PCA analysis to assess the efficiency of Epidural Electrical Stimulation in a SCI patient

Julia Heiniger, Tobias Bodenmann, Nerea Carbonell
BIOENG - 404 Analysis and Modeling of Locomotion, EPFL, Switzerland

I. Introduction and dataset description

In the United States (U.S.) alone, the annual incidence of spinal cord injury (SCI) is estimated at around 18 000 new cases each year. Consequently, the number of people with SCI living in the U.S. is around 300 000 [1], [2]. At the scale of the world between 250 000 and 500 000 people suffer from SCI each year [3]. The leading cause of the injury are vehicle crashes, followed by falls [2].

SCI disrupts the communication between supraspinal centers and spinalcircuits generating locomotion, leading to partial or even complete loss of essential neurological functions below the injury site [4]. As a consequence, individuals suffering from SCI are left to deal with possibly permanent disabilities, significantly compromising their quality of life. Fortunately, clinical advances have managed to decrease morbidity and improve rehabilitation outcomes [1], [4], [5]. Recent studies have even achieved to restore walking in people with SCI using epidural electrical stimulation (EES) [6].

The hereby presented work aims at characterizing and comparing gait patterns in healthy subjects and individuals with neurological disorders using PCA analysis. The analyzed datasets display different walking tasks performed by (1) healthy volunteers and (2) subjects with SCI using targeted spinal cord stimulation within the scope of the STIMO clinical trial.

1) In a first turn, two healthy subjects were asked to walk for 2 minutes on a flat treadmill at 2 km/h (condition 1). Following this first condition, the same subjects were then asked to continue walking on the flat treadmill for 4 minutes. For this condition the speed was increased every minute by 1 km/h, starting from 2 km/h, and eventually finishing at 5 km/h (condition 2). Finally, this speed ramp protocol was repeated for the subjects after inclining the treadmill by 20% (condition 3). For all three conditions hip, knee, ankle and toe markers were recorded to reconstruct the 3D kinematics. Additionally, EMG activity of the main flexor and extensor muscles of each leg were measured (Iliopsoas, Rectus Femoris, Vastus Lateralis, Tibialis Anterioris, Semitendinosus, Medial Gastrocnemius, Soleus).
2) For the second dataset, closed-loop EES combined with robot assisted rehabilitation training was used with the aim of restoring walking in individuals with chronic SCI. First, the participant was asked to walk on a flat treadmill without EES at 1km/h (condition 4). Then, the participant underwent a speed ramp from 0.8 to 1 km/h (condition 5) and from 0.8 to 2 km/h (condition 6). For this dataset, EMG activity of the same seven leg muscles (and additionally, right biceps femoris), together with hip, knee and ankle markers were recorded.

II. Methods

A. Gait event detection

For each gait cycle the initial contact (IC) and the toe-off (TO) were detected (see fig. 1 and 2). After studying the videos and comparing them to the marker data, the TO was identified as the maximum of the difference of the marker positions for ankle and knee in the

z-direction. To find those maxima, the MATLAB function findpeaks was used. Using several constrains false identified peaks were sorted out. First, the mean of the peaks was calculated. Then, the findpeaks function was applied for a second time using the two arguments: $'MinPeakProminence' = 0.75 \cdot mean$ and $'MinPeakDistance' = 30$. The assumption was that a gait cycle is at least 0.3s long. The last constraint was, after each right TO follows first a left TO before a next right TO could occur, and they have to be at least 0.2s apart. If more than one TO was between TOs of the other foot, the mean location was selected or the highest peak (if the peaks were more than 25mm apart). This approach led to a very precises detection for the TO of healthy individuals. As long as the subjects with SCI walks properly, the TO detection is robust, but if the subject struggles and makes for example two left steps in a row, the detection fails.

Since we study walking, IC can be found between the TO of one foot and the next TO of the other foot. When comparing the data and the videos, it could be determined that the IC always takes place when the hip reaches the minimum in the vertical between the two TOs. Thus, a threshold function of the z-coordinate of the HIP marker was used to detect the ICs. Therefore, only the signal 0.1 second after the TO and 0.1 second before the next TO was considered. The IC was then defined as the mid timepoint where the signal is below the threshold which was defined as $\frac{9}{10}$ of the minimum. Again, the approach is very precise for the healthy walking but less accurate for the patient.

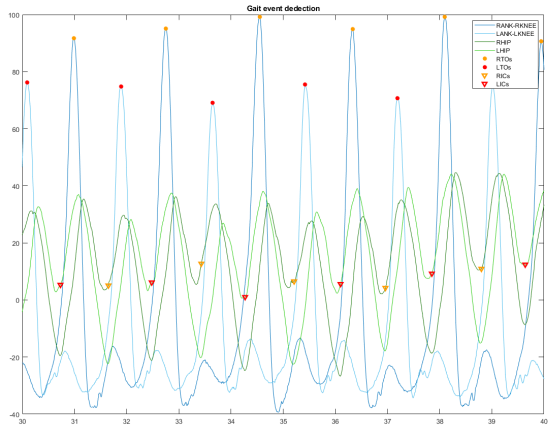

Figure 1: Gait event detection of the data set AML_02_3.mat

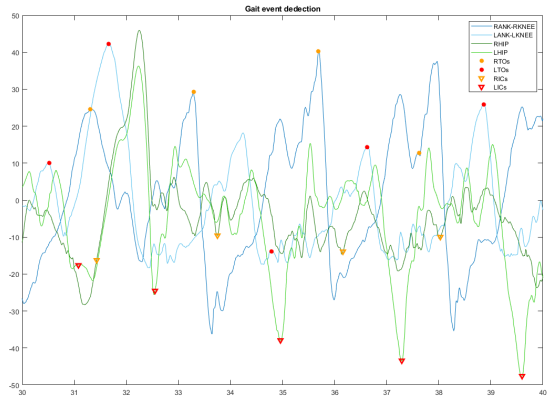

Figure 2: Gait event detection of the data set DM002_TDM_08_1kmh.mat

B. *Kinematic feature extraction*

The kinematic parameters extracted for each gait cycle were chosen by their suitability to differentiate healthy from SCI subjects based on the video recordings of their respective locomotion. Furthermore, we made sure to chose features that cover different groups of gait parameters in order to have a good overall representation of the gait.

1) *Gait cycle duration (left and right):* A full gait cycle begins with the initial contact of one foot and ends with the next initial contact of the same foot. Gait cycle duration, then measures the time for the subject to complete such a gait cycle. This feature is important for spatio-temporal analysis of

gait, and might allow to obtain quantitative information about the severeness of the injury and the effectiveness of the rehabilitation strategy. Healthy subjects display a rhythmic and symmetric gait, which should present constant and reasonable gait cycle duration values. This feature was implemented by calculating the difference between successive initial contacts for right and left foot, which were obtained in 2A.

2) *Step length (left and right):* Another classical measure for gait timing is the distance between the point of initial contact of one foot and the point of initial contact of the opposite foot. In healthy gait, left and right step length should be very similar. Therefore, this feature allows to investigate the symmetry of the gait. The step length ($step$) was calculated by taking the difference between the ankle coordinate (x_r) of one foot at initial contact (IC) and the ankle coordinate of the other foot (x_l) at the following initial contact. To this, the distance traveled by the treadmill, i.e the speed of the treadmill (v_t) multiplied by the time difference of the two contacts, was added. Ankle instead of toe coordinates were chosen, because the STIMO dataset did not provide any toe coordinates.

$$step_r = \Delta x + v_t \cdot \frac{\Delta t}{f}$$

$$\Delta x = x_r(IC_r) - x_l(IC_l), \quad \Delta t = IC_r - IC_l$$

3) *Stride Length (left and right):* Stride length was calculated analogically to the step length; but by considering the distance travelled between the two initial contacts of the same foot. Initially, the aim of using those two features was to use the step length mean and the stride length variance among gait cycles.

4) *Step height (left and right):* Step height has previously shown potential to help identifying gait abnormalities [7]. For the present case, SCI subjects lack the control and force to properly push off at the end of the stance phase, resulting in lower step height than in healthy subjects. This feature was implemented by taking the difference between maximum and minimum value of the vertical ankle coordinate during a gait cycle.

5) *Max knee angle (left and right):* This feature (θ_{knee}) was implemented by calculating the maximal angle between shank and thigh throughout the gait cycle. We decided to look at the knee angle, since based on the videos this joint differed the most among the two groups. Furthermore, as we do not have the toe coordinates for the SCI subject, calculating the ankle angle would have been more difficult and probably less precise in the end.

$$\theta_{knee} = 180° - tan(\frac{thigh}{shank})$$

$$shank = x_{knee} - x_{ankle}$$

$$thigh = x_{hip} - x_{knee}$$

6) *Correlation of knee and ankle oscillation (left and right):* This feature allows to also assess the intra-limb coordination. It was implemented by calculating the correlation coefficients of the ankle and knee coordinates throughout an entire gait cycle.

7) *Interlimb coordination (left and right):* This feature allows to investigate inter-limb coordination. It was calculated by taking the time difference between initial contact of one foot and the terminal contact of the opposite foot.

8) *Endpoint velocity (left and right):* Investigating the endpoint trajectory is another useful feature for gait characterization. This feature (vel_{ank}) was implemented by calculating the maximum derivative of the ankle coordinate in the direction of motion (x_anik). The ankle coordinates were used, as the toe coordinates were not available for the SCI dataset.

$$vel_{ank} = max(\frac{d}{dt}x_{ank})$$

9) *Whole limb angular velocity (left and right):* This feature served as another velocity measure based on the derivative of the angular position of the thigh. The thigh angle was considered with respect to the ground in the sagittal plane.
10) *Gait stability:* As a measure of stability ($stab$) the following implementation was chosen: A center of mass (COM) was defined as being the mid-point between both hip markers (y_{hip}). Then for each gait cycle the lateral displacement from the mid-line was integrated (using the matlab function trapz) after having subtracted its mean. Accordingly to this definition, healthy balanced gait should result in a gait stability value of zero.

$$stab = trapz(COM(t) - \overline{COM}_{cycle})$$

$$COM(t) = y_{rhip}(t) - y_{lhip}(t)$$

11) *Ratio between forward and lateral movement (left and right):* This features serves as another measure of gait stability including also the forward movement. This feature was calculated by taking the difference between maximal and minimal ankle coordinate value in the forward and in the lateral direction and then taking their ratio.

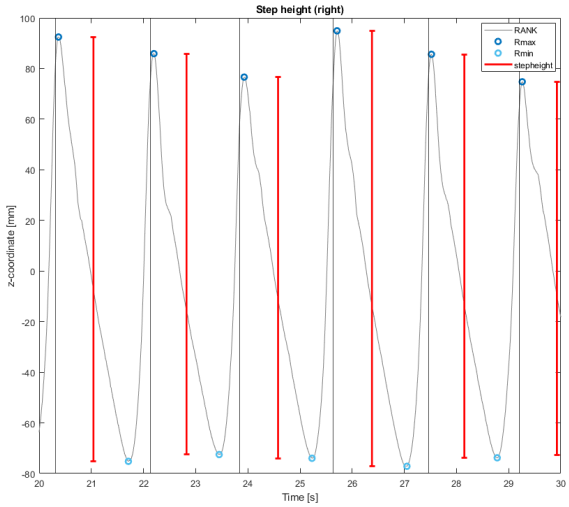

Figure 3: Example of the step height algorithm implementation

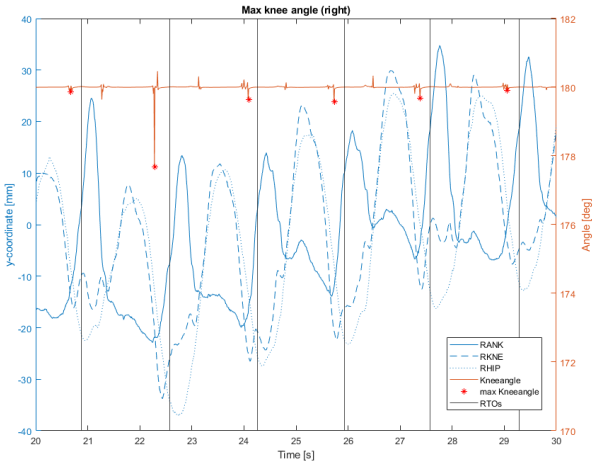

Figure 4: Example of the maximum knee angle algorithm implementation

C. EMG preprocessing

EMG raw signal is noisy and difficult to extract information from, if not preprocessed. From the multiple preprocessing strategies, three were compared for this project:

- **Butterworth low-pass filter.** A 4th order low-pass filter was used twice, with 10Hz as cut-off frequency.
- **Moving RMS.** A moving root-mean-square is computed, using a sliding window of 250 samples, corresponding to 125ms. This window size was found to be the best for the given sampling frequency.
- **Linear envelope.** First, the signal is band-passed between 10 and 2000Hz, then rectified (absolute value). High-frequency content is removed with a 30Hz high-pass filter and a bandstop filter was applied (between 30Hz to 70Hz). Finally, the upper RMS envelope of the transformed signal is computed, using a sliding window of 25ms.

It should be noted that for the first two approaches, the raw signal was first centered by subtracting the mean, and then rectified to its absolute value. The three methods are depicted in Figure 5, for all recorded muscles of the left and right legs. When comparing the preprocessing approaches to the raw data, the linear envelope is more robust against noise and represents better the

muscle activity. Hence, it will be the method used for EMG feature extraction. From this figure it can also be appreciated that the Tibial Anterioris (TA), the Medial Gastrocnemius (MG), the Soleus (Sol) and the Semitendinosus (ST) muscles are those which follow a clear pattern of activation and deactivation, leading to a clear linear envelope. Only those four muscles will be included in the study.

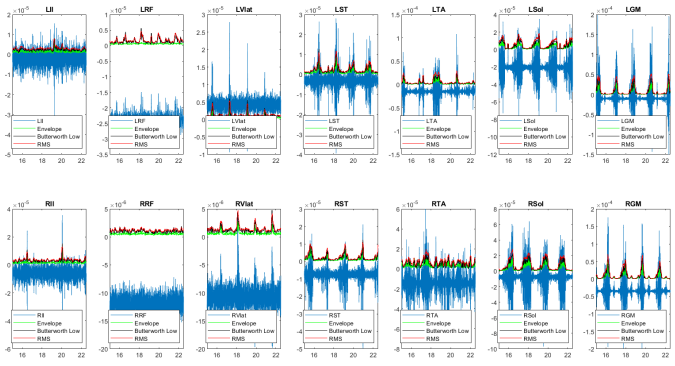

Figure 5: Comparison of the three EMG preprocessing methods for all muscles presents in a healthy dataset.

D. EMG feature extraction

Electromyography (EMG) is often used together with kinematic signals to link the activation of each muscle to the different phases of the gait cycle [8] or the different gait patterns. From this data, the following features were computed for each gait cycle:

Time domain

1) *Mean activity:* For each gait cycle, the mean of the linear envelope was computed to estimate the mean activation of the muscle during the gait cycle.
2) *RMS:* The Root Mean Square (RMS) reflects the muscle's physiological activity during contraction [9], as it represents the square root of the average power of the EMG signal for a given period of time. Here, RMS is computed for each gait cycle using raw data.
3) *Burst duration:* First, the Teager-Kaiser energy (TKE) operator was used to detect onset and offset times from the EMG signal. This method had previously shown better accuracy in finding onset times than conventional threshold methods. A special preprocessing of the signal is performed, as suggested in [10]. Once onset and offset are detected for each gait cycle,- corresponding to the beginning and end of bursts-, they are subtracted pairwise and divided by the sampling frequency to obtain the burst duration in seconds.

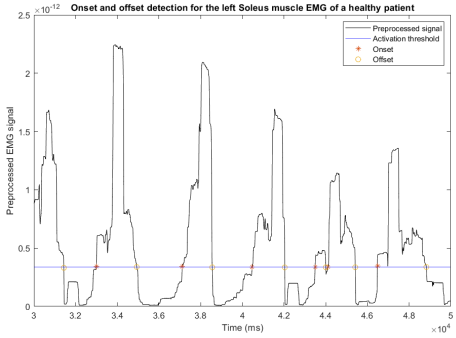

Figure 6: Example of the burst detection algorithm implemented on healthy data (AML_02_1.mat) for the left Soleus muscle.

Frequency domain

All the extracted EMG features until now were in the time domain. Frequency content of the signal may provide distinct information, which can help for the PCA analysis. Power spectral density can be used as a method for frequency domain analysis, but computing it along the whole gait cycle is not recommended, as it would only be valuable if the muscle was in steady state of activation during the whole period [11]. The common solution is to compute the power spectrum of the signal using a smooth overlapping moving window. Here, the Welch method was used for the power spectra calculation, with a Hamming window of 100 samples (50ms) and a 50% overlapping. Previous work found the median power frequency and the total power at low frequencies to be very helpful features to differenciate between gait abnormalities [12]. Thus, both are used here as EMG features.

1) *Power at low frequencies:* The total power at low frequencies is calculated as a sum of the first 10% of the power spectrum.

2) *Median power frequency:* As PCA requires feature rescaling, the exact frequency at which the power spectrum reaches its median is not relevant. Here, after computing the power spectrum for each gait cycle, the index corresponding to the median power is used as a feature.

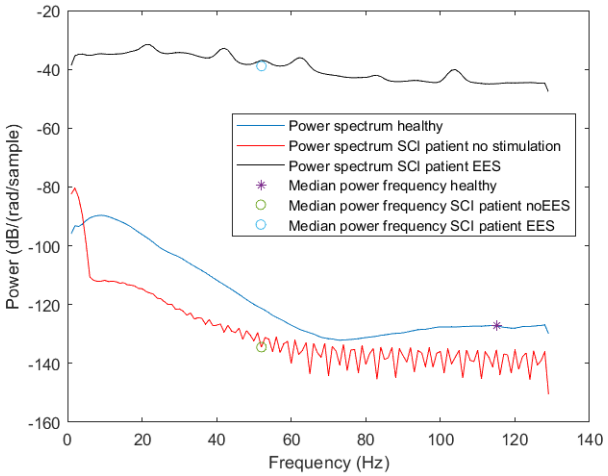

Figure 7: Power spectral density and median power frequency for three studied conditions. As it can be observed, the power for EES stimulation conditions is very high compared to the other two conditions, whereas the median power frequency allows to discriminate between healthy and SCI data.

E. PCA implementation

Principal Component Anlaysis (PCA) is a method for data dimensionality reduction, which projects the data into uncorrelated features. In the present case, it is used to visualize the data in a new feature space and discriminate between the different subject conditions (both for the healthy and the SCI subjets).

Before performing PCA analysis, all features were standardized to avoid biasing the principal components towards the features with higher variance. PCA analysis was computed using two different input structures:

- Using the result of each gait cycle as a sample.
- Using the mean and variance of the computed features over 5 gait cycles.

Using 5 cycles as an input sample was tested to increase the robustness of the features, even if it reduced the sample size. Also, this approach allows to use the variance of the computed parameters as an extra feature and see if the results improve. The complete features tested for the PCA can be found at Appendix A.

The *pca* function from Matlab was implemented, which returns the principal component coefficients (loadings), a squared matrix in where each column corresponds to a principal component (in variance descending order) and each row to the weight of the input features in the principal component. As the principal components are eigenvectors of the features covariance matrix, each one of them is associated to an eigenvalue. By dividing the eigenvalue by the sum of eigenvalues we can obtain the relative variance explained by each component. The *pca* function also provides the explained variance for each principal component.

When implementing PCA, the size of the input matrix was N x D, where N is the total number of detected gait cycles for the nine datasets, and D the number of computed features for right and left legs. The input matrix size was different for the two different approaches used.

III. Results and Discussion

A. Final approach and selected features

Different PCA implementations were tested, using one cycle or 5 as imput samples and including different computed features. Finally, a single cycle per sample gave the best results, when including features present in Table I. As it can be observed, only Soleus and Tibialis Anterioris muscles were kept, as adding other muscles did not improve PCA performance. Those were chosen to have at least two antagonist muscles which are active in different gait phases. Left and right legs were both used for the analysis, as their correlation and coordination is important for gait analyisis, so by having all parameters for both legs PCA could better find differences in gait patterns for the different datasets.

The results of the PCA using the final selected features, but using the mean and variance of five gait cycles as an unique sample can be found in

Features from kinematic data
R stride length
L stride length
R step height
L step height
R max knee ankle
L max knee ankle
R corr KA
L corr KA
R endpoint vel
L endpoint vel
R ratio fl
L ratio fl
gait stability
R interlimb coord
L interlimb coord
R limb ang vel
L limb ang vel
Features from EMG data
R RMS TA
L RMS TA
R RMS Sol
L RMS Sol
R burst dur TA
L burst dur TA
R burst dur Sol
L burst dur Sol
R medfreq TA
L medfreq TA
R medfreq Sol
L medfreq Sol

Table I: Final selected features for the PCA analysis

the Appendix B. As it can be expected, it is harder to discriminate between the different conditions when using this approach, particularly between the healthy cluster and the non-stimulated patient.

A total of nine different datasets were used for the PCA (Fig. 12). The first three datasets correspond to the recordings of one healthy subject that went through conditions 1 to 3 as explained in the introduction. The following three datasets represent the same experimental protocol for another healthy subject. Finally, the last three datasets represent conditions 4 to 6 for a STIMO participant with SCI.

The final features were selected to be the most independent possible, while covering most of the important parameters for the analysis of locomotion. For instance, gait cycle duration was correlated to inter-limb coordination, step length to stride length, and mean muscle activity and low-frequency power to the RMS of the EMG signal. From those features, stride length, interlimb coordination and RMS were kept.

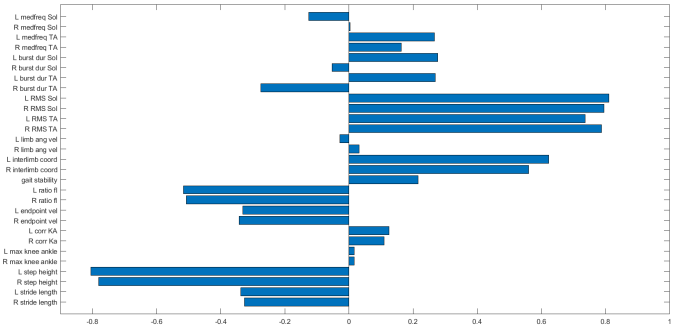

Figure 8: Loadings for all features from the first principal component.

B. Characterization of locomotion description

By comparing the feature loadings for the first principal component of the PCA to the videos for the different subjects, it is possible to characterize the gait for the different conditions.

The SCI patient walking without stimulation shows toe-walking, with approximately regular step length. Movements are short but smooth. He cannot push his body forward, so most of the force and the movement is carried out by the body weight support and the movement of the treadmill below the patient's feet. From the video, it can be expected that the muscles are poorly activated and that the stance and swing phases are fictive (the patient is not using the gravity to recover energy during walking as he cannot sustain his own weight). Also, the knee flexion is reduced, which leads to a reduced step height.

For the SCI patient with EES, movements are more stiff. This stiffness is possibly due to spasticity and co-contraction from the stimulation. He uses his muscles for generating the gait patterns instead of letting himself be carried by the weight support, as shown previously. As the patient has to re-learn how to walk, the gait is not regular and the movements exaggerated. For example, the knee angles show more flexion compared to both

the healthy gait or the absence of stimulation condition, and the hip shows wide lateral motion. Also, the patient looses stability compared to the non-stimulation condition due to the fact that he is not constantly holding the sidebars of the treadmill. Because of muscle weakness and fatigue, the patient sometimes drags his feet and needs to hold the sidebar to re-start walking again.

Having described the two SCI walking conditions, it is reasonable that step height and EMG RMS features for two antagonist muscles (responsible for plantar and dorsal flexion of the ankle) are the ones that best explain the variance in the data. Step height is affected by the lack of stance-swing phases and the differences on knee flexion. The patient without stimulation cannot use his muscles to elevate his leg, while when receiving stimulation his walking pattern goes towards the recovery of the stance-swing switching, but the stiff movements are very irregular. For the latest condition there is a combination between gait cycles with exaggerated knee flexion and ankle elevation, as well as dragging of the feet. As already observed in animal studies, before regaining the ability to step, the activation of the extensor muscle is not enough to lift the foot, resulting in dragging. Eventually, the activity of the extensor is regained and the foot is overlifted to avoid dragging [13]. The present studied patient shows a combination of both.

RMS of the EMG signal is a measure of the muscle activation during the gait cycle. For healthy gait, the pattern of activation is very clear, with an alternation of TA and Sol during the different gait phases. Without EES, the patient can only rely on residual motor function, and the activation of the muscles is too weak. When applying EES, the recorded muscle activity corresponds to the patterns of stimulation at the spinal cord and it is expected that there is some co-activation of antagonist muscles as the activation patterns might not be as selective as the natural human network. From the results in Figure 11, it seems that muscle activity for the EES data is five orders or magnitude above than for the rest of the conditions. Those results suggest that this feature may not accurately discriminate between the physical conditions of the patient, and only differ depending of the stimulation applied.

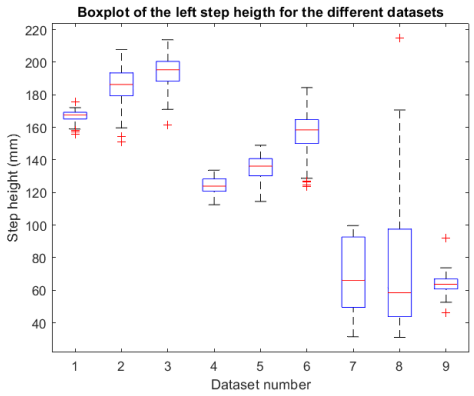

Figure 9: Boxplot of the left step height in function of the different included datasets.

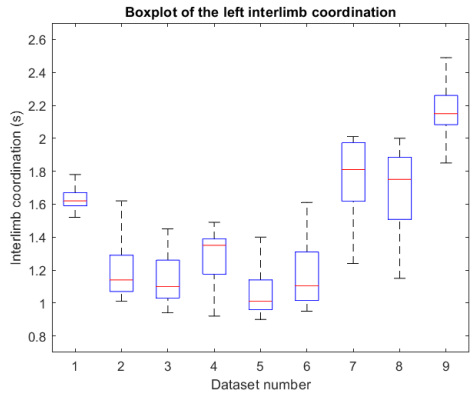

Figure 10: Boxplot of the left interlimb coordination feature in function of the different included datasets.

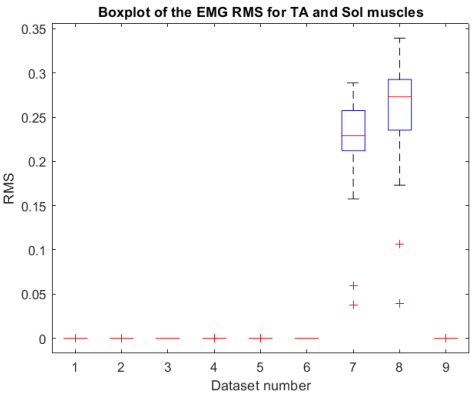

Figure 11: Boxplot of the left Soleus muscle EMG signal RMS in function of the different included datasets.

C. PCA for discrimination of abnormal gait patterns

As seen in Fig. 12, PCA allows to distinguish the different analysed gaits. In Fig. 13, the centroid of each dataset in the space spanned by component 1 and 2 was computed. Then, distances between the groups can be evaluated, which is a measurement for the similarity between clusters. The complete results can be found in tab. II. As already noticed in section III-B, the gait of the patient without EES is closer to normal gait than with EES.

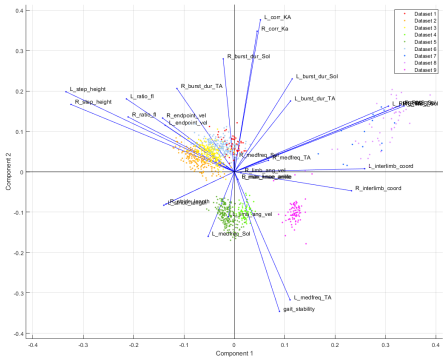

Figure 12: Data projection into the two first components of the PCA.

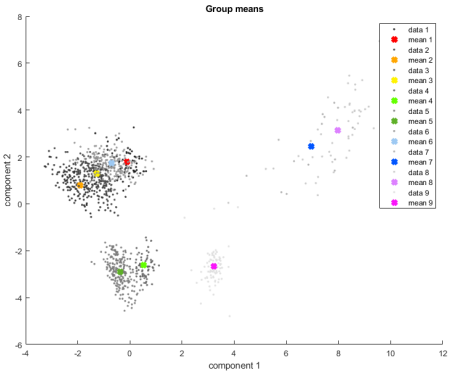

Figure 13: Computed Group means. The associated distances between the different groups can be found in Tab. II

To understand if using the two principal components was enough, the explained variance by all components was assessed, and the results are in Appendix C. PC1 only explains 20% of the variance in the data and PC2 about 17%. This suggests using more components can also be interesting. For instance, PC 3 has also been used to visualize clusters distribution and the results are found in Figure 14. Those three components explain about half the variance in the data. Using only components 2 and 3, the non-stimulated cluster is closer to only two clusters of healthy data, while there is no clear cluster separation between the EES data and the other healthy data clusters.

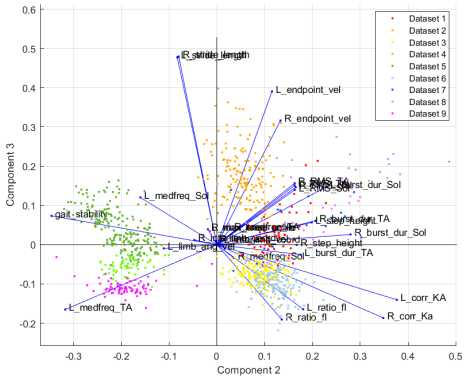

Figure 14: Data projection into the second and third components of the PCA.

D. Limitations of the gait analysis method

As already explained, the videos show a dragging walking behaviour for the patient, which is why in the present study the division between stance and swing phases of the gait cycle was not performed. However, the two phases are very important for gait analysis, and ignoring this division may have a negative effect in our analysis. For instance, for features related to muscle activation, knowing if the activation happens during stance or swing phases is very relevant.

Furthermore, the kinematic feature gait stability used in this work is highly modulated by the body weight support mechanism of the injured subject. Given that this mechanism stabilizes the subject the difference in the values between healthy and injured individual does not vary significantly. Nevertheless, in theory we believe that this feature would be an important parameter to distinguish healthy from disordered gait, especially because when looking at the videos, the difference in gait stability is immediately visible.

Dataset	1	2	3	4	5	6	7	8	9
1	0	2.06	1.26	4.45	4.7	0.59	7.1	8.19	5.56
2	2.06	0	0.81	4.19	4.01	1.54	9.02	10.15	6.19
3	1.26	0.81	0	4.31	4.3	0.73	8.31	9.42	5.99
4	4.45	4.19	4.31	0	0.95	4.54	8.2	9.4	2.69
5	4.7	4.01	4.3	0.95	0	4.67	9.1	10.31	3.61
6	0.59	1.54	0.73	4.54	4.67	0	7.69	8.78	5.91
7	7.1	9.02	8.31	8.2	9.1	7.69	0	1.21	6.36
8	8.19	10.15	9.42	9.4	10.31	8.78	1.21	0	7.5
9	5.56	6.19	5.99	2.69	3.61	5.91	6.36	7.5	0

Table II: Euclidian distances of the mean values of the different dataset in the two first component space

Additionally, it remains to say that the experimental conditions between healthy and injured subject are not the same, which could possibly bias direct comparison between the two. While for the healthy subject, the treadmill speed is much higher reaching up to 5 km/h, for the STIMO participant this speed only reaches 2 km/h at most. Moreover, condition 3 includes inclination of the treadmill, however, there is no comparable condition for the SCI subject. Therefore, potential differences in the gait analysis between the two groups, could also appear due to the different experimental protocols.

Moreover, the gait cycle extraction, as well as the majority of the kinematic feature calculations, were based on the ankle coordinates. This is due to the fact, the toe coordinates were not recorded for the STIMO participant. Nevertheless, for specific features (i.e. endpoint velocity, ankle angle etc.) having a toe marker might have let to better distinguishable results between the two groups. Comparing the videos, it seems that the toe movements differ significantly more than the ankle movements between healthy and SCI subjects.

Also, because of the stimulation pattern, EMG frequency content might be altered for the SCI patients, compared to the healthy. This feature selection might not be a fair way for discriminating between gait conditions. However, it can still allow the discrimination between the healthy gait and the SCI patient without stimulation.

Last but not least, given that the entire PCA analysis is based on the gait cycle extraction method, a quantitative analysis of the latter could be done in a future work. To do so, one could look at the video and note the exact frames where initial and terminal contact are occurring and then calculate the difference to our implementation. These results, might then give a better idea on possible further improvements of the method.

E. Contribution and future work

The results suggest that PCA allows identifying the gait parameters which are more relevant for a specific gait disorder, as well as discriminating between different gait patterns. Depending on the abnormal parameters, it is possible to gain an insight on the motor control circuits affected by the condition.

This type of study is not only useful to monitor SCI patients, but has the potential to be used in many gait disorders. For instance, PCA analysis has been used to monitor locomotion recovery in stroke patients [14], or for early stage Parkinson diagnosis, as at this point there are no existing specific biomarkers [15].

In the future, those type of studies could be standardized, using validation methods for the feature extraction and automatic feature selection approaches. This would allow the comparison of results between studies, groups and laboratories, as well as to enable patients to follow their rehabilitation treatment in different centers if needed.

IV. CONCLUSIONS

The present work proves PCA to be a very powerful tool for the study of gait disorders and the assessment of gait restoration. Without this type of analysis, evaluating the progress of rehabilitation and fine tuning the treatment individually for each patient would not be possible.

REFERENCES

[1] G. Courtine and M. V. Sofroniew, "Spinal cord repair: advances in biology and technology," *Nature Medicine*, vol. 25, no. 6, pp. 898–908, Jun. 2019. [Online]. Available: https://www.nature.com/articles/s41591-019-0475-6

[2] U. N. S. C. I. S. Center, "Spinal cord injury facts and figures at a glance," 2020. [Online]. Available: https://www.nscisc.uab.edu/Public/Facts%20and%20Figures%202020.pdf

[3] W. H. Organization. (2013) Spinal cord injury. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/spinal-cord-injury

[4] E. M. Moraud, M. Capogrosso, E. Formento, N. Wenger, J. DiGiovanna, G. Courtine, and S. Micera, "Mechanisms Underlying the Neuromodulation of Spinal Circuits for Correcting Gait and Balance Deficits after Spinal Cord Injury," *Neuron*, vol. 89, no. 4, pp. 814–828, Feb. 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0896627316000106

[5] F. B. Wagner, J.-B. Mignardot, C. G. Le Goff-Mignardot, R. Demesmaeker, S. Komi, M. Capogrosso, A. Rowald, I. Seáñez, M. Caban, E. Pirondini, M. Vat, L. A. McCracken, R. Heimgartner, I. Fodor, A. Watrin, P. Seguin, E. Paoles, K. Van Den Keybus, G. Eberle, B. Schurch, E. Pralong, F. Becce, J. Prior, N. Buse, R. Buschman, E. Neufeld, N. Kuster, S. Carda, J. von Zitzewitz, V. Delattre, T. Denison, H. Lambert, K. Minassian, J. Bloch, and G. Courtine, "Targeted neurotechnology restores walking in humans with spinal cord injury," *Nature*, vol. 563, no. 7729, pp. 65–71, Nov. 2018.

[6] A. Rowald, S. Komi, R. Demesmaeker, E. Baaklini, S. D. Hernandez-Charpak, E. Paoles, H. Montanaro, A. Cassara, F. Becce, B. Lloyd, T. Newton, J. Ravier, N. Kinany, M. D'Ercole, A. Paley, N. Hankov, C. Varescon, L. McCracken, M. Vat, M. Caban, A. Watrin, C. Jacquet, L. Bole-Feysot, C. Harte, H. Lorach, A. Galvez, M. Tschopp, N. Herrmann, M. Wacker, L. Geernaert, I. Fodor, V. Radevich, K. Van Den Keybus, G. Eberle, E. Pralong, M. Roulet, J.-B. Ledoux, E. Fornari, S. Mandija, L. Mattera, R. Martuzzi, B. Nazarian, S. Benkler, S. Callegari, N. Greiner, B. Fuhrer, M. Froeling, N. Buse, T. Denison, R. Buschman, C. Wende, D. Ganty, J. Bakker, V. Delattre, H. Lambert, K. Minassian, C. A. T. van den Berg, A. Kavounoudias, S. Micera, D. Van De Ville, Q. Barraud, E. Kurt, N. Kuster, E. Neufeld, M. Capogrosso, L. Asboth, F. B. Wagner, J. Bloch, and G. Courtine, "Activity-dependent spinal cord neuromodulation rapidly restores trunk and leg motor functions after complete paralysis," *Nature Medicine*, vol. 28, no. 2, pp. 260–271, Feb. 2022.

[7] A. Manickam and M. D. Gardiner, "Gait assessment in general practice," *Australian Journal for General Practitioners*, vol. 50, no. 11, pp. 801–806, November 2021. [Online]. Available: https://www1.racgp.org.au/ajgp/2021/november/gait-assessment-in-general-practice

[8] A. Stefano, J. Burridge, V. Yule, and R. Allen, "Effect of gait cycle selection on emg analysis during walking in adults and children with gait pathology," *Gait Posture*, vol. 20, no. 1, pp. 92–101, 2004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0966636203000997

[9] T. Fukuda, J. Echeimberg, J. Pompeu, P. Lucareli, S. Garbelotti Junior, R. Gimenes, and A. Apolinário, "Root mean square value of the electromyographic signal in the isometric torque of the quadriceps, hamstrings and brachial biceps muscles in female," *Applied Research*, vol. 10, pp. 32–39, 01 2010.

[10] S. Solnik, P. DeVita, P. Rider, B. Long, and T. Hortobágyi, "Teager-kaiser operator improves the accuracy of emg onset detection independent of signal-to-noise ratio," *Acta of bioengineering and biomechanics*, vol. 10, pp. 65–68, 2008.

[11] W. Rose, "Kaap686 mathematics and signal processing for biomechanics electromyogram analysis," 2011.

[12] R. E. Singh, K. Iqbal, G. White, and J. K. Holtz, "A review of emg techniques for detection of gait disorders," in *Artificial Intelligence*, M. A. Aceves-Fernandez, Ed. Rijeka: IntechOpen, 2019, ch. 2. [Online]. Available: https://doi.org/10.5772/intechopen.84403

[13] M. Capogrosso, T. Milekovic, D. Borton, F. Wagner, E. M. Moraud, J. B. Mignardot, N. Buse, J. Gandar, Q. Barraud, D. Xing, E. Rey, S. Duis, Y. Jianzhong, W. K. D. Ko, Q. Li, P. Detemple, T. Denison, S. Micera, E. Bezard, J. Bloch, and G. Courtine, "A brain-spine interface alleviating gait deficits after spinal cord injury in primates," *Nature*, vol. 539, pp. 284–288, 11 2016.

[14] I. Milovanović and D. B. Popović, "Principal Component Analysis of Gait Kinematics Data in Acute and Chronic Stroke Patients," *Computational and Mathematical Methods in Medicine*, vol. 2012, p. 649743, Feb. 2012, publisher: Hindawi Publishing Corporation. [Online]. Available: https://doi.org/10.1155/2012/649743

[15] D. Lukšys, D. Jatužis, R. Kaladytė-Lokominienė, R. Bunevičiūtė, A. Sawicki, and J. Griškevičius, "Differentiation of gait using principal component analysis and application for parkinson's disease monitoring," in *2018 International Conference BIOMDLORE*, 2018, pp. 1–4.

APPENDIX A.
COMPLETE LIST OF FEATURES USED AS INPUTS FOR THE PCA.

Features from kinematic data	
Feature Name	Feature description
R cycle duration	Gait cycle duration for the right foot
L cycle duration	Gait cycle duration for the left foot
R step length	Right gait cycle step length
L step length	Left gait cycle step length
R stride length	Right gait cycle stride length
L stride length	Left gait cycle stride length
R step height	Right gait cycle step height
L step height	Left gait cycle step height
R max knee ankle	Right gait cycle knee angle
L max knee ankle	Left gait cycle knee angle
R corr KA	Gait cycle correlation between the right knee and ankle
L corr KA	Gait cycle correlation between the left knee and ankle
R endpoint vel	Endpoint velocity based on the right ankle coordinate
L endpoint vel	Endpoint velocity based on the left ankle coordinate
R ratio fl	Ratio between forward and lateral movement based on the right ankle coordinate
L ratio fl	Ratio between forward and lateral movement based on the left ankle coordinate
Gait stability	Gait stability (lateral) based on hip markers
R interlimb coord	Interlimb coordination between right initial contact and left terminal contact
L interlimb coord	Interlimb coordination between left initial contact and right terminal contact
R limb ang vel	Whole limb angular velocity of the right thigh
L limb ang vel	Whole limb angular velocity of the left thigh

Features from EMG data	
Feature Name	Feature description
R RMS TA	RMS for the right EMG of the TA muscle
L RMS TA	RMS for the left EMG of the TA muscle
R RMS ST	RMS for the right EMG of the ST muscle
L RMS ST	RMS for the left EMG of the ST muscle
R RMS MG	RMS for the right EMG of the MG muscle
L RMS MG	RMS for the left EMG of the MG muscle
R RMS Sol	RMS for the right EMG of the Sol muscle
L RMS Sol	RMS for the left EMG of the Sol muscle
R Mean TA	Mean activity of the right TA muscle
L Mean TA	Mean activity of the left TA muscle
R Mean ST	Mean activity of the right ST muscle
L Mean ST	Mean activity of the left ST muscle
R Mean MG	Mean activity of the right MG muscle
L Mean MG	Mean activity of the left MG muscle
R Mean Sol	Mean activity of the right Sol muscle
L Mean Sol	Mean activity of the left Sol muscle
R burst dur TA	Mean burst duration for the right TA muscle
L burst dur TA	Mean burst duration for the left TA muscle
R burst dur ST	Mean burst duration for the right ST muscle
L burst dur ST	Mean burst duration for the left ST muscle
R burst dur MG	Mean burst duration for the right MG muscle
L burst dur MG	Mean burst duration for the left MG muscle
R burst dur Sol	Mean burst duration for the right Sol muscle
L burst dur Sol	Mean burst duration for the left Sol muscle
R lowfreq pow TA	Signal power at low frequencies for the right TA muscle
L lowfreq pow TA	Signal power at low frequencies for the left TA muscle
R lowfreq pow ST	Signal power at low frequencies for the right ST muscle
L lowfreq pow ST	Signal power at low frequencies for the left ST muscle
R lowfreq pow MG	Signal power at low frequencies for the right MG muscle
L lowfreq pow MG	Signal power at low frequencies for the left MG muscle
R lowfreq pow Sol	Signal power at low frequencies for the right Sol muscle
L lowfreq pow Sol	Signal power at low frequencies for the left Sol muscle
R medfreq TA	Median power frequency for the right TA muscle
L medfreq TA	Median power frequency for the left TA muscle
R medfreq ST	Median power frequency for the right ST muscle
L medfreq ST	Median power frequency for the left ST muscle
R medfreq MG	Median power frequency for the right MG muscle
L medfreq MG	Median power frequency for the left MG muscle
R medfreq Sol	Median power frequency for the right Sol muscle
L medfreq Sol	Median power frequency for the left Sol muscle

Note: Those are the features tested for the final approach (1 gait cycle per sample). For the 5 cycles per sample approach, the mean and variance along the cycles of the described features was used instead.

APPENDIX B.
RESULTS FOR THE PCA USING 5 GAIT CYCLES' MEAN AND VARIANCE AS FEATURES.

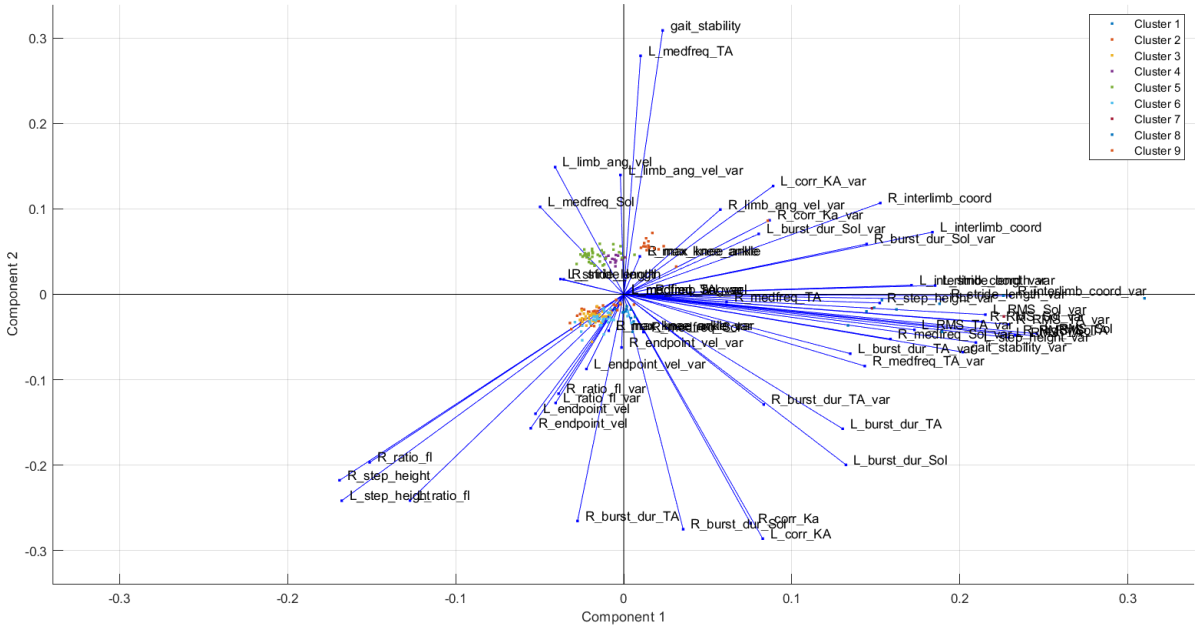

Figure 15: Data projection into the two first components of the PCA using the mean and variance over 5 gait cycles as a single sample.

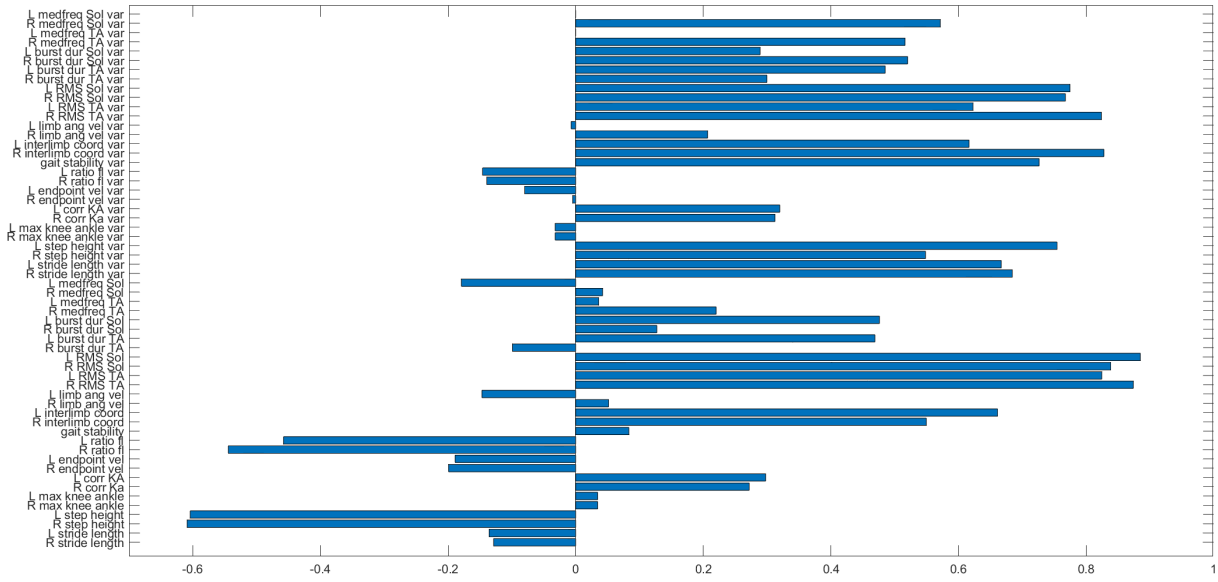

Figure 16: Feature loading for the 5 cycles/sample approach, using for each feature its mean and variance.

APPENDIX C.
EXPLAINED VARIANCE BY ALL COMPONENTS IN THE PCA ANALYSIS.

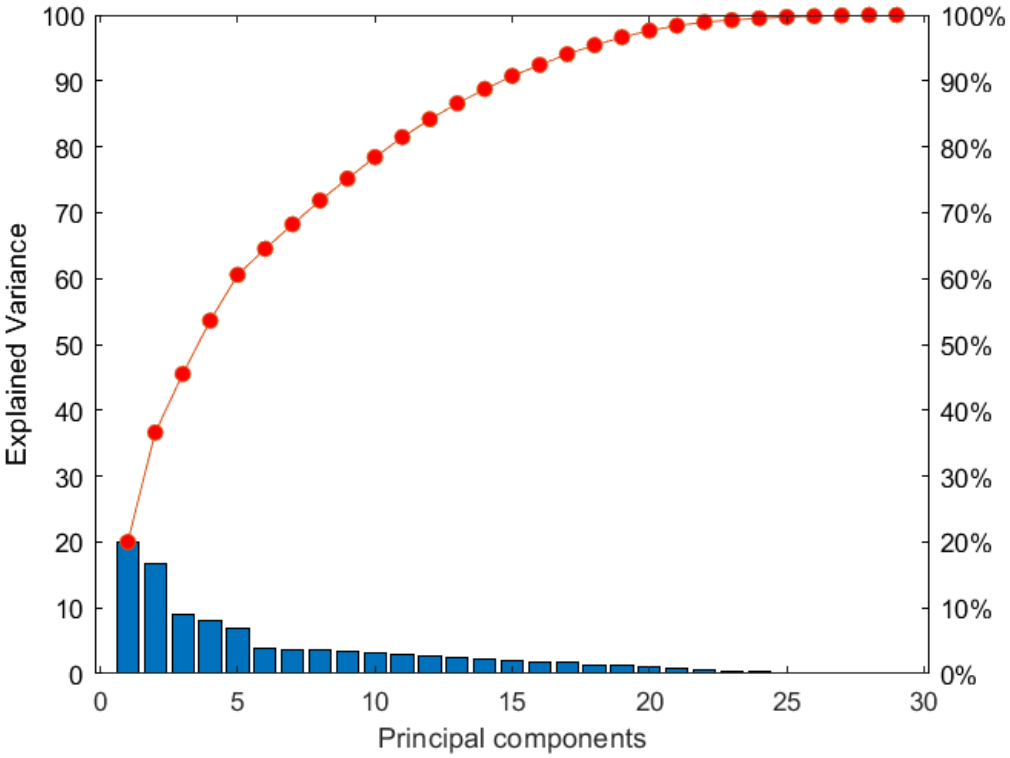

Figure 17: Explained data variance by all principal components for the final feature selection