463 | 0.387 | 0.484 | 0.449 | 0.487 | 0.649 | **0.660** |

doi:10.1371/journal.pcbi.1005092.t001

Temporal alignment of the spatial positions along the path for different repetitions and participants is necessary to avoid spurious spatial variance of the paths. The natural alignment of two movement paths is the one that matches their acceleration signatures. In other words, spatial positions along the paths at which similar accelerations are experienced should correspond to the same times. Thus, each individual spatial trajectory was aligned using the time warping predicted from the TMS$_p$-results of the previous section. Every sample path was represented by 30 equidistant sample points in time at which values were obtained by fitting a three-dimensional B-splines with 10 equidistantly spaced knots to each trajectory.

As an exemplary study, we analyze how spatial paths depend on obstacle height. We do this separately for each distance, so that we perform five separate analyses, one for each obstacle placement. In each analysis we have 10 participants with 10 repetitions for each of 3 obstacle heights. The three-dimensional spatial positions along the movement path, $y_{ijh} \in \mathbb{R}^{30 \times 3}$, depend on participant $i = 1,\ldots,10$, repetition $j = 1,\ldots,10$ and height $h = 1,2,3$.

We are interested in understanding how the space-time structure of movement captured in the 30 by 3 dimensions of the trajectories, $y_{ijh}$, varies when obstacle height is varied. We seek to find a low-dimensional affine subspace of the space-time representation of the movements within which movements vary, once properly aligned. That subspace provides a low-dimensional model of movement paths on the basis of which we can analyze the data.

We identify the low-dimensional subspace based on a novel factor analysis model. In analogy to principal component analysis (PCA), $q$ so-called *loadings* are estimated that represent dominant patterns of variation along movement trajectories. In contrast to PCA, the factor analysis model does not only model the residual-variance of independent paths around the mean, but also allows one to include covariates from the experimental design, for example by taking the repetition structure of participants and systematic effects of obstacle height into account. In other words, the proposed factor analysis model is a generalization of PCA suitable for addressing the question at hand while obeying the study design.

The idea is to use the mean movement trajectory, $\boldsymbol{\theta} \in \mathbb{R}^{30 \times 3}$, of one condition, the lowest obstacle height, as a reference. The movement trajectories $y_{ijh}$ (30 time steps and 3 Cartesian

coordinates; participant $i$; repetition $j$; and experimental condition with height $h$) are then represented through their deviation from the reference path. We estimate the hypothesized low-dimensional affine subspace in which these deviations lie. That subspace is spanned by the $q$ orthonormal $(30 \times 3)$-dimensional columns of the loadings matrix $W \in \mathbb{R}^{(30 \times 3) \times q}$. We assume a mixed-effect structure on the weights for the loadings that takes into account both the categorical effect of obstacle-height and random effects of participant and repetition. This amounts to a statistical model

$$\boldsymbol{y}_{ijh} = \boldsymbol{\theta} + (X_h \beta + \sum_{l=1}^{3} Z_{i,g_l(j,h),l}) W^\top + \boldsymbol{\varepsilon}_{ij}, \tag{5}$$

where $X_h \in \mathbb{R}^{1 \times 2}$ represents the covariate design that indicates obstacle heights: $S$: $X_1 = (0,0)$; $M$: $X_2 = (1,0)$; and $T$: $X_3 = (0,1)$. The parameters, $\beta \in \mathbb{R}^{2 \times q}$, are the weights for the loadings that account for the systematic deviation of obstacle heights, $M$ and $T$ from the reference height, $S$. $g_l$ is the factor that describes the $l$th level random effects design (participant, participants' reaction to obstacle-height change, and repetition). $Z_{i,g_l(j,h),l} \in \mathbb{R}^{1 \times q}$ are independent latent Gaussian variables with zero-mean and a covariance structure modeled with three $q \times q$ covariance matrices, each describing the covariance between loadings within a level of the random-effect design. $\boldsymbol{\varepsilon}_{ij} \in \mathbb{R}^{30 \times 3}$ is zero-mean Gaussian noise with diagonal covariance matrix $\Lambda$ with one variance parameter per dimension.

The loading matrix $W$ is identifiable in a similar way as for usual PCA. Firstly, the scaling of $W$ is identified by the assumption that $W^\top W$ is the $q$-dimensional identity matrix. Secondly, the rotation of $W$ is identified by the assumption that the total variance of the latent variables for a single curve

$$\sum_{l=1}^{3} \text{Var}(Z_{i,g_l(j,h),l}),$$

is a diagonal matrix. This identifies the loading matrix $W$ with probability 1.

The models were fitted using maximum likelihood estimation by using an ECM algorithm [52] that had been accelerated using the SQUAREM method [53].

Fig 9 shows the percentage of explained variance for various values of $q$ across different obstacle distances. For $q > 8$, the average percentage of variance explained by the ninth loading ranges from 1% to 2% and the combined percentage of variance explained by the loadings beyond number eight remains under 3%. From this perspective, $q = 8$ seems like a reasonable choice, with the loadings explaining 97.1% of the variance, meaning that the error term $\boldsymbol{\varepsilon}_{ij}$ should account for the remaining 2.9%. In the following, all results are based on the model with $q = 8$.

**Modeling obstacle height.**   The fitted mean trajectories for the three different obstacle heights at distance $d = 30.0$ cm can be found in Fig 10. A striking feature of the mean paths is the apparent linear scaling of elevation, but also of the lateral excursion with height. (The difference in the frontal plane, not shown, was very small, but follows a similar pattern). This leads to the hypothesis that the scaling of the mean trajectory with obstacle height can be described by a one-parameter regression model in height increase ($X_1 = 0$, $X_2 = 7.5$, $X_3 = 15$) rather than a more generic two-parameter ANOVA model.

We fitted both models for every obstacle distance and performed likelihood-ratio tests. The p-values can be found in Table 2. They were obtained by evaluating twice the difference in log likelihood for the two models using a $\chi^2$-distribution at $q = 8$ degrees of freedom. We see that no p-values are significant, so there is no significant loss in the descriptive power of the linear
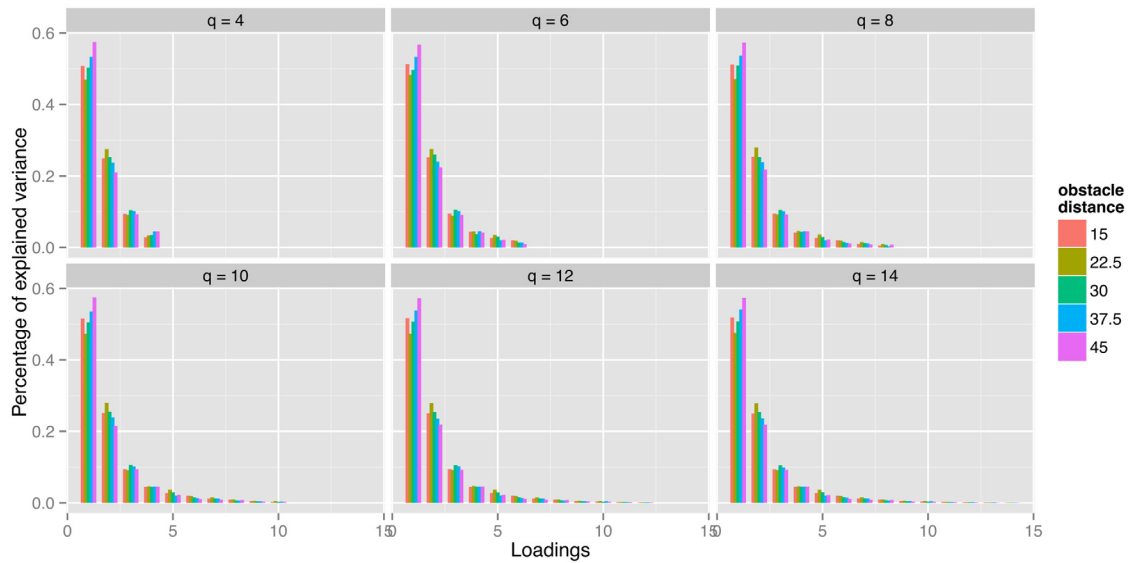
**Fig 9. Percentage of variance explained by individual loadings under different total number of loadings *q* and obstacle distance.**

doi:10.1371/journal.pcbi.1005092.g009
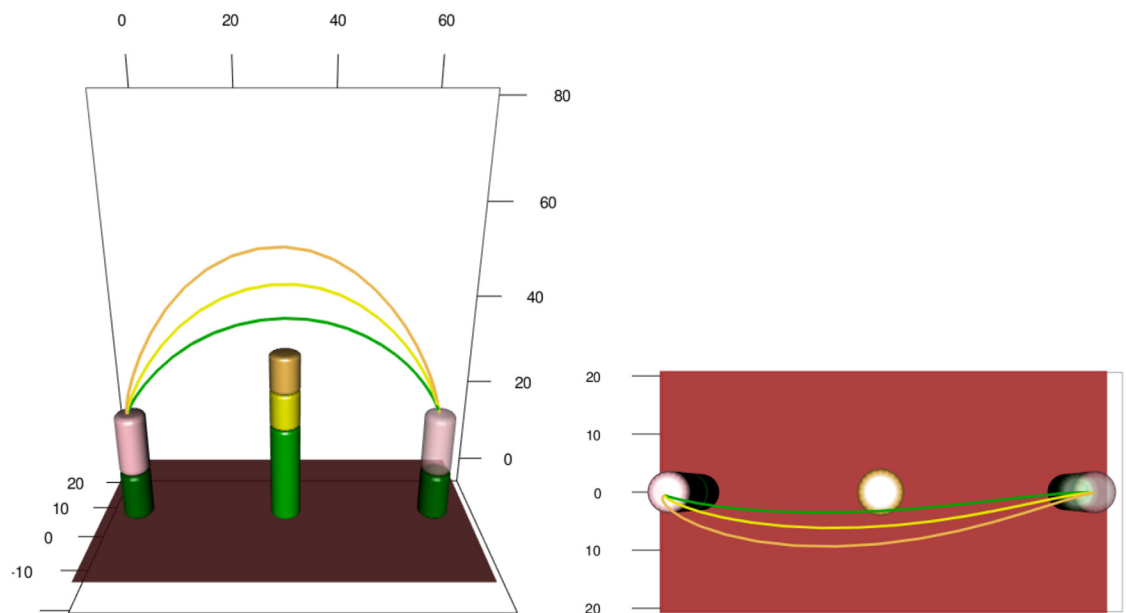


**Fig 10. Mean paths for the three obstacle heights (green: small; yellow: middle, orange: tall) at obstacle distance *d* = 30.0 cm.**

doi:10.1371/journal.pcbi.1005092.g010

**Table 2. p-values for the hypothesis of linearly amplified path changes in obstacle height increase factor.**

| Obstacle distance | 15.0 cm | 22.5 cm | 30.0 cm | 37.5 cm | 45.0 cm |
|---|---|---|---|---|---|
| p-value | 0.478 | 0.573 | 0.093 | 0.764 | 0.362 |

doi:10.1371/journal.pcbi.1005092.t002

scaling model compared to the ANOVA model. The remaining results in the paper are all based on the model with a regression design.

**Discovering the time-structure of variance along movement trajectory.** A strength of our approach to time warping is that we can estimate variability more reliably. In addition to observation noise, we model three sources of variation in the observed movement trajectories: individual differences in the trajectory, individual differences caused by changing obstacle height, and variation from repetition to repetition. The variances described from these three sources are independent of obstacle height. In Figs 11 and 12, two spatial representations of the mean movement path for the medium obstacle height are shown. The five distances of the obstacle from the starting position are shown in the five rows of the figures. The three columns show variance originating from individual differences in the trajectory, individual differences caused by changing obstacle height, and variation from repetition to repetition (from left to right). Variance is illustrated at eight equidistant points in time along the mean path by ellipsoids that mark 95% prediction for each level of variation.

Note the asymmetry of the movements with respect to obstacle position, both in terms of path and variation. This asymmetry reflects the direction of the movement. Generally, variability is higher in the middle of the movement than early and late in the movement. Individual differences caused by change in obstacle height (middle column) are small and lie primarily along the path. That is, individuals adapt the timing of the movement differently as height is varied. Individual differences in the movement path itself (left column) are largely differences in movement parameters: individuals differ in the maximal elevation and in the lateral positioning of their paths, not as much in the time structure of the movements. Variance from trial to trial (right column) is more evenly distributed, but is largest along the path reflecting variation in timing.

These descriptions are corroborated by the comparisons of the amounts of variance explained by the three effects in Fig 13. The obstacle distance of 45, in which the obstacle is close to the target lead to the largest variance in movement trajectory, with most of the increase over other conditions coming from repetition and individual differences caused by change in obstacle height. This suggests that this experimental condition is more difficult than the others, and perhaps much more so for the tall obstacle. Apart from this condition, we see that the largest source of variation is individual differences in movement trajectory. The second largest source of variation is repetition. Individual differences caused by change in obstacle height were are systematically the smallest source of variance.

**Trajectory focal points.** A final demonstration of the strength of the method of analysis is illustrated in Fig 14, which shows the mean paths for all obstacle distances and all obstacle heights. For each obstacle height, the paths from different obstacle distances intersect both in the frontal and the vertical plane. These focal points occur approximately at the same distance along the imagined line connecting start and end position. This pattern is clearly visible in the front view of the mean trajectories in Fig 14. Due to the limited variation of the path in the horizontal plane (Fig 14 top view), this effect is less clear in the horizontal plane. This pattern may reflect a scaling law, a form of invariance of an underlying path generation mechanism.

## Discussion

We have proposed a statistical framework for the modeling of human movement data. The hierarchical nonlinear mixed-effects model systematically decomposes movements into a common effect that reflects the variation of movement variables with time during the movement, individual effects, that reflect individual differences, variation from trial to trial, as well as measurement noise. The model amounts to a nonlinear time-warping approach that treats all sources of variances simultaneously.
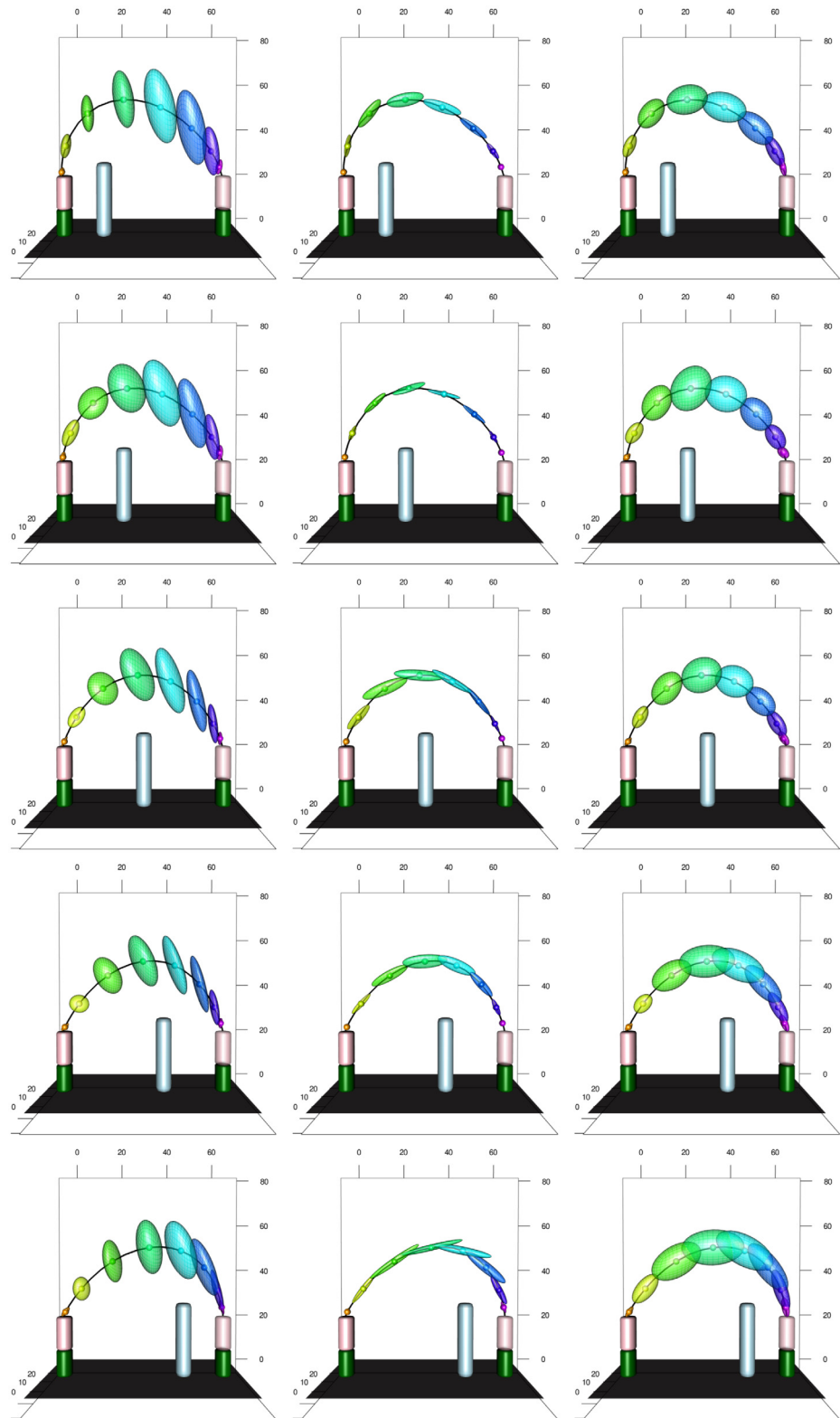
**Fig 11. Illustration of the experimental setup for the medium height obstacle at all obstacle distances (rows) with the mean trajectory plotted.** Along the trajectory eight equidistant points (in percentual warped time) are marked, and at each point 95% prediction ellipsoids are drawn. The three columns represent the random effects tied to participant, subjective reaction to height change and repetition, respectively.
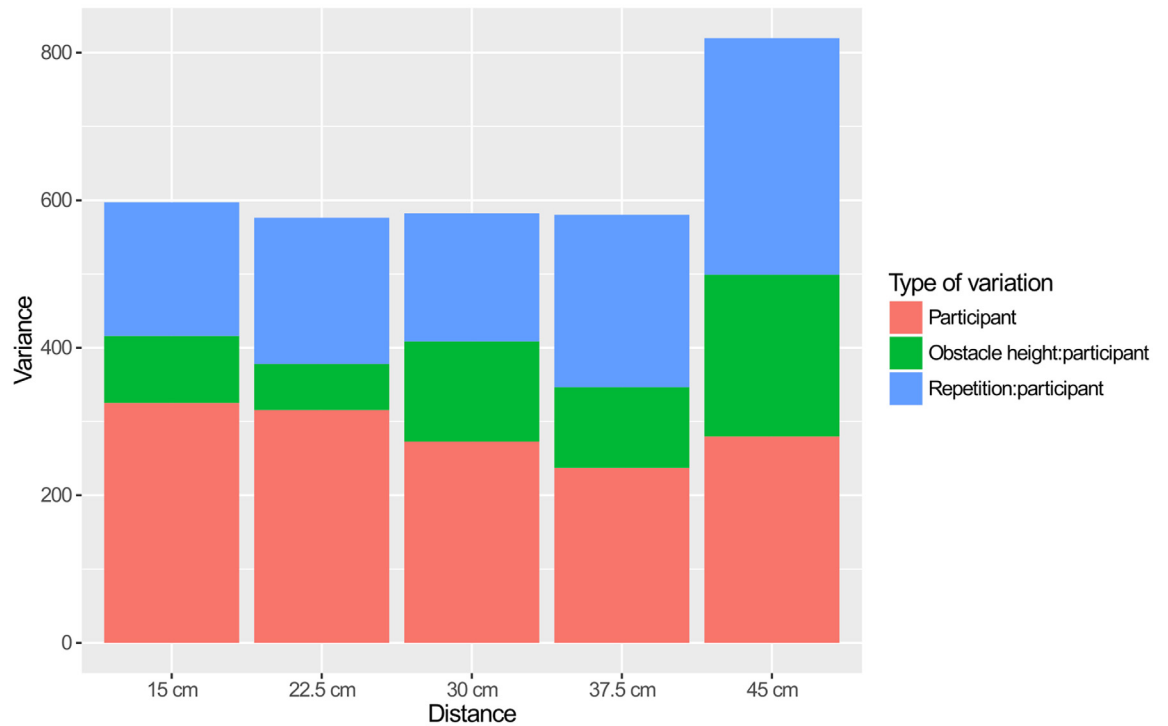
doi:10.1371/journal.pcbi.1005092.g011
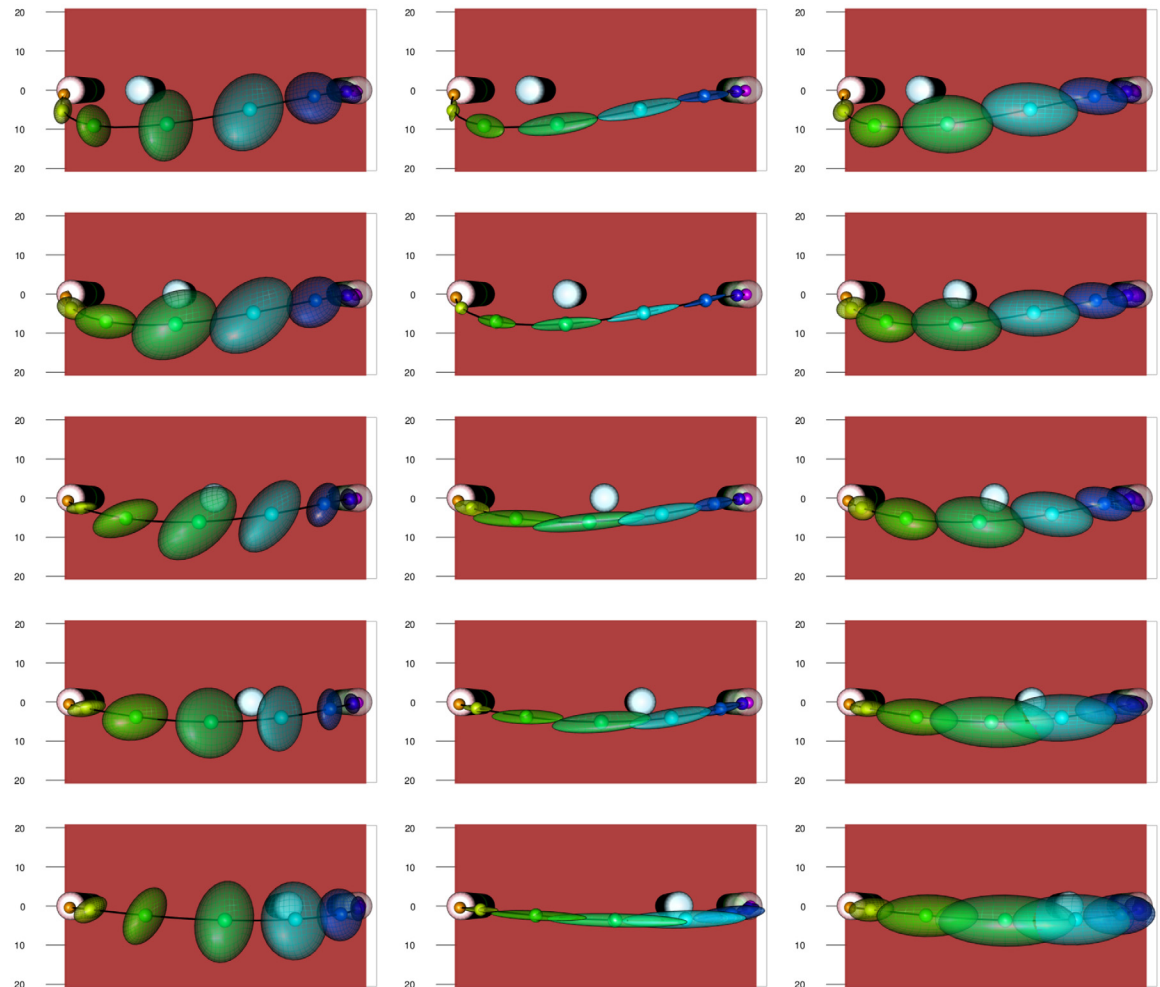
**Fig 12. Top-view of the experimental setup from the top for the medium height obstacle at all obstacle distances with the mean trajectory plotted.** Along the trajectory eight equidistant points (in percentual warped time) are marked, and at each point 95% prediction ellipsoids are drawn. The ordering is the same as in Fig 11.

We have outlined a method for performing maximum likelihood estimation of the model parameters, and demonstrated the approach by analyzing a set of human movement data on the basis of acceleration profiles in arm movements with obstacle avoidance. The quality of the estimates was evaluated in a classification task, in which our model was better able to determine if a sample movement came from a particular participant compared to state-of-the-art template-based curve classification methods. These results indicate that the templates that emerge from our nonlinear warping procedure are both more consistent and richer in detail.

We used the nonlinear time warping obtained from the acceleration profiles to analyze the spatial movement trajectories and their dependence on task conditions, here the dependence on obstacle height and obstacle placement along the path. We discovered that the warped movement path scales linearly with increasing obstacle height. Furthermore, we separated the variation around the mean paths into three levels: individual differences of movement trajectory, individual differences caused by change in obstacle height, and trial to trial variability. This combination of models uncovered clear and coherent patterns in the structure of variance. Individual differences in trajectory and variance from trial to trial were the largest sources of

**Fig 13. Amount of variance explained by random effects.**

variance, with individual differences being primarily at the level of movement parameters such as elevation and lateral extent of the movement while variance from trial to trial contained a larger amount of timing variance. We documented a remarkable property of the movement paths when obstacle distance along the path is varied at fixed obstacle height: all paths intersect at a single point in space.
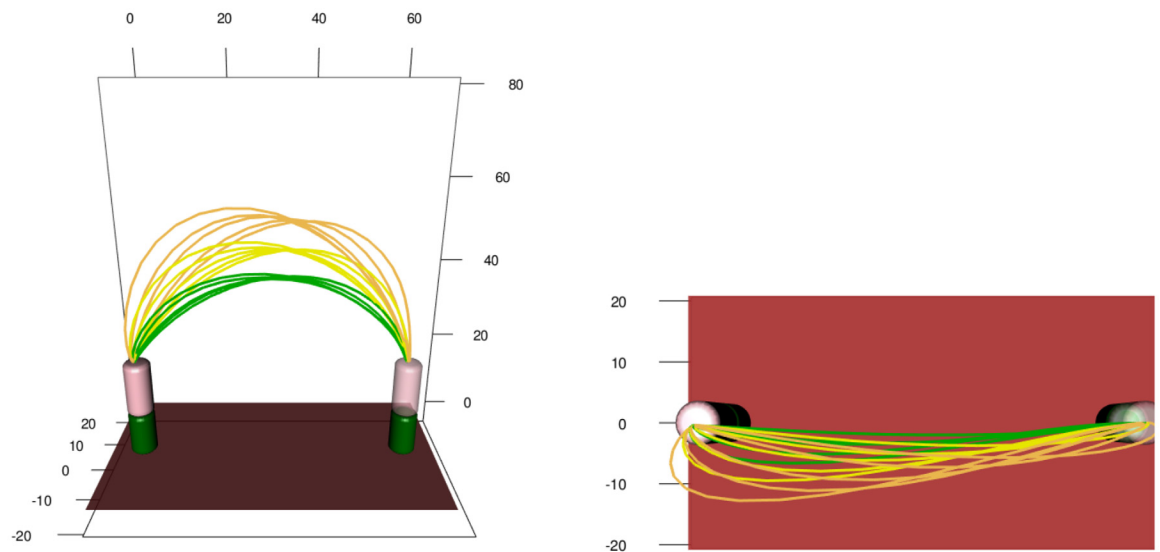


**Fig 14. Front and top view of the mean trajectories for the 15 obstacle avoidance setups.** Green lines correspond to low obstacles, yellow to medium obstacles and orange to tall obstacles.

We believe that the approach we describe enhances the power of time series analysis as demonstrated in human movement data. The nonlinear time warping procedure makes it possible to obtain reliable estimates of variance along the movement trajectories and is strong in extracting individual differences. This advantage can be leveraged by combining the non-linear time warping with factor analysis to extract systematic dependencies of movements on task conditions at the same time as tracking individual differences, both base-line and with respect to the dependence on task conditions, as well as variance across repetitions of the movement.

Recent theoretical accounts have used the analysis of variance across repetitions of movements to uncover coordination among the many degrees of freedom of human movement systems [17]. Differences in variance between the subspace that keeps hypothesized relevant task variables invariant and the subspace within which such task variables vary support hypotheses about the task-dependent structure of the underlying control systems. Because variance is modulated in time differently across the two subspaces, a more principled decomposition of time dependence and variance from trial to trial would give such analyses new strength. Because this application requires the extension of the proposed method to multivariate time series, it is beyond the scope of this paper. Together with the considerable practical interest in identifying individual differences, these theoretical developments underscore that the method proposed here is timely and worth the methodological investment.

## Supporting Information

**S1 Text. Supporting information contains a primer on model building using the proposed framework, an example analysis with accompanying R code, a simulation study that evaluates the proposed algorithm for maximum likelihood estimation, and the different parameter grids used for cross-validation in the paper.**
(TEX)

**S1 Fig. Twelve acceleration functions corresponding to repetitions of a signature being written.**
(TIFF)

**S2 Fig. The aligned acceleration functions from S1 Fig (a), along with the predicted warping functions (b).**
(TIFF)

**S3 Fig. Densities of the integrated square estimation errors ($L^2$ errors) for the common and participant-specific mean functions in the simulation study.** The left panel shows results for ordinary least square (OLS) estimation and the right panel shows the results for the proposed model and estimation algorithm. Both models were fitted using the correctly specified spline model for the mean. Note that the density is displayed on squareroot scale.
(TIFF)

**S4 Fig. Densities of the difference between the estimated and the true participant-specific warping parameters across participants.**
(TIFF)

**S5 Fig. Densities of the estimated variance parameters in the simulated experimental setups.** Dashed red lines indicate the true values of the parameters.
(TIFF)

## Author Contributions

**Conceived and designed the experiments:** BG GS.

**Performed the experiments:** BG.

**Analyzed the data:** LLR BG CI BM.

**Wrote the paper:** LLR BG GS CI BM.

## References

1. Morasso P. Three dimensional arm trajectories. Biological Cybernetics. 1983; 48(3):187–194. doi: 10.1007/BF00318086 PMID: 6639982

2. Soechting JF, Terzuolo CA. Organization of arm movements in three-dimensional space. Wrist motion is piecewise planar. Neuroscience. 1987; 23(1):53–61. doi: 10.1016/0306-4522(87)90270-3 PMID: 3683869

3. Morasso P. Spatial control of arm movements. Experimental Brain Research. 1981; 42(2):223–227. doi: 10.1007/BF00236911 PMID: 7262217

4. Lacquaniti F, Terzuolo C, Viviani P. The law relating the kinematic and figural aspects of drawing movements. Acta psychologica. 1983; 54(1):115–130. doi: 10.1016/0001-6918(83)90027-6 PMID: 6666647

5. Viviani P, McCollum G. The relation between linear extent and velocity in drawing movements. Neuroscience. 1983; 10(1):211–218. doi: 10.1016/0306-4522(83)90094-5 PMID: 6646421

6. Bennequin D, Fuchs R, Berthoz A, Flash T. Movement timing and invariance arise from several geometries. PLoS Computational Biology. 2009; 5(7):e1000426. doi: 10.1371/journal.pcbi.1000426 PMID: 19593380

7. Erlhagen W, Schöner G. Dynamic field theory of movement preparation. Psychological Review. 2002; 109(3):545–572. doi: 10.1037/0033-295X.109.3.545 PMID: 12088245

8. Georgopoulos AP. Current issues in directional motor control. Trends in Neurosciences. 1995; 18:506–510. doi: 10.1016/0166-2236(95)92775-L PMID: 8592761

9. Schwartz AB. Useful signals from motor cortex. The Journal of Physiology. 2007; 579(3):581–601. doi: 10.1113/jphysiol.2006.126698 PMID: 17255162

10. Harpaz NK, Flash T, Dinstein I. Scale-invariant movement encoding in the human motor system. Neuron. 2014; 81(2):452–462. doi: 10.1016/j.neuron.2013.10.058

11. Zhang W, Rosenbaum DA. Planning for manual positioning: the end-state comfort effect for manual abduction–adduction. Experimental Brain Research. 2008; 184(3):383–389. doi: 10.1007/s00221-007-1106-x PMID: 17763845

12. Cutting JE, Kozlowski LT. Recognizing friends by their walk: Gait perception without familiarity cues. Society, Bulleting of the Psychonomic. 1977; 9(5):353–356. doi: 10.3758/BF03337021

13. Pollick FE, Kay JW, Heim K, Stringer R. Gender recognition from point-light walkers. Journal of Experimental Psychology: Human Perception and Performance. 2005; 31(6):1247. PMID: 16366787

14. Lu J, Wang G, Moulin P. Human identity and gender recognition from gait sequences with arbitrary walking directions. Information Forensics and Security, IEEE Transactions on. 2014; 9(1):51–61. doi: 10.1109/TIFS.2013.2291969

15. Schöner G, Kelso JAS. Dynamic pattern generation in behavioral and neural systems. Science. 1988; 239:1513–1520. doi: 10.1126/science.3281253 PMID: 3281253

16. Scholz JP, Schöner G. The Uncontrolled Manifold Concept: Identifying Control Variables for a Functional Task. Experimental Brain Research. 1999; 126:289–306. doi: 10.1007/s002210050738 PMID: 10382616

17. Latash ML, Scholz JP, Schöner G. Toward a new theory of motor synergies. Motor Control. 2007; 11:276–308. doi: 10.1123/mcj.11.3.276 PMID: 17715460

18. Ramsay JO, Silverman BW. Functional Data Analysis. 2nd ed. Springer; 2005.

19. Pinheiro JC, Bates DM. Mixed effects models in S and S-PLUS. Springer; 2000.

20. Lindstrom MJ, Bates DM. Nonlinear Mixed Effects Models for Repeated Measures Data. Biometrics. 1990; 46(3):673–687. doi: 10.2307/2532087 PMID: 2242409

21. Grimme B, Lipinski J, Schöner G. Naturalistic arm movements during obstacle avoidance in 3D and the identification of movement primitives. Experimental Brain Research. 2012; 222(3):185–200. doi: 10.1007/s00221-012-3205-6 PMID: 22996050

22. Raket LL. pavpop version 0.10; 2016. Available from: https://github.com/larslau/pavpop/.

23. Grimme B. Analysis and identification of elementary invariants as building blocks of human arm movements. International Graduate School of Biosciences, Ruhr-Universität Bochum; 2014. (In German).

24. Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition. Acoustics, Speech and Signal Processing, IEEE Transactions on. 1978; 26(1):43–49. doi: 10.1109/TASSP.1978.1163055

25. Bruderlin A, Williams L. Motion signal processing. In: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques. ACM; 1995. p. 97–104.

26. Troje NF. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. Journal of Vision. 2002; 2(5):371–387. doi: 10.1167/2.5.2 PMID: 12678652

27. Berndt DJ, Clifford J. Using Dynamic Time Warping to Find Patterns in Time Series. In: KDD workshop. vol. 10. Seattle, WA; 1994. p. 359–370.

28. Gavrila D, Davis L, et al. Towards 3-D model-based tracking and recognition of human movement: a multi-view approach. In: In International Workshop on Automatic Face- and Gesture-Recognition. IEEE Computer Society; 1995. p. 272–277.

29. Giese MA, Poggio T. Morphable models for the analysis and synthesis of complex motion patterns. International Journal of Computer Vision. 2000; 38(1):59–73. doi: 10.1023/A:1008118801668

30. Niennattrakul V, Ratanamahatana CA. Inaccuracies of shape averaging method using dynamic time warping for time series data. In: Computational Science–ICCS 2007. Springer; 2007. p. 513–520.

31. Petitjean F, Ketterlin A, Gançarski P. A global averaging method for dynamic time warping, with applications to clustering. Pattern Recognition. 2011; 44(3):678–693. doi: 10.1016/j.patcog.2010.09.013

32. Rønn BB. Nonparametric maximum likelihood estimation for shifted curves. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2001; 63(2):243–259. doi: 10.1111/1467-9868.00283

33. Gervini D, Gasser T. Nonparametric maximum likelihood estimation of the structural mean of a sample of curves. Biometrika. 2005; 92(4):801–820. doi: 10.1093/biomet/92.4.801

34. Rønn BB, Skovgaard IM. Nonparametric maximum likelihood estimation of randomly time-transformed curves. Brazilian Journal of Probability and Statistics. 2009; 23(1):1–17. doi: 10.1214/08-BJPS004

35. Beath KJ. Infant growth modelling using a shape invariant model with random effects. Statistics in medicine. 2007; 26(12):2547–2564. doi: 10.1002/sim.2718 PMID: 17061310

36. Cole TJ, Donaldson MD, Ben-Shlomo Y. SITAR—a useful instrument for growth curve analysis. International Journal of Epidemiology. 2010; 39:1558–1566. doi: 10.1093/ije/dyq115 PMID: 20647267

37. Raket LL, Sommer S, Markussen B. A nonlinear mixed-effects model for simultaneous smoothing and registration of functional data. Pattern Recognition Letters. 2014; 38:1–7. doi: 10.1016/j.patrec.2013.10.018

38. Ilg W, Bakir G, Franz M, Giese M. Hierarchical spatio-temporal morphable models for representation of complex movements for imitation learning. In: Nunes U., de Almeida A., Bejczy A., Kosuge K., and Machado J., In Proceeding of The 11th International Conference on Advanced Robotics. vol. 2; 2003. p. 453–458.

39. Marron JS, Ramsay JO, Sangalli LM, Srivastava A. Functional data analysis of amplitude and phase variation. Statistical Science. 2015; 30(4):468–484. doi: 10.1214/15-STS524

40. Wolfinger R. Laplace's approximation for nonlinear mixed models. Biometrika. 1993; 80(4):791–795. doi: 10.1093/biomet/80.4.791

41. Robinson GK. That BLUP is a Good Thing: The Estimation of Random Effects. Statistical Science. 1991; 6(1):15–32. doi: 10.1214/ss/1177011933

42. de Boor C. A Practical Guide to Splines. Revised ed. Springer-Verlag, Berlin; 2001.

43. Billingsley P. Convergence of probability measures. vol. 493 of Wiley Series in Probability and Statistics. John Wiley & Sons; 1999.

44. Flash T, Hogan N. The coordination of arm movements: an experimentally confirmed mathematical model. The Journal of Neuroscience. 1985; 5(7):1688–1703. PMID: 4020415

45. Stein ML. Interpolation of Spatial Data: Some Theory for Kriging. Springer; 1999.

46. Arribas-Gil A, Romo J. Robust depth-based estimation in the time warping model. Biostatistics. 2012; 13(3):398–414. doi: 10.1093/biostatistics/kxr037 PMID: 22058080

47. Dimeglio C, Gallón S, Loubes JM, Maza E. A robust algorithm for template curve estimation based on manifold embedding. Computational Statistics & Data Analysis. 2014; 70:373–386. doi: 10.1016/j.csda.2013.09.030

48. Kurtek SA, Srivastava A, Wu W. Signal estimation under random time-warpings and nonlinear signal alignment. In: Advances in Neural Information Processing Systems; 2011. p. 675–683.

49. Tucker JD. fdasrvf: Elastic Functional Data Analysis; 2014. R package version 1.4.2. Available from: http://CRAN.R-project.org/package=fdasrvf.

50. Čencov NN. Statistical decision rules and optimal inference. American Mathematical Society; 2000.

51. Tucker JD, Wu W, Srivastava A. Generative models for functional data using phase and amplitude separation. Computational Statistics & Data Analysis. 2013; 61:50–66. doi: 10.1016/j.csda.2012.12.001

52. Meng XL, Rubin DB. Maximum likelihood estimation via the ECM algorithm: A general framework. Biometrika. 1993; 80(2):267–278. doi: 10.1093/biomet/80.2.267

53. Varadhan R, Roland C. Simple and globally convergent methods for accelerating the convergence of any EM algorithm. Scandinavian Journal of Statistics. 2008; 35(2):335–353. doi: 10.1111/j.1467-9469.2007.00585.x