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


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
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).

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 (HUFSCHEIDT 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 foot-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

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