# GymSkill: A Personal Trainer for Physical Exercises

Andreas Möller, Luis Roalter, Stefan Diewald,
Johannes Scherr, Matthias Kranz
*Technische Universität München*
*Munich, Germany*
*andreas.moeller@tum.de, roalter@tum.de, stefan.diewald@tum.de,*
*scherr@sport.med.tum.de, matthias.kranz@tum.de*

Nils Hammerla, Patrick Olivier, Thomas Plötz
*Culture Lab*
*Newcastle University*
*Newcastle, United Kingdom*
*nils.hammerla@ncl.ac.uk, p.l.olivier@ncl.ac.uk,*
*thomas.ploetz@ncl.ac.uk*

***Abstract*—We present GymSkill, a personal trainer for ubiquitous monitoring and assessment of physical activity using standard fitness equipment. The system records and analyzes exercises using the sensors of a personal smartphone attached to the gym equipment. Novel fine-grained activity recognition techniques based on pyramidal Principal Component Breakdown Analysis (PCBA) provide a quantitative analysis of the quality of human movements. In addition to overall quality judgments, GymSkill identifies interesting portions of the recorded sensor data and provides suggestions for improving the individual performance, thereby extending existing work. The system was evaluated in a case study where 6 participants performed a variety of exercises on balance boards. GymSkill successfully assessed the quality of the exercises, in agreement with the professional judgment provided by a physician. User feedback suggests that GymSkill has the potential to serve as an effective tool for motivating and supporting lay people to overcome sedentary, unhealthy lifestyles. GymSkill is available in the Android Market as 'VMI Fit'.**

***Keywords*-activity recognition, skill assessment, health, mobile, quantitative time-series analysis**

## I. Introduction

Encouraging people to exercise more is key to maintaining or regaining personal health but, unfortunately, difficult to achieve in practice. One barrier to exercising is that lay people often are insufficiently knowledgeable about effective and safe physical exercises. Maintaining a long-term exercise regime requires high levels of motivation and time demand, which often conflicts with people's busy lifestyles. It is well established that access to a personal trainer has a significant impact on both adherence to a physical exercise program, and the quality of the exercise undertaken [1]. Personal trainers continuously monitor the exercises and both provide individualized advice and motivate the trainee. They also play an important role in rehabilitation, e.g. exercise programs for muscle recovery after surgery, where the need for advice regarding effectiveness and safety is even greater.

Unfortunately, the availability of specialized trainers is limited and in most cases simply too expensive to provide over extended periods of time, and where financial factors are not a barrier, personal privacy preferences can be (i.e., the perception of the potential for embarrassment). Furthermore, where exercise equipment is involved, the use of a personal trainer significantly decreases the likelihood of inappropriate and even dangerous use of certain pieces of fitness equipment in either gym environments or at home.

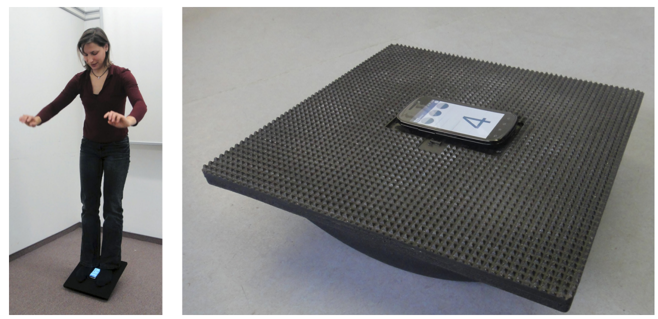


Figure 1. GymSkill: A smartphone-based personal trainer for monitoring and assessing physical exercises using fitness equipment. Left: A subject training with GymSkill on a rocker board. Right: the phone with the GymSkill application running placed on the board.

We present GymSkill, a smartphone-based personal trainer for ubiquitous monitoring and assessment of physical exercises performed using standard fitness equipment. Figure 1 illustrates a typical use-case for GymSkill in which a trainee is performing exercises on a balance board. The system utilizes the embedded sensing capabilities of a phone placed on the balance board (accelerometer and gyroscope) to record the exercises. The quality of recorded data is automatically analyzed, i.e., the skill of the trainee is assessed. The system provides basic situated (auditive and visual) feedback during exercising and, moreover, performs retrospective automatic assessments of the *quality* of the performed exercises [2]. It provides a global quantitative judgment of physical exercises in the form of an aggregated skill metric, which is the basis for competitive evaluations of physical exercises. It is, thus, ideal for tracking individual progress over the course of a long-term training program.

In addition to this summary feedback, GymSkill also analyzes physical exercises at a fine-grained level. By means of a multi-variate time-series analysis procedure, the system highlights critical portions of exercises that exhibit *quality breakdowns*, i.e., where improvements are needed or dangerous situations have been detected. Consequently, GymSkill not only provides information about the quality of an exercise session, but also potential reasons for quality differences. This is a unique capability that distinguishes it from existing personal health and fitness systems. This combination of global and local analysis directly corresponds to two key functions of a human personal trainer who both provides a thorough reactive analysis of individual exercises and keeps track of the overall training progress.

Our technical contributions are twofold. First, we present an integrated framework for *recording and analyzing physical exercises* that utilizes a smartphone attached to a piece of fitness equipment. As a concrete example, we focus on balance boards in particular; these are widely used in both physical training and rehabilitation [3]. GymSkill and the underlying analysis methods are, however, general approaches and not limited to balance board exercises. The smartphone app provides the 'infrastructure' for recording and analyzing sensor data as well as the user interface for progress tracking and cueing within a training program, and for the automated feedback on quality of the exercises.

Second, we present new approaches for activity recognition that go beyond the state-of-the-art by *quantitatively analyzing* the quality of sensor data. Given a segmentation of the sensor data into exercises, provided by standard activity recognition procedures [4], the overall quality of the exercises is assessed by comparing the empirical distribution of recorded data with expected behavior using Kullback-Leibler divergence (KLd) along with other statistical properties. For the subsequent fine-grained analysis of sensor data streams, a new approach is developed utilizing a pyramidal Principal Component Breakdown Analysis (PCBA) technique. This unsupervised analysis procedure highlights structural changes in multi-variate time-series data. The combination of global and local analysis provides an effective assessment of the overall quality of recorded time-series data in terms of measuring the deviation from, for example, a 'gold standard' template, and identifies reasons for quality differences.

In combination with a repository of pre-defined equipment-specific exercises, GymSkill enables personalized, ubiquitous training. We evaluated the system in a practical case-study, in which 6 trainees over a period of 5 consecutive days, with different levels of experience of fitness equipment, performed exercises on standard rocker boards (a common form of balance board). In total, the feedback from users, and comparison of assessments provided by a professional physician and the system, show that GymSkill has the potential to mimic key functions of a personal trainer.

## II. Background and Related Work

### A. Automated Personal Health and Fitness (PHF) Systems

Technical resources to support physical activity have evolved in research and commercial use. Medical devices and sensors, such as oximeters and heart rate monitors, formerly reserved to professional medical use, can now easily be connected to mobile phones. Sports devices like GPS watches, heart rate sensors and foot pods allow users to monitor and keep track of their activities. For home use, dedicated hardware platforms (e.g. the Nintendo Wii + balance board) encourage users to physical activity through fun and social commitment, but do not focus on medically correct exercising. Further examples of commercial PHF systems are activity loggers like activPal, or FitBit, which monitor activity data throughout the day. Many applications supporting physical activity can be found in smartphone app stores. They e.g. track running or cycling activities, and often provide exercising and workout instructions. However, assessing the skill level and providing targeted tips how to improve is not covered yet.

Several research approaches for sensor-augmented physical training devices have been presented for outdoor and indoor activities, such as skiing [5] or tracking free-weight exercises with accelerometers in a glove and on the waist [6]. In [7], a sensor-augmented balance board that supports the improvement of the equilibrium sense and muscular training [8] is described. Sensor data from the board was recorded and visual feedback on correct performance decreased the necessary amount for supervision. In the SESAME project (SEnsing for Sports And Managed Exercise)[1], mobile sensor-based approaches for coaching performance improvement of athletes are investigated.

### B. Quantitative Analysis of Human Activities

A common setting in activity recognition (AR) is that triaxial accelerometers or gyroscopes are worn on the body or embedded into objects of daily use. The recorded multivariate sensor streams undergo frame-wise analysis to infer the activities that were performed by the subject. Often, simple yet effective methods suffice to obtain good recognition accuracies, rendering information about *what* subjects are doing readily available. However, so far relatively little work has been invested into a further, detailed analysis of these segmented activities. Extracting their characteristics, i.e., how *well* these activities were performed, would be beneficial to a variety of applications.

One exemplary domain for quality analysis of human activities is the quantification of human motor performance, which is important for a number of disciplines and domains. The acquisition of motor abilities is a well established research field in the wider biological and physiological research context. However, still a thorough theoretic foundation for *motor skill assessment* has not yet been developed. Consequently, only few technical systems exist that directly assess human motor performance. Settings often are rather constrained, e.g. the acquisition of surgery skills [9], and the assessment of professional athletes' skills (e.g. in tennis

[1]SESAME. http://www.sesame.ucl.ac.uk/

[10]). So far, no generalization technique has been developed for applying approaches across application boundaries.

Another domain of quantitative analysis of human movement is gait analysis, which is so far almost exclusively restricted to computer vision based approaches [11]. These systems provide an in-depth analysis of peoples' behavior but the methods employed are not straightforward to adopt for the analysis of other activities and environments. Reasons for this lie either in restrictions of sensor-resolution or in dependencies on domain knowledge, which makes generalization difficult to achieve. Approaches for sensor data analysis from real-life environments that are independent of excessive domain knowledge therefore remain open research problems in this novel field of activity recognition.

## III. Automatic Assessment of Physical Exercises

Success in encouraging people to adopt healthier lifestyles by promoting regular physical exercises requires lowering the barrier for people to engage in sport activities they find embarrassing or are not used to. In addition to this 'persuasion', constant monitoring and timely intervention if necessary is required for health and safety reasons.

GymSkill is an automatic recording and analysis system for easy-to-use and correct assessment of physical exercises. The employed algorithm can basically be used for monitoring and analyzing virtually any kind of exercises that are based on reoccurring movements with quality constraints related to smoothness and efficiency criteria. For the sake of clarity, we focus on balance board[2] exercising as an example, which trains, e.g., ankles and the equilibrium sense, and contributes to the overall fitness.

Using digital technologies for constantly and automatically monitoring and analyzing fitness programs and for tracking the progress of an exercising person represents the conceptual equivalent to a personal trainer. It allows faster improvements, as targeted training becomes possible. Individual feedback addresses problem areas or exercises that need particular improvement. That way, the training progress is adapted to individual needs. Compared to its human counterpart, a digital personal trainer has the advantages of ubiquitous and permanent availability along with negligible costs. Furthermore, it helps preserving the privacy and dignity by allowing for exercising without supervision and in a more comfortable environment than a gym. Arguably, automated skill assessment and individualized feedback also increases and maintains motivation [12], which is crucial for effectiveness, since training needs to be done regularly.

We position our quality assessment approach on top of raw sensor data acquisition, data processing and activity detection. In previous work, analysis algorithms have already been described for sensing systems to automatically detect the activities performed by the wearer, and to provide statistics about these activities of interest. The latter usually corresponds to summaries of e.g. cycled or ran distances, estimates of energy expenditure and heart rate curves. However, although desirable, so far the automatic, expert-like judgment of the quality of physical exercises at a fine-grained level has not yet been addressed.

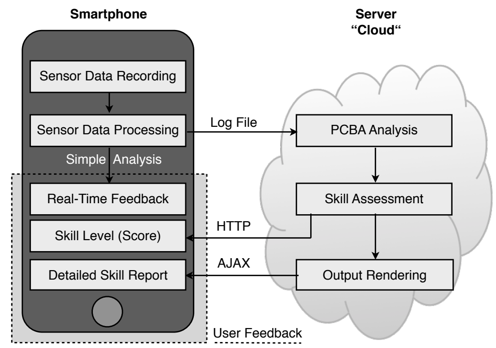



Figure 2. Overview of the GymSkill system: The smartphone records sensor data during exercising, which are processed by a server to generate a skill assessment. Besides simple real-time feedback (cueing), the sophisticated exercise analysis can be reviewed later in different levels of detail.

### A. System Overview

GymSkill currently consists of two parts: a smartphone application and an analysis component running on a server (see Figure 2). During the exercises, the smartphone (attached to the fitness equipment) records acceleration and orientation data and the user already gets immediate feedback based on an initial analysis like the remaining number of repetitions or whether the optimal deflection angle was exceeded. This feedback is presented in well-readable graphics (readable while exercising) and as audio feedback. The logged information is sent to a server, where the automatic analysis is performed in terms of a complex 'after-the-fact', i.e., retrospective assessment of the exercise correctness, which serves as indicator for the user's skill. The skill level (score) is calculated and sent back to the smartphone, indicated as visual feedback in different steps between 'thumbs up' and 'thumbs down'. Additionally, more sophisticated graph visualizations are available, which allow the review of the exercise over time and help the user to identify problems.

### B. Smartphone-Based Infrastructure

GymSkill is implemented as Android application[3] and consists of an exercise database, the sensor recording functionality and the skill assessment presentation. During the exercise, the smartphone (app running) is placed on top of the balance board so that it can record all of its movements. The user can then either work through a complete training plan, or select single exercises for individual improvement. The built-in training plan has been composed by a sports medicine specialist and consists of 20 exercises of increasing

[2]see, e.g., www.thera-band.com/store/products.php?ProductID=17

[3]https://market.android.com/details?id=de.tum.ei.vmi.fit

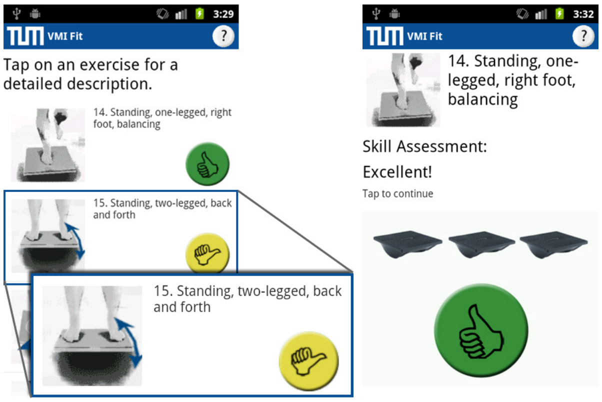



Figure 3. User interface of the GymSkill application. *Left:* Skill assessment after exercising, based on the evaluated sensor information. *Right:* In the exercise list, exercises that need further training can easily be identified.

difficulty. During the performance, incorrect movements (e.g. excessive displacement of the board) are signalled. The number of repetitions is shown on the display in large readable numbers, and a sound notification signals exercise completion (particularly useful for exercises types where glancing at the phone is difficult). Sensor data (acceleration, magnetic field, and orientation) are logged on the device and after each exercise completion submitted to a server for the skill assessment.

### C. Automatic Exercise Assessment

To assess exercise quality, we estimate global quality measures that cover important aspects of the performed motion. Advised by an expert clinician, the following attributes were defined as necessary for an automated assessment system:

**Smoothness and continuity of movement:** For continuous exercises, as they are typical for gym-based training, it is important to maintain smooth motion. In order to remain relatively independent of the particular exercise and to avoid the excessive use of prior knowledge, a novel local assessment approach has been developed (next section).

**Global motion quality:** Each exercise requires the user to perform particular motion sequences. The assessment on how well these motions were performed is crucial for the assessment of the quality of the performed task.

**Usage of board's degrees of freedom:** If a task requires the user to fully displace the board along at least one degree of freedom, the fraction to which he uses this opportunity while avoiding extreme postures (e.g., touching the ground) provides a valuable measure for exercise performance.

The goal of the automated assessment is to estimate measures for the aforementioned aspects and to combine them into a single performance criterion or metric. Aiming at transferability of the method, the amount of parameters and prior knowledge used is limited as much as possible.

Before the actual analysis, the recorded orientation values (azimuth, pitch, roll) are normalized to a common value range with zero-mean. Deviation from mean translates into

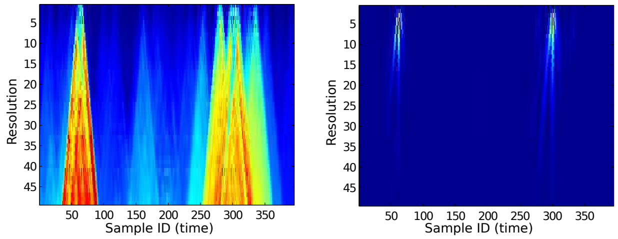
**Input:** seq. $\boldsymbol{S} = S_1 \dots S_L$ with $S_i \in \mathbb{R}^n$ (e.g. $n = 3$ for $a, p, r$ orientation), max. window length $W$, reconstr. perc. $p$
**Output:** matrix $\boldsymbol{P}[L \times W]$ containing localized reconstruction errors for all $L$ samples at $W$ levels
**for all** $w = 2 \dots W$ **do**
  input dim. $s = n \times w$
  extract all analysis windows $\{\boldsymbol{\mathcal{W}}\}$ with length $w$ using sliding window proc. with shift = 1, dim. = $s$:
    $\boldsymbol{S} \mapsto \{\boldsymbol{\mathcal{W}}_i\}_{i=1 \dots L-(w+1)}$, with $\boldsymbol{\mathcal{W}}_i \in \mathbb{R}^s$
  estimate
    PCA($\{\boldsymbol{\mathcal{W}}_i\}$) = $\{(\lambda_j, \mathbf{v}_j)\}, j = 1 \dots w, \lambda_j \geq \lambda_{j+1}$
  estimate target dimensionality $d$ for reconstr. quality $p$:
    $d = k : \left(\sum_{i=1}^{k} \lambda_i / \sum_{i=1}^{w} \lambda_i\right) \leq p$
  **for** $pos = 1; pos < L - (w+1); pos++;$ **do**
    extract analysis window $\boldsymbol{\mathcal{W}} \in \mathbb{R}^s$ at position $pos$
    project to $d$-dim. sub-space: $\boldsymbol{\mathcal{W}} \mapsto \boldsymbol{\mathcal{W}}^d \in \mathbb{R}^d$
    reconstruct original window: $\boldsymbol{\mathcal{W}}^d \mapsto \boldsymbol{\mathcal{W}}' \in \mathbb{R}^d$
    calc. reconstr. error, assign to position $pos$ at level $w$
      $P^w_{pos} = ||\boldsymbol{\mathcal{W}}' - \boldsymbol{\mathcal{W}}||$
    normalize pyramid range (across dataset)
  **end for**
**end for**
**return** $\boldsymbol{P}$

Algorithm 1: Principal Component Breakdown Analysis

$[-180°, +180°]$ and is mapped to $[-1, +1]$. For both pitch and roll, the calibrated 0-positions are taken as idle positions while the calibrated maximal displacement angles are mapped to $[-1, 1]$, i.e., $[\alpha_{min}, \alpha_{max}] \rightarrow [-1, +1]$.

*Local Analysis:* In addition to an overall analysis of quality of exercises, GymSkill aims for unveiling interesting and informative portions of the recorded sensor data streams. When analyzing re-occurring (ideally smooth) movements, 'interesting' refers to sections where the sensor data appear unusual compared to the rest. This is e.g. the case when a participant hesitates or gets stuck while exercising. Unlike standard techniques for time-series analysis (e.g. [13], [14]) our approach processes sensor data of arbitrary dimensionality. This is crucial since flattening sensor data to one-dimensional sequences (e.g. using the Euclidean norm) can destroy potentially important information of the original signals. Our basic assumption is that sensor data for a particularly analyzed movement should share certain (unknown) statistical properties. Unusual portions of a sequence violate this assumption and can thus be identified as such.

Algorithm 1 describes the local quality assessment algorithm – *Principal Component Breakdown Analysis (PCBA)*, based on a PCA of a sensor data sequence utilizing local neighborhoods. Using a sliding window technique, analysis windows of length $w$ are extracted and a PCA model is learned and applied to project all frames to a lower-dimensional sub-space. Its dimensionality is determined by the analysis of the eigenvalue spectrum. The target dimensionality is chosen based on a pre-defined threshold for reconstruction quality (typically 95% of the variance shall be preserved). Using the lower-dimensional projection, the orig-




(a) PCA-based assessment (b) Wavelet scalogram

Figure 4. Local skill assessment using PCBA. Yellow and red areas represent quality breakdowns in exercises. [best viewed in color]

inal frames are reconstructed. The resulting reconstruction errors are used as a measure for the quality of the underlying movement, which is sample-wise assigned to the original sequence. PCA models will require more eigenvectors to preserve a certain amount of variance of the original data if the underlying signal is less regular. By fixating the target dimensionality and implicitly analyzing the modeled variance, an effective quality assessment is gained.

For unsupervised analysis the 'correct' frame length, i.e., the size of the neighborhood that needs to be analyzed for discovering potential characteristic breakdowns, needs to be known. Unfortunately, this information is typically not available for practical applications. To overcome this dilemma we employ a multi-scale approach by performing quality assessment on a pyramidal adjacency representation of sensor values with increasing frame lengths, which is comparable to the general idea of Wavelet analysis or the approach presented in [14]. In contrast to the latter, however, the PCA based approach focuses on self similarity and breakdowns w.r.t. global characteristics. Figure 4(a) shows the pyramidal representation of an exemplary exercise segment where characteristic breakdowns (large values indicated in yellow and green) can clearly be seen. For example at around $t \approx 60$ and $t \approx 300$ the participant hesitated in her circular rocker board exercise, which results in substantial increases of the PCA-based reconstruction errors at different scales. For comparison the –far less informative– Wavelet scalogram of the same data is shown as well (Figure 4(b)).

As a measure for continuity the mean of the reconstruction error along a specific scale $c$ is derived for feedback generation, denoted as $D_c$:

$$D_c = 1/W \sum_{w=1}^{W} \boldsymbol{P}_{c,w}, \tag{1}$$

where $\boldsymbol{P}_{c,w}$ represents the element in row $c$ and column $w$ of the matrix $\boldsymbol{P}$ (see Algorithm 1).

*Global Analysis:* To derive the global motion quality, the motion axis, which shows the largest energy during the exercise (providing the dominant signal) is estimated. If the training device offers $d$ degrees of freedom, the following analysis is performed for the $d$ most dominant axes. Using standard Kaplan-Meier estimation [15] the empirical distribution function is derived and then integrated to form

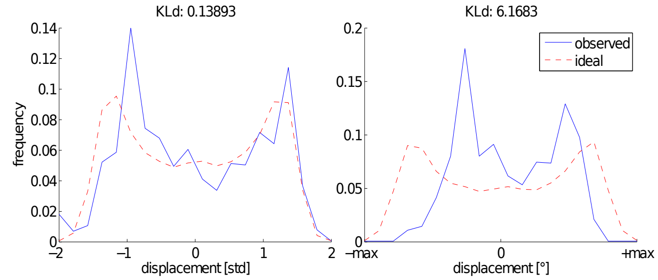



Figure 5. The same empirical distribution (blue, solid line) in comparison with an ideal distribution (red, dotted line) for the normalized (left) and un-normalized case (right). Although the *shape* is similar (left) there are significant differences in how far the distributions span the available space.

the empirical cumulative distribution function (ECDF). This ECDF is compared to a gold-standard template, represented as an ideal distribution function, in two different settings.

For rocking tasks, the ECDF of a sigmoid with an amplitude of $0.8 \times \alpha_{max}$ with added Gaussian noise is used as the ideal distribution. For balancing tasks, a normal distribution with a variance of $0.1 \times \alpha_{max}$ is employed. The parameters of both distributions are motivated by insights of a physician. For other cases where no such prior knowledge is available, the performance of a very skilled athlete or professional can be used to estimate the ideal behavior empirically.

In the first setting, the two normalized distributions (zero-mean, unified variance) are compared using standard Kullback-Leibler divergence (KLd):

$$D_g = \int_{-\infty}^{+\infty} P^n(x) \log \frac{P^n(x)}{\mathrm{ECDF}^n(x)} dx, \tag{2}$$

where $P^n$ and $\mathrm{ECDF}^n$ correspond to the normalized ideal and empirical distributions.

Second, the un-normalized distributions are compared (see Figure 5). The KLd in this setting gives insights into how well the degrees of freedom on the board are utilized during the exercise (for rocking tasks) or how well the subject can keep his balance (for balancing tasks):

$$D_a = \int_{-\alpha_{max}}^{+\alpha_{max}} P(x) \log \frac{P(x)}{\mathrm{ECDF}(x)} dx. \tag{3}$$

For translating these distances into suitable metrics, the logarithms of $D_g$, $D_a$ and $D_c$ (cf. Equation 1) are transformed with the sigmoid function:

$$\forall i \in \{g, a, c\}: \quad M_i = \frac{1}{1 + \exp\left(\log\left(D_i \times c_i\right)\right)} \tag{4}$$

where $c_g, c_a, c_c \in \mathbb{R} > 0$ are parameters that determine the sensitivity of the assessment and the impact of each single attribute. All estimated measures $M_g$, $M_a$ and $M_c$ lie between 0 and 1 where 1 corresponds to an ideal performance. The (normalized) number of extreme postures (i.e. touching the floor) and the average over all measures results in the final performance score $P$, which is translated into the thumb symbol representation shown in Figure 3.

| Id | Position | Description |
|---|---|---|
| 1 | Sitting | Back and forth, left foot |
| 2 | Sitting | Left and right, left foot |
| 3 | Sitting | Back and forth, right foot |
| 4 | Sitting | Left and right, right foot |
| 5 | Sitting | Back and forth, both feet |
| 6 | Sitting | Left and right, both feet |
| 7 | Standing, supported | Back and forth, left foot |
| 8 | Standing, supported | Left and right, left foot |
| 9 | Standing, supported | Back and forth, right foot |
| 10 | Standing, supported | Left and right, right foot |
| 11 | Balancing | Eyes open |
| 12 | Balancing | Eyes closed |
| 13 | Balancing | Left foot only |
| 14 | Balancing | Right foot only |
| 15 | Standing, free | Back and forth, both feet |
| 16 | Standing, free | Left and right, both feet |
| 17 | Standing, free | Back and forth, left foot |
| 18 | Standing, free | Left and right, left foot |
| 19 | Standing, free | Back and forth, right foot |
| 20 | Standing, free | Left and right, right foot |

Table I
SET OF EXERCISES USED FOR CASE-STUDY ON BALANCE BOARDS

*Automated Feedback:* The measures $M_g, M_a, M_c$, along with the mean and variance of the ECDF, are furthermore used to produce automated textual feedback for the exercising user. Simple rules combine multiple performance aspects in a single condition, which triggers a specific textual cue:

$$M_1 > t_i \;\wedge\; M_2 < t_j \;\wedge\; \ldots \rightarrow \text{Textual cue}, \qquad (5)$$

where $t_i$ corresponds to a real-valued threshold between 0 and 1. This way, expert knowledge can be translated easily into a rule-set specific for each exercise. Both positive feedback about improvement of certain performance aspects, as well as constructive criticism is provided to the user. Furthermore, the parameters $c_{\{g,a,c\}}$ (equation 4) control the sensitivity of the feedback and can be adapted easily, e.g., for novice vs. advance users, without the need to change the definition of the rules for feedback generation.

## IV. CASE STUDY

To evaluate the assessment procedure, we conducted a case study in which subjects performed balance board exercises. We afterwards used the GymSkill system to generate automatic exercise assessments of these data, providing both overall (global) and fine-grained (temporally localized) skill analysis as described in the previous sections. In addition, we surveyed subjects to get qualitative feedback on the practicability of the digital personal trainer for physical exercises. This included aspects like user satisfaction, potential long-term training motivation and the usability of GymSkill.

### A. Participants

Six subjects (1 female, 5 males) aged between 25 and 33 years (average: 29, SD=3.4) participated in the study. Four participants reported a rather sedentary lifestyle with less than two hours of sports per week; two of them indicated to do 6 to 10 hours of sports per week. Four participants were new to balance board training, two had experience with similar training devices.

### B. Procedure

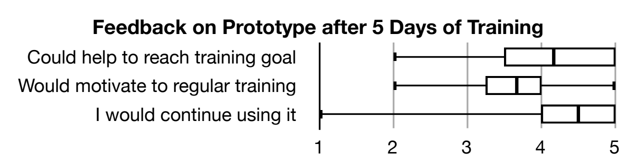
Participants trained for a period of 5 days and performed a set of 20 exercises twice a day (morning and afternoon), cf. Table I. With six subjects participating in the study 1,200 exercises were recorded. A rocker board (see Figure 1) was used as training device; the application itself was running on an Android phone. The smartphone was placed in the middle of the board and fixed with a rubber mat so that it would not slide off the rocker board during the exercises. For system evaluation the exercises were also video-recorded.

Qualitative feedback on the GymSkill prototype was gathered using questionnaires. After the first set of exercises subjects were asked to give first feedback. After the last exercise set subjects answered a second, more comprehensive questionnaire. We asked for their opinion on the training effect of the application, usability issues and training fun.

### C. Ground Truth

The video footage was reviewed and analyzed by an physician in light of quality assessment criteria: i) *regularity* of the movement, ii) *angles* of the back and forth movement (maximally possible deflection exhausted?), if the edge of the board *touched the floor* (should be avoided), iii) *tempo*, iv) *body posture* (e.g. leaning). On this basis, we judged the automatically generated analysis results regarding plausibility and correctness. Additional factors, such as compensating body movements or support by hands, being likewise a clue for correct exercise execution, can not be detected with the smartphone's built-in sensors. Four randomly chosen sets of exercises (i.e. 80 exercise runs in total) were assessed for each participant: two early ones from the first two days, and two late ones from the last two days.

We intentionally did not provide Gymskill's analyses to participants in this first case study in order to gain information on their improvements without feedback. We will use this ground truth for later evaluations of Gymskill's




| Most attractive features | Rating | SD |
|---|---|---|
| Individual feedback | 5.0 | 0.0 |
| Exercise suggestions | 4.8 | 0.4 |
| Overall assessment | 4.2 | 1.2 |

| Most attractive features | Rating | SD |
|---|---|---|
| History/Training diary | 4.0 | 1.1 |
| Exercise grading | 3.8 | 1.5 |
| Exercise replay | 3.8 | 0.4 |

Figure 6. *Top:* User feedback on GymSkill after 5 days of use. GymSkill can help to reach training goals faster, motivate in the long term, and participants would personally use it. *Bottom:* Results on most attractive features and live feedback. Answers based on a Likert scale (5=fully agree).

effectivity in skill improvements. Applying the automated analysis methods to the data recorded in the user study (by taking the average over all estimated performance measures as described above) revealed that the general improvement of the participants was only moderate. This indicates a potential for Gymskill's feedback to support faster improvements.

### *D. Automated analysis and individual feedback*

In order to demonstrate the suitability of our automated feedback to highlight sources of problems, we focus on the performance of subject 1 for two selected trials (start and end of the study), as shown in Figure 7.

Figure 7 (a) shows the detailed analysis of the performance of subject 1 at the beginning of the study. The different graphs show the assessment of different criteria of the performed exercise. The subject reaches the maximum angle possible with the calibrated board, i.e., touches the ground 3 times and shows difficulties with the motion which manifests in four different aspects: i) At certain points there are breakdowns in the structure of the motion as it is identified by the PCBA; ii) The subject does not follow a harmonic motion as can be seen in the angle distribution, normalized regarding standard deviation; iii) The subject does not utilize the whole degree of freedom available on the board; iv) The mean of the recorded board positions does not lie close to the calibrated zero-position, which leads to a unsuitable posture. The automated textual feedback highlights these shortcomings to support the visual feedback.

During the course of the study, subject 1 has shown significant improvement as can be seen in Figure 7 (b). The recorded motion is more continuous compared to the recording from the beginning of the study. Furthermore, the overall movement is much closer to the ideal sinusoid-like motion. However, there is still room for improvement as not the full range of angles is utilized during the exercise.

### *E. User Experiences*

After the study, participants were asked to judge GymSkill's potential in terms of exercising motivation, functionality and usability. All answers of the questionnaire were given on a 5-point Likert scale (1=fully disagree, 5=fully agree).

*1) Motivation:* Participants answered that GymSkill could help to reach a defined training goal faster with an average of 4.2, standard deviation (SD) =1.3. The potential to motivate regular training in the long run was evaluated with an average of 3.7 (SD=1.0). Participants confirmed that they would continue using the application with 4.0 (SD=1.5). The distribution of answers can be seen in the upper diagram in Figure 6. Overall, subjects agreed with the statement that training with GymSkill is 'fun' with 3.8 (SD=1.2).

*2) Requested Features:* Participants appreciated most the individualized feedback GymSkill provides after each exercise (5.0, SD=0.0), followed by suitable exercise suggestions (4.8, SD=0.4) for faster improvements. The ability to see a training history of results was evaluated with 4.0 (SD=1.1), the ability to replay individual exercises with 3.8 (SD=0.4). For the full results, see the lower diagram in Figure 6.

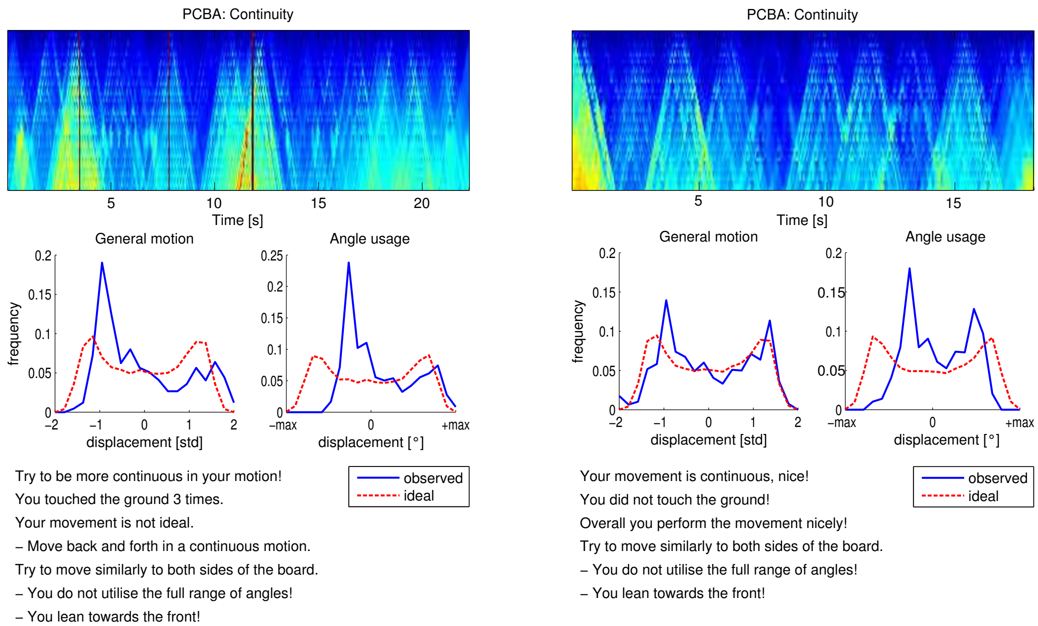
*3) Live Feedback:* Live feedback during exercising was well perceived by the participants. The counting of repetitions was considered helpful (4.7, SD=0.8). The same holds for warning when the board was deflected too much. This warning was given visually and acoustic; users here slightly preferred the sound (3.6, SD=0.8) over the visual feedback (3.3, SD=1.4). Users did not feel distracted by GymSkill (2.0, SD=0.9), e.g., because they had to look downwards on the display. However, the glance frequency was reduced over the course of the study. While subjects stated on day 1 to pay attention to the visual feedback during exercising with 4.3 (SD=0.5), they responded to the same question with 3.5 (SD=0.8) at day 5.

*4) Usability:* Training with GymSkill was evaluated positively. Overall handling (phone placement on the board etc.) was rated with 4.0 (SD=1.1), the interface of the application itself with 5.0 (SD=0.0). The readability of the display when placed on the board was evaluated with 3.8 (SD=1.3). The phone on the balance board obviously did not restrict or negatively affect the users while exercising (1.8, SD=1.3).

## V. Discussion

Encouraging people to exercise more often is important but difficult to achieve in practice. We presented GymSkill, a personal trainer for physical exercises using fitness equipment as it is commonly used in gyms. Utilizing smartphones and novel activity recognition algorithms, GymSkill serves as ubiquitous assessment system for regular exercises.

The results of our case study, where a cohort of participants used GymSkill over the course of a training program, are highly encouraging. The system proved suitable for assessing the overall quality of exercises, thereby replicating a physician's judgment. Furthermore, it unveiled typical exercising errors like deficient smoothness in movement or not using the available freedom of motion. Participants' feedback regarding the usability, suitability, and effectiveness of the digital personal trainer was overwhelmingly positive, which gives evidence for GymSkill's high potential for effectively motivating and supporting people to maintain physical exercises on a regular basis. Although GymSkill was deliberately designed as a standalone ubiquitous assessment system, it is possible to incorporate additional (wearable) sensors, e.g. on upper limbs, to extend the criteria accessible to the analysis. Linking GymSkill to –closed circuit– video capturing facilities represents another option, which would allow even more detailed post-exercise feedback. Such video streams could be annotated automatically to enable visual feeback in addition to the textual cues already generated by the system. GymSkill proved as a valuable complement to professional instructions, qualifying for a range of scenarios




(a) Subject 1 at start of study

(b) Subject 1 at end of study

Figure 7. Detailed analysis of subject 1 performing task 2 at the beginning (a) and the end of the user study (b). The two PCBAs highlight differences in continuity of the performed motion, showing a much better motion towards the end of the study that omits any extreme postures (touching the ground). The comparison of the the observed and the ideal angle distributions indicate that initially (a) the subject had trouble performing the overall motion (General motion) and did not utilise the full range of available angles (Angle usage). Towards the end of the study (b) the overall motion is much closer to the ideal, although not the full available range of angles is used. Throughout this study this subject has shown this systematic error and would therefore greatly benefit from the automated feedback highlighting this issue. [best viewed in color]

from physiotherapy, rehabilitation, fostering physical activity for elderly or faster progress in exercising.

## References


[1] R. W. Jeffery, R. R. Wing, C. Thorson, and L. R. Burton, "Use of personal trainers and financial incentives to increase exercise in a behavioral weight-loss program," *Jrnl. of Consulting and Clinical Psych.*, vol. 66, no. 5, pp. 777–783, 1998.

[2] A. Möller, J. Scherr, L. Roalter, S. Diewald, N. Hammerla, T. Plötz, P. Olivier, and M. Kranz, "Gymskill: Mobile exercise skill assessment to support personal health and fitness," in *9th Intl. Conf. on Pervasive Computing (Pervasive 2011), Video*, San Francisco, CA, USA, 2011.

[3] T. McGuine and J. Keene, "The effect of a balance training program on the risk of ankle sprains in high school athletes," *The Amer. Jrnl. of Sports Med.*, vol. 34, no. 7, p. 1103, 2006.

[4] L. Atallah and G. Yang, "The use of pervasive sensing for behaviour profiling – a survey," *Pervasive and Mobile Computing*, no. 5, pp. 447–464, 2009.

[5] F. Michahelles and B. Schiele, "Sensing and monitoring professional skiers," *IEEE Pervasive Comp.*, pp. 40–46, 2005.

[6] K. Chang, M. Chen, and J. Canny, "Tracking free-weight exercises," in *Proceedings of the 9th international conference on Ubiquitous computing*. Springer, 2007, pp. 19–37.

[7] M. Kranz, W. Spiessl, and A. Schmidt, "Designing ubiquitous computing systems for sports equipment," in *Proc. of the 5th IEEE Intl. Conf. on Pervasive Computing and Comm. (PerCom 2007)*, 2007, pp. 79–86.

[8] M. Kranz, P. Holleis, W. Spiessl, and A. Schmidt, "The therapy top measurement and visualization system - an example for the advancements in existing sports equipments," *Int. Journal of Comp. Science in Sport*, vol. 5, no. 2, pp. 76–80, December 2006.

[9] R. King, L. Atallah, B. Lo, and G. Yang, "Development of a wireless sensor glove for surgical skills assessment," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 13, no. 5, pp. 673–679, 2009.

[10] A. Ahmadi, D. Rowlands, and D. James, "Towards a wearable device for skill assessment and skill acquisition of a tennis player during the first serve," *Sports Technology*, vol. 2, no. 3-4, pp. 129–136, 2009.

[11] R. Best and R. Begg, "Overview of movement analysis and gait features," *Computational Intelligence for movement sciences. Neural networks and other emerging techniques.*, pp. 1–69, 2006.

[12] R. Ryan, C. Frederick, D. Lepes, N. Rubio, and K. . Sheldon, "Intrinsic motivation and exercise adherence," in *Int. Journal of Sport Psychology*, vol. 28, 1997.

[13] T. Ishigaki, T. Higuchi, and K. Watanabe, "Spectrum Classification for Early Fault Diagnosis of the LP Gas Pressure Regulator Based on the Kullback-Leibler Kernel," in *Proc. Int. Conf. Machine Learning for Signal Processing*, 2006.

[14] L. R. Olsen *et al.*, "Multiscale spectral analysis for detecting short and long range change points in time series," *Comput. Stat. Data Anal.*, vol. 52, pp. 3310–3330, 2008.

[15] D. Cox and D. Oakes, *Analysis of survival data*. Chapman and Hall, 1984.