Parameterizing Running Kinematics

Project background

By Brian V. Kloster & Kristoffer Iversen

The present study was undertaken in industrial collaboration with Setio (Setio, Aalborg, Denmark). Setio is a small start-up company what offers running style analysis for both inexperienced and experienced runners. Setio's value proposition is to bring the running laboratory to the running community with a specialist in individual running techniques.
At present the energies and forces in the Setio system can only be calculated roughly based on simple mechanical models, Newton's law and accelerations. Setio came to Aalborg University with a desire to establish a new and better understanding of e.g. kinematics, energies and forces connected to different running techniques. At the moment Setio bases their running analysis on a learning application with three major components: a 30fps video analysis with fourteen evaluation points in the sagittal plane, a physiotherapist communication module for individual running optimization, and accelerometers placed on the lower back and the right foot. With these three components, they are able to track their customers cadence, acceleration, deceleration, stance phase and vertical displacement, and are able to measure and asses their customers style of running (SetioIVS 2017). The system has been developed and tested since 2013 and is already launched commercially.

By applying the AnyBody modeling system and developing a parametric running model it potentially might be possible for Setio to extend their opportunities to adjust and optimize their customers individual running techniques. A computer simulation of their customers running technique could provide insight into changes in kinematics. However, at present Setio's experimental input is inadequate to drive a musculoskeletal model. Due to this, a different kind of model is needed in order to drive the models with lesser experimental input or furthermore, some inputs which are easily measured.

Parameterizing Running Kinematics

Brian V. Kloster & Kristoffer Iversen
Department of Health Science and Technology, Aalborg University

Abstract

The aim of the present study was to develop a parametric running technique model with a minimum set of parameters. Full-body kinematic trajectories of 90 running trials was used to drive musculoskeletal models in the AnyBody Modeling System for simulation of a variety of running motions. A mixed terms Fourier series (FS) representation of the anatomical joint angles along with the translational and rotational components of the pelvis was performed. The coefficients of these FS served as the primary parameters of the running simulation model. All FS coefficients in company with the subject anthropometry were exposed to a Principal Components Analysis (PCA) which revealed that 90% of the variance could be explained by the first 12 principal components (PCs). Transforming only these 12 PCs back to its original domain allowed for recreation of the running trials. New running techniques were generated by multiplying the standard deviation of different PCs with different factors, adding them to the mean values of the 12 first principal components, and importing them back into AMS for a pseudo generated population of different running techniques.

Keywords: Running Technique, Running Kinematics, Musculoskeletal Modeling, AnyBody Modeling System, Principal Components Analysis, Fourier Series, Parameterization

1 Introduction

Computer simulation of human movement provides non-invasive means to gain insight into a wide range of parameters as coordinated motion of various body parts and estimations of muscle and joint forces which are otherwise impossible or impractical to measure *in vivo* (Pandy 2001, Zajac & Winters 1990). However, some kind of experimental inputs, e.g. kinematic data captured with a motion-capture system (MoCap) or ground reaction forces obtained with a force plate, are necessary to accurately drive the musculoskeletal models for motion simulation (Pandy 2001). A lot can be learned from laboratory experiments, but it can be time consuming, possibly invasive, and there is a growing consensus that computer simulation of human movement can provide more quantitative explanations of how the neuromuscular and musculoskeletal systems interact to produce motion as explained by Pandy (2001). Parameterizing motions compatible with subject anthropometry, environment constraints and the overall purpose of the motion would decrease the need for experimental inputs. A parameterization in this manner could provide the means to simulate an enormous amount of motion variations, which are an effective way to explore the space of possible movement combinations to produce a specific motion (van de Panne 1996). A parameterization of a motion in a quantitative and objective manner could in general be of potential use in

rehabilitation programs as well as achieving general insight of the motion in question.

In the present study the purpose was to parameterize running. Running has become an extremely popular physical activity over the past decades (Fredericson & Misra 2007, Paluska 2005, van Mechelen 1992), and despite a substantial effort to reduce the increasing injury rate (RRIR), no significant decline has been visible during the last three decades (Daoud et al. 2012). Today the primary reason for discontinuation of running - as a physical activity - is sustaining an injury due to poor running technique (Bredeweg 2014). Different studies have assessed the factors contributing to the RRIR (Lieberman et al. 2015, Bredeweg 2014, Daoud et al. 2012, Derrick 2004). The causes seem multifactorial, but kinematics is suggested to influence the overall RRIR (Daoud et al. 2012). The following factors are suspected to be causes of severe running-related injuries: incorrect foot strike patterns (Daoud et al. 2012), over-extended knee contact angles (Derrick 2004) and over-striding (Lieberman et al. 2015). Potentially, the application of a parameterized running model could be of help to understand e.g. tibial stress when running. Tibial stress fractures are among the most common running related injuries (Milner et al. 2006, Taunton et al. 2002, Crossley et al. 1999). A parameterized model could provide the means for investigating tibial stress by easily varying the running technique without the need for new experimental inputs. The role of upper-body motion contributions to lower extremity joint kinematics and external force generation when running are not well known, and may cause implications in the design of a musculoskeletal model for computer simulation of running (Miller et al. 2009). Earlier studies indicate that arm swings are important in maintaining balance (Mann 1980), increases the vertical displacement of the center-of-mass (CoM) (Hinrichs et al. 1987) and showed that suppressed arm swing causes changes in peak vertical and peak lateral ground reaction forces (Miller et al. 2009). This implicates that a full-body representation of the kinematics are necessary for a model to be as representative of normal human running as possible.

Principal Components Analysis (PCA) is a statistical tool for potential data reduction and pattern recognition (Liwicki et al. 2013, Deluzio & Astephen 2007). PCA has earlier been used to recognize patterns in both healthy and pathological gait cycles (Muniz & Nadal 2009, Deluzio & Astephen 2007, Deluzio et al. 1997) and Chau (2001) has emphasized the usefulness of PCA in a review of analytical techniques for gait data (Chau 2001). Earlier studies using PCA to investigate running techniques have revealed distinct running patterns between competitive and recreational runners in terms of pelvic tilt, knee flexion and subtalar eversion (Clermont et al. 2017). A similar approach revealed gender differences in gait patterns for patients with knee osteoarthritis (Phinyomark et al. 2016). The approach in these previous studies has been to expose raw data points to PCA.

The objective of this study was to develop a parameterized running model with a minimum set of parameters, obtained from kinematic data, which could be used to simulate various running techniques.

2 Methods

2.1 Data Acquisition

A total of 90 running trials from both treadmill and overground running were gathered from previous studies (Skals et al. 2014, Kersting & Ferdinands 2005). The running trials contained

different people running at a variety of velocities and directions. Full-body kinematic trajectories was captured with MoCap systems (Qualisys AB, Gothenburg, Sweden, C-Motion Inc, Germantown, USA & Vicon Inc, New York, USA) at a sample rate ranging from 120 - 250 frames per second.

2.2 Musculoskeletal models in the AnyBody Modeling System

The computational models were prepared in the AnyBody Modeling System v. 6.0.4 (AMS) (AnyBody Technology A/S, Aalborg, Denmark), a musculoskeletal simulation software package. The mathematical background of AMS is explained by Damsgaard et al. (2006). The preparation of each model was based on importing a .c3d-file containing the spatial trajectories of the MoCap markers into AMS. The precooked *MoCap Model Runner* which comes with the AnyBody Managed Model Repository v. 1.6.3, was used to set up the models. Each model was individually prepared in terms of adjusting the placement of the predetermined marker set in AMS to the markers from the MoCap data. The placing of markers at mid humerus, mid femur and mid tibia were characterized as uncertain locations, and they were included in the parameter optimization problem along with the anthropometrical parameters: thigh length, shank length, foot length, pelvis width, trunk height, upper arm and lower arm length. Solving the kinematics in AMS was based on the local optimization method for parameter identification of determinate and over-determinate mechanical systems. The best fit positions between the predetermined marker set in AMS and the recorded kinematic trajectories was resolved, as explained by Andersen et al. 2009, 2010. A forward kinematics driver strategy was used, as the translational and rotational position of pelvis and all anatomical angles from pelvis and outwards were driven. In some cases, the marker protocol was insufficient to provide the necessary kinematic information concerning the neck and wrist joints to drive the model. Due to this, a *AnyKinEqSimple*-driver which provided motion with constant position for flexion and abduction was added to ensure that the model behaved predictably.

Performing data processing of running trials through AMS resulted in a time series of anatomical joint angles representing the motion. All running trials were re-oriented to face the same way. As the majority of the running trials only contained a sufficient amount of time steps adequate to account for one half of a stride cycle, some processing of the data was accomplished to account for the missing half of the stride cycle. The time vector was extended in order to represent a full stride cycle. The horizontal component of the pelvis position was assumed constant throughout the motion, converting all the running trials to treadmill running. The vertical and the lateral components of the pelvis position, as well as the three rotational components of the pelvis, was extended to include the last half of the stride cycle. Joint angle trajectories from one side of the body was duplicated and applied to account for the opposite side of the body in the missing half of the stride cycle.

2.3 Model Parameterization

A Fourier series representation (FS) with mixed sine and cosine terms was chosen to reflect the cyclic nature of the data and the assumed frequency given by the step frequency. The FS coefficients served as the primary parameters of the running simulation model. It was hypothesized that choosing parameter functions

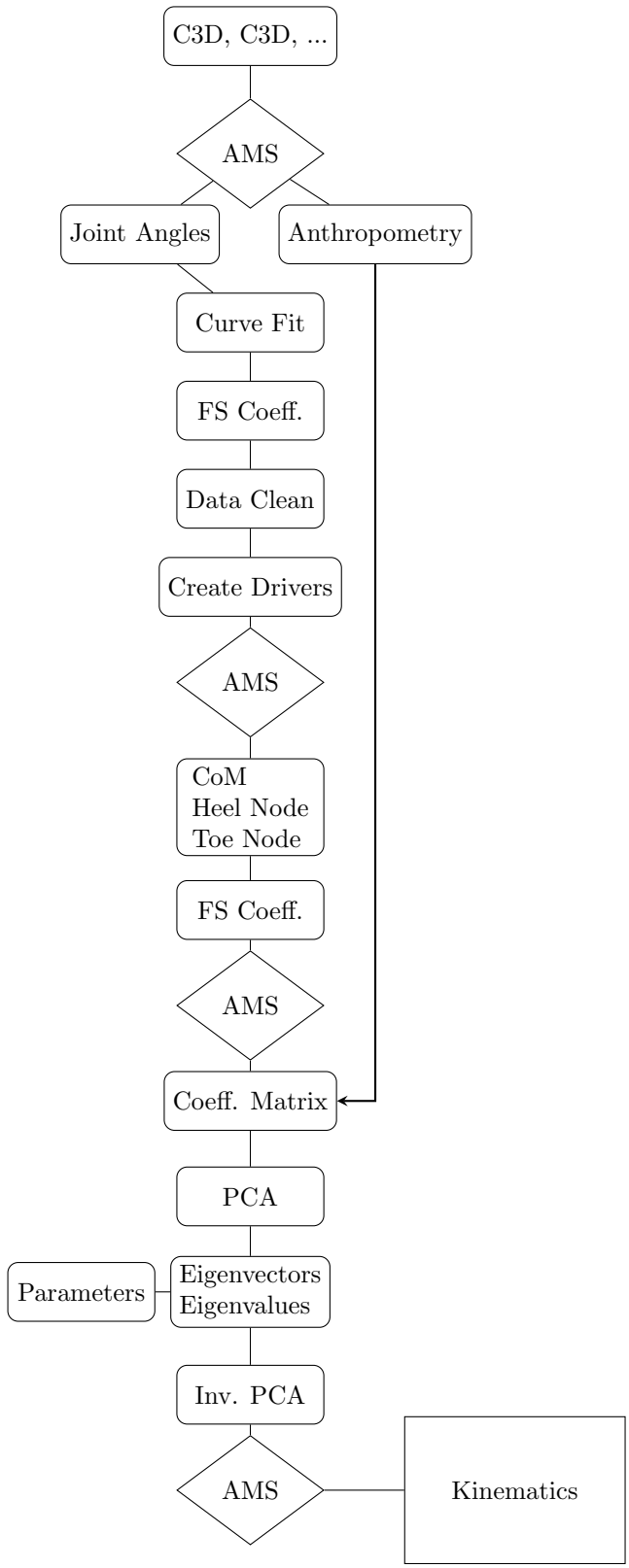


Figure 1: **Methodology flow chart** - Overview of the process of developing the parameterized running model. Beginning with importing the motion capture data as .c3d files into the AnyBody Modeling System (AMS). From AMS the anatomical joint angles and the subjects anthropometry were extracted.

Figure 1 *(previous column)*: A Fourier series representation (FS) of the joint angles was performed and the FS coefficients was processed to avoid unintended differences (Data Clean). The FS coefficients was imported into AMS to serve as the primary parameters to drive the models. Center-of-mass (CoM), heel nodes and toe nodes was extracted, represented as FS coefficients and imported back into AMS. All FS coefficients for one subject was arranged in a row vector along with the subject anthropometry and combined for a matrix exposed to principal components analysis (PCA). Inverse PCA was performed on the principal components representing the majority of the variance and imported back into AMS as drivers. The result is a parametric running model which could simulate running kinematics.

which are compatible with the cyclic nature of the fitted data, would result in a minimum number of parameters to obtain the desired fit. An add-on application, Curve fitting, in MATLAB (MATLAB v. R2016b and Statistics Toolbox 8.1, The MathWorks, Inc., Natick, Massachusetts, United States) was used to obtain the FS coefficients. The number of FS terms was selected based on the value of the correlation coefficient. The convergence criterion was set to ensure that the increase of the correlation coefficient, with each new term, was less than 5%. The FS drivers were implemented in AMS as an *AnyKinEqFourierDriver* class, which adds a FS as a driver function, and takes the following form:

$$Pos_i = \sum[(A_i Cos(\omega_i t)) + (B_i Sin(\omega_i t))] \quad (1.0)$$

With A_i and B_i denoting the FS cosine and sine coefficients, respectively and ω_i denoting the assumed frequency specified in hertz. To ensure the periodicity of the fitted curve and ensure that ω was not included as a variable in the curve fitting, the time period (T) for a full stride cycle was located and ω calculated as:

$$\omega = \frac{2\pi}{T} \quad (2.0)$$

2.4 Data Cleaning

A problem regarding the FS representation was encountered along the data processing pipeline. As almost all the running trials had different starting postures, a FS representation of two identical running techniques would return similar amplitudes but dissimilar phase shifts. A PCA would detect this as data differences, even though the running styles may be similar. An important insight into motions with a cyclic nature in general, as two identical motions would appear different.

Due to the nature of the mixed terms FS, the phase shift did not appear explicitly in the function, but was hidden implicit in the difference between the A_i and B_i coefficients. Therefore, the FS was converted to a FS with only sine and phase shift (ϕ) terms.

First, the original form of the function:

$$A_o + A_1 Cos(\omega t) + B_1 Sin(\omega t) + A_2 Cos(2\omega t)$$

$$+B_2 Sin(2\omega t) + ... + A_i Cos(i\omega t) + B_i Sin(i\omega t) \tag{3.0}$$

In effect we combined the coefficients A_i and B_i into an amplitude and phase form. A transformation possible since any sine wave can be expressed as a cosine wave with a phase shift or vice versa.

$$= A_0 + a_1 Sin(\phi_1) Cos(\omega t) + a_1 Cos(\phi_1) Sin(\omega t)$$

$$+a_2 ... + a_i \tag{4.0}$$

The relationship between A_i and B_i and the ϕ_i and a_i was derived by expanding the cosine wave with a phase shift using trigonometrical identities.

$$= a_0 + a_1[Sin\phi_1 Cos(\omega t) + Cos\phi_1 Sin(\omega t)] + a_2 ... + a_i \tag{5.0}$$

$$= a_0 + a_1 Sin(\omega t + \phi_1) + a_2 Sin(2\omega t + \phi_2)$$

$$+ ... + a_i Sin(n\omega t + \phi_i) \tag{6.0}$$

From this we got:

$$A_i = a_i \sin(\phi_i) \tag{7.0}$$

$$B_i = a_i \cos(\phi_i) \tag{8.0}$$

$$\frac{A_i}{B_i} = Tan\phi_i \tag{9.0}$$

Hence the result shown above, with all the coefficients for cosine, denoted A_i being replaced with sine functions, denoted a_i and the coefficients for sine, denoted B_i being replaced with ϕ_i.

Therefore, we could extract a_i and ϕ_i and manipulate the phase shift to be exactly half a period:

$$\phi_i = \arctan(\frac{A_i}{Bi}) \tag{10.0}$$

$$a_i = \frac{A_i}{\sin(\phi_i)} \tag{11.0}$$

and convert it back to the mixed cosine/sine format with the use of equation 7.0 and 8.0.

This conversion enabled manipulation of ϕ directly in order that the right heel strike would be corrected to t=0 at (x,z) = (0,0), before converting the FS back to mixed terms format, and implementing them in AMS. To avoid other unintended differences to be detected by the PCA some other adjustments were implemented in the data cleaning process. As the different degrees of freedom (DoF) in the same running trial had slightly different fundamental frequencies, an average frequency was used for all DoFs in a trial. All models were re-oriented to face the same direction, and the motion was made symmetrical by averaging the right and left sides and implementing a phase shift of exactly half a period between sides.

Having represented the anatomical joint angles with a FS, a normalization of the Y-coordinates of the models average pelvis position was made to account for subjects who ran in a

skew direction. Replacing the y-component of the pelvis position with a center of mass (CoM) measure allowed us to manipulate the amplitude of CoM to have a mean value of 0, by defining the A_0 coefficient to be 0.0. Equally the X- and Z-components of the pelvis position were replaced by equivalent CoM measures.

All models had a common pelvis origin, which consequently led to some models penetrating the floor or fly through the air due to their varying heights. Vectors between heel nodes and the origin of the global coordinate system (GCS) were extracted, expressed as step height, and recreated as FS. These coordinates replaced the hip flexion/extension angle and served as a driver in the model. This ensured that the models would always hit their feet exactly on the ground. Equally, a vector between the toe node and the origin of the GCS was made to replace the knee flexion/extension angle and was expressed as the step length. A parametric model with a minimum set of parameters was desired, and therefore some segments was excluded from the model due to their relatively low mass. Based on Newton's second law it was hypothesized that changes in wrist flexion, wrist abduction, elbow pronation and neck extension angles would not affect the forces significantly when running.

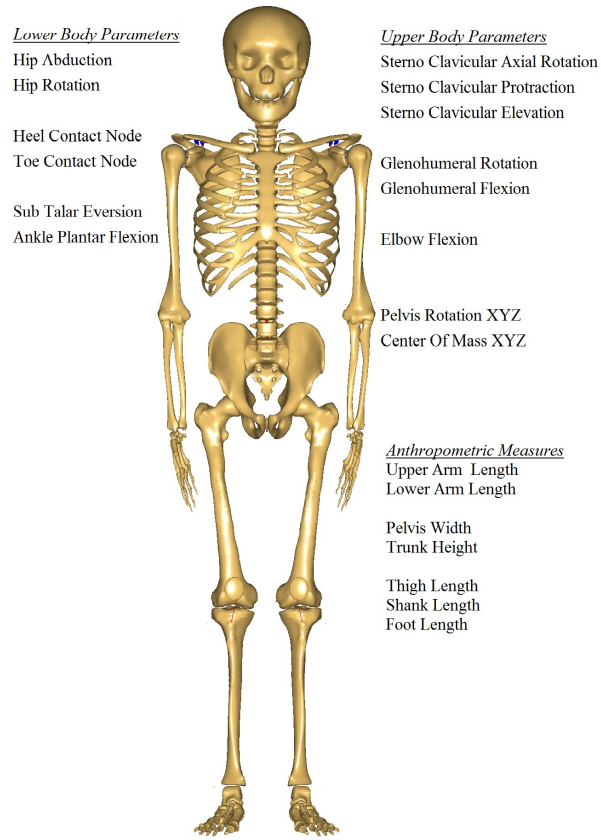


Figure 2: ***Overview of included parameters*** - *The included parameters in the principal component analysis, consisting of three rotational pelvis components, three center-of-mass components, twenty anatomical joint angles, foot height, step length, and seven anthropometric measures.*

2.5 Statistical Procedure

PCA was done in Python 3.6.1 (The Python Software Foundation, http://www.python.org). Frequency (ω) and all fitted FS coefficients from each of the processed running trials were arranged in a row vector along with the respective subject's anthropometry. All row vectors combined served as the coefficient matrix with the dimensions of 90 X 197 (subjects X FS coefficients) which was exposed to PCA. All columns were investigated to ensure all FS coefficients were normally distributed before conducting the PCA. A total of 18 parameters were included in the PCA along with seven anthropometric measures which are listed in figure 2. Due to the bilateral symmetry of the model was all left side coefficients neglected from the PCA.

With the use of inverse PCA, the first number of principal components adequate to explain >90% of the variance were transformed back to its original domain. By multiplying the standard deviation of these principal components with some factor and adding it to the mean values allowed for generation of a pseudo population of different running techniques.

3 Results

The FS representation of the anatomical joint angle time series from the processed MoCap-data succeeded to drive the running simulation models in AMS.

Table 1 summarizes the results of the mean anthropometry characteristics for all subjects ± the standard deviation, extracted from the motion optimization sequence in AMS.

Subject anthropometry characteristics		
	Mean	**Std. Dev.**
Pelvis width*	15.7cm	1.4cm
Thigh lengths	41.9cm	3.6cm
Shank lengths	42.7cm	2,0cm
Foot lengths	25.8cm	2.8cm
Trunk height	60.3cm	3.5cm
Upper-arm length	29.7cm	1.9cm
Lower-arm length	25.6cm	1.6cm

Table 1: **Anthropometry characteristics** - Mean anthropometry characteristics ± standard deviations (Std. Dev.). Note that the right and left side is mirrored.
*Pelvis width measured as the distance between hip joints.

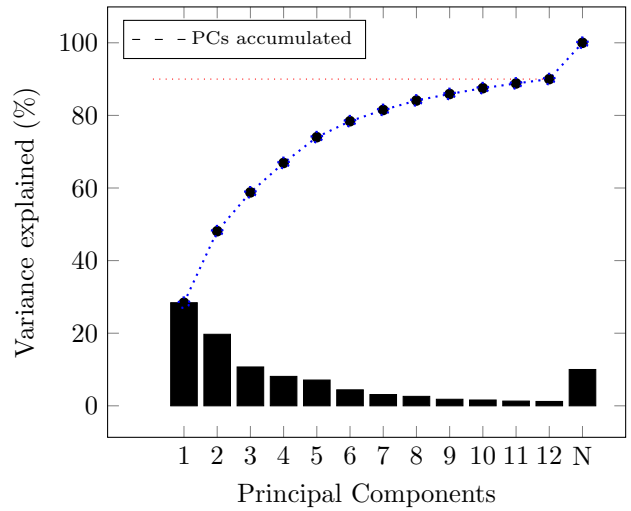


Figure 3: **Principal components** - Variance explained (%) versus principal components (PC) in bar diagram. The blue dotted line represents the cumulative explained variance (%) versus PCs, and the red dotted line illustrates when 90% are reached.

The highest eigenvalues of the first three PCs		
	FS coefficient	**Eigenvalue**
PC1	Frequency (ω)	-0.920
PC2	A0 Right Glenohumeral Rotation	0.782
PC2	A1 Right Glenohumeral Flexion	0.250
PC3	A1 Right Elbow Flexion	0.781
PC3	A0 Right Glenohumeral Rotation	0.402

Table 2: **Highest eigenvalues of the thirst three principal components** - The table provides an overview of the highest eigenvalues associated with the first three principal components (PC).

Figure 3 provides an overview of the explained variance distribution by the principal components. PCA revealed that >50% of the variance were contained in the first three principal components alone. The accumulated variance explained by the first twelve principal components were adequate to account for 90 % of the variance included in the dataset as illustrated by the red dotted line in figure 3.

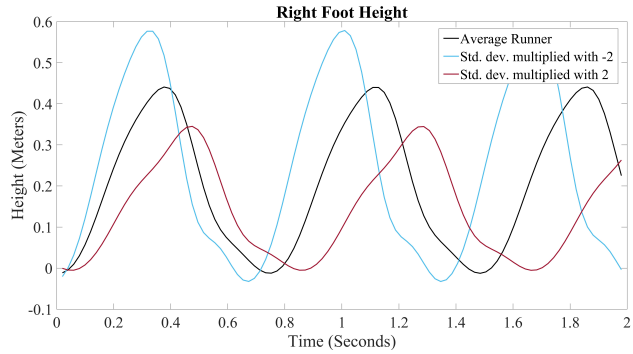
A closer inspection of eigenvalues of the principal components is provided in table 2. The results revealed that the parameter associated with the highest eigenvalue of the first principal component was the frequency (ω) coefficient (-0.920). This eigenvalue turned out to be excessively higher than the second highest eigenvalue. The highest eigenvalues associated with the second principal component proved to be the A_0 FS coefficient for the right glenohumeral external/internal rotation (0.782) and the A_1 coefficient for the right glenohumeral flexion/extension angle (0.250). Concerning the third principal component the highest eigenvalue belonged to A_1 right elbow flexion (0.781) and the second highest to A_0 right glenohumeral external/internal rotation (0.402).

Transforming only the first twelve principal components back to its original space through inverse PCA and implementing them as drivers in AMS succeeded to drive the running simulation models. Generating new running techniques within principal components space succeeded as well.

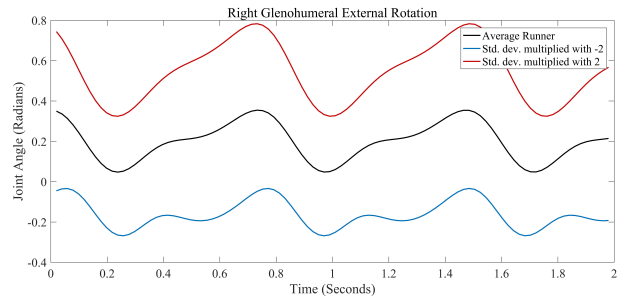
Figure 4 (a) illustrates the hight of the right foot during three new generated running techniques. The black graph represents the most 'average runner' generated by the mean values of the first twelve principal components. The red graph rep-

resent a running technique where two times the standard deviation (std. dev.) has been added to the mean of the first principal component and the remaining eleven principal components is the mean values. Likewise, does the blue graph represent a pseudo generated running style where negative two times the std. dev. of the first principal component has been added to the mean values.

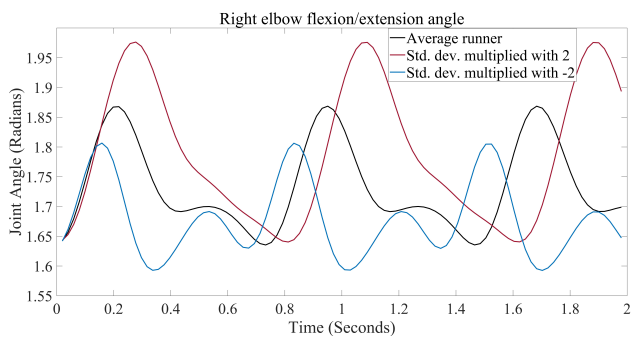
Generating the new running techniques by multiplying only the first principal component (figure 4a) with a negative factor resulted in faster execution of one stride cycle compared to the average runner (0.06sec faster), in consensus with the highest eigenvalue being the frequency. Likewise, did a multiplication with a positive factor result in a slower execution of one stride cycle compared to the average runner (0.11sec slower). Also, changes in the amplitude of the foot height were present. The changes resulted in the foot height being 13.6cm higher and 9.6cm lower compared to the average runner, for the negative and positive factor, respectively (figure 4a). Similar factors was multiplied to the std. dev. of the second principal component in figure 4b. This proved to represent a change in the offset of the glenohumeral rotation, corresponding to an increase of 0.39 radians for the positive factor and a decrease of 0.37 radians for the negative factor, which is in general consensus with A_0 being the coefficient with the highest eigenvalue and representing the offset. Figure 4c represents the third principal component being manipulated to generate new running techniques. The results of this turned out to be an increase in the right elbow joint angle when the multiplication factor was positive and a decrease when the factor was negative, compared to the average runner.



(a)



(b)



(c)

Figure 4: **Pseudo Generated Running Styles** - (a) illustrates the height of the right foot during running. (b) illustrates the right glenohumeral rotation and (c) illustrates the right elbow flexion/extension angle. In all three figures does the black graph illustrate the average runner (mean values for all 12 principal components), the red and blue graph illustrates a pseudo generated runner where the standard deviation (std. dev.) of the first, second or third principal component has been multiplied by 2 and negative 2, respectively. The x-axis represents time and the y-axis for (a) represents the change in foot height (in meters), where the y-axis for (b) and (c) represents the changes in joint angle (in radians).

4 Discussion

In the present study, a method for developing a parameterized model for simulation of varying running techniques was proposed. Exposing the coefficients from the FS representation of the joint angles along with the subjects anthropometry to PCA, resulted in twelve principal components being adequate to account for the majority of the variance in the data set (90%). Re-transforming only these twelve principal components back to its original space allowed for generating the running techniques as well as generating new pseudo running techniques. The new generated running techniques reflected the results of the eigenvalues and changes in the FS coefficients associated with the highest eigenvalues were present in the new running techniques. The models basically allows for generating an enormous amount of combinations to produce different running styles, which would be extremely time consuming to acquire in the laboratory with MoCap-data.

Earlier Studies have used PCA for data reduction and pattern recognition (Muniz & Nadal 2009, Deluzio & Astephen 2007, Deluzio et al. 1997, Chau 2001) in gait studies. However, these studies exposed the raw data points to PCA. In the present study, a FS representation of the raw data points was applied to reduce the data, such that an even more compact representation of the data was achieved. This allowed for an model with a minimum set of parameters representing the full body during the entire running motion.

The highest eigenvalue of the first principal component proved to be the frequency. Earlier studies have suggested that stride frequency during running is associated with velocity (Nilsson et al. 1985) and anthropometric dimensions of the runner (Van der Walt & Wyndham 1973). This result indicate that the subject group was heterogeneous in anthropometric characteristics which is in consensus with the extracted anthropometry measures from the motion optimization sequence in AMS. This could also be an indicator of the velocity of the running trials being varying between subjects, which was not known in advance. The highest eigenvalues of the second and third principal components proved to be associated with upper-body kinematics which is in agreement with an visual inspection of the musculoskeletal models. The role of upper-body contributions to lower-body joint kinematics are not well known (Miller et al. 2009), but the result from this study suggest that some variability in the upper-body is present between subjects which might be interesting for further investigation.

A larger sample size would increase the likelihood of representing a broader population of people with different running techniques. As all running trials were captured in laboratory settings it is conceivable that some subjects might have found it hard to achieve their preferred technique for continuous running due to the limited amount of space. Likewise, as the majority of the MoCap data only contained one half of a stride cycle it is not entirely certain that exactly that one stride cycle is representative for that subjects running technique. However, the process of developing the model allows for easily adding new running trials to the existing sample to expand the sample size.

Due to the limited amount of time frames were all running trials preprocessed to be symmetrical by averaging the left and right sides and shifted by half a period. Earlier studies have shown high variability between subjects in running asymmetry. Considering kinetics the asymmetry may range from 3.0% to 54.0%(Zifchock et al. 2006), whereas kinematic asymmetries range

from 3.0% for the knee flexion/extension joint angle at touchdown to 25.59% for ankle plantar flexion velocity at touchdown (Karamanidis et al. 2002). Therefore, these variations should be considered for a model to be as representative of normal human running as possible. However, if a method for identifying these bilateral kinematic asymmetries is present, the nature of FS drivers allows for adjusting offsets (A_0 coefficients) and motions individually, and thereby account for these asymmetries.

Expanding the model to include ground reaction forces and moments (GRF&Ms) would provide an even more detailed tool for investigating different motion combinations. We suggest that a continuation of the development of a parameterized running model should include GRF&Ms prediction in AMS, which is already an establish method (Fluit et al. 2014).

The present method for developing a parameterized running model should not be limited to running gait but should be applicable to other motions. Other sports related movements such as rowing and kayaking or motions of daily living e.g. gait or squatting to a chair could be parameterized as well. In practice, models of such motions allows for generating an enormous amount of combinations to produce the motion in question, which otherwise would be extremely time consuming to acquire with MoCap-data. This could provide the means to simulate and predict a variety of outcomes based on changes in kinematics.

References

Andersen, M. S., Damsgaard, M., MacWilliams, B. & Rasmussen, J. (2010), 'A computationally efficient optimisation-based method for parameter identification of kinematically determinate and over-determinate biomechanical systems', *Computer methods in biomechanics and biomedical engineering* **13**(2), 171–183.

Andersen, M. S., Damsgaard, M. & Rasmussen, J. (2009), 'Kinematic analysis of over-determinate biomechanical systems', *Computer Methods in Biomechanics and Biomedical Engineering* **12**(4), 371–384.

Bredeweg, S. W. (2014), Running related injuries: the effect of a preconditioning program and biomechanical risk factors, PhD thesis, University of Groningen.

Chau, T. (2001), 'A review of analytical techniques for gait data. part 1: fuzzy, statistical and fractal methods', *Gait & posture* **13**(1), 49–66.

Clermont, C. A., Osis, S. T., Phinyomark, A. & Ferber, R. (2017), 'Kinematic gait patterns in competitive and recreational runners', *Journal of Applied Biomechanics* pp. 1–26.

Crossley, K., Bennell, K. L., Wrigley, T. & Oakes, B. W. (1999), 'Ground reaction forces, bone characteristics, and tibial stress fracture in male runners.', *Medicine and Science in Sports and Exercise* **31**(8), 1088–1093.

Damsgaard, M., Rasmussen, J., Christensen, S. T., Surma, E. & De Zee, M. (2006), 'Analysis of musculoskeletal systems in the anybody modeling system', *Simulation Modelling Practice and Theory* **14**(8), 1100–1111.

Daoud, A. I., Geissler, G. J., Wang, F., Saretsky, J., Daoud, Y. A. & Lieberman, D. E. (2012), 'Foot strike and injury rates in endurance runners: a retrospective study', *Med Sci Sports Exerc* **44**(7), 1325–34.

Deluzio, K. & Astephen, J. (2007), 'Biomechanical features of gait waveform data associated

with knee osteoarthritis: an application of principal component analysis', *Gait & posture* **25**(1), 86–93.

Deluzio, K. J., Wyss, U. P., Zee, B., Costigan, P. A. & Serbie, C. (1997), 'Principal component models of knee kinematics and kinetics: normal vs. pathological gait patterns', *Human Movement Science* **16**(2), 201–217.

Derrick, T. R. (2004), 'The effects of knee contact angle on impact forces and accelerations.', *Medicine and Science in Sports and Exercise* **36**(5), 832–837.

Fluit, R., Andersen, M. S., Kolk, S., Verdonschot, N. & Koopman, H. (2014), 'Prediction of ground reaction forces and moments during various activities of daily living', *Journal of biomechanics* **47**(10), 2321–2329.

Fredericson, M. & Misra, A. K. (2007), 'Epidemiology and aetiology of marathon running injuries', *Sports Medicine* **37**(4-5), 437–439.

Hinrichs, R. N., Cavanagh, P. R. & Williams, K. R. (1987), 'Upper extremity function in running. i: center of mass and propulsion considerations', *International Journal of Sport Biomechanics* **3**(3), 222–241.

Karamanidis, K., Arampatzis, A. & Brüggemann, G.-P. (2002), Symmetry and reproducibility of kinematic parameters during various running techniques, *in* 'ISBS-Conference Proceedings Archive', Vol. 1.

Kersting, U. G. & Ferdinands, E. (2005), 'Impact forces, rearfoot motion and the rest of the body in heel-toe running', *In Proc. Of the International Society of Biomechanics Conference* .

Lieberman, D. E., Warrener, A. G., Wang, J. & Castillo, E. R. (2015), 'Effects of stride frequency and foot position at landing on braking force, hip torque, impact peak force and the metabolic cost of running in humans', *Journal of Experimental Biology* **218**(21), 3406–3414.

Liwicki, S., Tzimiropoulos, G., Zafeiriou, S. & Pantic, M. (2013), 'Euler principal component analysis', *International journal of computer vision* **101**(3), 498–518.

Mann, R. V. (1980), 'A kinetic analysis of sprinting.', *Medicine and Science in Sports and Exercise* **13**(5), 325–328.

Miller, R. H., Caldwell, G. E., Van Emmerik, R. E., Umberger, B. R. & Hamill, J. (2009), 'Ground reaction forces and lower extremity kinematics when running with suppressed arm swing', *Journal of biomechanical engineering* **131**(12), 124502.

Milner, C. E., Ferber, R., Pollard, C. D., Hamill, J. & Davis, I. S. (2006), 'Biomechanical factors associated with tibial stress fracture in female runners', *Medicine and science in sports and exercise* **38**(2), 323.

Muniz, A. & Nadal, J. (2009), 'Application of principal component analysis in vertical ground reaction force to discriminate normal and abnormal gait', *Gait & posture* **29**(1), 31–35.

Nilsson, J., Thorstensson, A. & Halbertsma, J. (1985), 'Changes in leg movements and muscle activity with speed of locomotion and mode of progression in humans', *Acta Physiologica* **123**(4), 457–475.

Paluska, S. A. (2005), 'An overview of hip injuries in running', *Sports Medicine* **35**(11), 991–1014.

Pandy, M. G. (2001), 'Computer modeling and simulation of human movement', *Annual review of biomedical engineering* **3**(1), 245–273.

Phinyomark, A., Osis, S. T., Hettinga, B. A., Kobsar, D. & Ferber, R. (2016), 'Gender differences in gait kinematics for patients with knee osteoarthritis', *BMC musculoskeletal disorders* **17**(1), 157.

SetioIVS (2017), 'Runtimizer'.
URL: *http://www.runtimizer.com/*

Skals, S., Rasmussen, K., Bendtsen, K. & Andersen, M. S. (2014), Validation of musculoskeletal models driven by dual microsoft kinect sensor data, *in* '13th International Symposium on 3D Analysis of Human Movement, Lausanne, Switzerland', pp. 14–17.

Taunton, J. E., Ryan, M. B., Clement, D., McKenzie, D. C., Lloyd-Smith, D. & Zumbo, B. (2002), 'A retrospective case-control analysis of 2002 running injuries', *British journal of sports medicine* **36**(2), 95–101.

van de Panne, M. (1996), 'Parameterized gait synthesis', *IEEE Computer Graphics and Applications* **16**(2), 40–49.

Van der Walt, W. & Wyndham, C. (1973), 'An equation for prediction of energy expenditure of walking and running.', *Journal of Applied Physiology* **34**(5), 559–563.

van Mechelen, W. (1992), 'Running injuries', *Sports Medicine* **14**(5), 320–335.

Zajac, F. E. & Winters, J. M. (1990), Modeling musculoskeletal movement systems: joint and body segmental dynamics, musculoskeletal actuation, and neuromuscular control, *in* 'Multiple muscle systems', Springer, pp. 121–148.

Zifchock, R. A., Davis, I. & Hamill, J. (2006), 'Kinetic asymmetry in female runners with and without retrospective tibial stress fractures', *Journal of biomechanics* **39**(15), 2792–2797.

Concluding Remarks

The proposed method for developing a parameterized running kinematics simulation model succeeded to generate various different running techniques. The model allows for generating an enormous amount of body part motion combinations to produce the running motion, which otherwise would have been very time consuming to gather in a laboratory. Setio can use this model to generate running techniques similar to their customers running technique and extract a wide range of kinematic measures. Further studies should investigate if the model is able to produce fairly accurate estimations of ground reaction forces and moments for the analysis of kinetics and metabolism during running. To extend the value of this tool for Setio, the project should be extended further to link the data from Setio's accelerometers to the models in the AnyBody Modeling System. This would provide an very effective tool for analyzing their customers running technique in terms of parameters as kinematics, kinetics and/or metabolism.

Parameterizing Running Kinematics Appendix

Contents

1 Musculoskeletal Modeling

Computer-aided engineering (CAE) is an integrated approach in a variety of engineering fields, and today it is inconceivable to design and build without the use of CAE (Lund et al. 2012). Computer aided methods has reached many other fields as well and is constantly providing important information in e.g. the field of biomechanics (Pandy 2001). Due to the various advances in research, and a growing belief that computer aided methods can provide quantitative explanations of how the neuromuscular and musculoskeletal systems interact to produce movement (Pandy 2001), the field of human musculoskeletal modeling has evolved over the last decades (Lund et al. 2012, Pandy 2001). Simulation of human movement has the benefits of providing non-invasive means to gain insight into joint kinematics and kinetics of the human body during various motions and loading conditions (Jung et al. 2016).

In musculoskeletal modeling the body is usually approached in a mechanical perspective, which means that bones are modeled as rigid bodies and muscles are modeled as actuators. This approach permits the human system to be analyzed by the means of multi-body dynamics (Damsgaard et al. 2006). Analyzing human motion through a biomechanical approach, are driven by equations of motion, these equations provide the relationship between a motion and the forces acting on the human system. It implies solving equations of a generalized version of Newtons second law (The vector sum of the forces acting on an object is equal to the mass of that object multiplied by the acceleration of the object), but is a challenging task as multi-body systems exhibit notoriously complex behavior (Otten 2003). Simulation of human movement is indeed a complicated task due to the complex morphology of the human body (Damsgaard et al. 2006), and the accuracy of computer simulation relies strongly on the quality of the experimental inputs as kinematic trajectories from motion capture, external forces obtained by force plates and electromyography (Jung et al. 2016). A fundamental problem exists as the amount of muscles in the human body exceeds the number of degrees of freedom ($_n$DOF), which implies that there exists infinitely many muscle recruitment patterns to balance a given load, a problem often referred to as the redundancy problem of the muscle recruitment (Damsgaard et al. 2006). Due to the complex morphology of the human body and a lacking understanding of the central nervous systems muscle activation pattern, some general simplified assumptions of the human systems function is necessary when simulating human motion (Damsgaard et al. 2006). Due to this, advanced modeling systems as OpenSIM (Simbios, Stanford, USA) (Delp et al. 2007) and The AnyBody Modeling System (AMS) (AnyBody Technology A/S, Aalborg, Denmark) are necessary for building, validating and studying these types of models (Seth et al. 2011).

In order to perform the most comprehensive analyzes, the most accurate simulation of human motion is needed. To achieve the most accurate simulation of human motion, the musculoskeletal models needs to be subject specific. Estimation of body segment parameters (BSPs) can provide some accuracy to the musculoskeletal model (Contini et al. 1963). Inertial properties as segment masses, center of mass and moments of inertia are difficult to measure in vivo, and therefore these measures relies strongly on cadaver studies as Clauser et al. (1969). BSPs, as the ratio between total body mass to segment mass, is usually based on cadaver studies in computer based methods, where we can determine the BSPs generally and subsequently scale it to be subject-specific, to achieve the most accurate simulation of the motion in question (Contini et al. 1963). Basic properties as the length and joint location of different segments are uncertain, but they are possible to estimate by running a parameter optimization in association with motion capture input from a .c3d file. Assuming that the marker positions on a segment are well known, it is possible to estimate the basic properties quite accurately by reducing differences between the motion capture markers from the .c3d file and the positions of the corresponding nodes on each respective segment (Damsgaard et al. 2006). The joint angles derivations from the entire motion are then extracted and saved to an input folder which consequently can be used to drive the optimized musculoskeletal model.

1.1 Data Acquisition for Musculoskeletal Modeling

As mentioned in the previous section, the accuracy of computer simulation of human motion relies on the quality of the experimental input (Jung et al. 2016). In the present study, kinematic trajectories captured with motion capture

was used to prepare the computational models. None of the experimental input data was captured by the authors but gathered from previous studies (Kersting & Ferdinands 2005, Skals et al. 2014). Due to this, at least a certain degree of standardization should be sought in order to make sure that the marker protocol would be appropriate to drive the musculoskeletal models. An example of an utilized marker protocol is illustrated at figure 1.

Each marker protocol should at least include a set of three markers on each rigid segment to account for the entire motion (Cappozzo et al. 1996). Each unconstrained rigid segment is considered to have six variables which describes its pose, three independent translational degrees of freedom and three independent rotational degrees of freedom (position and orientation, respectively). One marker is sufficient to account for three degrees of freedom, two markers adds two additional degrees of freedom, and a third marker account for the last degrees of freedom (Cappozzo et al. 1996). The three markers have to be placed in a non-colinear manner to account for all the degrees of freedom. In addition to this, certain reservations has to be made in order to minimize that the markers placed on the skin moves relatively to the underlying skeleton, often referred to as soft tissue artifact (Peters et al. 2010). Soft tissue artifact can produce data which is noisy or distorted and consequently result in a poor estimation of the pose. Therefore, markers should be placed in bony landmarks, as these areas exhibit the least soft tissue artifacts (Cappozzo et al. 1996). All markers should be visible for as many of the cameras as possible throughout the entire motion. If one or more markers is out of sight it is referred to as a marker dropout. A marker dropout can in some cases be fixed by applying a interpolation function as a linear interpolation function (Robertoson et al. 2014). If it is assumed that the data was separated by such a small time increment, that the motion between two consecutive time points was in fact linear, then the linear interpolation function would be sufficient. However, if the missing time period is too long, applying interpolation functions as a linear interpolation function might not reflect the actual nature of the kinematic trajectory of the marker, and wont be able to recreate the actual motion (Robertoson et al. 2014).

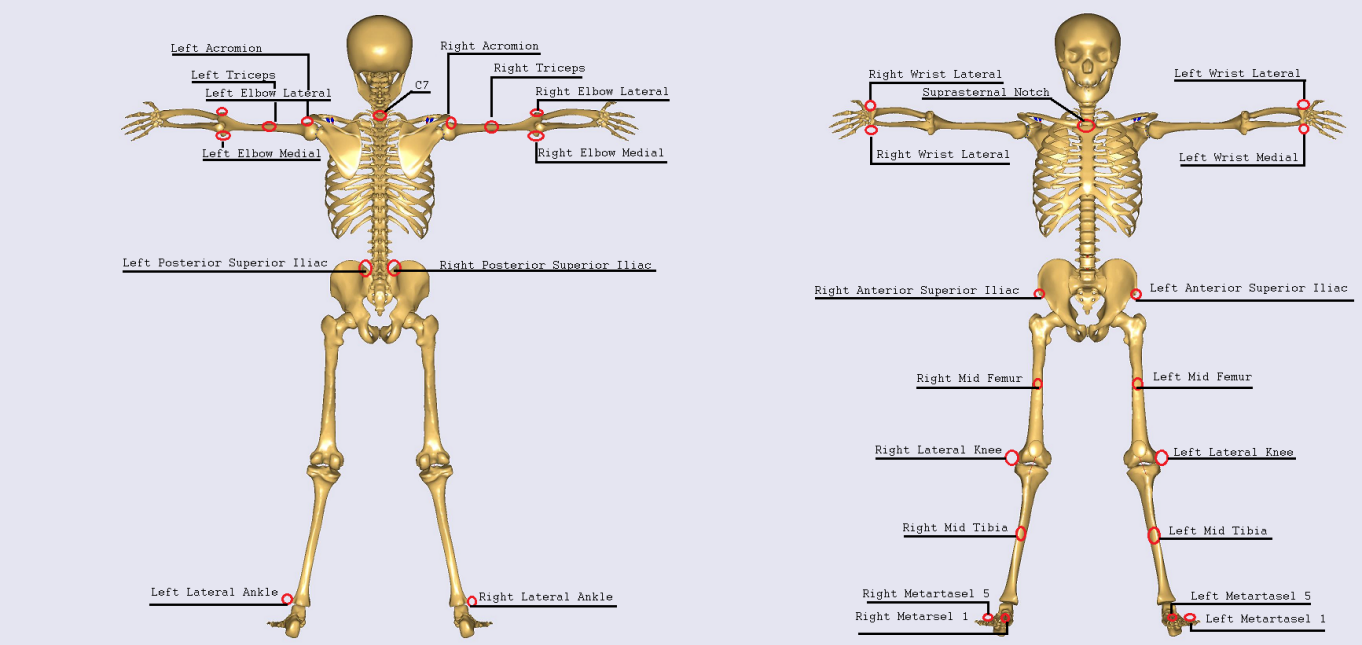


Figure 1: ***Marker Protocol*** *- Example of one of the used marker protocol in the present study. The red circles represent the markers placement on the model.*

Providing that the marker protocol is adequate to estimate the motion, the next step is ensuring that the data is available, or can be converted to a format which can be imported into the chosen simulation software. A .c3d file provides the means for storing both processed and unprocessed 3d coordinates and numeric data as EMG, force plate data and marker sets (Cramp 2001) and is supported by the most major motion capture companies as e.g. Qualisys (Qualysis AB, Gothenburg, Sweden) and Vicon (Vicon Motion systems, USA). Many motion capture software systems

has an embedded export function for .c3d files but other open-sourced softwares as Motion kinematic & kinetic analyzer (MOKKA) both reads and writes .c3d files as well.
When designing a musculoskeletal model for computer simulation, the model should be specific to answer the research question. The role of upper-body motion contributions to lower extremity joint kinematics and external force generation when running are not well known, and may have implications in the design of a musculoskeletal model (Miller et al. 2009). Earlier studies have indicated that arm swings are important in maintaining balance (Mann 1980), increased the vertical range of motion of the center of mass (Hinrichs et al. 1987) and showed that suppressed arm swing during running causes changes in peak vertical and peak lateral ground reaction forces (Miller et al. 2009), suggesting that a full-body representation of the kinematics was needed for the model to be as representative of normal human running as possible.

1.2 The AnyBody Modeling System

AMS is a simulation software initiated at the University of Aalborg, and intended for the analysis of musculoskeletal systems of humans and other creatures as rigid-body systems (Damsgaard et al. 2006). In the present study AMS served as the software for preparing the computational models. AMS was designed to meet four overall purposes, as explained by Damsgaard et al. (2006):

1. A tool which allowed users to construct models from scratch or use or modify existing models to suit different purposes.

2. The system should facilitate model exchange and cooperation on model development, and it should allow models to be scrutinized.

3. If possible, it should have sufficient numerical efficiency to allow ergonomic design optimization on inexpensive computers.

4. The system should be capable of handling body models with a realistic level of complexity.

(Damsgaard et al. 2006).
Programming in AMS is based on a text-based declarative language (AnyScript) which allows users to create objects from predefined templates. The system comes with the AnyBody managed model repository (AMMR) which includes precooked full body models containing bones, joints and over 1000 individual muscles. The models in the AMMR can be changed to fit the users own needs for a variety of purposes. Since the startup of AMS, the system has been used in numerous research areas including e.g. sports biomechanics (Dupré et al. 2016, Ji et al. 2016), work place ergonomics (Madeleine et al. 2011, Blab et al. 2016) and gait analysis (Al-Munajjed et al. 2016). For a more detailed overview we refer to AnyBody-TechnologyA/S (2016).

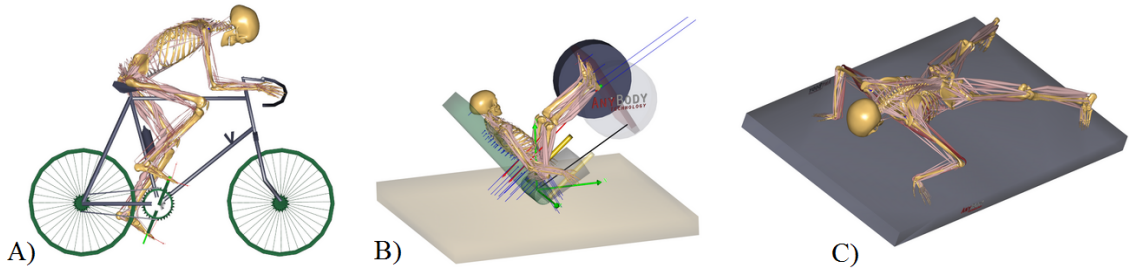


Figure 2: ***Examples of applications of the AnyBody Modeling System*** *- A few examples of models in the AnyBody Modeling System. A) illustrates a model of human cycling. B) illustrates a model seated in a leg press machine with conditional contact between the feet, back and the surfaces of the leg press. C) illustrates a model of a human doing push-ups on the floor.*

1.3 Processing Of Kinematics In The Anybody Modeling System

It is important to identify model parameters, e.g. location of marker coordinates, segment lengths and joint rotation axes, when processing subject specific musculoskeletal models from motion capture data(Andersen et al. 2010). To identify these parameters a local optimization-based method is used (Andersen et al. 2010). This method offers a computationally efficient way to determine parameters for determined and over-determined systems. Andersen et al. (2009, 2010) introduced the local optimization-based method that are used in the AnyBody Modeling systems. The position analysis for a determined system is based on the number of constraint equations being equal to the number of system unknowns (m=n) (Andersen et al. 2009). Subject to holonomic constraints, this can be formulated as a set of m equations.

$$\Gamma = \Gamma(q, (t), d, t) = 0 \tag{1.0}$$

Where the independent constraint equations in equation 1.0 can be solved numerically, and the velocity and acceleration can be derived by differentiation of this equation (1.0). However, if the system is over-determined, no general solution exist to equation 1.0. In this case Andersen et al. (2009) splits equation 1.0 into two sets:

$$\Gamma(q, (t), d, t) = [\frac{\Psi(q(t), d, t)}{\Phi(q(t), d, t)}] \tag{2.0}$$

Where $\Phi = \Phi$(q,t) must be solved exactly and $\Psi = \Psi =$ (q,t) should be solved as well as possible. The solution to equation 2.0 can be obtained at n discrete time steps by following the optimization problem as explained by Andersen et al. (2009):

$$\begin{gathered} q_i^* = argminH(\Psi(q_i, \hat{d}, t_i)) \\ q_i \\ \text{Subject to } \Phi(q_i, \hat{d}, t_i) = 0 \end{gathered} \tag{3.0}$$

The objective function H is formulated such that violations on $\Psi(q_i, \hat{d}, t_i)$ are allowed (Andersen et al. 2009). In equation 2.0 the optimal constant parameters are assumed to be known. However, an optimization problem can be formulated to produce the optimal system coordinates at each time step and optimal constant parameters throughout the time series:

$$\begin{gathered} Min \sum_{n=1}^{N} H(\Psi(q(t_i), d, (t_i)) \\ q(t_1), q(t_2), ..., q(t_n), d \\ \text{Subject to } \Phi(q(t_k), d, (t_k) \end{gathered} \tag{4.0}$$

Further, to solve the optimization problem the search direction specifying approximation to the optimal solution and step length must be determined (Andersen et al. 2009).

2 Fourier Series

A Fourier Series (FS) is a particular type of infinite series. It is a mathematical way to represent a function as a series of weighted sine and cosine terms (Davis 2012). The first thing to notice about a FS representation of a function is that the function must be periodic. Basically, FS decomposes any periodic function or signal as a sum of possibly infinite series of oscillating sine and cosine functions. The FS has a time based pattern, measures every possible cycle

and returns the amplitude, offset, and rotation speed for each cycle (Bracewell 1999). A FS can be denoted:

$$f(x) \sim a_0 + \sum_{n=1}^{\infty}[a_n Cos(\frac{n\pi}{L}x) + b_n Sin(\frac{n\pi}{L}x)] \tag{5.0}$$

and the coefficients: a_0, a_n and b_n can be derived as noted in equation 6.0, 7.0 and 8.0, respectively.

$$a_0 = \frac{1}{2L}\int_{-L}^{L} f(x)dx \tag{6.0}$$

$$a_n = \frac{1}{L}\int_{-L}^{L} f(x)Cos(\frac{n\pi}{L}x)dx \tag{7.0}$$

$$b_n = \frac{1}{L}\int_{-L}^{L} f(x)Sin(\frac{n\pi}{L}x)dx \tag{8.0}$$

with n = 1,2,3...
As the function converges it can approximate any given continuous function by each term being increasingly insignificant. According to Joseph Fourier, we should assume that the period P is equal to 2L (Bracewell 1999).

Lets now look at an example. A periodic step function is illustrated in figure 3, and we can use a FS representation to mathematically model that function. As the example in figure 3 is an odd function, a sine function is in phase

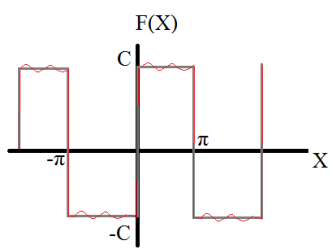


Figure 3: ***Periodic step function*** - *The grey line represents the step function and the red line represent the Fourier series representation of that step function.*

with the the step function and therefore using only sine terms would approximate the step function better than the cosine terms. The red graph in figure 3 illustrates the FS representation of the step function. As more terms is added to the FS, the red line will increasingly approximate the step function. Lets start by solving a_0 using equation 6.0. As one cycle goes from -π to π in the step function, it means that the total period is equal to 2π. We can then break the step function up from -π to 0 and from 0 to π:

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{0}(-C)dx + \int_{0}^{\pi}(C)dx = 0 \tag{9.0}$$

$$a_0 = 0$$

We can now move on and calculate a_n using equation 7.0.

$$a_1 = \frac{1}{\pi}\int_{-\pi}^{0}(-C)Cos(\frac{1\pi}{\pi}x)dx + \int_{0}^{\pi}(C)Cos(\frac{1\pi}{\pi}x)dx \tag{10.0}$$

$$a_1 = 0$$

As we are dealing with an odd function we can now see that

$$a_1 = a_2 = a_3 = ... = a_n = 0$$

Next step is calculating b_n using equation 8.0.

$$b_1 = \frac{1}{\pi}\int_{-\pi}^{0}(-C)Sin(\frac{1\pi}{\pi}x)dx + \int_{0}^{\pi}(C)Sin(\frac{1\pi}{\pi}x)dx \tag{11.0}$$

$$b_1 = \frac{1}{\pi}(-C)(-2) + (C)(2) = \frac{4c}{\pi} \tag{11.1}$$

Now we should solve b_2 which means that we can substitute n with 2:

$$b_2 = \frac{1}{\pi}\int_{-\pi}^{0}(-C)Sin(\frac{2\pi}{\pi}x)dx + \int_{0}^{\pi}(C)Sin(\frac{2\pi}{\pi}x)dx \tag{11.2}$$

$$b_2 = 0$$

Using this method we can calculate the desired number of a and b coefficients for the FS representation (Bracewell 1999).

2.1 Troubleshooting with the Fourier Series Phase Shift

In the present study we encountered a problem regarding the FS representation. A FS with mixed sine and cosine terms was chosen to represent the cyclic nature of the data, and therefore the phase shift is implicit in the difference between the cosine and sine terms. Lets assume that two running trials were exactly similar but captured at two different starting points, then the FS would return similar amplitudes but dissimilar phase shifts. Due to this we corrected all data such that heel strike for the right foot always would happen to t=0 at (x,z) = (0,0). Enabling the possibility to manipulate the phase shift involved converting the mixed Cos/Sin FS to a FS with only sine terms and a phase shift (ϕ), and then converting it back to its original form. Beginning with the original form:

$$A_o + A_1Cos(\omega t) + B_1Sin(\omega t) + A_2Cos(2\omega t) + B_2Sin(2\omega t) + ... + A_nCos(n\omega t) + B_nSin(n\omega t) \tag{12.0}$$

With A_i and B_i denoting the FS coefficients and ω denoting the frequency.

$$= A_0 + a_1Sin(\phi_1)Cos(\omega t) + a_1Cos(\phi_1)Sin(\omega t) + ... \tag{12.1}$$

$$= a_0 + a_1[Sin\phi_1Cos(\omega t) + Cos\phi_1Sin(\omega t)] + a_2... + a_n \tag{12.2}$$

$$= a_0 + a_1Sin(\omega t + \phi_1) + a_2Sin(2\omega t + \phi_2) + ... + a_nSin(n\omega t + \phi_n) \tag{12.3}$$

All the coefficients for cosine (A_i) are replaced with sine functions, denoted a_i and the coefficients for sine (B_i) are replaced with a phase shift, denoted ϕ_i. This is basically possible as any cosine-wave can be formulated as a sine-wave with a phase shift and vice versa (Bracewell 1999). By applying trigonometrical identities, it is possible to expand the relationship between A_i and B_i, such that the coefficients are combined into a phase shift and an amplitude (Bracewell 1999):

$$A_i = a_i \sin(\phi_i) \tag{13.0}$$

$$B_i = a_i \cos(\phi_i) \tag{14.0}$$

$$\frac{A_i}{B_i} = Tan\phi_i \tag{15.0}$$

So now we can extract a_i and ϕ_i and manipulate the phase shift:

$$\phi_i = \arctan(\frac{A_i}{Bi}) \tag{16.0}$$

$$a_i = \frac{A_i}{\sin(\phi_i)} \tag{17.0}$$

And convert it back to the mixed Cos/Sin format by using equation 13.0 and 14.0.

3 Principal Components Analysis

Principal Components Analysis (PCA) is data transformation method which is intended to shed light on certain aspects of a data set. As many other transformations PCA transforms data into another representation with a new set of basis vectors. PCA is a standard tool in modern data analysis, and is used in a wide range of fields (Shlens 2014). PCA is based on a linear transformation algorithm where the variance in the data set is in focus, and can be expressed as a translation and rotation (Moeslund 2001). The purpose of this transformation has four overall goals:

1. Extract the most important information from the data set.
2. Compress the size of the data by reducing dimensions with a minimum loss of information.
3. Simplify the description of the data set.
4. Analyze the structure of the observations and the variables.

(Smith et al. 2002, Abdi & Williams 2010)

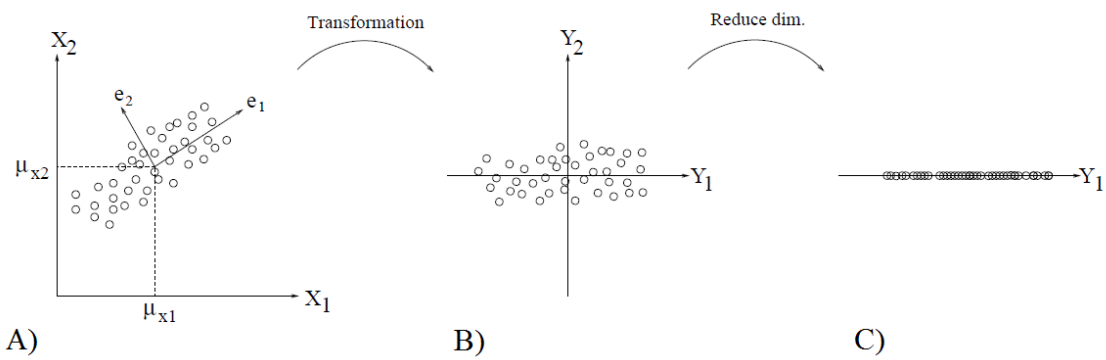


Figure 4: ***Principal Component Analysis - Two-Dimensional Example (courtesy of Moeslund (2001)).*** *Figure A) displays the input data denoted X_i. Figure B) displays the data after the transformation into a new representation where the input data now is denoted Y_i. Figure C) displays the main part of the variance kept in Y_1 after the second variable Y_2 is ignored. - In summary PCA finds directions in the input data with the highest variance and ignores directions with the least variance (Moeslund 2001).*

Figure 4 illustrates a two-dimensional example of PCA. The first figure A) displays a cloud of data points sampled from some distribution which happens to lie in a two-dimensional space. PCA basically finds directions of maximal variance. In the case in figure 4, the majority of the variance would lie in a diagonal direction, denoted e_1 which is now the new basis vector and termed the first principal component (PC). As mentioned above, PCA is a linear transformation algorithm and is therefore constrained by the basis vectors being mutually orthogonal. In this two-dimensional example, it means that the direction with the second highest variance is e_2 (Moeslund 2001).

In figure B) the data set has been transformed into a new representation where the ith sample is denoted Y_i, in this new reference frame we can note that the variance is greater along the Y_1-axis. The spatial relationship of the data points has not been altered but remains unchanged in this process (Holland 2008). The transformed data (y) calculated as:

$$y = A(x - \mu_x) \tag{18.0}$$

Where A denotes e_i as row vectors, x denotes the input data and μ_x is the mean of the data set (Moeslund 2001). The most significant direction, in terms of variance, is denoted Y_1 and the least significant is denoted Y_2. Figure C) displays how the dimensionality of the data set is reduced by ignoring the least significant PC. The process of analyzing which PCs should be kept and which should be excluded is referred to as PCA. In the case of a two-dimensional example it might not seem as a lot can be gained by reducing the dimensionality. However, imagine a data set with a higher dimensionality, but with the main variance represented by only a minor number of PCs, then a compact representation would be obtained (Moeslund 2001).

3.1 Eigenvectors and Eigenvalues

The rows of matrix A from equation 18.0, which contains the new basis vectors $[e_1,\ e_2,\ e_3\ \ldots\ \ e_n]$, are termed the eigenvectors of the variance-covariance matrix of the original data (Holland 2008). The eigenvectors indicates a direction with a certain variance, and the eigenvalues denotes the amount of variance in that direction. The elements in the diagonal of a matrix of the PCs, are know as the eigenvalues. The eigenvalues are basically the variance explained by each PC, and decreases monotonically from the first PC to the last (Holland 2008).

References

Abdi, H. & Williams, L. J. (2010), 'Principal component analysis', *Wiley interdisciplinary reviews: computational statistics* **2**(4), 433–459.

Al-Munajjed, A. A., Bischoff, J. E., Dharia, M. A., Telfer, S., Woodburn, J. & Carbes, S. (2016), 'Metatarsal loading during gait—a musculoskeletal analysis', *Journal of biomechanical engineering* **138**(3), 034503.

Andersen, M. S., Damsgaard, M., MacWilliams, B. & Rasmussen, J. (2010), 'A computationally efficient optimisation-based method for parameter identification of kinematically determinate and over-determinate biomechanical systems', *Computer methods in biomechanics and biomedical engineering* **13**(2), 171–183.

Andersen, M. S., Damsgaard, M. & Rasmussen, J. (2009), 'Kinematic analysis of over-determinate biomechanical systems', *Computer Methods in Biomechanics and Biomedical Engineering* **12**(4), 371–384.

AnyBody-TechnologyA/S (2016), 'Anybodytech'.
URL: *https://www.anybodytech.com/downloads/publications/*

Blab, F., Avci, O., Daub, U. & Schneider, U. (2016), New approaches for analysis in ergonomics: From paper and pencil methods to biomechanical simulation, *in* '16. Internationales Stuttgarter Symposium', Springer, pp. 821–835.

Bracewell, R. (1999), *The Fourier Transform and its Applications*, 3. edition edn, McGraw-Hill Science/Engineering/Math.

Cappozzo, A., Catani, F., Leardini, A., Benedetti, M. G. & Della Croce, U. (1996), 'Position and orientation in space of bones during movement: experimental artefacts', *Clinical biomechanics* **11**(2), 90–100.

Clauser, C. E., McConville, J. T. & Young, J. W. (1969), Weight, volume, and center of mass of segments of the human body, Technical report, DTIC Document.

Contini, R., Drillis, R. J. & Bluestein, M. (1963), 'Determination of body segment parameters 1', *Human Factors* **5**(5), 493–504.

Cramp, E. (2001), 'www.c3d.org'.
URL: *https://www.c3d.org/about.html*

Damsgaard, M., Rasmussen, J., Christensen, S. T., Surma, E. & De Zee, M. (2006), 'Analysis of musculoskeletal systems in the anybody modeling system', *Simulation Modelling Practice and Theory* **14**(8), 1100–1111.

Davis, H. (2012), 'Fourier series and orthogonal functions'.

Delp, S. L., Anderson, F. C., Arnold, A. S., Loan, P., Habib, A., John, C. T., Guendelman, E. & Thelen, D. G. (2007), 'Opensim: open-source software to create and analyze dynamic simulations of movement', *IEEE transactions on biomedical engineering* **54**(11), 1940–1950.

Dupré, T., Mortensen, K. R., Lysdal, F. G., Funken, J., Müller, R., Braun, M., Krahl, H., Kersting, U., Potthast, W. & Pforzheim, A. C. (2016), 'Hip joint load and muscle stress in soccer inside passing'.

Hinrichs, R. N., Cavanagh, P. R. & Williams, K. R. (1987), 'Upper extremity function in running. i: center of mass and propulsion considerations', *International Journal of Sport Biomechanics* **3**(3), 222–241.

Holland, S. M. (2008), 'Principal components analysis (pca)', *University of Georgia* .

Ji, Z., Wang, H., Jiang, G. & Li, L. (2016), 'Analysis of muscle activity utilizing bench presses in the anybody simulation modelling system', *Modelling and Simulation in Engineering* **2016**.

Jung, Y., Jung, M., Ryu, J., Yoon, S., Park, S.-K. & Koo, S. (2016), 'Dynamically adjustable foot-ground contact model to estimate ground reaction force during walking and running', *Gait & posture* **45**, 62–68.

Kersting, U. G. & Ferdinands, E. (2005), 'Impact forces, rearfoot motion and the rest of the body in heel-toe running', *In Proc. Of the International Society of Biomechanics Conference* .

Lund, M. E., de Zee, M., Andersen, M. S. & Rasmussen, J. (2012), 'On validation of multibody musculoskeletal models', *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* **226**(2), 82–94.

Madeleine, P., Samani, A., De Zee, M. & Kersting, U. (2011), *Biomechanical assessments in sports and ergonomics*, INTECH Open Access Publisher.

Mann, R. V. (1980), 'A kinetic analysis of sprinting.', *Medicine and Science in Sports and Exercise* **13**(5), 325–328.

Miller, R. H., Caldwell, G. E., Van Emmerik, R. E., Umberger, B. R. & Hamill, J. (2009), 'Ground reaction forces and lower extremity kinematics when running with suppressed arm swing', *Journal of biomechanical engineering* **131**(12), 124502.

Moeslund, T. B. (2001), Principal component analysis - an introduction, Technical report, University of Aalborg, Laboratory of Computer Vision and Media Technology.

Otten, E. (2003), 'Inverse and forward dynamics: models of multi–body systems', *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **358**(1437), 1493–1500.

Pandy, M. G. (2001), 'Computer modeling and simulation of human movement', *Annual review of biomedical engineering* **3**(1), 245–273.

Peters, A., Galna, B., Sangeux, M., Morris, M. & Baker, R. (2010), 'Quantification of soft tissue artifact in lower limb human motion analysis: a systematic review', *Gait & posture* **31**(1), 1–8.

Robertoson, G., Caldwell, G., Hamill, J., Kamen, G. & Whittlesey, S. (2014), *Research Methods in Biomechanics*, Human Kinetics.

Seth, A., Sherman, M., Reinbolt, J. A. & Delp, S. L. (2011), 'Opensim: a musculoskeletal modeling and simulation framework for in silico investigations and exchange', *Procedia Iutam* **2**, 212–232.

Shlens, J. (2014), 'A tutorial on principal component analysis', *arXiv:1404.1100* .

Skals, S., Rasmussen, K., Bendtsen, K. & Andersen, M. S. (2014), Validation of musculoskeletal models driven by dual microsoft kinect sensor data, *in* '13th International Symposium on 3D Analysis of Human Movement, Lausanne, Switzerland', pp. 14–17.

Smith, L. I. et al. (2002), 'A tutorial on principal components analysis', *Cornell University, USA* **51**(52), 65.