Feature selection of postural summary statistic scores based on principal component analysis in Parkinson's disease

L. Rocchi[1], L. Chiari[1], A. Cappello[1] & F. B. Horak[2]
[1]*Department of Electronics Computer Science & Systems, University of Bologna, Italy*
[2]*Neurological Sciences Institute, Oregon Health & Science University, Portland, Oregon, U.S.A.*

Abstract

The present study addresses the challenge of identifying the features of the centre of pressure (CoP) trajectory that are most sensitive to postural performance, to progress in the process of transforming CoP data into useful information and to promote standardization in quantitative posturography. For this purpose, we singled out the features that characterize postural sway in subjects with Parkinson's disease (PD) with and without levodopa (*on* and *off* states). The feature selection was performed using principal component analysis. Results suggest that CoP in subjects with PD can be primarily characterised by four parameters, in both *off* and *on* states, describing: the size of the path over the support surface; the principal sway direction; and the shape and bandwidth of the power spectral density plot. The similarity of the present results with a previous study that considered young healthy subjects, allows us to define more confidently the minimum set of measures to recommend for specific applications as optimal descriptor of CoP sway in quiet stance.
Keywords: principal component analysis, posture, Parkinson's disease.

1   Introduction

Body posture is the output of complex interactions between central nervous system control mechanisms and the musculo-skeletal actuators acting against the support surface. Because of its complexity body posture is challenging to

measure with simple methods, but simple methods are necessary in neurological, orthopaedic and clinical practice. Quantitative posturography, i.e. the measurement of forces exerted against ground from a force platform during quiet stance is commonly used to quantify postural steadiness both in research and in the clinic. It typically focuses on the properties of the center of pressure (CoP). The CoP is the point location of the ground reaction force vector [1] and reflects the sway of the body and forces used to maintain the centre of gravity within the support base [2]. This single variable reflects both the balance controlling process and movements of the center of mass of the entire body and thus provides a single global measure of postural control [3]. However CoP analysis potentially generates a dataset (the sway-measures) that can be difficult to manage [4]. The most common CoP analysis, particularly in clinical practice, estimates statistical properties of CoP trajectory, considered as a stationary signal, in the time and frequency domains [2]. These sway-measures are usually referred to as postural *summary statistic scores* [4, 5].

In a previous study [4] we addressed the challenge of identifying the features of the CoP trajectory that are most sensitive to postural performance, with the aim of avoiding redundancy and allowing a straightforward interpretation of the results. In this light, after principal component analysis (PCA) for feature selection considering a healthy young population, we identified four CoP summary statistic scores as the most characterizing the CoP trajectory in the horizontal plane. The selected postural summary statistic scores identified the size of the CoP path over the support surface, the principal sway direction, the shape and bandwidth of the power spectral density of the CoP signal. On the basis of the results we suggested guidelines for the choice of postural measures to use, to the aim of promoting standardization in quantitative posturography.

In the present study we extended the feature selection of postural summary statistic scores to quiet stance data of subjects with Parkinson's disease (PD). PD is a neurodegenerative disease with main symptoms related to the motor system, and associated with a reduced production of dopamine, a neurotransmitter in the basal ganglia. In particular, postural problems in patients with PD become increasingly severe as the disease progresses, despite dopamine replacement therapy with levodopa [6, 7]. In previous studies [8, 9] we quantified postural sway in subjects with PD and determined the effects of PD and levodopa treatment on postural control during quiet stance. We concluded that abnormality in postural sway of subjects with PD may be clearly detected by quantitative posturography. In addition we found that CoP measures are highly sensitive both to the disease and to the treatment with levodopa that increases postural sway abnormalities. The results verified that quantitative posturography may be a useful adjunct to clinical measures in patients with PD.

The aim of the present study is to progress in the process of transforming CoP data into useful information and to promote standardization in quantitative posturography. For this purpose, we intend to single out the features that characterize postural sway in a pathological population with and without a specific treatment and to compare the results with previous results on healthy young subjects.

2   Methods

2.1 Subjects

Twenty-one subjects with Parkinson's disease were included in the study, after giving informed consent in accordance with the Oregon Health & Science University, Internal Review Board regulations for human subject studies. All subjects with PD could stand independently and were responsive to levodopa, the medication typically given to subjects with PD. Medical examination and history showed they did not have other pathological conditions that could affect postural control. Subjects with PD were tested in two different states: in the practical *off* state (with a levodopa wash-out of at least 12 hours) and in the *on* state, approximately one hour after the assumption of levodopa in the patients' usual dosage. In both states, clinical scores for the subjects with PD were obtained using the Unified Parkinson's Disease Rating Scale (UPDRS) – Motor Section (III).

2.2 Experimental session

To measure postural sway, subjects stood on a dual force plate platform, with one foot on each force plate, for 3 sequential trials in the *off* state, and 3 sequential trials in the *on* state. Subjects were instructed to maintain an upright standing position, with crossed arms, eyes open with gaze straight ahead at an art poster, and feet in a free position. Four vertical forces were recorded from each force plate at 480 Hz for 60 seconds, and data were filtered by a 30th order low pass FIR digital filter (cut off frequency 10 Hz). The total body CoP was computed from the vertical forces. CoP migration in the horizontal plane was represented by the coordinates in the anteroposterior or mediolateral directions (mono-dimensional time-series), and by the bidmensional time-series, computed from the coordinates mentioned above.

The first of the 3 trials in each state was performed to allow subjects to get comfortable with the motor task. Then, the first valid trial of the following two was considered for further analyses.

2.3 Postural summary statistic scores

Fourteen different postural summary statistic scores were computed from the CoP time series. Ten of these were computed directly from the bidimensional time-series. They quantify the major properties of the CoP time-series in the time and frequency domains. The remaining four sway-measures also characterize the bidimensional CoP migration. They estimate planar characteristics, such as the area covered by the CoP and the principal sway direction [2, 4, 8, 10]. A list and a brief description of the sway-measures computed in the present study are reported in Table 1.

Summary statistic scores are frequently applied in clinical practice, being easy to compute and relatively straightforward to interpret [11, 12].

Table 1:   Postural summary statistic scores: acronyms and brief descriptions.

Acronym	Description
MD	mean distance from centre of COP trajectory, [mm]
RMS	root mean square of COP time series,[mm]
RANGE	range of COP displacement, [mm]
MV	mean velocity (total COP trajectory length / T°), $[mm/s]^{-1}$
MF	mean frequency, i.e. number, per second, of loops that have to be run by COP, to cover total trajectory equal to total COP trajectory (MF= MV / 2π·MD), [Hz]
TP	total power, $[mm^2]$
f50	median frequency, frequency below which 50% of TP is present, [Hz]
f95	95% power frequency, frequency below which 95% of TP is present, [Hz]
CF	centroidal frequency, frequency at which spectral mass is concentrated, [Hz]
FD	frequency dispersion, unitless measure of variability of frequency content of power spectral density (zero for pure sinusoid, increases with spectral bandwidth to one)
DEV-AP	angular deviation from AP sway, [deg]
CCA	area of 95% confidence circumference, $[mm]^2$
CEA	area of 95% confidence ellipse, $[mm^2]$
SA	sway area, computed as area included in COP displacement per unit of time, $[mm^2/s^{-1}]$

°T: duration of trial, [s]

2.4 Feature selection by means of principal component analysis

The PCA procedure was applied to the summary statistic scores, considering separately the population in the levodopa *off* and *on* states. Principal components (PCs) were calculated from the correlation matrix, since the summary statistic scores were very different in value and variance [13]. In other words, PCs were constructed as weighted averages of the normalized original variables (the summary statistic scores).

Several methods have been proposed for determining the number of PCs that should be kept for further analysis, such as dropping PCs whose eigenvalues are less than one [14] or retaining just enough PCs to account for a pre-set percentage of the data variation [13]. These criteria often lead to the same result. In the present study, we retained the minimum number of PCs that accounted for at least 90% of the total variance. The number *m* of PCs considered defines the dimension of the reduced dataset (feature extraction).

After the PCA was completed, we performed a procedure aimed at making the *m* PCs more meaningful for interpretation. As first step, we provided an intuitive connotation to each of the *m* PCs, considering the minimum and maximum values along each of them (and the corresponding raw data) [4].

As in the previous study by Rocchi *et al.* [4], we introduced a feature selection procedure, to avoid the adoption of new, possibly misleading measures. Thus we selected a subset of summary statistic scores, with the aid of the *m* retained PCs. Among several possible criteria for the selection [13, 15], we chose

the one already proposed in our previous study [4], thus we associated one summary statistic score with each of the m PCs, on the basis of the higher correlation with the PC itself. This criterion, appropriate for highly correlated original variables, as summary statistic scores are, was chosen to obtain results completely homogeneous and comparable with the ones of the previous study.

3   Results

Results of PCA and feature selection are presented in the following considering the summary statistic scores in PD subjects in the *off* and in the *on* states. Summary statistic scores that primarily determine each PC are presented, and their relevance quantified by means of the correlation coefficient r. In addition, qualitative interpretation of PCs is provided by representation of raw data (as the CoP migration or as the power spectral density –PSD- of the CoP) of specific trials, identified by the extreme values along each PC [4].

3.1 Feature selection: *off* state

The first 4 PCs computed from summary statistic scores in the *off* state account for the 94.7% of the variation of the original 14-dimension data set (see table 2 and Figure 1).

- The first PC describes the size of CoP oscillation; in fact it is determined mainly by measures describing the amount of sway (see values of r in Table 2). Figure 1.a supports this interpretation showing CoP of trials at the opposite boundaries of PC_1. One among the measures describing sway size should be selected from PC_1.

- PC_2 describes spectral properties of CoP. Table 2 shows that original measures most involved are measures in the frequency domain, with the only exception of the mean velocity (MV) and the sway area (SA). MV is actually highly related with the spectral measures of the CoP signal [2, 16], and SA describes characteristics already detected by PC_1. The frequency measures determining PC_2 regard the bandwidth of the power spectral density (PSD). Figure 1.b shows the signal in the upper panel to have a percentage of the power more toward the lower frequency than the signal represented in the lower panel. One among the frequency measures listed in Table 2 should be selected from PC_2.

- PC_3 detects again spectral properties of the signal principally. In particular the frequency dispersion (FD) correlates the most with PC_3 (positive correlation) together with the median frequency f50 (negative correlation). Figure 1.c highlights their influence in the determination of PC_3; in fact the upper panel of Figure 1.c shows a PSD plot characterized by peaks in concentrated area of the bandwidth and shifted toward higher frequency compared to the signal represented in the lower panel. DEV-AP also correlates with PC_3. However, FD is the sway measure to select, because the content of f50 is already detected by PC_3, and DEV-AP could be selected from PC_4.

- In fact DEV-AP is the only summary statistic score to be highly influential in PC4 as shown in Table 2 and Figure 1.d, where direction of sway in the lower panel CoP is clearly more toward the anteroposterior course than CoP in the upper panel.

Table 2:   Off state: Correlation coefficients (only |r|>0.4 are listed) between postural summary statistic scores and the 4 PCs that account for 94.6% of total variation of original 14-dimension data set. In brackets the% of variance explained by each PC.

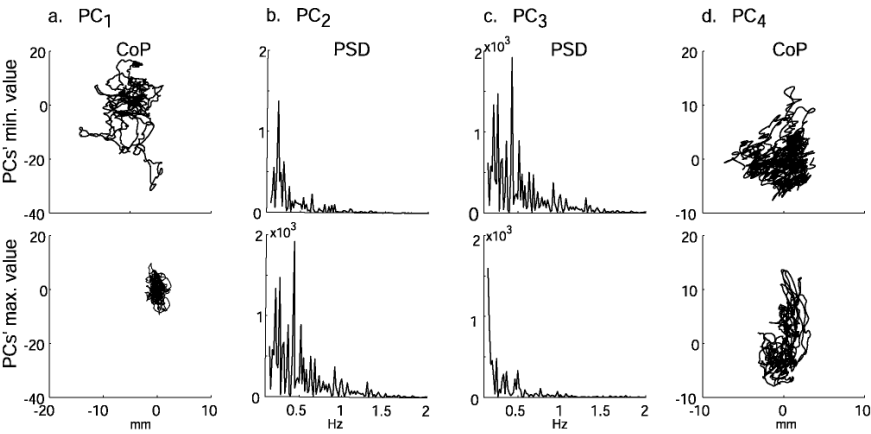| PC$_1$ (55.2%)	r	PC$_2$ (22.9%)	r	PC$_3$ (10.7%)	r	PC$_4$ (5.8%)	r
RANGE	-0.9	MV	0.89	FD	0.79	DEV-AP	0.65
RMS	-0.96	CF	0.8	DEV-AP	0.61		
MD, TP	-0.95	f95	0.71	f50	-0.56		
CCA	-0.94	MF	0.7				
CEA	-0.91	SA	0.56				
SA	-0.76	f50	0.47				
MF, f50, f95	0.6						
CF	0.54						
FD	-0.48						
MV	-0.41						

Figure 1:   PD subjects in the *off* state: description of each of the first 4 PCs (94.6% of total variance), considering the minimum and maximum values along each of them (and the corresponding raw data, in the time or frequency domain).

3.2 Feature selection: *on* state

Results from PCA on subjects with PD in the *on* state lead to the selection of the first 3 PCs that can account for the 93.2% of the variation of the original 14-dimension data set (see Table 3 and Figure 2).

Table 3:   On state: Correlation coefficients (only |r|>0.4 are listed) between postural summary statistic scores and the 3 PCs that account for 93.2% of total variation of original 14-dimension data set. In brackets the% of variance explained by each PC.

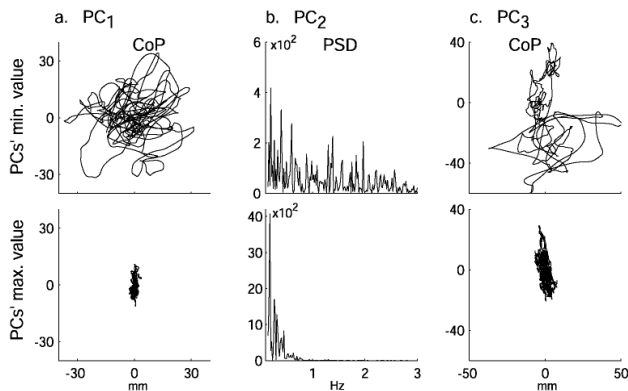| PC$_1$ (53.9%)	r	PC$_2$ (30%)	r	PC$_3$ (9.3%)	r
CEA, CCA, MD	-0.97	CF	-0.97	FD	0.74
RMS, RANGE	-0.96	MF	-0.96	DEV-AP	-0.53
TP	-0.95	f50	-0.9		
SA	-0.94	f95	-0.88		
DEV-AP, MV	-0.7	MV	-0.6		
		FD	0.51		

Figure 2:   PD subjects in the *on* state: description of each of the first 3 PCs (93.2% of total variance), considering the minimum and maximum values along each of them (and the corresponding raw data, n the time or frequency domain).

- PC$_1$ describes the size of CoP oscillation, as it did in the *off* state data set. It is determined mainly by that summary statistic scores that characterize the amount of sway. Figure 2.a exemplify the interpretation of PC$_1$ and confirms the choice of one of the sway-amount measure. They all describe the same characteristic, and they present in fact a high inter-correlation [2, 16]. The DEV-AP and MV, that contains different information than the sway-amount scores, also correlates with PC$_1$ and should be considered too, if not individuated by any of the following PCs.

- Also in the *on* state data, PC$_2$ describes spectral properties of CoP, as presented by the differences between PSD functions of the upper and lower panels of Figure 2.b. Table 3 shows that the frequency summary statistic scores that determine PC$_2$ are mainly the ones that describe the frequency band width of the CoP (from CF to f95 in the 2nd column of Table 3). MV is significant in PC$_2$, and it correlates with the spectral measures, thus just one of the last may be selected, causing a small loss of information.

- PC$_3$ is primarily determined by the frequency dispersion FD, as shown in Table 3. PC3 is determined also by DEV-AP, as shown by the negative correlation highlighted in the 3rd column of Table 3, and underlined by CoP trajectories represented in Figure 2.c. The variability associated with this PC explains cases of sway trend moving toward the mediolateral direction, combined with decreasing FD. In this case, both DEV-AP and FD should be considered as descriptor of this PC, because they are not already identified by previous PC and, on the contrary DEV-AP is relevant also in PC1, and FD in PC2, but both of them are not selected from them.

4   Discussion

In this study, we expanded the feature selection approach, based on principal component analysis, already proposed in a previous study on a healthy young population [4], to investigate the postural sway-measures considering a population with impairment to the motor system. For this purpose we considered subjects with Parkinson's disease, with a dual aim: - to proceed in the process of transforming data into information; - to promote the standardization in quantitative posturography.

We identified the features that characterize the postural sway in the PD population before and after the assumption of levodopa (*off* and *on* states). The feature selection in the *off* and *on* states on 14 postural sway-measures allowed us to identify the two subsets that explain the greatest part of the variability in the population in the two conditions. Interestingly the PCs that explain more than the 90% of the total variation were very similar in the *off* and in the *on* states, and the results were analogous to the results achieved on the young population in [4].

4.1 Postural sway features of PD subjects in *off* and *on* states

In the *off* state the selection of 4 PCs allows the explanation of more than 90% of the original dataset variation. In the *on* state, 3 PCs are sufficient to this purpose. This difference may be interpreted considering the levodopa to cause a coupling effect of postural control characteristics, and hence an increase of correlation among sway-measures. The feature selection procedure applied to the postural measures in the *on* and *off* states highlighted the following distinctive properties of the COP trajectory, in common to the two states:

- size of the path travelled by the COP over the support surface, estimated by RMS, MD, RANGE, CEA, SA and CCA;
- relevant frequencies that characterise the power spectral density curve: f50, f95, MF and CF
- a unitless measure of the frequency dispersion, estimated by FD, a parameter related to the shape of the PSD curve;

In the *off* state, PC$_4$ leads to the selection of the principal sway direction, estimated by DEV-AP. In the *on* state DEV-AP is not straightforward related to a single PC, but it is particularly relevant in PC$_1$ and PC$_3$, even if not univocally

selected by any of them. This really significant information could be lost if DEV-AP is not considered. Hence also in the *on* state, the feature selection procedure leads to the selection of 4 from the summary statistic scores, even of only 3 PCs are necessary to explain more than 90% of the original variation.

The few differences identified by the PCs selection in *off* and *on* states explicitly agree with differences in the CoP pattern due to levodopa, investigated in previous studies [8, 17]. Indeed, differences in motor performance due to levodopa assumption in the subjects included in this study were confirmed by clinical tests performed before posturography: UPDRS (mean±SD) is 49.3±16.7 in the *off* state and 29.1±14.9 in the *on* state. In particular PC_1 in the on state describes variability of data associated with sway travelled increased and shifted toward the mediolateral direction. This result is in complete accordance with the previous ones, where + one of the main described effect of levodopa on the CoP was that of increasing CoP sway and moving it toward the lateral direction. The coupling between increase and direction of sway path may be the main reason why just 3 PCs, instead of 4, are necessary to describe the 90% of the original variability. At the same time it corroborates the necessity of choosing DEV-AP as a descriptor of CoP features during quiet stance. Therefore, even if characteristics of sway in *off* and *on* states are different, their main descriptors are the same. In fact, considering the present results and selection criteria explicated in [4] (high measure reliability, low inter-correlation and dependence to anthropometric factors [5] the present study guides to the selection of the following summary statistic scores, as the most typical of quantitative posturography in PD subjects in off and on state: RMS; f95 or CF; FD; DEV-AP.

4.2 Postural features of PD subjects compared to healthy young subjects

We were actually surprised by the similarity of extracted PCs in PD subjects and in healthy young subjects from [4]. In fact The PCs extraction led to homogeneity between the sway-measures selection in the two populations. The guidelines that were introduced in [4] recommended RMS, f95,DEV-AP, FD as selected features. Because PCA points out the major factors behind the variability of a dataset, we did not expect, a priori, the same set of measures to be selected in every case. Different measures could be the most sensitive to different specific conditions. The homogeneity of results considering different population, and homogeneity within the same population in different treatment conditions, allow us to define more confidently the minimum set of measures to recommend for specific applications, as optimal descriptor of CoP sway in quiet stance, and to stimulate standardization in quantitative posturography.

4.3 Future developments

Even if in the literature it has already been observed that some of the selected summary statistic scores may well discriminate between pathological and control subjects [8, 18, 19], the present study did not aim at identifying features that best discriminate between different groups of subjects, or between subjects in different conditions. To this aim, further investigations are necessary to select

measures characterized by high specificity. For this purpose procedures of feature selection for classification will be developed and the results then compared with ones obtained within homogeneous populations, with the aim of identifying features both optimally descriptive and specific of quantitative posturography.

References

[1]   Winter, D.A., Human balance and posture control during standing and walking, Gait. Posture., 3 (1995) 193-214.
[2]   Prieto, T.E., Myklebust, J.B., Hoffmann, R.G., Lovett, E.G., Myklebust, B.M., Measures of postural steadiness: differences between healthy young and elderly adults, IEEE Trans. Biomed. Eng, 43 (1996) 956-966.
[3]   Chiari, L., Bertani, A., Cappello, A., Classification of human strategies in human postural control by stochastic parameter, Hum Mov Sci, 19 (2000) 817-842.
[4]   Rocchi, L., Chiari, L., Cappello, A., Feature selection of stabilometric parameters based on principal component analysis, Med. Biol. Eng Comput., 42 (2004) 71-79.
[5]   Chiari, L., Rocchi, L., Cappello, A., Stabilometric parameters are affected by anthropometry and foot placement, Clin. Biomech. (Bristol. , Avon.), 17 (2002) 666-677.
[6]   Bonnet, A.M., Loria, Y., Saint-Hilaire, M.H., Lhermitte, F., Agid, Y., Does long-term aggravation of Parkinson's disease result from nondopaminergic lesions?, Neurology, 37 (1987) 1539-1542.
[7]   Lim, E., A walk through the management of Parkinson s disease, Ann. Acad. Med. Singapore, 34 (2005) 188-195.
[8]   Rocchi, L., Chiari, L., Horak, F.B., Effects of deep brain stimulation and levodopa on postural sway in Parkinson's disease, J. Neurol. Neurosurg. Psychiatry, 73 (2002) 267-274.
[9]   Rocchi, L., Chiari, L., Cappello, A., Gross, A., Horak, F.B., Comparison between subthalamic nucleus and globus pallidus internus stimulation for postural performance in Parkinson's disease, Gait. Posture., 19 (2004) 172-183.
[10]   Oliveira, L.F., Simpson, D.M., Nadal, J., Calculation of area of stabilometric signals using principal component analysis, Physiol Meas., 17 (1996) 305-312.
[11]   Hufschmidt, A., Dichgans, J., Mauritz, K.H., Hufschmidt, M., Some methods and parameters of body sway quantification and their neurological applications, Arch. Psychiatr. Nervenkr., 228 (1980) 135-150.
[12]   Diener, H.C., Dichgans, J., Bacher, M., Gompf, B., Quantification of postural sway in normals and patients with cerebellar diseases, Electroencephalogr. Clin. Neurophysiol., 57 (1984) 134-142.
[13]   Jolliffe, I.T., Principal component analysis, Springer-Verlag, New York, 1986.

[14] Jolliffe, I.T., Discarding variables in a principal component analysis, I: Artificial data, Appl. Statist, 21 (1972) 160-173.
[15] McCabe, G.P., Principal Variables, Technometrics, (1984) 137-144.
[16] Maurer, C., Peterka, R.J., A new interpretation of spontaneous sway measures based on a simple model of human postural control, J. Neurophysiol., 93 (2005) 189-200.
[17] Faist, M., Xie, J., Kurz, D., Berger, W., Maurer, C., Pollak, P., Lucking, C.H., Effect of bilateral subthalamic nucleus stimulation on gait in Parkinson's disease, Brain, 124 (2001) 1590-1600.
[18] Yamamoto, R., Kinoshita, T., Momoki, T., Arai, T., Okamura, A., Hirao, K., Sekihara, H., Postural sway and diabetic peripheral neuropathy, Diabetes Res. Clin. Pract., 52 (2001) 213-221.
[19] Maki, B.E., Holliday, P.J., Topper, A.K., A prospective study of postural balance and risk of falling in an ambulatory and independent elderly population, J. Gerontol., 49 (1994) M72-M84.