Online Detection of Freezing of Gait with Smartphones and Machine Learning Techniques

Sinziana Mazilu*, Michael Hardegger, Zack Zhu,
Daniel Roggen, Gerhard Tröster
Wearable Computing Lab
Swiss Federal Institute of Technology Zürich
*Corresponding author: sinziana.mazilu@ife.ee.ethz.ch
http://www.wearable.ethz.ch

Meir Plotnik, Jeffrey M. Hausdorff
Neurodynamics & Gait Research Laboratory
Tel Aviv Sourasky Medical Center, Israel
http://www.tasmc.org.il/e/

Abstract—Freezing of gait (FoG) is a common gait deficit in advanced Parkinson's disease (PD). FoG events are associated with falls, interfere with daily life activities and impair quality of life. FoG is often resistant to pharmacologic treatment; therefore effective non-pharmacologic assistance is needed.

We propose a wearable assistant, composed of a smartphone and wearable accelerometers, for online detection of FoG. The system is based on machine learning techniques for automatic detection of FoG episodes. When FoG is detected, the assistant provides rhythmic auditory cueing or vibrotactile feedback that stimulates the patient to resume walking.

We tested our solution on more than 8h of recorded lab data from PD patients that experience FoG in daily life. We characterize the system performance on user-dependent and user-independent experiments, with respect to different machine learning algorithms, sensor placement and preprocessing window size. The final system was able to detect FoG events with an average sensitivity and specificity of more than 95%, and mean detection latency of 0.34s in user-dependent settings.

I. INTRODUCTION

A. Freezing of Gait in Parkinson's Disease

FoG is a gait impairment common among patients with PD. According to a survey of 6620 PD patients by Macht et al. [1] 47% of the subjects reported regular freezing (28% experienced FoG daily). FoG is associated with falls ([2], [3]) and has substantial clinical and social consequences ([4], [5]). FoG is defined as a "brief, episodic absence or marked reduction of forward progression of the feet despite the intention to walk" [6]. Patients describe FoG as a feeling of having the feet glued to the ground and being temporarily unable to reinitiate gait. Episodes last between a few seconds and up to one minute [7]. While FoG can appear everywhere, it happens most often during turns, before gait initiation, in tight quarters such as doorways and in stressful situations ([5], [8], [9]). Treatment of PD patients with Levodopa reduces the FoG frequency during the ON state of medication, but like most gait deficits in PD patients, FoG is often resistant to pharmacological treatment [2].

A common non-pharmacological therapy for FoG is rhythmical auditory cueing [10]. In recent years, the option of using rhythmic auditory stimulation (RAS), e.g. with a metronome that provides a rhythmic ticking sound has gained support. RAS supports the patient to return to a more normal gait pattern. Unfortunately, the effectiveness of RAS wears off with time, so permanent cueing is not advised ([11], [12], [13]). For this reason, context-aware cueing systems are proposed where the auditory signal only starts in response to the occurence of FoG. Patient interviews in a preliminary study by Bächlin et al. [14] suggest that context-aware cueing may help to overcome freezing and reinitiate gait. The challenge in context-aware cueing is to reliably detect FoG episodes online using unobtrusive wearable sensor systems.

B. Mobile Phones as Wearable Assistants

A context-aware cueing system should be wearable and unobtrusive to the users as they have to wear it during daily-life activities. Recently, smartphones have evolved into a standard equipment in daily life due to their unobtrusive design. In addition, they are relatively cheap and offer high computational power. Therefore, smartphones are an interesting alternative to dedicated hardware in medical applications requiring wearable assistants. Users do not have to buy additional hardware as they often already possess a smartphone. Smartphones were, for example, used as assistants in fitness monitoring [15], heart rate monitoring [16], gait recognition [17], or to promote wellbeing [18]. Another example is the *iPhone* application *iFall* by Sposaro and Tyson [19], which uses the internal acceleration sensor for fall detection. In addition, modern smartphones offer a large number of internal sensors, including accelerometers, gyroscopes and magnetometers. A survey on mobile phone sensing is presented in [20]. To our knowledge, mobile phones have not yet been used in the context of online detection of FoG in PD.

Here, we propose the use of smartphones as a wearable device for FoG detection and perhaps treatment. This has several significant merits: *(1)* Economical – a FoG detection system on a smartphone will be economically beneficial compared to dedicated hardware [14]. *(2)* User friendliness – subjects with PD are typically older adults, a population that usually assimilates new technologies more slowly. However, it may hold that either this or the next generation of elderly people will be sufficiently familiar with smartphone technologies to adapt its use as a wearable assistant. *(3)* Tele-Medicine aspects – one of the future uses of automated detection of FoG

provides the treating physician with information on the FoG symptom burden. Having the FoG-detection device built in on a tele-communication platform has obvious advantages with respect to transfer of data from the patient to the clinic. *(4)* Social – a patient will feel more comfortable moving around with a smartphone, rather than a dedicated device that may draw unnecessary attention from their social environment.

C. Machine Learning for Analysis of Motion Data

FoG stands out as a typical motion pattern that is visually distinguishable from normal gait and has a unique frequency range. For example, when using wearable devices, motion patterns can be analyzed using acceleration features of body-mounted sensors. If sufficiently discriminative features are used, different motion patterns have different feature-space representations. In simple tasks where a low number of features is discriminative, decision boundaries can be set manually. This however becomes very tedious if the feature space is of high dimensionality.

Machine learning, on the other hand, offers methods for automated setting of decision boundaries, even in higher-dimensional problems. In contrast to manual thresholds, these boundaries are optimal in terms of decision accuracy for a set of training data. When sufficient training data is available, machine learning will outperform manual thresholding. Examples for successful application of machine learning methods to acceleration data are particularly abundant in the field of activity recognition [21], where activities such as running, walking or opening a door were detected using machine learning techniques with features computed from motion data. FoG is not an intentional movement but can be seen as a specific activity in the context of activity recognition. Thus, the analytical techniques applied for activity classification may be applicable to FoG detection.

D. Contributions

The goals of this work are *(1)* to improve the FoG-detection performance by using machine learning techniques and *(2)* to deploy the final system on a smartphone as an unobtrusive and inexpensive wearable assistant.

To reach this goal, we performed the following steps:

- We evaluated several machine learning algorithms on a real FoG dataset [14] in terms of FoG-detection accuracy and FoG-detection latency (the delay between a FoG start and its detection by the system).
- We performed experiments to optimize the FoG-detection accuracy and latency with respect to sensor placement and sensory-data window size. As a result of this optimization we propose a combination of machine learning algorithm, sensor location and window size that will achieve favorable results in terms of performance and detection latency, while considering the system's wearability and minimizing computational costs.
- Based on the previous evaluation steps, we built a system for online FoG detection using a smartphone and wearable sensors. The resulting Android application uses external accelerometers for FoG detection, but could easily be extended to use the internal smartphone sensors instead. The system provides auditory feedback whenever FoG is detected.

II. RELATED WORK

Han and collegues [22] made a first attempt to detect FoG episodes in PD patients by monitoring body acceleration using a 3-axis accelerometer. Freezing appearances were detected offline, by analyzing the differences in the recorded signals between freezing and normal gait. Moore et al. [23] analyzed offline accelerometer data from the left shank collected in a study with 11 PD patients. They observed high frequency components in the 3–8Hz band of the leg movement during FoG episodes that were not apparent in normal gait or volitional standing. Their algorithm obtained up to 89% accuracy and sensitivity for FoG detection. Delval and collegues [24] collected data from PD patients wearing goniometers, monitorized while walking on a treadmill. They obtained a sensitivity of $75 - 83\%$ and specificity above 95% for FoG detection by analyzing the frequency representation of knee joint signals. However, in all of these studies the FoG detection was done offline, so it is not actively helping the PD patients.

Bächlin et al. [14] developed a system for online FoG detection based on the algorithm of Moore [23]. The system contained three 3-axial accelerometers and a wearable computer. It was able to detect FoG episodes in user-dependent settings with a sensitivity of 88.6%, a specificity of 92.4% and a maximum latency of 2s. Whenever FoG was detected, the system provided a metronome ticking sound as feedback to the patient. The FoG-detection results were promising, but there is space for improvement. Also, manual adjustment of algorithm parameters was necessary to achieve optimal results. With machine learning, a patient-specific FoG-detection model can be built automatically and without need for manual parameter optimization.

Another online FoG-detection system based on the algorithm by Moore was presented by Jovanov et al. [25]. Using a 3-axis accelerometer and a wearable computer they detected FoG with an average latency of 332ms and maximum latency of 580ms. However, they do not provide information regarding the detection accuracy.

That machine learning techniques are helpful when analyzing motor fluctuations of PD patients was shown by Bonato et al. [26] and Patel et al. [27]. However, to the best of our knowledge, no one applied machine learning to the problem of FoG detection so far.

III. SYSTEM OVERVIEW

In this section we present our system for online detection of FoG episodes (Fig. 1). First, we describe the hardware components and explain how data from sensors is collected and processed in real-time. In the second part, we describe the online FoG-detection application that runs on top of our wearable system.

Fig. 1.   Online FoG-detection system.

Fig. 2.   The classification process for the Android online FoG detection application.

A. Mobile Platform

The proposed wearable assistant is composed of: *(1)* up to three external sensors, which could easily be replaced with the internal sensing platform of the mobile phone and *(2)* a smartphone as the wearable computer.

Similar to [14] we make use of the NTMotion:AccGyro sensors described in [28]. From the available sensor data, we utilize the wirelessly transmitted 3-dimensional acceleration data. The physical dimensions of the sensor node are 25 x 44 x 17 mm^3 and the weight is approximately 22g. On-board is a 300mAh Li-ion battery that lasts 6 hours per charge. Although the sensors are capable of sampling at up to 256Hz, we set the sampling rate to 64Hz for the purposes of this work, to match the FoG dataset sampling rate.

In contrast to the dedicated hardware in [14], we utilize the Nexus One smartphone as a wearable computer. Specifically, we implemented an application on top of the Android platform that acts as the hub of the system. In addition to gathering internal accelerometer readings, our application communicates with multiple NTMotion sensors simultaneously via Bluetooth. We ported an existing sensor reading acquisition software – Java Bluetooth Gateway[1] – for the Android platform. As a modular component, the sensor communication and packet parsing functionalities were implemented separately as a helper class which sends acceleration readings to our Android application. Together with the data of the phone's internal accelerometer, these readings are piped into a queue for window-based classification.

The contents of the queue are read at fixed time intervals (e.g. 1s) for preprocessing of the raw acceleration data. The stream is segmented into windows. Features are computed for each window and sensor axis. For this work, we compute 4 time-domain features (mean, variance, standard deviation and entropy), the signal energy and the two features proposed in [14] and [23] for FoG detection. See Table I for further details.
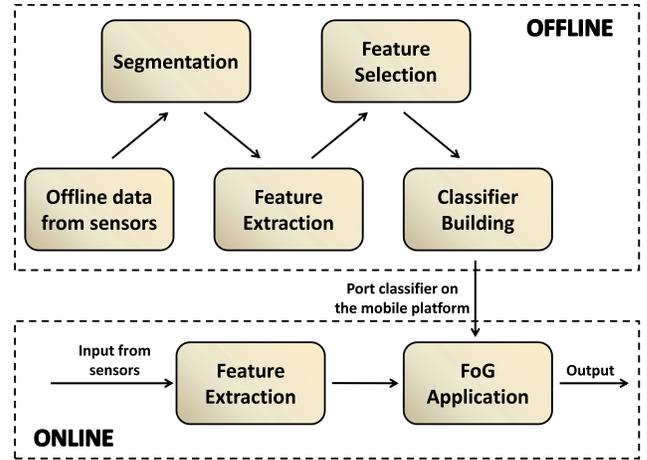
B. Online FoG Application

The online FoG-detection application has two components (Fig. 2): A FoG-detection classifier, built offline on a base station, and the real-time FoG-detection app on the smartphone. The two parts are explained as follows:

- **Offline**: The FoG-detection classifier is trained offline with previously collected data (from the same patient or from different patients). Data for offline training must be labeled, i.e. we need to know for each window whether it is a member of the FoG class or not. As a first (optional) step, we perform a feature selection to choose the most discriminative features for distinguishing FoG from normal gait. For this work, we used Correlation-based Feature Subset Selection as described in [29]. Only the selected features are then used for classifier training with supervised machine learning techniques from the Weka data mining suite [30]. The classifier is serialized and ported to the smartphone using the Weka serialization API. The data structure representing the classifier is converted to byte format and stored in a file, which is then copied to the mobile-phone application resources. Since this building and serialization process is performed offline on a computer, there are no additional costs in terms of latency and computational power for the wearable system.
- **Online**: In the online phase of the process, the Android app uses the deserialized classifier built in the previous step to detect FoG events in real-time from sensor data. We use a modified version of the Weka API for Android[2] to deserialize and use the classifier. The FoG-detection application calls the sensor communication module to get the last recorded data, preprocesses it to extract the necessary features and passes them to the classifier. When a FoG event is detected, the application provides feedback to the patient. Depending on the patient's preferences

---

[1]http://code.google.com/p/javabtgateway/

[2]https://github.com/rjmarsan/Weka-for-Android

TABLE I
USED FEATURES WITH SYMBOLS AND BRIEF DESCRIPTIONS

Feature	Description
Mean	The DC component (average value) of the signal in the window
Standard Deviation	Mean deviation of the signal compared to the average in the window
Variance	The square of the standard deviation
Entropy	Measure of the distribution of frequency components
Energy	Sum of the squared discrete FFT-component magnitudes of the signal, divided by the window length for normalization
Freeze Index	Power of the freeze band (3–8Hz) divided by the power in the locomotor band (0.5–3Hz) as used in the FoG-detection algorithm from [23]
Power	The sum of the power in the freeze and locomotion band – this feature was used by Bächlin et al. to distinguish volitional standing from FoG [14]

and favorable response in terms of gait rehabilitation, this can be an RAS signal (metronome ticking sound) or vibrotactile feedback (i.e. the smartphone vibrates).

The application respects the patient's privacy by not saving collected sensor data. It is built as an Android service, meaning that the user can start and stop the application anytime. It runs in the background and therefore does not interfere with the use of other applications on the smartphone.

IV. SYSTEM EVALUATION

A. Dataset

We tested our system on the DAPHNet dataset [14], the result of a study carried out by the Laboratory for Gait and Neurodynamics, Department of Neurology, Tel Aviv Sourasky Medical Center (TASMC). The DAPHNet dataset contains data collected from 10 PD patients that experienced regular FoG in daily life. Data was recorded using three 3D-acceleration sensors attached to the shank (above the ankle), the thigh (above the knee) and to the lower back of each subject. The sensors recorded at 64Hz and transmitted the acceleration data via a Bluetooth link to a wearable computing system that was located at the lower back of the subjects.

Patients completed two sessions of 10-15 minutes each. Both recording sessions consisted of three walking tasks:

- Walking back and forth in a straight line, including several 180-grade turns
- Random walking in a reception hall space, including a series of initiated stops and 360 degrees turns. Subjects should stop or turn in different directions
- Walking simulating activities of daily living (ADL), which included entering and leaving rooms, walking to the kitchen, getting something to drink and returning to the starting room with a cup of water

Motor performances varied strongly among the participants. While some subjects maintained regular gait during nonfreezing episodes, others walked slowly and very unstable. Overall, 8h20min of data were recorded. To label FoG episodes in the data set, synchronized video recordings were analyzed by physiotherapists. The start of a FoG event was considered when the gait pattern (i.e., alternating left-right stepping) was arrested, and the end of a FoG was defined as the point in time when the pattern was resumed. In total, 237 FoG episodes were identified (23.7±20.7 per patient). The duration of FoG episodes was between 0.5s and 40.5s (7.3±6.7s). 50% of the FoGs lasted for less than 5.4s and 93.2% were shorter than 20s.

B. Experiments and Evaluation

We evaluated different sensor placements, window lengths and machine learning algorithms in terms of detection accuracy and latency in user-dependent and user-independent experiments.

For supervised machine learning we tested the following algorithms: Random Trees (RT), Random Forests (RF), decision trees and pruned decision trees (C4.5), Naive Bayes (NB), Bayes Nets (BN), k-nearest neighbor with one neighbor (KNN-1) and two neighbors (KNN-2), Multilayer perceptron (MLP), boosting (AdaBoost) and bagging with pruned C4.5 trees as base classifiers.

The reference for all our evaluations is the video annotation provided by physiotherapists in the DAPHNet dataset. The detection performance is based on window evaluation, i.e. for each window the classifier output is compared to the reference annotation. Windows that are correctly labeled as FoG are counted as *True Positive (TP)*. We define as *False Positives (FP)* the FoG detections in episodes where physiotherapists did not identify FoG. *False Negatives (FN)* are windows where the system failed to detect FoG during FoG episodes in the reference. The remaining windows are correctly labeled as no FoG and are therefore *True Negatives (TN)*.

We measure the *Sensitivity* ($Sens = \frac{TP}{TP+FN}$), which represents the proportion of correctly detected FoG windows to the total of reference FoG windows. The *Specificity* ($Spec = \frac{TN}{TN+FP}$) measures the proportion of correctly detected no-FoG windows to all reference no-FoG windows. Additionally we report *F1-measure* and *area under the curve (AUC)* in the ROC space [31] as performance measures to evaluate our system. The F1-measure takes into account the precision ($Prec = \frac{TP}{TP+FP}$) and recall rate (identical to specificity) for each class (in our case the FoG class and the *null-* or no-FoG class).

	1s window				4s window			
Classifier	**Sens (%)**	**Spec (%)**	**F1 (%)**	**AUC (%)**	**Sens (%)**	**Spec (%)**	**F1 (%)**	**AUC (%)**
Random Forest	**97.76**	**99.75**	**98.35**	**99.82**	**99.54**	**99.96**	**99.75**	**99.98**
C4.5	93.47	99.38	95.94	97.40	98.43	99.84	99.04	99.35
Näive Bayes	48.06	98.66	73.62	93.08	41.87	99.73	71.13	95.31
MLP	77.46	97.29	82.94	95.84	84.99	98.63	91.17	98.41
AdaBoost with C4.5	**98.35**	**99.72**	**98.37**	**99.85**	**99.69**	**99.96**	**99.78**	**99.98**
Bagging with C4.5	97.57	99.59	97.60	99.84	99.41	99.91	99.53	99.97

Furthermore, we evaluated the algorithms in terms of latency, which is the time delay between the start of a FoG episode in the reference and the start of a detected FoG episode by the application.

V. EVALUATION RESULTS

A. User Dependent

In user-dependent experiments, both training and testing data are from the same patient. We performed 10-fold cross validation using feature selection and classification methods enumerated in Subsection IV-B for each patient data from the DAPHNet dataset. We report comparative results on average performance measures on the whole dataset for RF, C4.5, NB, MLP, boosting and bagging methods, in case of window lengths of 1s and 4s. The 4s-window results are presented for comparison with the state-of-the-art FoG detection system in [14], which also uses 4s windows. Best results were obtained with boosting of pruned decision trees and RF. The use of such large windows increases the latency of online FoG detection, therefore we investigated shorter windows as well and report results for window length of 1s. Detection performance was only slightly decreased for 1s windows (98.35% sensitivity and 99.72% specificity for boosting of pruned decision trees).

Compared to [14], that report results of 88.6% sensitivity and 92.4% specificity for patient-dependent experiments with window length of 4s, we obtained better performances with all tested machine learning algorithms, except Näive Bayes and Bayes Nets. This was expected, as machine learning techniques offer more possibilities to explore the data properties (i.e. computation of more features, selection of most discriminative features, automatic setting of decision boundaries). The results are slightly biased due to the random selection of training and testing data in 10-fold cross-validation, which may lead to selection of subsequent and therefore correlated samples to both subsets. Since FoG episodes of a single user are anyway correlated and have similar feature-space representation, we assume that this bias is small. However we will address the issue in future work.

According to the results presented in Table II the best performances were obtained by AdaBoost (with pruned C4.5 as base classifier) and Random Forest classifiers. Comparing the ensemble methods, we observed that boosting classifiers obtain slightly better performances than bagging classifiers. However

all the ensemble methods tested (boosting, bagging, Random Forests) obtain better results than the single classifiers.

All results were obtained using 10-11 features on average (10.3 features for 1s windows and 11.6 features for 4s windows, selected from the 63 features computed in the preprocessing step). The most discriminative features, regardless of sensor location, were the mean and standard deviation as time-domain features, and the physiological features – freeze index and power computed as in [14].

For all further experiments we selected C4.5 and RF as classification algorithms. Pruned C4.5 was chosen because of its simplicity and still good detection results and RF for its very good detection performances.

B. Latency Results

As mentioned before, we refer to latency as the time between a FoG episode starts and the time when the system detects it. Here, we only discuss the latency which is inherent to the machine learning algorithm with corresponding window size. We neglect further delays caused by sensor data transmission, preprocessing and feedback generation, assuming that these contributions are small.

Fig. 3 shows a section of the acceleration signals of Patient 02 with the ground truth annotation (FoG or normal gait) and the FoG-detection results of our system. We observe that the system detects FoG episodes shortly after the FoG event starts.

The system's latency depends on the classification algorithm chosen and on the sampling window length used. Table III depicts the latency results for patient-dependent experiments with 1s and 4s window lengths and two types of classification algorithms: pruned C4.5 and RF. There is a trade-off between performance of the detection algorithm, in terms of sensitivity and specificity, and the detection latency. While RF obtains better performances for FoG detection than C4.5 trees, the detection latency is higher. Fig. 4 shows that with C4.5 most FoG episodes are detected shortly after their start. For RF classifiers, FoGs are typically detected after approximately half the window length. However, the differences are small and the maximum latency for 1s windows in all experiments was only 0.718s. We further discuss the latency results and the relation to window size in Section VI.

C. User Independent

The online FoG-detection system was also evaluated using leave-one-patient-out cross validation. The classifiers were
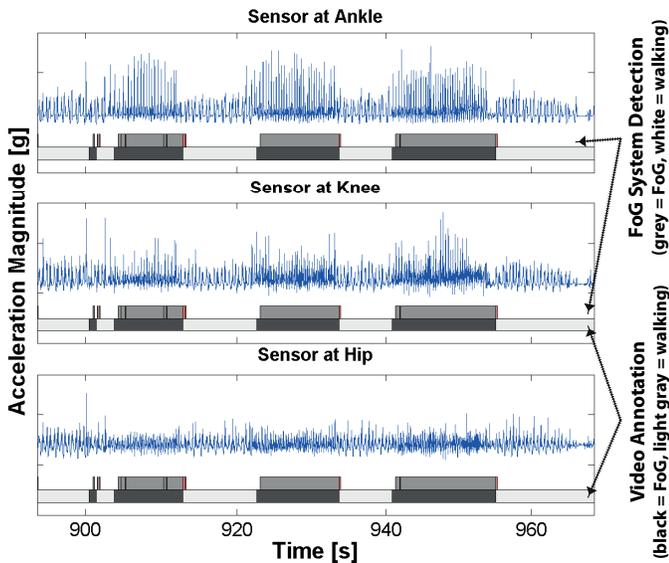
Fig. 3. A signal extract from Patient 02 using data from ankle, knee and hip sensors, together with the ground-truth labels and the FoG-detection-system labels.

TABLE III
LATENCY RESULTS FOR PATIENT DEPENDENT EXPERIMENTS.

Window	Classifier	Mean (s)	Std (s)	Max (s)
1s	C4.5	**0.235**	0.175	**0.578**
	RF	0.346	0.169	0.718
4s	C4.5	**1.085**	0.731	**2.016**
	RF	1.653	0.59	2.047

Fig. 4. Histogram indicating the correlation of detection latency between the FoG start and the moment of detection.

trained on features selected from $N-1$ subjects and performance was tested on the remaining subject. We report results for RF classifiers only, because of lack of space. As in the previous set of experiments, we performed feature selection on the data before training the classifiers. Again, mean, standard deviation, freeze index and power were the most frequently selected features. For 1s-window experiments, we report 62.05% sensitivity and 95.15% specificity. For 4s-windows, the results slightly improved – 66.25% sensitivity and 95.38% specificity. In [14], the proposed system obtained 73.1% sensitivity and 81.6% in the same 4s-window-based evaluation. However, the algorithm from [14] allowed for 2s tolerance after the start of a FoG episode, which explains the slightly better sensitivity results.

The comparatively poor results for user-independent FoG detection are a result of the large variability in motor performance, caused by different walking styles among subjects of the DAPHNet study. In this case, training classifiers on general data does not always result in good performances when tested on a specific subject. For some patients the general classifier worked well, e.g. for Patient 09 where window-based sensitivity was 98.19% and specificity 91.17% with 26 out of 27 FoG events detected (mean latency of 0.41s), for others the general classifier failed (e.g. for Patient 01 where sensitivity was 20.53% only). Bächlin et al. approached
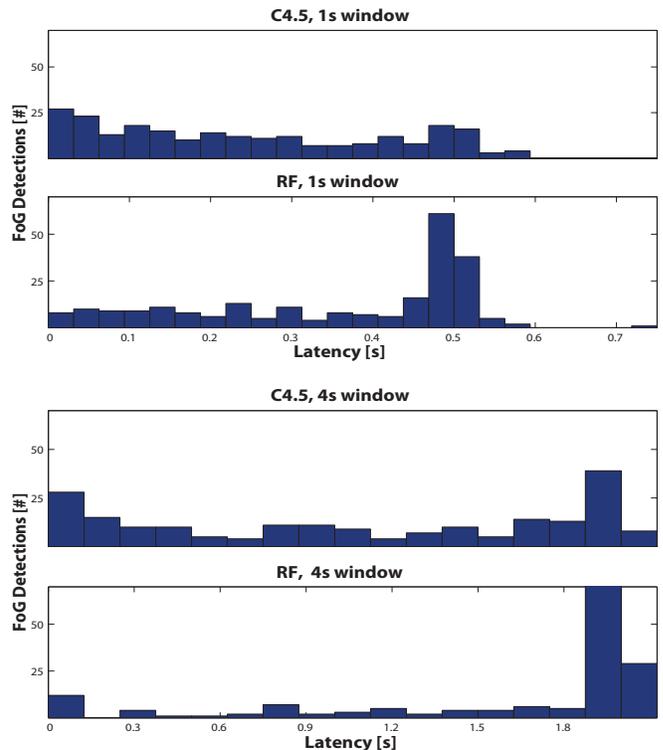
the problem by dividing the patients in two groups (smooth walkers and intensified stepping walkers) with individually optimized thresholds and thus improved detection results in user-independent experiments [14]. In future, we plan to apply transfer learning techniques [32] to improve the user-independent performance of the system.

VI. PERFORMANCE OPTIMIZATION

A. Sensor Placement Characterization

In this subsection, we analyze the FoG-detection performance with respect to different sensor placements and orientations. Our goal is to determine the best sensor location taking into account the trade-off between wearability of the system and the detection performance. There are visually obvious differences between the signals from different sensors. For example the signal of the sensor at the hip is more damped than the signal of the sensor at the ankle. Still, the leg motion is visible at both locations (Figure 3).

From all 63 features, only a subset of features extracted from a single sensor and a combination of sensor axes was used as input to feature selection and classification for this experiment. Table IV shows results for 12 combinations of three sensor positions (ankle, knee, and hip) and four combinations of the sensor axes (horizontal forward axis x, vertical axis y, horizontal lateral axis z, and features extracted from all the three axes together x & y & z). Evaluation results in terms of sensitivity and specificity are given for 1s windows in patient-dependent experiments, for C4.5 and RF classifiers. We

	x	y	z	x & y & z
Ankle				
Sens	92.65%	89.18%	91.85%	98.21%
Spec	99.26%	98.84%	99.02%	99.76%
Knee				
Sens	89.36%	87.71%	88.58%	97.94%
Spec	98.86%	98.80%	98.80%	99.73%
Hip				
Sens	88.81%	83.68%	90.77%	98.63%
Spec	98.77%	98.26%	99.02%	99.83%

Fig. 5.   Online FoG-detection performance and detection latency vs. window length, for RF classifiers.

present and discuss the RF results only, but the observations also apply for C4.5.

As expected, the best results were achieved when using features collected from all three axes of the sensors. When using data from a single sensor, best results in terms of both sensitivity and specificity were obtained for the hip sensor. However, the performances for the three positions did not differ much. Taking into account only a single axis resulted in slightly lower detection performance.

These findings are promising as they indicate that *(1)* a single sensor is sufficient for FoG detection and *(2)* this sensor can be placed at a convenient location for the patient. Even the use of internal sensors of a smartphone placed in the pockets of a patient is possible, making the system even more wearable and patient-friendly.

B. Performance vs. Window Length Optimization

The latency of the online FoG-detection system is a function of the window length used for feature extraction. We analyzed the potential of window length optimization by plotting *(1)* the FoG-detection performance in terms of minimum between sensitivity and specificity versus window length, and *(2)* the detection latency versus window length. We measured average sensitivity, specificity and latency for C4.5 and RF classifiers in user-dependent settings, for window sizes from 0.5s to 8s, in steps of 0.5s. Further we detail only the results obtained with RF classifier (Fig. 5), C4.5 results having similar properties.

When increasing the window up to a maximum length of 3s the detection performance increases. For larger windows, noise in the computation of features is reduced, leading to better discrimination between FoG events and normal gait. For window lengths above 3s the window size has no significant influence on the detection performance. The mean latency of FoG detection increases linearly with the size of the windows. A good trade-off between latency and detection performance is reached for window lengths of 1s.

VII. DISCUSSION

Following the detailed algorithm evaluation presented in the previous sections, we chose C4.5 and RF classifiers to be used in the smartphone application. Pruned C4.5 classifiers are small, easy and quick to port to the smartphone. Also,
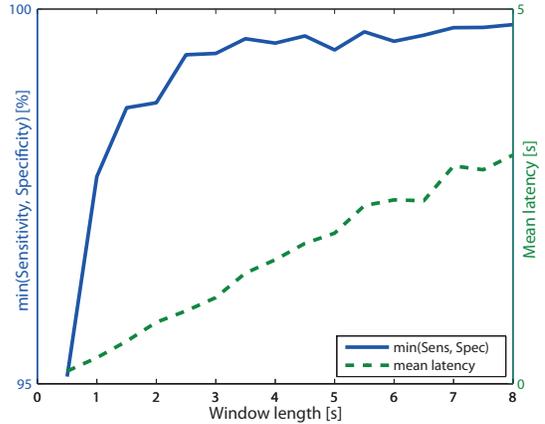
they obtain good results in terms of FoG-detection accuracy and latency. On the other hand, RF is more complex, but offers even better FoG-detection performance.

For feature extraction we chose a window size of 1s, which is a satisfying trade-off between detection accuracy and latency. Regarding the system components, experiments showed that a single accelerometer is sufficient for FoG detection and various locations (ankle, knee and hip) result in similar performances. Therefore, the current system uses only a single body-worn acceleration sensor together with the smartphone as the wearable computer. In future, the possibility of replacing external sensors with smartphone sensors will be investigated.

The algorithms presented in this paper worked much better in user-dependent experiments. Patients using this system will therefore need to record a short training set with expert supervision for labeling of FoG episodes. The classifier is then trained offline and without any additional user input.

The system provides RAS or vibrotactile feedback whenever FoG is detected. To reduce the number of single false positive windows, which would cause a start of cueing even though no FoG is present, we applied a median filter to the raw classifier output. This increases the detection latency, but is necessary to reduce the number of "false alarms". Using a 31-sample median filter, we detected 100% of the 237 FoG episodes in the DAPHNet dataset with only 9 false alarms. For a 15-sample median filter 58 false positive FoG detections were found.

VIII. CONCLUSION

We proposed a system for online FoG detection with wearable accelerometers and a smartphone using machine learning techniques. The system provides feedback to the patients whenever a FoG event is detected. To our knowledge, this is the first time that machine learning algorithms were used to detect FoG episodes online. Our system was capable of detecting all FoG events (237 out of 237 events in the DAPHNet dataset) with an average window-based sensitivity of 98.35% and an average specificity of 99.72%. When com-

paring directly to the state-of-the-art system for online FoG detection [14], we obtain an average sensitivity of 99.69% and an average specificity of 99.96% compared to 88.6% sensitivity and 92.4% specificity for identical window size. Furthermore we were able to detect all FoG events with a mean latency of 0.34s (and a maximum of only 0.71s).

The analysis of different sensor locations showed promising results in making the system less intrusive and more wearable. The good results obtained with the sensor at the hip open the door to the use of the internal smartphone sensing platform for online FoG detection. Thus a mobile phone might be sufficient for assisting PD patients with FoG.

In the future we plan to build a user-independent FoG-detection algorithm that automatically adapts to the patient's specific gait using domain adaptation algorithms. Also, additional physiological and heuristic features might allow to better distinguish FoG from the many variations of Parkinsonian gait.

ACKNOWLEDGMENTS

REFERENCES

[1] M. Macht, Y. Kaussner, J. C. Möller, K. Stiasny-Kolster, K. M. Eggert, H.-P. Krüger, and H. Ellgring, "Predictors of freezing in Parkinson's disease: A survey of 6,620 patients," *Movement Disorders*, vol. 22, pp. 953–956, 2007.

[2] B. R. Bloem, J. M. Hausdorff, J. E. Visser, and N. Giladi, "Falls and freezing of gait in Parkinson's disease: A review of two interconnected, episodic phenomena," *Movement Disorders*, vol. 19, pp. 871–884, 2004.

[3] M. Latt, S. Lord, J. Morris, and V. Fung, "Clinical and physiological assessments for elucidating falls risk in Parkinson's disease," *Movement Disorders*, vol. 24, pp. 1280–1289, 2009.

[4] O. Moore, C. Peretz, and N. Giladi, "Freezing of gait affects quality of life of peoples with Parkinson's disease beyond its relationships with mobility and gait," *Movement Disorders*, vol. 22, pp. 2192–2195, 2007.

[5] N. Giladi and J. M. Hausdorff, "The role of mental function in the pathogenesis of freezing of gait in Parkinson's disease," *Journal of the Neurological Sciences*, vol. 248, pp. 173–176, 2006.

[6] J. G. Nutt, B. R. Bloem, N. Giladi, M. Hallett, F. B. Horak, and A. Nieuwboer, "Freezing of gait: moving forward on a mysterious clinical phenomenon," *The Lancet Neurology*, vol. 10, pp. 734–744, 2011.

[7] J. D. Schaafsma, Y. Balash, T. Gurevich, A. L. Bartels, J. M. Hausdorff, and N. Giladi, "Characterization of freezing of gait subtypes and the response of each to levodopa in Parkinson's disease," *European Journal of Neurology*, vol. 10, pp. 391–398, 2003.

[8] D. Cowie, P. Limousin, A. Peters, M. Hariz, and B. L. Day, "Doorway-provoked freezing of gait in Parkinson's disease," *Movement Disorders*, 2011.

[9] J. Spildooren, S. Vercruysse, K. Desloovere, W. Vandenberghe, E. Kerckhofs, and A. Nieuwboer, "Freezing of gait in Parkinson's disease - The impact of dual-tasking and turning," *Movement Disorders*, vol. 25, pp. 2563–2570, 2010.

[10] T. Hashimoto, "Speculation on the responsible sites and pathophysiology of freezing of gait," *Parkinsonism & Related Disorders*, vol. 12, pp. 55–62, 2006.

[11] T. C. Rubinstein, N. Giladi, and J. M. Hausdorff, "The power of cueing to circumvent dopamine deficits: A review of physical therapy treatment of gait disturbances in Parkinson's disease," *Movement Disorders*, vol. 17, pp. 1148–1160, 2002.

[12] A. Nieuwboer, "Cueing for freezing of gait in patients with Parkinson's disease: A rehabilitation perspective," *Movement Disorders*, vol. 23, pp. 475 – 481, 2008.

[13] E. Cubo, S. Leurgans, and C. G. Goetz, "Short-term and practice effects of metronome pacing in Parkinson's disease patients with gait freezing while in the on state: randomized single blind evaluation," *Parkinsonism & Related Disorders*, vol. 10, pp. 507 – 510, 2004.

[14] M. Bächlin, M. Plotnik, D. Roggen, I. Maidan, J. M. Hausdorff, N. Giladi, and G. Tröster, "Wearable assistant for Parkinson's disease patients with the freezing of gait symptom," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, pp. 436–446, 2010.

[15] F. Buttussi and L. Chittaro, "MOPET: A context-aware and user-adaptive wearable system for fitness training," *Artificial Intelligence in Medicine*, vol. 42, pp. 153 –163, 2008.

[16] D. Albert, B. R. Satchwell, and K. N. Barnett, "Heart monitoring system usable with a smartphone or computer," Patent 20 110 301 435, 2011.

[17] C. Nickel, C. Busch, S. Rangarajan, and M. Mobius, "Using hidden markov models for accelerometer-based biometric gait recognition," in *Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on*, 2011.

[18] N. D. Lane, M. Mohammod, M. Lin, X. Yang, H. Lu, S. Ali, A. Doryab, E. Berke, T. Choudhury, and A. Campbell, "BeWell: A smartphone application to monitor, model and promote wellbeing," in *the 5th International Conference on Pervasive Computing Technologies for Healthcare*, 2011.

[19] F. Sposaro and G. Tyson, "iFall: An Android application for fall monitoring and response," in *Engineering in Medicine and Biology Society. Annual International Conference of the IEEE*, 2009.

[20] N. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. Campbell, "A survey of mobile phone sensing," *IEEE Communications Magazine*, vol. 48, pp. 140–150, 2010.

[21] S. J. Preece, J. Y. Goulermas, L. P. J. Kenney, D. Howard, K. Meijer, and R. Crompton, "Activity identification using body-mounted sensors - a review of classification techniques," *Physiological Measurement*, vol. 30, pp. 1–33, 2009.

[22] J. H. Han, W. J. Lee, T. B. Ahn, B. S. Jeon, and K. S. Park, "Gait analysis for freezing detection in patients with movement disorder using three dimensional acceleration system," in *Engineering in Medicine and Biology Society. Proceedings of the 25th Annual International Conference of the IEEE*, 2003.

[23] S. T. Moore, H. G. MacDougall, and W. G. Ondo, "Ambulatory monitoring of freezing of gait in Parkinson's disease," *Journal of Neuroscience Methods*, vol. 167, pp. 340–348, 2008.

[24] A. Delval, A. H. Snijders, V. Weerdesteyn, J. E. Duysens, L. Defebvre, N. Giladi, and B. R. Bloem, "Objective detection of subtle freezing of gait episodes in Parkinson's disease," *Movement Disorders*, vol. 25, pp. 1684–1693, 2010.

[25] E. Jovanov, E. Wang, L. Verhagen, M. Fredrickson, and R. Fratangelo, "deFog - a real time system for detection and unfreezing of gait of Parkinson's patients," in *31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009.

[26] P. Bonato, D. Sherrill, D. Standaert, S. Salles, and M. Akay, "Data mining techniques to detect motor fluctuations in Parkinson's disease," in *Engineering in Medicine and Biology Society. 26th Annual International Conference of the IEEE*, 2004.

[27] S. Patel, K. Lorincz, R. Hughes, N. Huggins, J. Growdon, D. Standaert, M. Akay, J. Dy, M. Welsh, and P. Bonato, "Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, pp. 864–873, 2009.

[28] D. Roggen, M. Bächlin, J. Schumm, T. Holleczek, C. Lombriser, G. Tröster, L. Widmer, D. Majoe, and J. Gutknecht, "An educational and research kit for activity and context recognition from on-body sensors," in *Proc. IEEE Int. Conf. on Body Sensor Networks (BSN)*, 2010.

[29] M. A. Hall and L. A. Smith, "Practical feature subset selection for machine learning," 1998.

[30] I. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*, J. Gray, Ed. Morgan Kaufmann Publishers, 2005.

[31] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, pp. 861–874, 2006.

[32] S. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, pp. 1345 –1359, 2010.