

Getting started - PMAalyzer

This introduction provides a quick overview of the interface and some of the most important features to consider for any PCA analysis. For a more detailed description of the software tool and a mathematical background on using kinematic PCA as feature extractor, the user can check the more extensive user's manual on the website:

<https://www.uibk.ac.at/isw/forschung/neurophysiologie-bewegungswissenschaft/software.html>

0 The PMAalyzer software and basic theory

The software was developed to facilitate and motivate research involving principal component analysis PCA on kinematic data. PCA can be used on 3D kinematic data (3D marker positions over time) to extract marker movements that correlate. Extracted movements consist of whole-body movements called *principal movements* PM_k that consist of one simultaneous movement of every marker (see example in Figure 1). The appearance of each PM_k is defined by the respective PCA output $PC_k = \text{eigenvectors}$, where k denotes the **PM-order**.

Note: The number of extractable features depends on the amount of input variables, i.e. 5 Markers - with x-y-z coordinates each - will allow to extract from one to a maximum of 15 PMs.

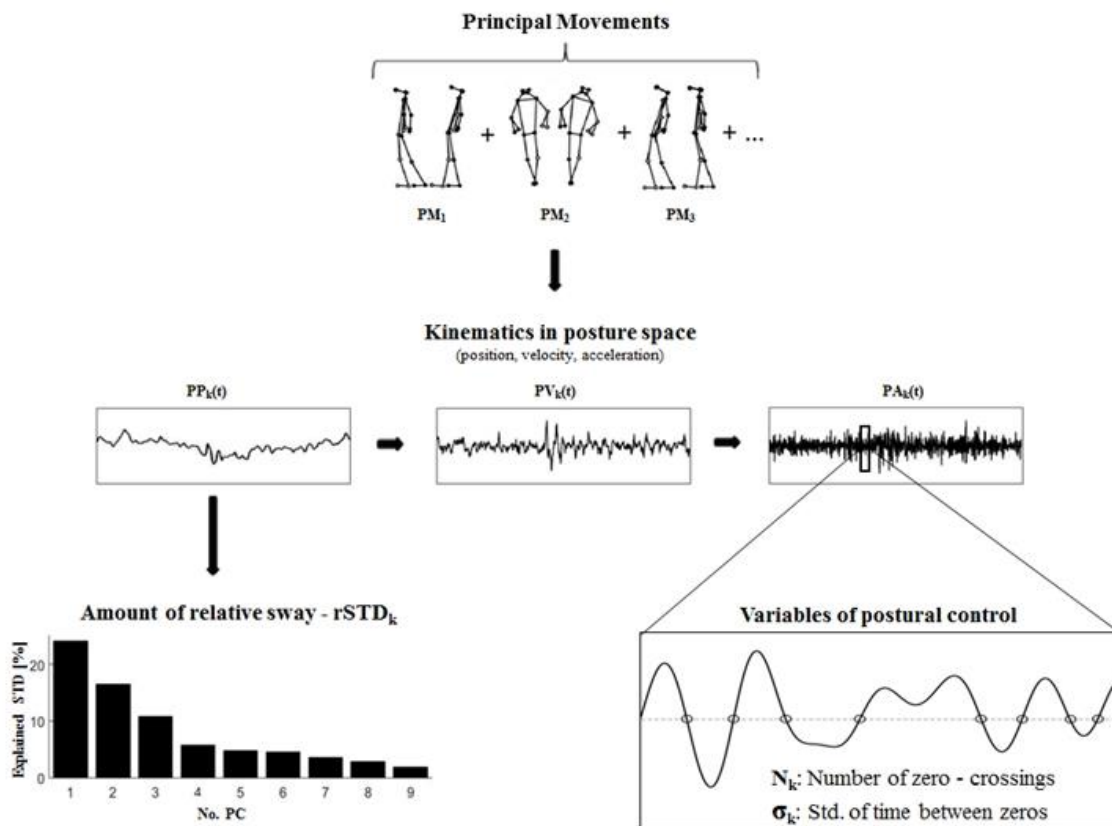


Figure 1: This figure shows the extraction of principal movements (PM_k) of a tandem stance (Haid et al., 2018). The original movement consists of an anteroposterior ankle sway (PM_1) combined with a mediolateral ankle sway (PM_2), knee flexion-extension (PM_3) and several other PMs that can be analyzed separately using the kinematic PCA approach. In this specific case three variables ($rSTD_k$, N_k and σ_k) were defined on the principal position (PP_k) and acceleration (PA_k) time-series in order to explore influences of aging on the neuromuscular control system.

Main features/variables: (Beginners/advanced/experts)

This software allows beginners to calculate these and many other variables effortlessly and export the results as Excel spreadsheet and basic plots (**EV**, **rVAR**, **rSTD**, **PP**, **PV**, **PA**). They can also experiment with the PM video options (2D or 3D), comparing original movements with PM movements and generating hypotheses based on the extracted PMs.

- **Overall eigenvalues EV_k :** The amount of explained overall variance is described by a the direct PCA output called **eigenvalues EV_k** . The first PM (PM_1) will describe the movement that generates the most overall variance (or overall marker movement) - highest EV; PM_2 will describe the movement generating the second most variance - second highest EV, etc.
- **Time-series (PP/PV/PA):** Each PM has respective time-series called **principal positions $PP_k(t)$** that describe the trial specific postural configurations over time. Analogously to conventional kinematics, the time-series can be differentiated to compute **principal velocities PV** and **principal accelerations PA_k** .
These time-series, displaying the kinematics of a trial, can be used to compute a variety of variables (see detailed user's manual and example in Figure 1).
- **Subject-specific relative eigenvalues:** Amongst the most commonly used variables are subject specific relative eigenvalues (**$rVAR_k$** and **$rSTD_k$**), as they contain information about the movement structure. Specifically, each **$rVAR_k$** and **$rSTD_k$** defines the relative amount of trial specific variance that can be explained with the respective PM_k (analogously to the EV_k for the overall variance) (Haid et al., 2018; Haid and Federolf, 2018). A variable related to the cumulated relative eigenvalues is the **residual variance RV**, that describes the dimensionality of a movement (Zago et al., 2017a, 2017b).
- **“Self-defined” variables:** Some examples of **variables on the PP/PV/PA** time-series are pre-implemented, such as the **mean, standard deviation, RMS, mean power** and **further variables** attributed to the neuromuscular control system (Haid et al., 2018; Promsri et al., 2018). See the complete user's manual for detailed description.

Advanced features: (Advanced and expert users)

More advanced users (familiar to some degree with MATLAB) can follow the instructions in the user's manual to:

- **Define variables on PP/PV/PA:** easily integrate any sort of variable computation on PP, PV and PA, by following the template in the function “optionsVariablesComp.m” in the folder “FunctionsToEdit”.
- **Customize the user interface**
 - pre-processing,
 - coordinate transformations,
 - normalization,
 - weighting,
 - define variables on PP/PV/PA,
 - plotting options,
 - video coloring options.

IMPORTANT NOTE: Only modifications and usage with non-commercial purpose allowed! (This software is property of the University of Innsbruck and is to be used exclusively for scientific research.)

1 Data preparation

Create a **folder with your data** (ASCII file) in the following format: “*PREFIXs*t*.extension*”.

Example: “tandem_s4t1.txt”.

- **PREFIX** = “tandem_”,
- **Subject number** = “s4”,
- **Trial number** = “t1”,
- **File extension** = “.txt”

Notes:

- The number of trials per subject read in must coincide.
- The extension specification in the interface must contain the “.” (**Examples:** “.txt”, “.csv”)
- The files are read in with the MATLAB function “importdata()”.
- Data may contain headers, frame specifications and gaps -> *Specifications needed in interface:*
 - *N^o header lines,*
 - *delete columns with frames,*
 - *no complete columns or rows should be missing for gap filling*
- Saved GUI data will be contained in an automatically created subfolder of the folder containing the data!

2 Start user interface and set options

NOTE: **Popup-menus** with **red font enable/disable** all necessary/unnecessary **items** that are **related** to the respective **main panel's** choice. The menus automatically handle incompatible selections.

2.1 Start PMA analyzer interface

- Enter folder containing the PMA analyzer tool. **Open** “PrincipalMovementAnalyzer.m” with MATLAB. Press “F5” key or click “Run”. The user interface will appear. (*Note: Opening the “PrincipalMovementAnalyzer.fig” will open the interface without loading the application -> error messages when computing*)
- **Copy** the **path to the folder containing the data** into the edit box marked with “PATH”.
 - String “PATH” may contain a backslash at the end or not
 - **No spaces, or dots at the end or beginning of the PATH allowed**

2.2 Main panel - Input data

Subpanel - Subjects/Trials:

- **Analysis description:** Name will be included in the subfolder that contains the output data.
- **Subjects/Trials:** Specify subjects/trials to read in (*Example: [2,5:7, 10] for s2, s5, s6, s7 and s10, or t2, t5, t6, t7, t10*)

Subpanel - Data information: Input data information.

- **Prefix:** see example: “tandem_”

- **file extension:** see example: “.txt”
- **Delimiter:** Typical data delimiters are “,” for comma, “\t” for tab.
- **Header lines:** Define integer values
- **sample-frequency:** Define integer values
- **No. markers post-processing:** Define integer values
 - *Number of markers AFTER PRE-PROCESSING*
- **Expected number of rows:** Define integer value. Just for console output

Subpanel - Special data treatment :

Note: After pre-processing the number of columns of the data must be dividable by 3 (x-y-z coordinates of markers only). Make sure to delete time-frames or other unnecessary columns. Gap-filling also requires only x-y-z of markers without any other columns!

- **Delete cols/rows/markers:** Select checkbox and specify numbers. (**Note:** Columns are deleted before markers).
- **Gap filling:** Gap filling is PCA based and preserves PC data structure.
- **Coordinate transformation:** GUI expects input in Cartesian (x-y-z) coordinates. Transformation into spherical or cylindrical coordinates may be selected.
- **Filter data:** Low pass - Butterworth filter available (**Note:** order and cut-off as selected in “Principal Component Analysis” subpanel.)
- **Pre-processing options:** Follow template in function “optionsPreProcessing.m” in software tool subfolder “FunctionsToEdit” (see detailed user’s manual). Data can be pre-processed at three time-points:
 1. Step 1: After Deleting cols/rows/markers
 2. Step 2: After gap-filling
 3. Step 3: After coordinate transformation and filtering.

2.3 Main panel - Computation and output

Subpanel - Principal component Analysis

- **PCA normalization options:** See detailed user’s manual on customization.
 - No normalization
 - Mean Euclidean distance
 - Height estimation (through specified markers)
 - Height estimation (through estimated markers)
 - Maximum range in x, y or z component.
- **PCA weighting options:** Customize in function “optionsWeights.m” (folder “FunctionsToEdit”). Follow template (see detailed user’s manual).
 - No weights
 - Standard plugin gate (39 markers)
 - Standard plugin gate (37 markers - two finger-markers missing)
- **Filtering:** Low-pass Butterworth is available.
 - Choose order and cut-off frequency.
 - Specifying more cut-off frequencies will compute the entire PMAalyzer computation several times, with each cut-off once.
- **Cross validation:** Computes a leave one out cross-validation by comparing the angles between the PC-eigenvectors of the used PC-basis and the PC-eigenvectors of the bases computed with one trial missing (see user’s manual).

- **Show Powerspectrum (Welch's PSD estimate):** Will compute the power spectral density estimate

Subpanel - Output/Save/Load

- **Save data:** Optionally, data can be saved and saved name can be edited. (**Note:** Information about the cut-off frequency and the number of PMs is included in the name.)
- **Compute selfdefined Variables:** rSTD, rVAR, CUM_rVAR and CUM_rSTD are always computed. Other available variables are optionally computed (on PP, PV and PA) when checkbox is selected.
- **Additional Excel - output:** All of the computed variables with scalar values can be exported to Excel file (file saved in subfolder "Computed data", see section 3.)

2.4 Main panel - Plots

Subpanel - PCA plots

- **EVs - overall:** Eigenvalue and cumulative eigenvalue plots as relative overall eigenvalues.
- **EVs (subject specific):** rVAR, rSTD, CUM_rVAR and CUM_rSTD plots. (**Note:** "No. figures" = number of figures for each trial, subjects split accordingly)
- **Scoreplots (PP/PV/PA):** Dynamic of manual limits
 - Dynamic y-limits: a specified proportion of the overall maximum/minimum
 - Manual: set limits manually

Subpanel - Personalized plots

- **Plot-option #:** Follow template in function "personlaizedPlots.m" in folder "FunctionsToEdit" to customize (see detailed user's manual).

2.5 Main panel - Videos

Video Dimensions: All videos available in 2D and 3D. (**Note:** Select respective "main panel option")

Subpanel - Subject/Trial/PC selection

- **Video folder name:** Name of subfolder containing the videos.
- **Video of subjects/trials/pc:** Numbers of the subjects/trials/pcs of which a video shall be created.
- **PC per figure:** Number of components in one video.

Subpanel - Video type selection:

- **Input Data-post processing:** Video of input data without PCA.
- **PM visualization:** Visualizes the PM itself (oscillation from minimum to maximum of first available subject)
- **PM trials/subjects:** Visualizes the manifestation of the PM over time of selected trials/subjects.

- **PM-combination:** Combines selected PMs and visualizes the respective manifestation of the combined PMs for selected subjects. (*Example: [2,3; 1:6; 5,9] will make three combined PM videos: 1. PM2 + PM3, 2. PM1-PM6 combined and 3. PM5+PM9 -> comma “,” adds a PM and semicolon “;” starts new combined movement*)
- **“Add: activity of PP”:** Add a barplot under the PM-video of a trail depicting the PP activity at each time-point (see user’s manual). (*Note: Only available for video type “PM trials/subjects”*)

Subpanel - Video settings:

- **Axes settings:** all six x-y-z permutations (y-z-x, z-y-x, ...) are available for plotting if data input is not in x-y-z format.
- **Coloring options:** Per default all markers are plotted in black markers without connection lines. Customize in function “optionsColoringVids.m” (folder “FunctionsToEdit”). Follow template (see detailed user’s manual).
- **Skip frames:** Skips time-frames while plotting (speed up when sample-frequency is high)
- **PM amplification:** One number will amplify all selected PMs with same factor. One number per selected PM will amplify every PM accordingly.
- **Visualization only options:** “Motion-range - Auto” will use maximum and minimum of first available subject for PM visualization. “Runthroughs” determines the number of oscillations from minimal to maximal amplitude. “N^o positions” determines the N^o distinguishable positions from minimal to maximal amplitude.

2.6 Main panel - Save/Load settings

Save: Saves all interface options in file “PMAalyzer_settings_struct.mat” in folder containing the data under the name specified in the edit box below. File is created automatically if not existent.

Load: Loads available interface options from file “PMAalyzer_settings_struct.mat” into the listbox. File is created automatically if not existent in folder containing the data.

Change to selected: Changes interface settings to settings selected in listbox.

Delete: Deletes settings selected in listbox from “PMAalyzer_settings_struct.mat”.

3 Output and results

3.1 Folder containing all data:

The folder will be created including the information in edit box “Analysis description”, used cut-off and number of PMs.

Subfolder - All plots: Created plots are saved in this subfolder

Subfolder - Computed data: Location of all computed data and the excel-output.

Subfolder - Videos: Location of created videos. (**Note:** Folder name editable in “Video folder name” edit box)

4 Loading data:

To load data, **copy the data to load into the folder containing the original input data** (data to load is contained in folder “Computed data” if saved during previous computation). **File names must coincide** with defined **name in interface** (**Example:** deleting the “*CO_*PC__” prefix is the easiest).

Alternatively the PATH in the interface can be changed to a path containing the data -> Names can also be altered in the interface. Note: The results will then be saved in subfolders of this new folder containing the loaded data!

- **“Load input data from:”** requires
 - DataReadin.mat

- **“Load PCA-data from PATH:”** requires **all above and**
 - DataPCA_TimeSeries.mat
 - DataPCA_Basic.mat
 - DataPCA_Expanded.mat

- **“Load PCA-variables from PATH:”** requires **all above and**
 - ComputedVariables