Feature selection of stabilometric parameters based on principal component analysis

L. Rocchi **L. Chiari** **A. Cappello**

Biomedical Engineering Unit, Department of Electronics, Computer Science & Systems, University of Bologna, Italy

Abstract—*This study addresses the challenge of identifying the features of the Centre of pressure (COP) trajectory that are most sensitive to postural performance, with the aim of avoiding redundancy and allowing a straightforward interpretation of the results. Postural sway in 50 young, healthy subjects was measured by a force platform. Thirty-seven stabilometric parameters were computed from the one-dimensional and two-dimensional COP time series. After normalisation to the relevant biomechanical factors, by means of multiple regression models, a feature selection process was performed based on principal component analysis. Results suggest that COP two-dimensional time series can be primarily characterised by four parameters, describing the size of the COP path over the support surface; the principal sway direction; and the shape and bandwidth of the power spectral density plot. COP one-dimensional time series (antero-posterior (AP) and medio-lateral (ML)) can be characterised by six parameters describing COP dispersion along the AP direction; mean velocity along the ML and AP directions; the contrast between ML and AP regulatory activity; and two parameters describing the spectral characteristics of the COP along the AP direction. On the basis of the results obtained, some guidelines are suggested for the choice of stabilometric parameters to use, with the aim of promoting standardisation in quantitative posturography.*

Keywords—*Posture, Stabilometric parameters, Feature selection, Principal component analysis, Normalisation*

Med. Biol. Eng. Comput., 2004, **42**, 71–79

1 Introduction

BODY POSTURE is the output of complex interactions between central nervous system control mechanisms (visual, vestibular and somatosensory systems, integration of afferent information, generation of motor output) and the musculo-skeletal actuators acting against the support surface. Because of its complexity, body posture is challenging to measure with simple methods, and yet simple methods, both in terms of time expenditure and data interpretation, are needed in neurological, orthopaedic and geriatric clinical practice, where balance impairments are commonly reported (HUFSCHMIDT *et al.*, 1980; DIENER *et al.*, 1984; MAKI *et al.*, 1994). Stabilometry, i.e. the measurement of forces exerted against the ground from a force platform during quiet stance, is commonly used to quantify postural steadiness both in research and in the clinic.

Typically, stabilometry focuses on the properties of the centre of pressure (COP) time series, representing the point location of the ground reaction force vector as it evolves on the horizontal plane (2D) or along two orthogonal axes, fixed with the platform (antero-posterior (AP) and medio-lateral (ML)) (KAPTEYN *et al.*, 1983). This single variable reflects both the balance controlling process and movements of the centre of mass of the entire body and thus provides a single global measure of posture control. However, COP analysis produces a potentially large dataset (stabilometric parameters) that can be difficult to manage.

The stabilometric parameters that are most commonly reported in the literature are those that describe the statistical properties of the COP trajectory, considered as a stationary signal, in the time and frequency domains (PRIETO *et al.*, 1996). Under this assumption, the number of stabilometric parameters that can be extracted from the COP is large, and many of the parameters are redundant, complicating interpretation of the dataset (NEWELL *et al.*, 1997).

The way to turn data into information is a common problem in the human movement analysis community (KAUFMAN and SUTHERLAND, 1996). In particular, to date, stabilometry undoubtedly suffers from several limiting factors including

(i) the absence of a definite 'normal pattern'
(ii) the lack of standardisation in the measurement protocols
(iii) the large number of highly coupled variables that are computed from the force platform recordings.

The present study moves from this latter evidence and addresses the open challenge of developing guidelines to

Correspondence should be addressed to Dr Lorenzo Chiari; email: lchiari@deis.unibo.it

Paper received 14 March 2003 and in final form 23 September 2003

MBEC online number: 20043847

identify the most relevant COP measures for quantifying postural steadiness. The number of COP-based parameters selected should be small enough to reduce computation, avoid redundancy and enable clear interpretation of the results, and yet large enough to quantify different aspects of postural control. The ideal parameters recommended by the guidelines should be uninfluenced by spurious sources of within- (e.g. non-stationarity, fatigue) and between-subject variability (e.g. anthropometry) that can abnormally inflate the signal components of the COP. At the same time, the parameters must be sufficiently sensitive to identify real and meaningful differences in posture control between subject groups.

A few efforts have been made, to date, in the direction of selecting a subset of variables computed from stabilometric recordings. In particular, some authors implemented dimension reduction procedures based on mono-dimensional statistical methods and sensitivity to visual condition and different subject groups (PRIETO *et al.*, 1996; BARATTO *et al.*, 2002).

The major aim of the present study was to design a multi-dimensional feature selection procedure for stabilometric parameters based on principal component analysis (PCA). The central idea of PCA is to reduce the dimensionality of a dataset of several interrelated measures (JOLLIFFE, 1986), here, the parameters computed from the COP trajectory. This reduction is achieved by transforming the parameters to a new set of variables (the principal components), so that the first few retain most of the variation (assumed to represent information) present in the original dataset. Indeed, it can be proven that the representation given by PCA is an optimum linear dimension reduction technique in the mean-square sense (JOLLIFFE, 1986). In addition, PCA facilitates the interpretation of the results because it extracts features that are directly related to the original data set.

In this regard, in the present work, to make the procedure also appropriate for clinical purposes, where traditional COP-based parameters are routinely used, we preferred to select a subset of the original dataset, rather than introduce new measures. To this aim, the principal components (PCs) were used to suggest suitable selections from the whole set of 37 parameters (JOLLIFFE, 1986). This was achieved by investigating separately the variability of the parameters computed from 2D and AP–ML representations.

2 Methods

2.1 *Experimental session*

Postural sway was measured on 50 healthy young adults (25 males and 25 females) without musculoskeletal or neurological disorders. The subjects' mean age was 25.7 years (SD 2.8, range 21–30 years). The experimental set-up consisted of two 50 s trials for each subject. Subjects stood on a strain-gauge force platform*, with eyes open, looking towards an achromatic target (a 5 cm diameter circle) 2 m away. To avoid any kind of learning or fatigue effect (TARANTOLA *et al.*, 1997), only the first trial was retained in the analysis.

The three force and three moment components were recorded from the force plate at 200 Hz. Subsequently, data were filtered at 8 Hz by a 30th -order low-pass FIR filter with zero-phase and down-sampled at 20 Hz. From the output signals of the platform, the two COP co-ordinates were computed in the AP and ML directions. Next, the 2D description of the migration of the whole-body COP was obtained by representing the AP as a function of the ML time series, and by computing the distances between each point in the (ML, AP) plane and the mean COP. We also recorded anthropometric and base of support measures for each subject. These measures were later used to evaluate and then remove the influences of biomechanical factors on parameters extracted from the COP data (see Section 2.3).

*Model 4060-08, Bertec Corporation, Columbus, Ohio, USA

2.2 *Stabilometric parameters*

Fifteen different parameters were computed from the COP course. Eleven of these were computed from each of the 1D (AP and ML) and 2D time series. These 11 parameters typically quantify the major properties of the COP time series in the time and frequency domains. The remaining four parameters were extracted from the 2D time series only, because they describe planar characteristics, such as the area covered by the COP and the principal sway direction (OLIVEIRA *et al.*, 1996). Thus we computed a total of 37 measures from the COP data. These measures are usually referred to as summary statistic scores (PRIETO *et al.*, 1996).

Summary statistic scores are frequently applied in clinical practice, being easy to compute and relatively straightforward to interpret (HUFSCHMIDT *et al.*, 1980; DIENER *et al.*, 1984). A full list and a brief description of the parameters computed in the present study are reported in Table 1.

2.3 *Parameter normalisation*

A previous work (CHIARI *et al.*, 2002), using robust regression analysis, revealed the dependence of most COP stabilometric parameters on biomechanical factors. This dependence can cause data misinterpretation when between-subject comparisons are performed. In particular, it was shown that the following set of biomechanical factors should be taken into account: height, weight, base of support area, maximum foot width and feet-opening angle. These were able to explain more than 80% of the variance in the overall set of 17 considered anthropometric and foot placement measurements (CHIARI *et al.*, 2002).

In the present work, the assessment of biomechanical influences on the parameters was performed by multiple regression analysis. In this phase we determined which of the five biomechanical factors were correlated with variations in each stabilometric parameter. To uncover this optimum subset, an iterative algorithm was applied (HINTZE, 2000). This algorithm fits all possible regression models and suggests the optimum solution(s) in terms of a balance between simplicity (as few regressors as possible) and fit (as many regressors as needed). After the optimum regression model was determined for each stabilometric parameter, the latter was normalised first by subtraction of the value predicted by the model and then addition of the mean value of the parameter across subjects to return the value of the normalised parameter to the original range (O'MALLEY, 1996; CHIARI *et al.*, 2002).

2.4 *Principal component analysis for feature extraction*

The PCA procedure was applied to stabilometric parameters after normalisation. The correlation matrix (instead of the covariance matrix) was used to estimate the PCs, because the parameters were very different in value and variance (JOLLIFFE, 1986). Parameters characterising the COP trajectory on the horizontal plane (2D parameters) were treated separately from those computed from the 1D time series (AP–ML parameters), under the assumption that the two groups share most of the information.

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 (KAISER, 1960; JOLLIFFE, 1972) or retaining just enough PCs to account for a pre-set percentage of the data variation (JOLLIFFE, 1986). In the present study, we adopted the last criterion and chose the

Table 1 Stabilometric parameters: acronyms and brief descriptions

Acronym	Description
Parameters computed from 2D, AP and ML COP displacements	
MD	mean distance from centre of COP trajectory, mm
RMS	root mean square of COP time series, mm
SP	sway path, total COP trajectory length, mm
RANGE	range of COP displacement, mm
MV	mean velocity (SP/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 *SP* ($MF = SP/(2\pi \cdot MD \cdot T^*)$), Hz
TP	total power, mm^2
*f*50	median frequency, frequency below which 50% of *TP* is present, Hz
*f*95	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)
Parameters computed from 2D COP displacements only	
|90-*MDir*|	angular deviation from AP sway, deg^{-1}
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

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.

After the PCA was completed, we performed a procedure aimed at making the m PCs more meaningful for interpretation. This operation was not easy, because a linear combination of original variables may not have a clear interpretation. Nevertheless, we tried to provide an intuitive meaning to each of the m PCs.

The first m PCs defined a new co-ordinate system, and each trial was identified by new co-ordinates in this m-dimension space (feature extraction process). Along each direction (i.e. PC), the two trials whose co-ordinates were the minimum and maximum were considered for qualitative interpretation of the corresponding PC, by means of consideration of the relevant features of their COP in the time and frequency domains.

2.5 *From feature extraction to feature selection*

As we did not wish to introduce new, possibly misleading, variables, we then sought a subset of m original variables, with the aid of the features extracted so far. Among several possible criteria (McCabe, 1984; Jolliffe, 1986), we chose the one that associated one stabilometric parameter with each of the m PCs, on the basis of the higher correlation with the PC itself. This approach is suggested when there are groups of highly correlated variables, such as stabilometric parameters (Prieto *et al.*, 1996), so that just one variable is selected from each group.

3 Results

3.1 *Multiple regression and parameter normalisation*

Results of multiple regressions are presented in Table 2. Not all the parameters are equally correlated with the biomechanical factors. For parameters that are weakly correlated, the normalisation does not have a strong effect. We highlight in the Table, and discuss in the following, only those stabilometric parameters that were heavily influenced by the biomechanical regressors ($r^2 > 0.2$).

The stabilometric parameters are usually dependent on either height or weight. These two anthropometric measurements are never co-present, because their joint occurrence represents a sort of redundancy. All the ML parameters (except mean velocity) are dependent on the base of support area, whereas the AP parameters are dependent on maximum foot width. The biomechanical factors that affect ML and AP parameters also influence the corresponding 2D parameters. Thus most of the 2D parameters show dependence on both base of support area and on maximum foot width. The feet-opening angle only affects frequency-domain parameters from ML time series.

3.2 *PCA and feature selection applied to 2D parameters*

The minimum number of PCs that can explain at least 90% of the total variation in the 2D stabilometric parameters is four. The first four PCs account for 90.95% of the total variation of the original 14-dimension dataset. The 2D parameters are actually 15 in number (see Table 1), but parameter *SP* was not included in the PCA, being directly proportional to *MV*. Table 3 lists the PC coefficients (i.e. eigenvectors of the correlation matrix).

Table 2 Multiple regression of stabilometric parameters against biomechanical factors. Only regressions with $r^2 > 0.2$ are reported. MFW-maximum foot width; BOS-base of support area; α-feet-opening angle

Parameter (r^2)	Regressors		
MD (0.21)	weight	MFW	BOS
RANGE (0.26)	weight	MFW	BOS
MV (0.23)	height	MFW	
*f*95 (0.23)	height	MFW	BOS
MD_{ML} (0.27)	weight	BOS	
RMS_{ML} (0.27)	weight	BOS	
$RANGE_{ML}$ (0.33)	weight	BOS	
MV_{ML} (0.26)	height	MFW	
TP_{ML} (0.32)	height	BOS	
$f50_{ML}$ (0.37)	weight	BOS	α
$f50_{AP}$ (0.34)	height	MFW	
CF_{AP} (0.29)	height	MFW	
CF_{ML} (0.30)	BOS	α	
FD_{ML} (0.24)	MFW	BOS	α
SA (0.32)	weight	MFW	BOS

Table 3 2D parameters: PC coefficients and correlation coefficients between parameters and the corresponding PC. Only values of $|r| > 0.4$ are shown, in brackets

	PC_1	PC_2	PC_3	PC_4
Cumulative % explained variance	53.58%	74.00%	84.32%	90.95%
MD	0.35 (0.97)	−0.05	−0.11	−0.02
RMS	0.35 (0.97)	−0.04	−0.10	−0.06
RANGE	0.35 (0.7)	−0.08	0.07	−0.05
MV	0.17 (0.48)	−0.37 (−0.63)	0.38 (0.46)	−0.05
MF	−0.23 (−0.63)	−0.23	0.49 (0.59)	−0.09
TP	0.32 (0.88)	−0.05	0.22	0.01
*f*50	−0.08	−0.52 (−0.87)	−0.23	0.24
*f*95	−0.21 (−0.59)	−0.38 (−0.64)	0.04	−0.41
CF	−0.20	−0.43 (−0.73)	−0.06	−0.26
FD	−0.02	0.39 (0.66)	0.38 (0.46)	−0.57 (−0.55)
|90-*MDir*|	−0.06	0.02	0.56 (0.67)	0.59 (0.57)
CCA	0.34 (0.93)	−0.05	−0.05	−0.12
CEA	0.35 (0.96)	−0.08	0.02	−0.03
SA	0.32 (0.87)	−0.21	0.16	−0.05

In the following, qualitative interpretations of PCs are presented, based on both the signs and values of PC coefficients and on COP characteristics of the subjects identified by co-ordinates at the extremes along each PC direction (Fig. 1). In addition, we present the results of the feature selection method that identifies a subset of the original 2D stabilometric parameters, needed to describe different components of the COP pattern (JOLLIFFE, 1986) and explaining almost the same amount of information as the four selected PCs. In this light, Table 3 reports the correlation coefficients r between each PC and the parameters, which helped to carry out the feature selection process.

PC_1 describes the size of the COP oscillation and, in fact, it is determined mainly by parameters describing the amount of sway. Fig. 1*a* illustrates this interpretation, showing the main differences between subjects at the opposite boundaries of PC_1. In particular PC_1 is highly correlated ($r > 0.93$) with a group of five parameters that all describe the size of the COP travelled path in the horizontal plane: *RMS*, *MD*, *RANGE*, *CEA* and *CCA* (see Table 1 for acronym definitions). Thus, on this basis, any one of these measures could be chosen as a representative parameter.

PC_2 describes relevant spectral properties, reflecting mainly the position of the power spectral density (PSD) plot, as proved by the most influential parameters in this PC. Fig. 1*b* shows the PSD of COP trajectories of the extreme subjects along this PC, with evident spectral differences. PC_2 is highly correlated with two measures in the frequency domain: *f*50 ($r = -0.87$) and *CF* ($r = -0.73$); one of them may be considered as a selected feature from this PC.

PC_3 is easily interpretable, identifying the principal direction of sway (AP or ML) in the horizontal plane, being mostly determined by $|90 - MDir|$ (displacement of the main direction of the COP trajectory away from the AP axes). Fig. 1*c* confirms that the characteristic that most distinguishes the COP trajectories is the sway direction (primarily ML for the subject represented in the upper panel and AP for the subject represented in the bottom panel). As expected from this qualitative interpretation of PC_3, $|90 - MDir|$ is the parameter most correlated with PC_3 ($r = 0.67$), and it is the representative parameter for this PC.

PC_4 is difficult to interpret, because it is strongly and oppositely determined by two parameters: $|90 - MDir|$ and frequency dispersion *FD*. This means that the variability associated with this PC explains cases of sway direction moving towards ML combined with decreasing *FD*. PC_4 is most correlated with $|90 - MDir|$ ($r = 0.57$) and *FD* ($r = -0.55$). Nevertheless, *FD* should be preferred to represent this component of the COP, as $|90 - MDir|$ has already been

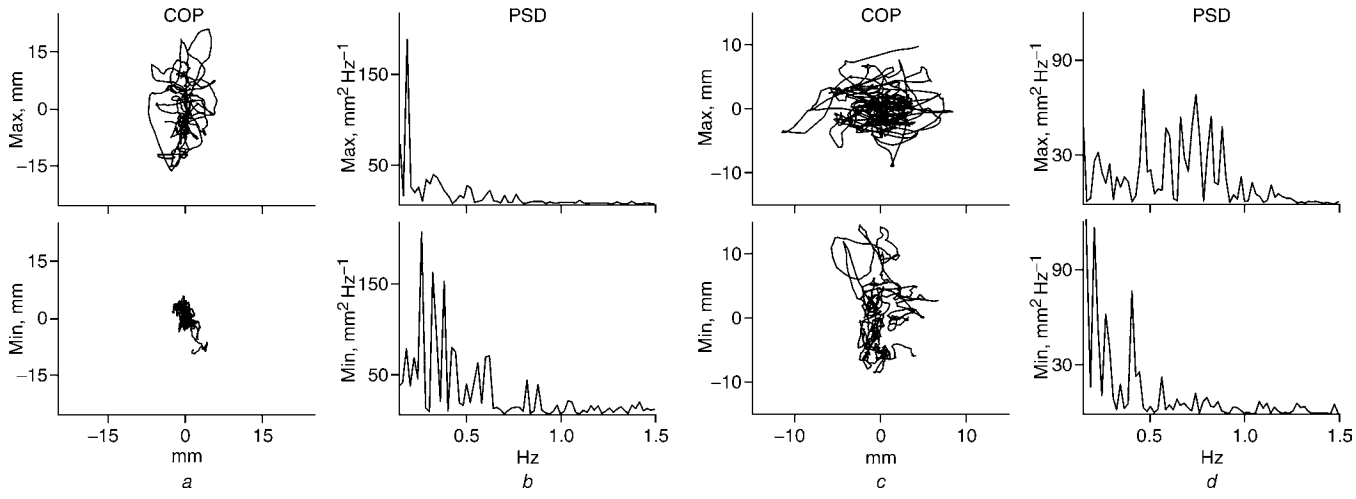

Fig. 1 *COP trajectories and power spectral density (PSD) of two subjects for each PC explaining 90.95% of variation of 2D stabilometric parameters. Two subjects are identified by highest (Max, upper panel) and lowest (Min, bottom panel) co-ordinate values along corresponding PC (a) PC_1; (b) PC_2; (c) PC_3; (d) PC_4*

addressed within PC_3. Hence we can conclude that PC_4 defines the shape of the PSD, as can be seen in the comparison between PSD plots in Fig. 1*d*.

3.3 *PCA and feature selection applied to AP–ML parameters*

Six PCs are required to explain 90% of the total variation of the original variables represented by the AP–ML parameters. The six PCs account for 90.02% of the total variation of the original 20-dimension dataset. Parameters computed from AP–ML time series are actually 22 in number (see Table 1), but parameters SP_{AP} and SP_{ML} were not included in the PCA, being directly proportional to MV_{AP} and MV_{ML}, respectively. PC coefficients are shown in Table 4.

Considering both the signs and values of the PC coefficients and the COP trajectories identified by co-ordinates at the extremes along each PC direction, interpretation of the six PCs follows, with the output of the feature selection procedure that identifies a subset of the AP–ML original stabilometric parameters performed using the correlation coefficients r, listed in Table 4.

PC_1 is primarily a descriptor of dispersion in the AP direction, as is well illustrated in Fig. 2*a* and highlighted by coefficients in the first column of Table 4, which principally weigh the AP component of distance parameters. PC_1 correlates almost equally with four parameters (r ranges from -0.73 to -0.77): $RANGE_{AP}$, TP_{AP}, RMS_{AP} and MD_{AP}. They all describe the amount of COP migration along the AP direction, and one of them should be considered as a descriptor of this PC.

PC_2 highlights postural activity that occurs principally in the ML direction, as shown in Fig. 2*b* by the comparison between ML COPs of subjects at the opposite boundaries along this PC. ML mean velocity and ML frequency parameters show the highest PC coefficients. MV_{ML} is the parameter most correlated with PC_2 ($r = -0.86$), followed by MV_{AP} ($r = -0.7$). Among the 12 parameters having $|r| > 0.4$ (see Table 4), nine are from the ML time series, denoting the importance of this component for PC_2.

PC_3 distinguishes between subjects with a predominance of ML activity and subjects with a predominance of AP activity. The upper panel of Fig. 2*c* shows a trial where postural activity is principally along the ML direction, whereas, in the bottom panel, it is the AP direction that is dominant. This is confirmed by the coefficients in the third column of Table 4; observe that their signs are opposite for the AP and ML parameters. Regarding r-values, the parameters principally involved in the identification of PC_3 are the ones describing the amount of sway in the AP direction, followed by the ones in the ML direction, with opposite signs of r for the AP and ML parameters. Even if the AP parameters are more correlated with PC_3, probably owing to biomechanical properties, one of them has already been suggested from PC_1, and so it seems preferable to select a parameter describing dispersion along the ML direction, to avoid loosing the information provided by PC_3.

PC_4, PC_5 and PC_6 highlight the spectral characteristics of the COP, especially in the AP direction, as evidenced by the coefficients in Table 4 and by the plots in Figs. 2d–f. Details follow.

PC_4 distinguishes traces with power concentrated at low frequencies (here below 0.5 Hz, as shown in the top panel of Fig. 2*d*) and traces with power more spread through the bandwidth. It is not surprising that the highest coefficient found in the fourth column of Table 4 is that corresponding to CF_{AP} (centroidal frequency of COP sway in the AP direction). PC_4 is mostly correlated to AP frequency parameters and principally with CF_{AP} ($r = -0.67$), followed by $f50_{AP}$ ($r = -0.63$) and $f95_{AP}$ ($r = -0.60$). One of them may be considered as the parameter characterising PC_4.

PC_5 is particularly related to FD_{AP} (see Table 4, fifth column), which measures how far the signal is from a pure sinusoid. By looking at Fig. 2*e*, it is possible to note that the PSD plot represented in the top panel presents an evident peak at around 0.3 Hz, whereas there is no evident peak in the bottom panel (with the exception of high power at very low frequency, related to a continuous component). Hence, PC_5 seems sensitive to the presence of a prevalent sinusoidal oscillation along the AP direction (shape of AP PSD). The parameters that primarily correlate with PC_5 are FD_{AP} ($r = -0.61$), $f95_{AP}$ ($r = -0.49$) and $f50_{AP}$ ($r = 0.43$). The relevance of FD_{AP} is significant, and it should be considered as the stabilometric parameter most representative of this PC.

Table 4 AP–ML parameters: PC coefficients and correlation coefficients between parameters and corresponding PC. Only values of $|r| > 0.4$ are shown, in brackets, with exception of PC_6 for which only minor correlation were found

	PC_1	PC_2	PC_3	PC_4	PC_5	PC_6
Cumulative % explained variance	35.67%	56.15%	70.21%	79.04%	85.86%	90.02%
MD_{AP}	−0.27 (−0.73)	0.03	−0.36 (−0.60)	−0.19	−0.01	−0.12
MD_{ML}	−0.26 (−0.69)	−0.20 (−0.41)	0.26 (0.44)	0.01	−0.22	−0.24
RMS_{AP}	−0.28 (−0.74)	0.03	−0.35 (−0.59)	−0.20	−0.01	−0.12
RMS_{ML}	−0.26 (−0.71)	−0.20 (0.41)	0.26 (0.44)	0.02	−0.22	−0.23
$RANGE_{AP}$	−0.29 (−0.77)	−0.06	−0.31 (−0.53)	−0.12	0.05	−0.03
$RANGE_{ML}$	−0.25 (−0.67)	−0.26 (−0.53)	0.21	0.09	−0.16	−0.16
MV_{AP}	−0.15 (−0.41)	−0.34 (−0.70)	−0.10	−0.11	0.05	0.42 (0.38)
MV_{ML}	−0.06	−0.43 (−0.86)	−0.09	0.15	−0.05	0.15
MF_{AP}	0.18 (0.48)	−0.23	0.32 (0.53)	0.20	0.07	0.34
MF_{ML}	0.24 (0.65)	−0.19 (−0.46)	−0.26 (−0.43)	0.17	0.09	0.22
TP_{AP}	−0.28 (−0.75)	−0.11	−0.22	0.01	0.04	0.35 (0.32)
TP_{ML}	−0.25 (−0.66)	−0.27 (−0.55)	0.06	0.11	−0.05	0.14
$f50_{AP}$	0.09	−0.21 (−0.43)	0.18	−0.47 (−0.63)	0.37 (0.43)	−0.16
$f50_{ML}$	0.25 (0.67)	−0.28 (−0.56)	−0.20	0.07	0.00	−0.19
$f95_{AP}$	0.21 (0.56)	−0.04	0.03	−0.45 (−0.60)	−0.42 (−0.49)	0.19
$f95_{ML}$	0.21 (0.55)	−0.26 (−0.53)	−0.20	0.04	−0.27	−0.23
CF_{AP}	0.21 (0.57)	−0.07	0.03	−0.50 (−0.67)	−0.32	0.18
CF_{ML}	0.23 (0.63)	−0.25 (−0.51)	−0.19	0.10	−0.18	−0.30
FD_{AP}	0.02	0.27 (0.55)	−0.18	0.30	−0.52 (−0.61)	0.13
FD_{ML}	−0.23 (−0.61)	0.19	0.21	−0.08	−0.24	0.20

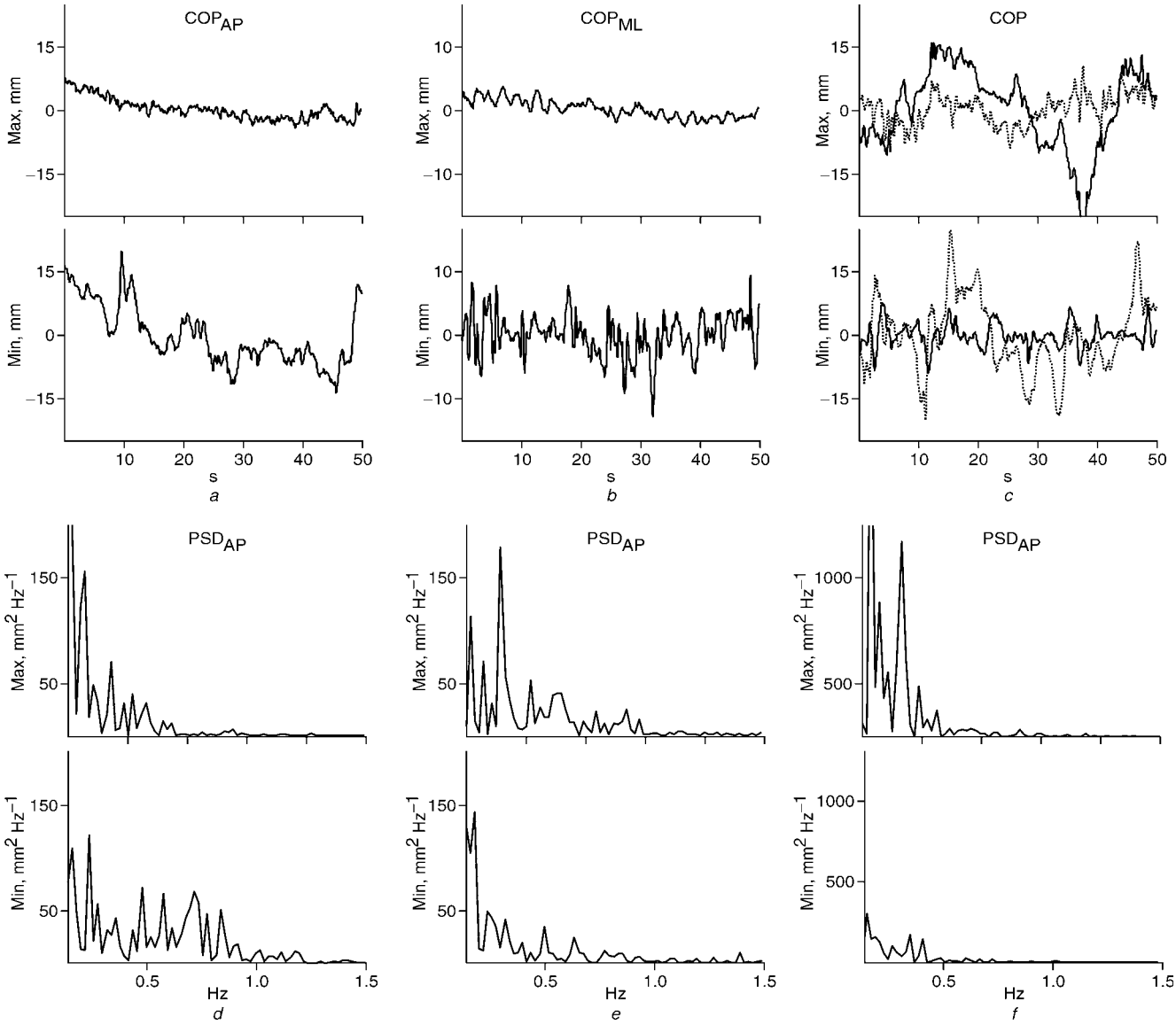

Fig. 2 *COP trajectories and power spectral density (PSD) of two subjects for each PC explaining 90.02% of variation of AP and ML stabilometric parameters. Two subjects are identified by highest (Max, upper panel) and lowest (Min, bottom panel) co-ordinate values along corresponding PC. (a) PC_1; (b) PC_2; (c) PC_3 (—)ML, (·····) AP; (d) PC_4; (e) PC_5; (f) PC_6*

PC_6 highlights the spectral properties of the AP component of COP, with little contribution from the ML component. In fact, in subjects oppositely situated along PC_6, shown in Fig. 2*f*, we notice a large difference in the AP PSD plot, with a larger amount of power in the top panel compared with the bottom panel. This interpretation is confirmed by the weight of parameters TP_{AP} and MV_{AP} in the definition of PC_6, which are also the most correlated ($r = 0.38$ and $r = 0.32$, respectively). The higher correlation for the velocity parameter rather than for the spectral one is not in contrast with the interpretation of PC_6. In fact, MV_{AP} and TP_{AP} are highly correlated with each other ($r = 0.71, p < 0.01$) and the information held by these parameters is very similar.

4 Discussion

4.1 *Suggested sets of parameters*

In this study, we developed a feature selection approach, based on principal component analysis, to investigate the redundancy of stabilometric parameters. An analysis performed on the two sets of 2D and AP–ML COP-based parameters allowed us to identify the two subsets that explain the greatest part of the variability in a population of healthy young subjects.

Based on the first PCs, the feature selection procedure applied to the 2D parameters highlighted the following distinctive properties of the COP trajectory:

(*a*) size of the path travelled by the COP over the support surface, estimated by *RMS*, *MD*, *RANGE*, *CEA* and *CCA*
(*b*) relevant frequencies that characterise the power spectral density curve: *f*50, *f*95 and *CF*
(*c*) principal sway direction, estimated by $|90 - MDir|$, reflecting the relative weight of the AP and ML components of the oscillation
(*d*) a unitless measure of the frequency dispersion, estimated by *FD*, a parameter related to the shape of the PSD curve that quantifies the degree of determinism in the COP displacements.

A valid choice among the parameters of group (*a*) could be provided by the area of the bivariate confidence ellipse *CEA*, which is the best estimate of COP sway area and may help to produce a report that facilitates visual inspection of the results. Nevertheless, owing to its common use, straightforward

definition and robustness to violation of the stationarity assumption, we would recommend choosing *RMS* from the parameters of group (*a*). *RMS* has been related to the effectiveness of the postural control system (PRIETO *et al.*, 1996). Regarding the parameters of group (*b*), variable independence (i.e. selected parameters should be uncorrelated with one another) suggests consideration of $f95$. In fact, it is the frequency-domain parameter that keeps all the off-diagonal elements of the parameter correlation matrix below 0.5.

Thus the selection takes us to the set of 2D parameters that we would propose for routine use in stabilometric practice, as listed in Table 5.

It is worth noting that $|90 - MDir|$ provides information about the AP and ML mutual relationship not achievable by any other 2D parameters. For this reason, it may be of particular interest in clinical applications, even if its use is still limited to date.

Similar considerations can be formulated about feature selection for the AP–ML parameters. In this case, distinctive properties emerging from the first PCs included

(i) dispersion of the AP time series, estimated by RMS_{AP}, MD_{AP}, and $RANGE_{AP}$
(ii) mean velocity along the ML direction MV_{ML}
(iii) contrast between the dispersions of the AP and ML time series. As the former are already accounted for by parameters in group (i), and to describe the different strategies that can take place for postural stabilisation (with the prevalence of AP or ML adjustments), the parameters to consider here are RMS_{ML}, MD_{ML} and $RANGE_{ML}$
(iv) relevant frequencies of the AP power spectral density curve, such as $f50_{AP}$, $f95_{AP}$ and CF_{AP}
(v) frequency dispersion of the AP power spectral density curve FD_{AP}
(vi) mean velocity along the AP direction MV_{AP}.

The guidelines already taken into account for the 2D parameters suggest selecting the features listed in Table 5.

This set explains the effectiveness (RMS_{AP}, RMS_{ML}) and the amount (MV_{ML}, MV_{AP}) of the regulatory activity taking place along the AP and ML directions, with a significant dominance of stability achieved in AP (RMS_{AP}) and control of workload achieved in ML (MV_{ML}). The relevance of MV_{ML} suggests a strong influence of ML control activity on postural adjustments (PRIETO *et al.*, 1996; ROCCHI *et al.*, 2002). Frequency-domain parameters portray only the AP power spectral density curve ($f95_{AP}$, FD_{AP}). Overall, this selection emphasises the prevalence of AP components in characterising the postural performance.

These results could be compared with those obtained by PRIETO *et al.* (1996), who investigated the redundancy of stabilometric parameters. They implicitly performed a feature selection process, on a set of parameters almost coincident with the one we used here, but the aim (the evaluation of the relative sensitivity of COP-based parameters to postural changes related to age) and the tools (Pearson correlation analysis and repeated measures ANOVA on 2D, AP and ML parameters altogether) of their analysis were quite different from the present ones. In this light, it is not surprising that the measures they recommend are somewhat different.

Table 5 Selected features, for 2D and AP–ML parameters, proposed for routine use in stabilometric practice

Selected features	
2D parameter set	AP–ML parameter set
RMS	RMS_{AP}
$f95$	MV_{ML}
$\|90\text{-}MDir\|$	RMS_{ML}
FD	$f95_{AP}$
	FD_{AP}
	MV_{AP}

4.2 *Comparison between 2D and AP–ML datasets*

The qualitative interpretation of PCs and the consequent feature selection process for 2D and AP–ML parameters offer parallel results, identifying the same COP properties, such as the size/dispersion of the oscillation, spectral attributes (shape and position of the power spectral density function) and the relative weight of AP and ML components. This similarity supports the validity of the PCs as interpreted so far, and it shows how information embodied in the parameters extracted from the AP and ML time series is implicitly present, even if more concise, in the parameters from the 2D time series. In fact, MV_{AP} and MV_{ML} do not find direct correspondence with any of the selected 2D parameters.

There is no a unique method to decide whether to estimate 2D or AP–ML parameters, but different considerations can influence this choice. It is noteworthy that the selected 2D features are more immune to biomechanical factors than the AP–ML parameters. This was confirmed by both the simple robust regression analysis proposed by CHIARI *et al.* (2002) and the multiple regression analysis presented in the current study (Table 2). These techniques demonstrated that the only parameter heavily influenced by the biomechanics is $f95$, which is correlated with height and two measures of base of support. When a normalisation procedure is not available, *CF* can be considered alternatively, as it represents a good compromise between dependence on the biomechanics and variable independence.

In contrast, AP–ML measures (if not normalised) are, in general, more influenced by biomechanical factors, as already reported in CHIARI *et al.* (2002) and confirmed by the results presented in Table 2, where r^2 is frequently higher than 0.2 for such measures. The more frequent dependence is on base-of-support measurements. For this reason, if no normalisation is undertaken, it is crucial at least to constrain foot position, in an attempt to limit inter-subject variability of base-of-support measurements. This suggestion is corroborated by the need to avoid cross-talk between AP and ML information, which can occur if the anatomical frame of the subject is not precisely aligned with the reference frame of the platform. On this matter, the value of 2D parameters is that they are more robust to imprecise orientation of the subject on the platform, the only exception being $|90 - MDir|$. Anyway, AP and ML measures can be more indicative of the true directional component of the sway. In fact, analysis of single COP components proved useful in predicting the risk of falling (MAKI *et al.*, 1994), sensitivity to changes in postural performance due to ageing (MAKI *et al.*, 1990) and to Parkinson's Disease (VIITASALO *et al.*, 2002), and in discriminating between predominant ankle or hip strategies (WINTER *et al.*, 1996).

4.3 *Future developments*

The target population of this study (healthy young subjects) does not allow us to predict whether the selected measures will be sensitive enough to changes related to age or pathologies affecting the postural control system. For this reason, future investigations will involve the analysis of composite populations of healthy young and healthy elderly subjects or healthy and pathological subjects. PCA points out the major factors behind the variability of a dataset, and therefore we do not expect, *a priori*, that the same set of parameters will be selected in every case. Different parameters could be the most sensitive to the specific conditions.

Nevertheless, in the literature, it has already been observed that some of the selected parameters may well discriminate between pathological and control subjects. In a previous study, *RMS*, $|90 - MDir|$ and f_{95} proved sensitive to Parkinson's Disease and to different treatments, including levodopa and deep brain stimulation (ROCCHI *et al.*, 2002). In another recent study, the amount of COP sway (here quantified by *RMS*) was found to detect postural abnormalities in diabetic subjects with peripheral neuropathy (YAMAMOTO *et al.*, 2001). Frequency dispersion was found to be sensitive to ageing (MAKI *et al.*, 1990), together with mean velocity in the AP direction (MAKI *et al.*, 1990; PRIETO *et al.*, 1996). This latter measure was also found to correlate with clinical scores characterising stroke patients (KARLSSON and FRYKBERG, 2000). Future results will allow us to define a minimum set of parameters that would be recommended for specific applications and to stimulate standardisation in the field.

The use of PCA for classification of movement patterns is well established (YAMAMOTO *et al.*, 1983; DELUZIO *et al.*, 1997). However, it is worth mentioning that, if the technique is optimum for dimension reduction, it does not mean that it is optimum for classification purposes (FUKUNAGA and KOONTZ, 1970). For this reason FUKUNAGA and KOONTZ (1970) proposed a data normalisation technique to use before PCA that may emphasise the differences between groups. The extension to composite populations may also take advantage of different techniques that have been proposed in the literature for dimension reduction, such as neural networks (CASTELLANO and FANELLI, 2000; ACCIANI *et al.*, 2003), second-order statistical methods (FODOR, 2002), higher-order statistical methods (independent component analysis, projection pursuit) (HYVARINEN, 1999) and methods based on wavelet transforms (COCCHI *et al.*, 2003).

In particular, independent component analysis may be a suitable technique for future developments of the present study. In fact, it may be seen as an extension of PCA, looking for statistical independence among components, instead of uncorrelation as PCA does. In addition, independent component analysis has already proved to be useful in the evaluation of physiological signals, such as EEG and EMG data (VIGARIO *et al.*, 2000; MCKEOWN and RADTKE, 2001).

The results of the present study and the number of the selected features could have interesting implications also for recently proposed approaches to stabilometric analysis, attempting to characterise the dynamic properties of the COP time series (COLLINS and DE LUCA, 1993; NEWELL *et al.*, 1997; CHIARI *et al.*, 2000*a*; *b*). These new approaches, based on a process model, are, by their nature, more selective and lead to a smaller set of parameters than simply treating the COP as a stationary signal and computing all its summary statistical properties. The models proposed so far have different dimensions in the parameter space, ranging from two (NEWELL *et al.*, 1997; CHIARI *et al.*, 2000*a*) to six (COLLINS and DE LUCA, 1993). Although these models may provide information different from that provided by traditional parameters, they could benefit from a deeper knowledge of the latter. In fact, once the distinctive properties of the COP time series are fully characterised, the choice between existing process models (or the definition of new ones) will be much easier, and their parameters will not be redundant and may be simpler to interpret, as well, for a clinical user.

Acknowledgments—This work was partially supported by the Research Program '*Evaluation of postural and locomotor ability in man for clinical applications*' funded by the Italian Ministry of the University and Scientific & Technological Research, 2000.

The authors would like to thank Dr JoAnn Kluzik for her comments regarding the manuscript and help in reviewing the English.

References

ACCIANI, G., CHIARANTONI, E., FORNARELLI, G., and VERGURA, S. (2003): 'A feature extraction unsupervised neural network for an environmental data set', *Neural Netw.*, **16**, pp. 427–436

BARATTO, L., MORASSO, P., RE, C., and SPADA, G. (2002): 'A new look at posturographic analysis in the clinical context: sway-density vs. other parameterization techniques', *Motor Control*, **6**, pp. 248–273

CASTELLANO, G., and FANELLI, A. M. (2000): 'Variable selection using neural-network models', *Neurocomputing*, **31**, pp. 1–13

CHIARI, L., BERTANI, A., and CAPPELLO, A. (2000*a*): 'Classification of human strategies in human postural control by stochastic parameter', *Hum. Mov. Sci.*, **19**, pp. 817–842

CHIARI, L., CAPPELLO, A., LENZI, D., and DELLA, C. U. (2000*b*): 'An improved technique for the extraction of stochastic parameters from stabilograms', *Gait Post.*, **12**, pp. 225–234

CHIARI, L., ROCCHI, L., and CAPPELLO, A. (2002): 'Stabilometric parameters are affected by anthropometry and foot placement', *Clin. Biomech.(Bristol, Avon)*, **17**, pp. 666–677

COCCHI, M., HIDALGO-HIDALGO-DE-CISNEROS, J. L., NARANJO-RODRÍGUEZ, I., PALACIOS-SANTANDER, J. M., SEEBER, R., and ULRICI, A. (2003): 'Multicomponent analysis of electrochemical signals in the wavelet domain', *Talanta*, **59**, pp. 735–749

COLLINS, J. J. and DE LUCA, C. J. (1993): 'Open-loop and closed-loop control of posture: a random-walk analysis of center-of-pressure trajectories', *Exp. Brain Res.*, **95**, pp. 308–318

DELUZIO, K. J., WYSS, U. P., ZEE, B., COSTIGAN, P. A., and SORBIE, C. S. (1997): 'Principal component models of knee kinematics and kinetics: normal vs. pathological gait patterns', *Hum. Mov. Sci.*, **16**, pp. 201–217

DIENER, H. C., DICHGANS, J., BACHER, M., and GOMPF, B. (1984): 'Quantification of postural sway in normals and patients with cerebellar diseases', *Electroencephalogr. Clin. Neurophysiol.*, **57**, pp. 134–142

FODOR, I. K. (2002): 'A survey of dimension reduction techniques'. LLNL Technical Report, UCRL-ID-148494

FUKUNAGA, K., and KOONTZ, W. L. G. (1970): 'Application of the Karhunen–Loeve expansion to feature selection and ordering', *IEEE Trans. Comput.*, **19**, pp. 311–318

HINTZE, J. L. (2000): 'NCSS user's guide' (NCSS, Kaysville, 2000)

HUFSCHMIDT, A., DICHGANS, J., MAURITZ, K. H., and HUFSCHMIDT, M. (1980): 'Some methods and parameters of body sway quantification and their neurological applications', *Arch. Psychiatr. Nervenkr.*, **228**, pp. 135–150

HYVARINEN, A. (1999): 'Survey on independent component analysis', *Neural Comput. Surv.*, **2**, pp. 94–128

JOLLIFFE, I. T. (1972): 'Discarding variables in a principal component analysis, I: Artificial data', *Appl. Statist.*, **21**, pp. 160–173

JOLLIFFE, I. T. (1986): 'Principal component analysis' (Springer-Verlag, New York, 1986)

KAISER, H. F. (1960): 'The application of electronic computers to factor analysis', *Educ. Psychol. Meas.*, **20**, pp. 141–151

KAPTEYN, T. S., BLES, W., NJIOKIKTJIEN, C. J., KODDE, L., MASSEN, C. H., and MOL, J. M. (1983): 'Standardization in platform stabilometry being a part of posturography', *Agressologie*, **24**, pp. 321–326

KARLSSON, A., and FRYKBERG, G. (2000): 'Correlations between force plate measures for assessment of balance', *Clin. Biomech.*, Bristol, Avon, **15**, pp. 365–369

KAUFMAN, K. R., and SUTHERLAND, D. H. (1996): 'Future trend in human motion analysis' in HARRIS, G. F., and SMITH, P. A. (Eds): 'Human motion analysis' (IEEE Press, Piscataway, NJ, 1996), pp. 187–215

MAKI, B. E., HOLLIDAY, P. J., and FERNIE, G. R. (1990): 'Aging and postural control. A comparison of spontaneous- and induced-sway balance tests', *J. Am. Geriatr. Soc.*, **38**, pp. 1–9

MAKI, B. E., HOLLIDAY, P. J., and TOPPER, A. K. (1994): 'A prospective study of postural balance and risk of falling in an ambulatory and independent elderly population', *J. Gerontol.*, **49**, pp. M72–M84

MCCABE, G. P. (1984): 'Principal variables', *Technometrics*, pp. 137–144

MCKEOWN, M. J. and RADTKE, R. (2001): 'Phasic and tonic coupling between EEG and EMG demonstrated with independent component analysis', *J. Clin. Neurophysiol.*, **18**, pp. 45–57

NEWELL, K. M., SLOBOUNOV, S. M., SLOBOUNOVA, E. S., and MOLENAAR, P. C. (1997): 'Stochastic processes in postural center-of-pressure profiles', *Exp. Brain Res.*, **113**, pp. 158–164

O'MALLEY, M. J. (1996): 'Normalization of temporal-distance parameters in pediatric gait', *J. Biomech.*, **29**, pp. 619–625
OLIVEIRA, L. F., SIMPSON, D. M., and NADAL, J. (1996): 'Calculation of area of stabilometric signals using principal component analysis', *Physiol. Meas.*, **17**, pp. 305–312
PRIETO, T. E., MYKLEBUST, J. B., HOFFMANN, R. G., LOVETT, E. G., and MYKLEBUST, B. M. (1996): 'Measures of postural steadiness: differences between healthy young and elderly adults', *IEEE Trans. Biomed. Eng.*, **43**, pp. 956–966
ROCCHI, L., CHIARI, L., and HORAK, F. B. (2002): 'Effects of deep brain stimulation and levodopa on postural sway in Parkinson's disease', *J. Neurol. Neurosur. Psych.*, **73**, pp. 267–274
TARANTOLA, J., NARDONE, A., TACCHINI, E., and SCHIEPPATI, M. (1997): 'Human stance stability improves with the repetition of the task: effect of foot position and visual condition', *Neurosci. Lett.*, **228**, pp. 75–78
VIGARIO, R., SARELA, J., JOUSMAKI, V., HAMALAINEN, M., and OJA, E. (2000): 'Independent component approach to the analysis of EEG and MEG recordings', *IEEE Trans. Biomed. Eng.*, **47**, pp. 589–593
VIITASALO, M. K., KAMPMAN, V., SOTANIEMI, K. A., LEPPAVUORI, S., MYLLYLA, V. V., and KORPELAINEN, J. T. (2002): 'Analysis of sway in Parkinson's disease using a new inclinometry-based method', *Mov. Disord.*, **17**, pp. 663–669
WINTER, D. A., PRINCE, F., FRANK, J. S., POWELL, C., and ZABJEK, K. F. (1996): 'Unified theory regarding A/P and M/L balance in quiet stance', *J. Neurophysiol.*, **75**, pp. 2334–2343
YAMAMOTO, S., SUTO, Y., KAWAMURA, H., HASHIZUME, T., KAKURAI, S., and SUGAHARA, S. (1983): 'Quantitative gait evaluation of hip diseases using principal component analysis', *J. Biomech.*, **16**, pp. 717–726
YAMAMOTO, R., KINOSHITA, T., MOMOKI, T., ARAI, T., OKAMURA, A., HIRAO, K., and SEKIHARA, H. (2001): 'Postural sway and diabetic peripheral neuropathy', *Diabetes Res. Clin. Pract.*, **52**, pp. 213–221

Authors' biographies

LAURA ROCCHI received her degree in Electronic Engineering in 1999, and the PhD in Biomedical Engineering in 2003, from University of Bologna, Italy, with a thesis titled "Quantitative techniques to characterize the postural behavior in normal and disease-altered conditions". She has currently a postdoctoral position at the University of Bologna. Her research interests are focused on motor and postural control, signal processing, data analysis, sensory integration models in balance control, signals modeling and the equipment used for movement analysis. She is involved in international collaborations, in particular with the Neurological Sciences Institute of OHSU (Portland, OR, USA) and with the Department of Biocybernetics of SAV (Bratislava, SK).

LORENZO CHIARI is currently Assistant Professor of Biomedical Engineering at the Department of Electronics, Computer Science and Systems of the University of Bologna. He received the Laurea degree in Electronic Engineering and the PhD in Biomedical Engineering from the University of Bologna, in 1993 and 1997, respectively. Subsequently he was a postdoctoral fellow and a research scientist at the same university. His main research activity, in cooperation with national and international scientific institutions, is in the area of biomechanics of human posture and movement, neuromuscular control of posture in health and disease, functional evaluation by movement analysis and the design and testing of motor rehabilitation equipment.

ANGELO CAPPELLO received the Laurea degree in Electronic Engineering from the University of Bologna, Italy, in 1977, and the PhD degree in Biomedical Engineering in 1988. Since 1981 he has been with the Department of Electronics, Computer Science and Systems of the University of Bologna, where he is presently a Professor in Biomechanical Engineering. He is a member of the board of the PhD program in Biomedical Engineering and a member of ISB and IEEE. His research activity, in cooperation with national and international scientific institutions, mainly concerns the application of measurement, modelling and control methods in orthopaedics and movement analysis in both physiological and assisted conditions. He is author of over 150 publications in peer-reviewed international journals, books and international conference proceedings.