 

# Feature Subset Selection for a Diagnostic Test of Human Balance

**Timo Tossavainen**

*Department of Computer Sciences, FI-33014 University of Tampere, Finland*

## Abstract

*We consider the problem of discriminating between patients with Ménière's disease and healthy control subjects on the basis of stabilograms measured under different sensory conditions using our virtual reality posturography system. In it, virtual reality techniques provide visual stimulation and a tilting force platform provides mechanical perturbations. We computed a set 156 parameters from each subject's measurements and selected a subset of these parameters for a simple 1 nearest neighbor classifier using the wrapper approach with the .632+-bootstrap error estimator. The results show that on the basis of the measurements we can obtain classification accuracy near 90%.*

*Keywords:*
Virtual reality; Posture; Balance; Classification; Feature subset selection; Bootstrap

## 1. Introduction

The central nervous system integrates sensory input and executes the constant corrective movements required for upright stance. Balance disorders, such as Ménière's disease and benign positional vertigo, impair the postural control process. They are common among the elderly and increase the risk of falls. One standard approach to detect balance disorders is to observe subjects maintain balance under different sensory conditions. For example, increased reliance on vision is typical in balance disorders. In Romberg's classical test, a test subject's stability in quiet stance with eyes open is compared to stability in quiet stance with eyes closed. Posturography uses force platform measurements to study balance. The force platform measures forces occurring under the feet of a test subject standing on it. A triangular arrangement of three force transducers under the platform measures vertical forces that are usually converted to the center of pressure (COP) on the platform and the total vertical force (vertical ground reaction force, VGRF) acting on the platform. Intuitively, VGRF is the instantaneous weight on the platform. A stabilogram (Figure 1) presents the COP and VGRF as a function of time. It summarizes movements of the body and the signals seem to have little regularity, particularly in the case of no external stimulation. There is currently no widespread consensus on how to analyze the signals [1]. Researchers usually compute different measures of stability, such as mean velocity, from stabilograms to evaluate a test subject's performance; see [1] for a review.

Our group works on a stimulation and measurement system for postural control and balance disorder research. We develop virtual reality (VR) techniques for visual stimuli and also use a mechanical tilting force platform for mechanical perturbations. Our system makes it easy to implement various tests to evaluate a test subject's balance under external

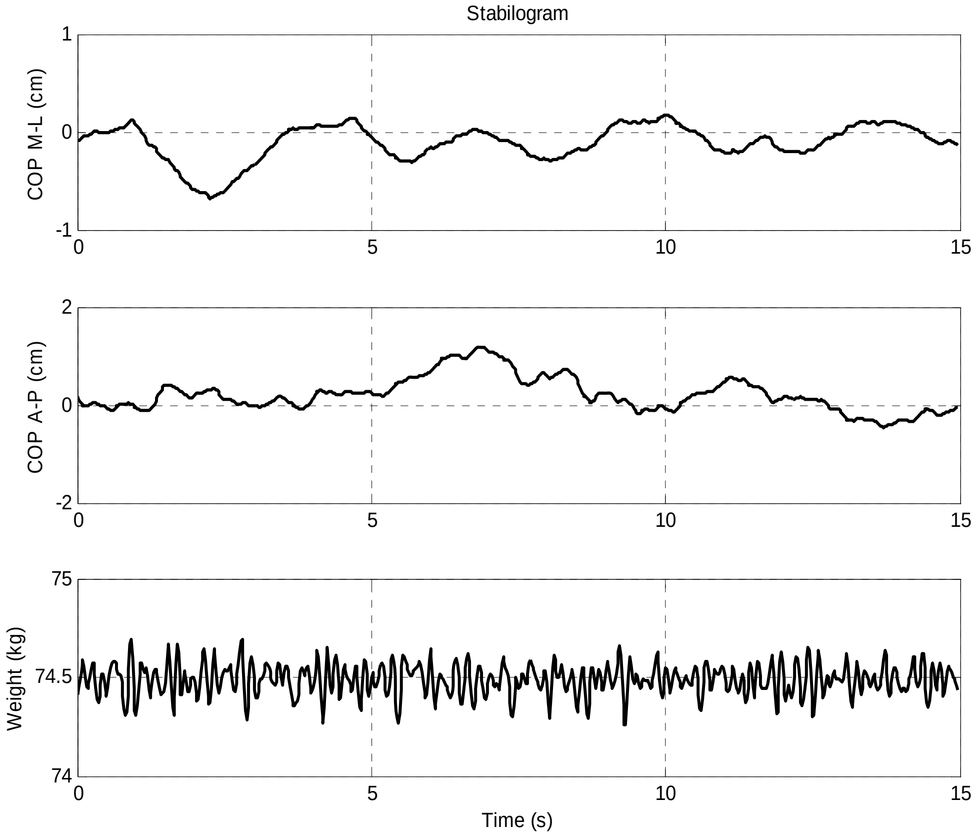


*Figure 1 - Stabilogram obtained from a Ménière patient in quiet standing with eyes open.*

stimulation. In the present study, we consider the detection of Ménière's disease based on a comprehensive balance test. The test setup consists of a series of tests where a subject is exposed to different visual and mechanical stimuli. Ménière's disease is difficult to detect as it is occurs in attacks and between the attacks the subject's balance may be close to normal. The posturographic measurements are difficult to interpret and it is not clear how to integrate the results from individual tests. We resort to a brute-force approach to select suitable features for a simple 1 nearest neighbor (1NN) classifier. The initial set of features comes from different models and theories of postural control from the literature and our own analysis. Our goals are two-fold: We wish to estimate the performance of the test setup and also to improve our understanding on how differences between the two groups are reflected in the stabilograms. The biggest problem with this approach is the scarcity of data coupled with abundance of possible models. Feature subset selection optimizes the error estimate over a large set of models and may overfit and also the choice of best subset has quite a lot of uncertainty. Thus, the subset with lowest estimated error may generalize poorly. We try to reduce this problem by cross-validating the selection procedure and by selecting a suitably small number of variables.

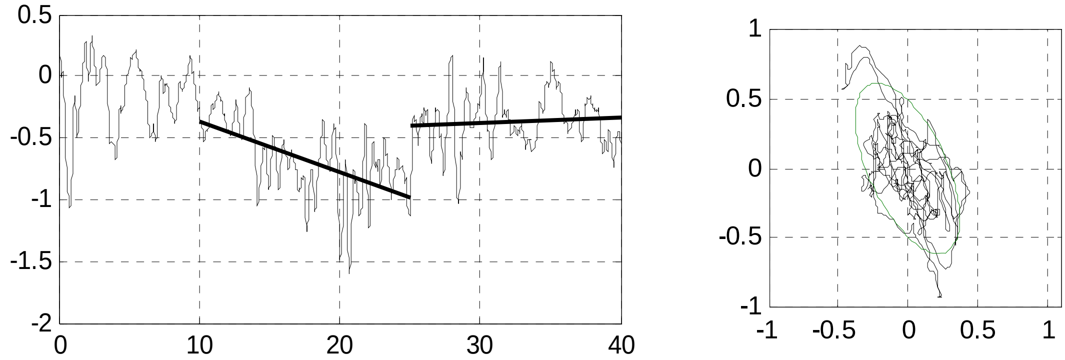


*Figure 2-Examples of parameters. Left: Parameters of lines fitted to COP M/L-position during cylinder rotation. Right: 95% confidence ellipse during counter-clockwise rotation of cylinder.*

## 2. Materials and Methods

We used data from 33 healthy control subjects (33 male, age mean 32.8, std. dev. 5.3 years) and 55 patients with Ménière's disease (19 male, 36 female, age mean 60.1, std. dev. 8.8 years) measured at the balance laboratory of the Tampere University Central Hospital hearing center. The data includes results from 6 different tests: quiet standing eyes open and eyes closed (15 s each), cylinder stimulus (2 sections of 30 s each), moving platform (15 s), and tunnel (60 s). During measurement with external stimulation, a safety harness was used to prevent falling. The data set contained a total of 77 Ménière cases, but we discarded cases with missing data from being unable to complete some of the tests.The visual stimuli were administered using a Virtual Research V8 head-mounted display (HMD). In the cylinder stimulus, the subject is standing inside a rotating cylinder formed by colored dots. The cylinder first rotates in the counter-clockwise direction for 30 s and then after a pause of 10 s in the clockwise direction for 30 s. Both rotations start and stop gradually with constant angular acceleration. The directions (CW and CCW) are analyzed separately. The moving platform test tilts the platform randomly; the subject is wearing the HMD but no visual stimulus is present. In the tunnel test, the subject is moving in a twisting tunnel. The platform also tilts first following the tunnels and then in the middle of the test inverts phase. The virtual reality stimuli are treated in more detail in [2].

The stabilograms were sampled at 50 Hz with a resolution of 16 bits. They were filtered bidirectionally using a $2^{nd}$ order Butterworth lowpass filter with a cutoff frequency of 10 Hz to remove noise. We computed a set of 156 real-valued parameters from each test subject's measurements. The parameters are mean velocity of COP on the platform (MV), 95% confidence ellipse area (CEA), low frequency band containing 90% of signal energy after DC-removal for each channel (FBX09, FBY09, FBM09), fraction of VGRF signal power below 1 Hz after DC-removal (VFPF1), parameters of lines fitted to result of cylinder measurement during 4 phases (slope, constant, and error of fit) and a set of 18 parameters per measurement from Collins and De Luca's diffusion analysis [3]. CEA and cylinder leaning are illustrated in Figure 2. The first three analysis methods are standard and the diffusion analysis is quite also quite popular. The CEA is computed by considering the COP points to be bivariate normally distributed; the computation of the ellipse was implemented using singular value decomposition. The FB09 and VGRF1 parameters were estimated using a simple periodogram power spectral density estimate. The VFPF1 parameter is unique to our test. In the difficult tests, moving platform and tunnel, test subjects tend to lean on the safety harness which results in slow and large fluctuations of weight on the platform. The parameter attempts to quantify this effect. The lines fitted to cylinder measurements parameterize leaning caused by the rotation. Subjects tend to lean in the direction of rotation, which can be interpreted as compensating for illusory self-motion in the opposite direction. A baseline correction was used when fitting the lines; we used the mean COP position during 1 s before the rotation starts (separate baselines for both rotation directions). The most significantly differing parameters from the full data set are given in Table 1 (outliers were removed based on 1.5 times interquartile range above and below upper and lower quartile; the number of outliers is indicated in the table). Due to lack of space we omit the results for the rest of the parameters.

The diffusion analysis considers COP movements a correlated random walk. The parameters are extracted from linear and log-log plots of mean squared displacements of COP in time. There are usually two scaling regions in the plots. They are thought to correspond to open-loop and closed-loop control. The 18 parameters are critical time $\Delta t_x$ and the displacement at the critical time $\Delta x^2$, scaling coefficients $D_{xs}$, $D_{xl}$, $H_{xs}$, $H_{xl}$ for linear and log-log short and long time regions and the same for y and r; see [3] for details. We computed the parameters from single measurements whereas in [3] they were averaged from 10 measurements. This probably affects the reliability of the results, but it would be impractical to repeat the whole test series 10 times for each test subject.

We used the wrapper approach [4] for feature subset selection; an estimate of the resulting classifier's performance is used to select a subset of variables. Obviously an exhaustive search through all subsets of variables is intractable, so a heuristic method has to be used. We used sequential forward floating selection search (SFFS) [5]. SFFS starts with the empty set of variables. The best new variable is added to the current selection. After addition, the variable whose removal most improves performance is removed if it results in improves on the best subset found so far of the same size. Variable removal is repeated until no improvement can be attained. The process is repeated until the desired number of variables has been included. The final selection is the best subset found during the search. Different search strategies for variable subset selection are discussed in [6]. We used a simple 1NN classifier that standardizes the variables and uses the Euclidean metric.

*Table 1-Variables with most significant differences*

| Test | Param. | Group | N | Out | Mean | Std | Min | Max |
|---|---|---|---|---|---|---|---|---|
| Tunnel | VFPF1 | Control | 33 | 1 | 0.41 | 0.18 | 0.11 | 0.84 |
| | | Ménière | 56 | 0 | 0.77 | 0.25 | 0.19 | 0.99 |
| Eyes Closed | VFPF1 | Control | 33 | 5 | 0.01 | 0.01 | 0.01 | 0.02 |
| | | Ménière | 76 | 7 | 0.03 | 0.02 | 0.01 | 0.09 |
| Platform | VFPF1 | Control | 33 | 0 | 0.46 | 0.18 | 0.07 | 0.79 |
| | | Ménière | 69 | 1 | 0.75 | 0.20 | 0.26 | 0.99 |
| Eyes Open | VFPF1 | Control | 33 | 1 | 0.02 | 0.01 | 0.00 | 0.04 |
| | | Ménière | 76 | 4 | 0.04 | 0.02 | 0.01 | 0.09 |
| Eyes Closed | $D_{ys}$ | Control | 33 | 2 | 18.29 | 9.58 | 4.27 | 41.26 |
| | | Ménière | 76 | 8 | 47.51 | 38.14 | 2.74 | 166.73 |
| Cyl. CW | VFPF1 | Control | 33 | 0 | 0.35 | 0.23 | 0.02 | 0.83 |
| | | Ménière | 67 | 1 | 0.63 | 0.18 | 0.26 | 0.98 |
| Eyes Closed | $D_{rs}$ | Control | 33 | 2 | 26.89 | 13.52 | 5.53 | 58.18 |
| | | Ménière | 76 | 10 | 62.31 | 46.55 | 4.61 | 225.00 |
| Eyes Closed | $\Delta r^2$ | Control | 33 | 0 | 43.30 | 23.07 | 8.17 | 93.29 |
| | | Ménière | 76 | 8 | 133.88 | 126.13 | 11.14 | 550.79 |
| Cyl. CCW | VFPF1 | Control | 33 | 0 | 0.35 | 0.22 | 0.02 | 0.79 |
| | | Ménière | 67 | 0 | 0.61 | 0.19 | 0.14 | 0.94 |
| Eyes Closed | $\Delta y^2$ | Control | 33 | 1 | 28.50 | 15.92 | 6.32 | 66.79 |
| | | Ménière | 76 | 10 | 80.05 | 71.36 | 6.73 | 295.60 |

We used the .632+-bootstrap [7], which is an advanced estimator of classifier error, as the evaluation function during the search. It uses the leave-one-out bootstrap estimator, ε(0), that estimates the performance of a classifier by building it on bootstrap samples and classifies samples left out of the bootstrap sample. The ε(0) is pessimistic as it does not use all the data; the .632-bootstrap tries reduce the bias by using a weighted average of ε(0) and resubstitution error. The .632+-bootstrap is an improvement of .632-bootstrap that takes overfitting into account. It adjusts the weighting between ε(0) and resubstitution error using estimated relative overfitting rate. We computed estimate from 500 bootstrap samples to keep the variance low during variable selection. After about 4-5 variables the search gets improvements that are close to the variance of the .632+-estimator resulting from the bootstrap samples and also starts to correspond to less than one case's contribution to the estimate, so we decided to select 5 variables. A subset containing fewer variables is also easier to interpret. The whole selection process was evaluated using leave-one-out cross-validation (LOOCV).

## 3. Results and Discussion

The LOOCV error of the selection process was 13.6%. Using all the data, SFFS selected a subset of 5 variables: eyes closed CEA, Δx2, VFPF1, and moving platform Dys, VFPF1. The .632+ bootstrap error estimate for the selection was 9.5%. The selected subset only contained features from two tests. Specifically, it did not contain information from tests using VR visual stimulation. However, the groups differ significantly in some of the variables obtained from the VR tests. The search procedure decided that the selected parameters generalize better according to .632+-bootstrap error estimates. The other parameters are probably still useful, for example, for differential diagnosis.

One interesting find here is that eyes closed VFPF1 is included in the parameters. The VFPF1 parameters are computed from the VGRF that is rarely analyzed and yet they seem to provide useful information for classification. The parameter detects slow fluctuations in the instantaneous weight on the platform. These fluctuations arise when the subject moves his center of gravity (COG) vertically, for example, by bending his knees or from the hip. One reason may be that the subjects lower their COG for better balance. Maintaining this slightly lower position would cause more vertical movements than standing with legs straight. It could also result from a different postural control strategy. In the case of the moving platform tests, VFPF1 mostly detects the use of extra support from the safety harness. The eyes closed CEA estimates how much the subject sways during quiet standing with eyes closed. Ménière patients being more visually dependent tend to sway more with eyes closed. The diffusion analysis parameters are more difficult to interpret. The parameter Δx2 can be roughly interpreted as how far the COP travels in one direction before turning back; for controls the distance is about 0.36 cm and for Ménière cases it is about 0.62 cm. The parameter Dys measures in a way how quickly the COP moves away from its current position in the anteroposterior direction. Here the value is greater for controls than for Ménière cases. This result may be an artifact from hanging on the safety harness as the extra support reduces the need to use ground reaction forces to maintain balance and probably also results in smaller COP movements.

## 4. Conclusion

We had measurements of 77 Ménière cases but 22 had incomplete data, because they were unable to complete all the tests and all control subjects were able to complete all the tests. Thus, if a test subject is unable to complete a test then there is reason to suspect a balance disorder. All of the subjects included in the variable selection had complete data. Taking this into account, we have shown that our test setup can reach near 90% classification accuracy with these cases. The generalizability of this result is probably reduced by the age and sex differences between the control and Ménière group and is also somewhat uncertain due to the small sample. The selected variable subset highlighted some of the differences between controls and Ménière patients. In the future, we will also consider finding differences between patient groups for differential diagnosis of different balance disorders, which is a much more difficult problem. Eventually we also have to consider handling missing data as many of the patients have difficulties completing the series of tests used in this study.

## Acknowledgements

The Finnish Work Environment Fund and Finnish Foundation for Advancement of Technology are thanked for financial support.

## References

[1] Baratto L, Morasso PG, Re C and Spada G. A new look at posturographic analysis in the clinical context: Sway density vs. other parameterization techniques. *Motor Control* 2002; 6(3): 248-273.

[2] Tossavainen T, Juhola M, Pyykkö I, Aalto H and Toppila E. Development of virtual reality stimuli for force platform posturography. *Int J Med Inform* 2003; 70:277-283.

[3] Collins JJ and De Luca CJ. Open-loop and closed-loop control of posture: A random walk analysis of center-of-pressure trajectories. *Exp Brain Res* 1993; 95:308-318.

[4] Kohavi R and John GH. Wrappers for feature subset selection. *Artif Intell* 1997 Dec; 97(1-2): 273-324.

[5] Pudil P, Novovicová J and Kittler J. Floating search methods in feature selection. *Pattern Recogn Lett* 1994; 15:1119-1125.

[6] Kudo M and Sklansky J. Comparison of algorithms that select features for pattern classifiers. *Pattern Recogn* 2000; 33:25-41.

[7] Efron B and Tibshirani M. Improvements on cross-validation: The .632+ bootstrap method. *J Amer Statistical Assoc* 1997 Jun; 92(438):548-560.

## Address for Correspondence

Timo Tossavainen
Department of Computer Sciences
FI-33014 University of Tampere
e-mail: tt@cs.uta.fi
http://www.cs.uta.fi/~tt