The mobile fitness coach: Towards individualized skill assessment using personalized mobile devices

Matthias Kranz [a], Andreas Möller [b,*], Nils Hammerla [c], Stefan Diewald [b], Thomas Plötz [c], Patrick Olivier [c], Luis Roalter [b]

[a] Luleå University of Technology, Department of Computer Science, Electrical and Space Engineering, Luleå, Sweden
[b] Technische Universität München, Distributed Multimodal Information Processing Group, Munich, Germany
[c] Newcastle University, Culture Lab, School of Computing Science, Newcastle upon Tyne, United Kingdom

ARTICLE INFO

ABSTRACT

We report on our extended research on GymSkill, a smartphone system for comprehensive physical exercising support, from sensor data logging, activity recognition to on-top skill assessment, using the phone's built-in sensors. In two iterations, we used principal component breakdown analysis (PCBA) and criteria-based scores for individualized and personalized automated feedback on the phone, with the goal to track training quality and success and give feedback to the user, as well as to engage and motivate regular exercising. Qualitative feedback on the system was collected in a user study, and the system showed good evaluation results in an evaluation against manual expert assessments of video-recorded trainings.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Regular physical activity is important to personal health and well-being, both for the individual and the society. Encouraging people to exercise more is key to maintaining or regaining personal health but, unfortunately, difficult to achieve in practice. One barrier to exercise is that lay people often are insufficiently knowledgeable about effective and safe physical exercises. Maintaining an exercise regime over longer periods of time requires high levels of motivation. It is well established that access to a personal trainer has a significant impact on both adherence and motivation to a program of physical exercise [1], and the quality of the exercise undertaken [2]. They continuously monitor the exercises and both provide individualized advice and motivate the trainee. Personal trainers also play an important role in rehabilitation, e.g., exercise programs for muscle recovery after surgery, for which there is a need for advice regarding effectiveness and safety. Unfortunately, this is 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).

Smartphones, being pervasive devices, are ideal to support and contribute to regular physical exercising. *Apps* for all purposes have turned the phone to a multi-functional device, far beyond its classic domain of application. They transform the phone to a platform for a variety of applications pervading everyday life: From reading news on the go, over location-based services, games, up to specialized leisure-time and 'hobby' apps for every flavor. Musicians can turn their phone into a guitar tuner, gourmets into a wine guide, and so on. This trend is also visible in the sports domain. Increasing processing

---

* Corresponding author.
  E-mail addresses: matthias.kranz@ltu.se (M. Kranz), andreas.moeller@tum.de (A. Möller), nils.hammerla@newcastle.ac.uk (N. Hammerla), stefan.diewald@tum.de (S. Diewald), thomas.ploetz@newcastle.ac.uk (T. Plötz), patrick.olivier@newcastle.ac.uk (P. Olivier), roalter@tum.de (L. Roalter).

power, integrated sensors, and the ability for rich, multimodal interaction do qualify smartphones not only as personal assistants, but also as personal fitness coaches for supporting individualized training and skill assessment.

The current generation of smartphones with their wealth of sensing, communication, and computing capabilities represents an ideal platform for replicating the services of personal trainers in a way that is accessible to and affordable for virtually everyone. Enabling expert assessment of physical exercises and/or rehabilitation in everyday life (sports) situations integrates mobile interaction further into the real world and has the potential to have a substantial positive impact on people's lives.

We report on our research on an integrated human-activity assessment system, which supports lay users in their everyday life exercising routine with automatically generated expert assessments. The presented system captures the complete training process from exercise descriptions, sensor data logging, activity recognition and on-top skill assessment. This allows the user for individualized, self-determined training supported by coach-quality feedback. Our approach uses balance board training as one representative example for physical exercise support, applicable to a wide range of users and scenarios, both for fitness training and rehabilitation. In an iterative process, we developed a smartphone application that uses integrated sensors for exercise skill assessment on multiple levels for a variety of exercises. We understand skill assessment as individualized, personalized and automated feedback that allows tracking training quality and success, with the claim to replicate the expert judgments of a professional human coach.

As a basis for our research, we performed an extensive review of both commercial applications and scientific research on training support. With the assistance of a sports medicine specialist, we developed an exercise set, exercise descriptions and training plans that we evaluated in an iterative design process with a focus group. Six users performed a 20-exercise set twice a day during one week and produced 1200 exercise records. This basis was manually assessed by an expert and used as a basis and evaluation of ground truths of our assessment algorithms.

In a first prototype, a novel unsupervised and automated activity analysis approach – principal component breakdown analysis (PCBA) – is used to generate informative and easy to interpret qualitative user feedback in the form of color maps. Based on this, a second iteration then goes further and generates more precise and targeted feedback on individual qualities of the performed exercises directly on the device.

Our assessment algorithms show high accuracy in practical evaluations compared to manual judgments provided by sports medicine experts. Qualitative user feedback indicated the acceptance of our system and a notable potential for maintaining long-term exercising engagement. The analysis provided to the user is consistent to the professional analysis of a personal coach. This was confirmed in the two studies presented with the aid of an external expert. Physical interaction with training devices is facilitated using NFC-augmented (Near Field Communication) balance boards.

We show for the first time that an automated approach utilizing sensing and computing capabilities of modern mobile phones can provide expert-level feedback on the quality of physical exercises as they are regularly performed for recreation or rehabilitation purposes. This approach has great potential for the generalizability of physical mobile interaction in sports, especially in the context of the aging society, but also in the more general domain of human skill assessment of everyday activities, which is relevant for a number of therapeutical applications, e.g. skill maintenance in Dementia or Parkinson's care. Our evaluation indicates that we can keep the motivation with coach-like feedback longer than with conventional approaches.

2. State of the art

Advances in the miniaturization of (inertial) sensing technology, along with the increasing availability of smartphones as a sensing platform, have given rise to many commercial appliances, apps, and general academic interest into people's physical activity. This allows for embedded interaction [3], that is to look at new opportunities and application areas that arise for interactive systems and at novel ways for human–computer interaction (HCI) enabled thereby.

Beyond straightforward applications in medicine, where physical activity has to be quantified reliably, many lifestyle, sports and professional assessment systems have been developed. This section first explores systems that employ inertial sensors to assess different aspects of physical activity. Subsequently, the landscape of (commercial and free) health and fitness applications on smartphones is summarized in a comparative study.

2.1. Automatic assessment of physical activity

Body-worn and pervasive sensors have been employed in a large variety of recognition scenarios, identifying different types of physical activities [4–8] with satisfying accuracy. Quantifying qualitative aspects of human motion, such as motor performance, has been of intense focus in medicine, where particularly the assessment of degenerative conditions such as Parkinson's disease is of interest [9,10]. Automated skill assessment for domains where less prior knowledge is accessible is a relatively novel application of pervasive computing. Here augmentation of physical training devices with sensors has been used as the basis for monitoring outdoor sports such as skiing [11] or tennis [12], as well as indoors, such as recognition and tracking of free-weight exercises with accelerometers in a glove [13]. In the gym, sensor data from balance board training [14,15] has been used to provide feedback on the performance quality. Furthermore, body-worn inertial sensors have been used for the assessment of (professional) athletes in sports such as snow-boarding [16], swimming [17] and

Table 1

The 15 health and fitness apps considered by this review can be classified into the categories *GPS tracker*, *workout planner* and *exercise book*. The last two columns indicate the number of reviews and average ratings in the Android market.

Category	Name	Developer	Version	Downloads	Rev.	★
GPS tracker	AndAndo	Javi Pacheco	1.37	50.000–250.000	6342	4.5
GPS tracker	Cardio trainer	WorkSmart labs inc.	3.3.0	>250.000	28326	4.5
GPS tracker	miCoach	Adidas	2.0	>1.000.000	11843	4.5
GPS tracker	RunKeeper	Fitness keeper inc.	2.4.1.10	>250.000	5364	4.6
GPS tracker	runtastic	runtastic	1.4	50.000–250.000	529	4
GPS tracker	Sports tracker	Endomondo	3.3.2	>250.000	5999	4.5
GPS tracker	Sports tracker	Sports tracking ltd.	1.8.5	>500.000	15963	4.5
Workout planner	Ab workout free	Daniel Miller	1.0	5000–10.000	12	4.5
Workout planner	Abs & core	PumpOne	1.0	100–500	5	4.5
Workout planner	Body fitness	Health team	1.2.0	50.000–250.000	662	4
Workout planner	C25K lite	Guy Hoffmann	1.1.2	10.000–50.000	232	4
Workout planner	JEFit	JeFit inc.	3.2.1208	>250.000	6342	4.5
Workout planner	Workout coach	Culleware	2.01	10.000–50.000	79	3.5
Exercise book	Fitness expert	R4z0	1.5	5000–10.000	32	3
Exercise book	Pilates	Saulius	1.0	5000–10.000	57	3
Exercise book	Yoga	SusaSoftX	1.4	50.000–250.000	293	4

running [18]. Beyond applications in sports, similar systems have been employed to assess professional skills, e.g. in surgery using a sensor-enhanced glove [19] or to assess metal filing skill [20], that have applications in training and evaluation.

UbiFit Garden [21] brings activity recognition to the smartphone. For a certain amount of physical activity, flourishing flowers appear on the phone's display as a motivational component. BALANCE [22] estimates the calorie expenditure in everyday life, contributing to long-term wellness management. Both smartphone solutions rely on additional sensors worn on the body.

A multitude of commercial health devices and sensors, such as oximeters and heart rate monitors, formerly reserved for professional use, are now available and can be connected to smartphones. GPS watches, pedometers and heart rate monitors, allow recording and tracking of physical activity. For home use, hardware platforms like Nintendo Wii or Microsoft Kinect encourage users to physical activity, without focus on correct execution. Activity loggers like activPal[1] or FitBit[2] monitor health-related data and help create an activity profile. However, those solutions build upon dedicated systems or external sensor hardware. Since we want to motivate smartphone solutions working the real world without the need for additional hardware, the state of the art in health and fitness assistants on the smartphone is investigated subsequently.

2.2. Comparative review of smartphone apps for health and fitness

We evaluated the design space of current (beginning of 2011) popular health and fitness smartphone apps in a comparative review. For our review, we considered applications from the *Health and Fitness* category in the Android Market. A comparison with other app repositories (iTunes App Store and Nokia OVI Store) revealed that the offerings on different platforms are very similar. We chose Android for our review, being the platform with the highest coverage in 2011 and the fastest growing one. The heterogeneity of Android devices makes the platform also very suitable for fitness support, as the user can select a device adequately for personal needs (e.g. a small and light one on the run, and larger-screen devices for indoor usage).

In order to reflect the highest quality and most satisfying apps available at investigation time, we used the *Most Popular* list in the Android Market and only considered apps with at least 3-star user ratings. We chose 16 apps as a representative sample for this review according the procedure described above. Except Workout Planner, which costs $1.99, all apps were available for free. Endomondo Sports Tracker, SportsTracking Sports Tracker, Adidas miCoach, RunKeeper Free, Cardio Trainer, AndAndo, JEFit, C25K, Daily Ab Workout Free, Body Fitness, Abs & Core, Workout Coach, Fitness Expert, Pilates, and Yoga (see Table 1). Some of them offer additional functionality in a pay version, but do not show a *qualitative* difference.

2.2.1. Heuristic evaluation

Quality assessments were created in a heuristic evaluation of the selected apps, performing a three-step analysis:

1. The core task functionality was tested (task-focused walkthrough).
2. Optional features were explored (explorative walkthrough).
3. The descriptions by the app developers in the Android Market were compared to the actual functionality.

---

[1] http://www.paltech.plus.com/products.htm.
[2] http://www.fitbit.com/.

Four heuristics were used to cover the most important aspects of mobile health and fitness support, which are explained below.

App utility and usability for regular training. The usability for daily application in context of the supported activities was evaluated. Factors observed in this area are the adequacy of interaction (size of controls, appropriate output, comprehensibility) and the customizability, i.e. whether the user can tailor the app to his needs. It was also assessed whether the app makes 'smart' use of data (reuses once entered information, infers information from known data to minimize explicit input).

Instructional quality of apps. It was examined how well the instructions serve to enable autonomous training guided by the mobile device. This includes the comprehensibility and extent of instructions, sufficient level of detail, etc.

Sensor data usage. We evaluated to what extent sensor information is recorded and used by training and fitness applications. This includes sensor data recording for later review (e.g. GPS traces, acceleration data), inference of activity types, assessment of exercising skills, and incorporation of recent developments of Pervasive Computing, e.g. location determination to adapt training to indoor and outdoor environments, context-sensitivity and multi-user interaction.

Motivation. It was examined how well an application is suited to generate and maintain long-term motivation and engagement, through e.g. diversity in the training experience, cooperative training (social incentive), and feedback provision on the training progress to maintain extrinsic motivation.

2.2.2. Results and discussion

We identified three categories of applications: GPS trackers, workout planners, and exercise books, each of which might include an option to connect, share or otherwise use the data in the context of social networks.

GPS trackers. Apps in this category (see Table 1) annotate outdoor activities like running or cycling with location information. After the training, the GPS traces can be reviewed. Further information from the phone's built-in sensors, such as accelerometer or magnetometer data, is usually not included. Some do, however, allow the connection to e.g. heart rate monitors and adapt training instructions to the heart rate (miCoach).

Workout planners. These apps accompany goal-directed workout such as bodybuilding or weight training. They typically contain exercises organized by body parts or muscle groups and log exercise performances (quantitative, not qualitative). While the functionality and usability of apps in this category greatly varies, none offer exercising monitoring in terms of quality or individual performance.

Exercise books. These apps provides a browsable compendium of exercises, with usually the least functionality, compared to the other genres, but the deepest background information on correct exercise performance and health-related issues.

Summarizing our findings of the evaluation, the following points can be concluded.

- The diversity of health and fitness apps is limited, despite a great number of apps in the respective category in the Android Market. Mobile fitness apps either focus on the recording part (which we called GPS loggers), or the instructional part (which we called exercise books). GPS loggers record location information so that a trace of the cardio activity becomes visible. A combination of exercise description and logging are workout planners, which often combine the weak parts of both sides: They just log *that* a certain exercise was performed, and instructions are not as detailed as in the exercise book category. The ideal would be combining the advantages from all three categories.
- GPS loggers have the highest popularity (according to the number of downloads) in the Android Market. The reason probably lies in the high popularity of running, cycling and similar activities and the variety of supported activities by these apps. It can also be seen as an indicator in favor of a comprehensive approach of supporting fitness in different situations and locations. Body Fitness already picked up this idea by integrating both workout and cardio into one app, as well as supplements like a BMI (Body Mass Index) calculator and recipes. However, apps do not support yet pervasive training (at home, in the office, outdoors, …) with activity recognition and contextually suited exercises.
- Multi-user support is not yet integrated, but could be enabled by the creation of user profiles on the device could enable multi-user support also for logging applications and workout planners.
- Few applications are compatible with sports hardware, such as heart rate monitors or wearable sensors. Neither do smartphones use the built-in sensors in current health and fitness applications, although almost all modern smartphones integrate accelerometers, magnetometers and gyroscopes. Device sensor usage could comfortably enable more functionality, without the need to buy, connect and synchronize hardware.
- While some apps allow uploading training information to a portal for later review, none provide individualized, immediate and thereby motivating feedback directly on the phone on the exercise just performed.
- The 'advice' provided by current systems such as e.g. the Adidas miCoach is on a more general level and requires educated interpretation and therefore is not really comparable to expert-like assessment of the individual exercises.
- Current health and fitness apps only partially focus on long-term motivation so far. Social connections, e.g. to beat a friend's results or to challenge a stranger, can be a motivational factor. The Facebook and Twitter integration offered by some apps is presumably rather motivated by promoting the app. Information on long-term usage is missing in any of the reviewed app's descriptions in the Market. If an app would explicitly address motivation, this could be an enormous benefit for advertisement. We believe that adding individualized and personalized feedback could increase the long-term motivation of the user. We contribute to this high-level goal by facilitating this feedback on mobile devices.

2.3. Lessons learned and recommendations

Based on this review, we deduce some general guidelines for future health and fitness apps. The evaluation results show great potential for improvement in the reviewed disciplines usability, instruction quality and fostering motivation. Sensing and context information hereby both play an important role.

Usability improvement. Unnecessary interaction with the device can be reduced e.g. by activity recognition. Research investigated activity detection with wearable sensors [4–6,23]. These techniques need to be tailored to smartphones, which are already equipped with a multitude of sensors. Evaluating acceleration and gyroscope data, could make the recognition of the performed activity possible, making a manual activity selection in the app obsolete. Also the activity-related calorie expenditure can thereupon be estimated. This results in a far more accurate determination of burnt calories after cardio training, as e.g. pauses or speed changes are taken into account.

Instructional quality. The quality of exercise instructions can be improved through well-founded information and physiological correctness. Moreover, the combination with skill assessment and targeted feedback could be a large step towards self-determined, autonomous training. In physiotherapy or rehabilitation, the *correct* exercise performance is important for the healing process. The presence of a doctor or physiotherapist is not always possible, and permanent supervision not comfortable. Accurate exercise assessment based on sensor information and individualized feedback supports correct exercise accomplishment. Mobile phone feedback could thus be a valuable extension – but no replacement – to physiological care.

Long-term motivation. We pointed out the importance of intrinsic motivation for upholding regular physical activity. The integration of feedback in sports and fitness applications adds to establish long-term motivation and engagement in various ways. First, a training assessment for singular exercises in the form of a score ('you reached 85/100 points') motivates beating this value and improving further to reach perfection. Second, a history of training results allows tracking improvement, acknowledging that regular training 'pays off'. More sophisticated skill assessment, which we report on, could even enable targeted feedback. Apps could identify aspects of exercises with potential for improvement or indicate body parts that need particular training. Based on this analysis, the app could suggest suitable exercises addressing 'sticking points' and create a tailored training plan. Training assistance would become more efficient and help to reach goals faster.

The integration of playful aspects and connection to social networks can further contribute to long-term motivation. Some applications already upload training results to Facebook, but real cooperative training needs to go further. Viewing friends' high scores, and beating competitors make physical activity more fun than solitary training. These apps, though, miss a professional assessment of the training quality. We try to close this gap by the presented research.

A step into this direction are apps such as Nike Training Club,[3] an iPhone app (workout planner) combined with a social network, advertising with long-term motivation, but it does not incorporate automated activity recognition. Based on skill assessment, apps could also propose individualized training goals so that they are challenging, but not frustratingly hard. This also has an impact on social sports apps, as then relative comparisons between users would be possible. This, we do argue, will further support long-term motivation.

Intelligent exercise assessment and monitoring is also relevant for elderly people, e.g. in an Ambient Assisted Living context. The smartphone application could be a reminder and a motivational factor to support physical activity, which is important for health risk reduction.

3. GymSkill: automated assessment for balance board training

With GymSkill [24,25], we present a smartphone application that addresses the shortcomings described above and introduces individualized exercise skill assessment fully based on integrated sensors. We describe the iterative design process, qualitative user feedback and quantitative validation of our assessment algorithm.

3.1. Balance board training

We chose balance board training as a representative discipline for physical exercising support, applicable to a wide range of young and old users, both for fitness training and rehabilitation [26,27]. With help of a sports medicine specialist, a set of 20 exercises was conceived and fed into the GymSkill application.

3.2. The GymSkill application

GymSkill is implemented as an Android application and consists of an exercise database, sensor data recorder and the skill assessment presentation (see Fig. 3). During the performance, the smartphone (app running) is placed on top of the balance board (see Fig. 1) so that it can record all of its movements. The user begins her training by selecting an individual exercise or a complete training plan (see Fig. 4). Each exercise is explained step by step in text and pictures. During the actual performance, the phone records accelerometer and magnetometer data as a basis for the assessment which is presented to the user after completion of each exercise. GymSkill is freely available in Google Play as a research app under the name 'GymSkill'.

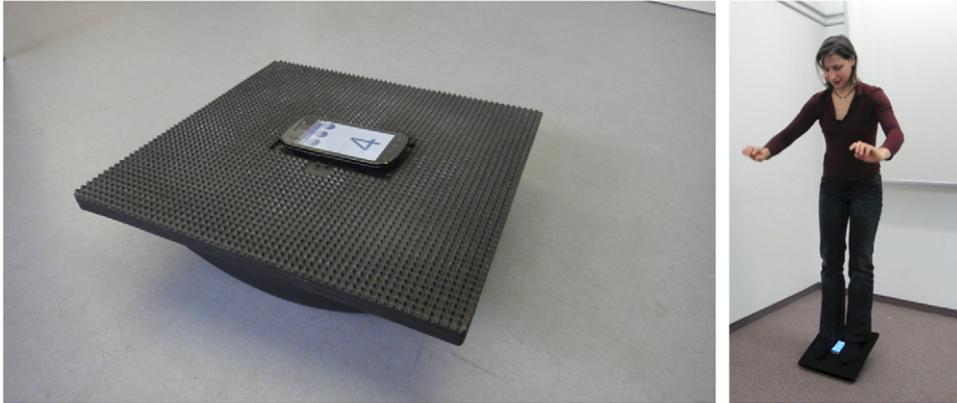3 http://www.nike.com/nikewomen/features/faqs?locale=en_GB.

Fig. 1. Left: GymSkill setup with smartphone placed on a balance board. Right: subject training with GymSkill.

3.3. Case study

A case study was conducted to collect initial sensor data for automatic exercise assessment, and to collect qualitative user feedback on satisfaction with the GymSkill application. We recruited six users (2 females, 4 males) aged from 25 to 33 years (average age: 29) to perform sessions of 20 balance board exercises twice a day for five days, thus collecting 1200 exercise records.

3.3.1. Ground truth collection

For all performed exercises performed in the study, we recorded accelerometer data as ground truth using a Nexus One smartphone that was fixed to the balance board using a rubber mat. The performances were recorded on video and their quality was assessed by an expert. A rating scheme of 11 individual aspects such as movement angles, regularity and tempo was used, adding up to 100 possible points.

3.3.2. Qualitative feedback

After the training sessions, participants were interrogated on the GymSkill application. Since the prototype used for this study was not yet able to provide real-time feedback, the questions focused on the handling of the application and its potential to motivate for regular training. The need for online feedback that was requested by ten participants at the end of this study, has been implemented in the second iteration presented later. A summary of the results can be seen in Fig. 2. Subjects stated that GymSkill could help to reach a training goal faster with an average of 4.2 on a 5-step Likert scale ($1 =$ fully disagree, $5 =$ fully agree), $SD = 1.3$. GymSkill's potential to maintain long-term motivation was confirmed with an average of 3.7 ($SD = 1.0$). Asked for the most attractive potential features of a personal fitness trainer like GymSkill, individual exercise feedback was named with 5.0 ($SD = 0.0$), followed by individual exercise suggestions (4.8, $SD = 0.4$). Subjects stated that they would use the application regularly with averagely 3.6 ($SD = 1.0$). The handling of the system (placing the phone on the board to record data) was evaluated as easy (average agreement: 4.0, $SD = 1.1$); placing the phone on the board to record data was apparently not perceived as problematic.

3.4. First iteration: Server-based automatic exercise assessment

In the first iteration, GymSkill was a two-part system consisting of the app itself and a server analysis component (see Fig. 3, left diagram).

The logged information is sent to a server, where the automatic analysis is performed in terms of a complex retrospective assessment of the exercise correctness, which serves as an indicator of the user's skill. The calculated skill level sent back to the mobile phone, 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 specific problems.

During training, simple feedback like the remaining number of repetitions or excessive displacement of the board is is signaled by a warning sound and visually.

3.4.1. Quality measures

In order to assess the quality of exercises, it is important to estimate global quality measures that cover the important aspects of the performed motion. Advised by an expert clinician, the following attributes were defined.
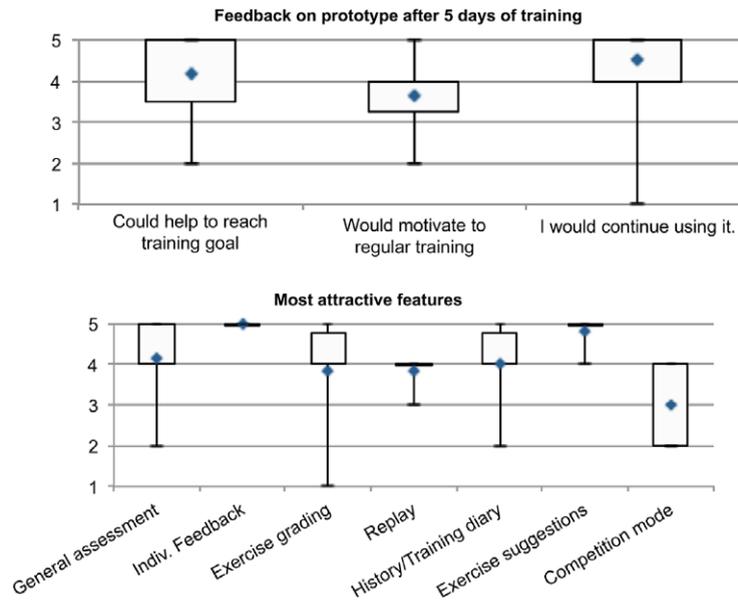
Fig. 2. Left: after 5 days of training with a GymSkill prototype, users believe the app could motivate and help to reach training goals faster. Right: individual feedback and suggestions of exercises were particularly attractive features. Answers given on a Likert scale (1 = fully disagree, 5 = fully agree). The box indicates the interquartile range (middle 50%), the dot the mean value, and the bars minima and maxima.

(a) Iteration 1: the smartphone records sensor data during exercising, which are processed by a server after the training to generate skill assessment.

(b) Iteration 2: data is processed on the phone for real-time feedback as well as sophisticated feedback addressing individual aspects after the execution.
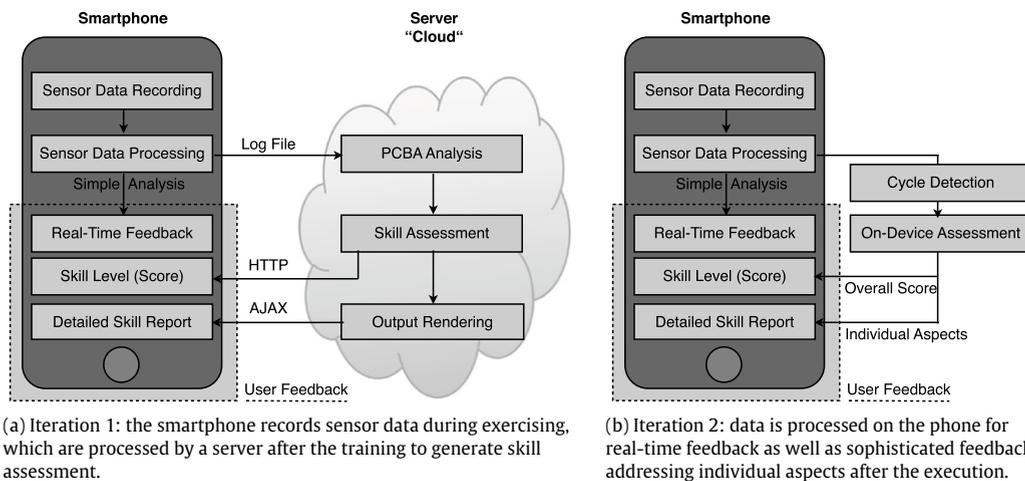
Fig. 3. Iterations of the GymSkill application.

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 us 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 $[-180°, +180°]$ 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., $[\alpha_{min}, \alpha_{max}] \rightarrow [-1, +1]$.

Fig. 4. User interface of the GymSkill application. *Left:* skill assessment after exercising, based on the evaluated sensor information. *Middle:* in the exercise list, exercises that need further training can easily be identified in the first version of the GymSkill prototype. *Right:* global score, detailed skill report on individual criteria and trends (small arrows) in the second iteration of GymSkill.
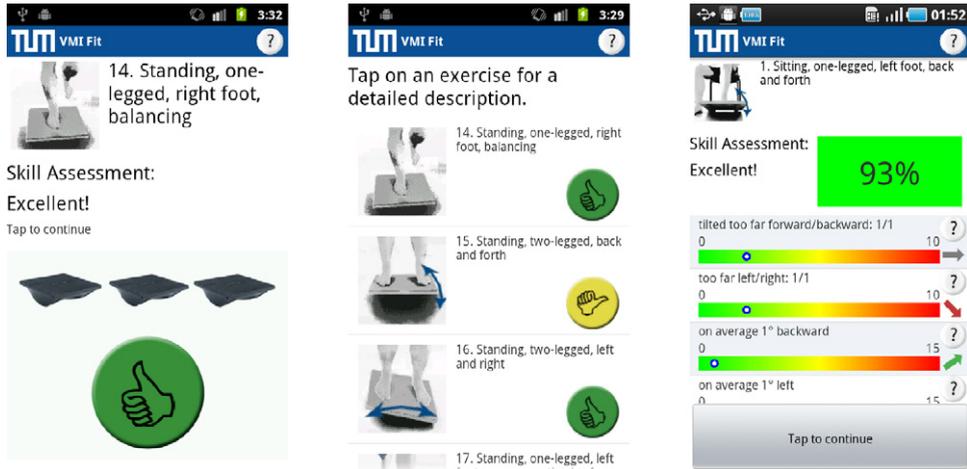

Algorithm 1 Principal Component Breakdown Analysis

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

3.4.2. PCBA analysis

Local analysis. In addition to an overall analysis of quality of physical exercises GymSkill aims for unveiling interesting and informative portions of the recorded sensor data streams. When analyzing re-occurring movements, which are smooth in the ideal case, the term 'interesting' refers to those 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. [28,29]) 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 will violate this assumption and can thus be identified as such.

Algorithm 1 describes the local quality assessment algorithm — *Principal Component Breakdown Analysis (PCBA)*. It is based on a PCA of a sensor data sequence utilizing local neighborhoods. By means of a sliding window technique all analysis windows of length w are extracted and a PCA model is learned. It is applied to project all frames to a lower-dimensional sub-space whose dimensionality is determined by the analysis of the eigenvalue spectrum. Given a pre-defined threshold for reconstruction quality (typically 95% of the variance shall be preserved) the target dimensionality for a particular frame-size is chosen. Using the lower-dimensional projection then the original frames are reconstructed. The resulting reconstruction errors are then used as measure for the quality of the underlying movement, which is sample-wise assigned to the original sequence. The idea behind this quality measure is that PCA models will require more eigenvectors to preserve a certain
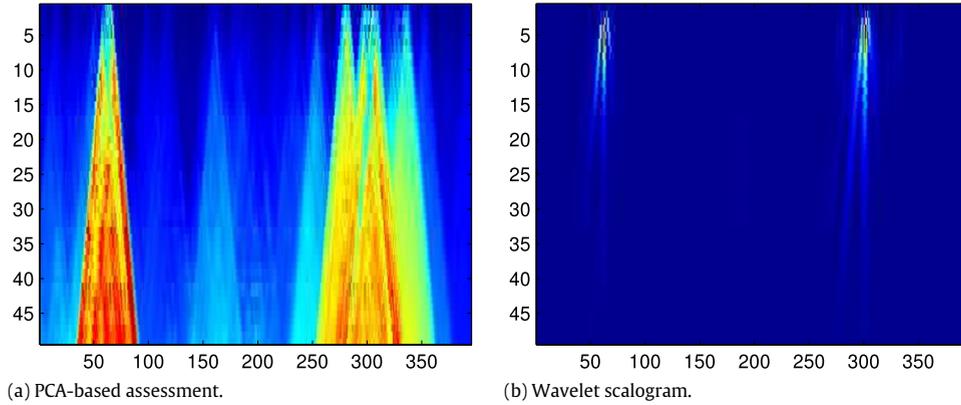
(a) PCA-based assessment.   (b) Wavelet scalogram.

Fig. 5. Local skill assessment using PCBA. Yellow and red areas represent quality breakdowns in exercises.

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 [29]. In contrast to the latter, however, the PCA based approach focuses on self similarity and breakdowns w.r.t. global characteristics. Fig. 5(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 (Fig. 5(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:

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

Global analysis. In order to derive the global motion quality, the motion axis which shows the largest energy during the exercise, i.e., the one that provides 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 subsequently. Using standard Kaplan–Meier estimation [30] the empirical cumulative distribution function is derived and then integrated to form the empirical 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 instead a normal distribution with a variance of $0.1 \times \alpha_{max}$ is employed. The parameters of both distributions are motivated by insights provided by an expert clinician. For other cases where no such prior knowledge is available instead 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):

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

where P^n and ECDF^n correspond to the normalized ideal and empirical distributions.

Second, the un-normalized distributions are compared. 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):

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

Fig. 6 illustrates the procedure.

Fig. 6. The same empirical distribution (blue, continuous 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. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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

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

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. Together with the (normalized) number of extreme postures (i.e. touching the floor), the average over all the measures is used to create the final performance score P. This score is then translated into the visual representation shown in the middle image of Fig. 4 (the thumb symbol).

Automated feedback

The measures M_g, M_a, M_c, along with the mean and variance of the EDF, 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:

$$M_1 > t_i \wedge M_2 < t_j \wedge \cdots \rightarrow \text{Textual cue}, \tag{5}$$

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 $c_{\{g,a,c\}}$ (Eq. (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.

3.5. Second iteration: Criteria-based on-device assessment

With the second iteration of GymSkill, we redesigned the previously described system in terms of comprehensibility, location- and time-independent training, and physical interaction with the training device, incorporating the gathered feedback from the users. The three most important additions are:

- The skill assessment is entirely performed on the mobile phone, with no need for an external server or Internet connection. This is a step towards ubiquitous exercising and training.
- Using a new approach for the assessment, we are able to address individual qualities of performed exercises more intuitively. While the PCBA renderings of the previous version provide a comprehensive overview on different training aspects at a glance, sub-scores now directly tell the user where to improve. We also address with this redesign the limitations of small screens on mobile phones.
- NFC (Near Field Communication) has proven to simplify interaction between phones and the real world [31,32] By the integration of NFC to GymSkill, we simplified the training setup. Placing the smartphone on an NFC-augmented rocker board launches the training application on the phone, loads the board's calibration settings, and the phone even remembers previous training sessions. The need for manual adjustments is minimized.
- We investigated both accelerometer and magnetometer data for their suitability for automated assessment.

In the following, we describe the implementation of the on-device skill assessment algorithm incorporated in this version. Based on the additional user feedback we gathered at the end of the first study, we iteratively developed a second, more sophisticated version of GymSkill. This version does, in comparison to the initial system, feature online feedback (which was missing and considered important for keeping the motivation by the users) and fine-grained feedback along several dimensions for each individual exercise.

3.5.1. Skill assessment implementation

The scoring scheme (with slight changes for tilting and balancing exercises) was developed in the second iteration based on the manual expert assessment. In the following, we list the individual aspects that are weighted and summed to make up the overall skill assessment score.

Touching the ground/balance. When the board is tilted to the maximal possible angle of deflection (known by initial calibration), it is assumed that the user touched the ground with the edge of the board. For each occurrence of this undesired effect, one point of the maximal possible score of 10 points is subtracted.

For balance exercises, a deviation from the neutral position by more than $\pm 5°$ (user-adjustable) results in subtraction of a point.

Repetition count/exercise length. The score in this category is the ratio of actual and required exercise repetitions.

In order to count repetitions and assess other cycle-based quality factors, we detect and count zero crossings in the orientation data. Under the assumption that we only look at one movement direction (which is the case for the exercise set considered in VMI Fit), in the ideal case, two zero crossings of the board's orientation make one repetition. In practice, when the user is unable to keep steady on one side, more than two zero crossings can occur in one repetition. This problem was addressed by filtering in the orientation and time domain with values obtained by experimentation. An interval of $\pm 2°$ degree was considered as zero crossing, and multiple zero crossings in an interval smaller than 3 samples (corresponding to 90 ms) were unified to one.

For balance exercises, the ratio of actual and required exercise length is used for this score.

Pace. This aspect scores the length of repetitions, where a periodic time between 0.9 and 2 s is considered as optimal, and periodic times shorter than 0.5 s or longer than 4.5 s are scored with 0. This sub-score is not used for balance exercises.

Smoothness. Both variances between repetition lengths and distortions within repetitions make up the smoothness value. Within repetitions, the variance of the amplitude and the DC component is regarded, calculated as average signal to distortion ratio. This is derived from the Fourier spectrum of a repetition by taking the ratio of the power in the first bin and the power in the next maximum five bins. For an (ideal) sine wave, all the power of the spectrum would reside in the first bin, otherwise overtones contribute to the signal. The distorting components were restricted to the first five overtones because finer ones cannot be consciously controlled by the user any more. However, further overtones can indicate e.g. tremors and, in future extensions, be a hint for the patient's general physical condition beyond exercising assessment. For balance exercises, the smoothness is expressed through the variance of angles during the performance.

Overall correctness. This aspect bundles further criteria, such as the average pitch and roll angle (i.e. the user tends to lean towards one direction), the number of repetitions where the board was tilted too far (the ideal is stopping slightly before the ground; the limits are user-adjustable), and average amplitudes. Grouping these criteria to one sub-score decreases their weight for the overall score, but incorporates features that correspond to the human understanding of correct exercise execution.

3.5.2. Evaluation of the assessment algorithm

The accuracy of our skill level scores was evaluated against the expert assessment on the recorded training set in an exercise-by-exercise comparison. We did not consider criteria from the expert score that could not be assessed using our algorithm (such as body posture).

The assessment was conducted twice, using magnetometer and accelerometer data as input, to compare the differences in their performance. The main characteristics of the scores are summarized in Table 2 for dynamic (that are exercises where the user actively generates motion) and Table 3 for static exercises (that are exercises where the user has to hold or keep a certain position).

For dynamic exercises, the assessment error was $<10\%$ in 77.86% and $<20\%$ in 94.27% of the cases (with magnetometer data). Using the accelerometer, the assessment error was $<10\%$ in 65.89% and $<20\%$ in 90.10%.

Static exercise assessments showed a $<10\%$ error in 89.58% of the cases and a $<15\%$ error in 98.96% (magnetometer). For accelerometer data, the results were slightly better with a $<10\%$ error in 90.63% of the cases and a $<15\%$ error in 97.92%.

The bias between automatic and expert score is generally higher for accelerometer data, which could possibly be fine-tuned by optimized sub-score weighting. Neither accelerometer nor magnetometer show significant advantages over each other.

Individual aspects. The accuracy differed among the criteria forming the overall score. While repetitions were detected very precisely with less than ± 2 points error in more than 96% of all cases, only about 75% of ground touches could be detected. The largest differences were observed in pace; about 68% of the automatically assessed exercises had an error smaller than 2 points.

Table 2
Match of assessment with expert score, dynamic exercises.

	Bias	$|\Delta| < 10$ points (%)	$|\Delta| < 20$ points (%)	Correlation
Magnetometer	−3.48	77.86	94.27	0.512
Accelerometer	−7.01	65.89	90.10	0.585

Table 3
Match of assessment with expert score, static exercises.

	Bias	$|\Delta| < 10$ points (%)	$|\Delta| < 15$ points (%)	Correlation
Magnetometer	0.08	89.58	98.96	0.764
Accelerometer	−0.99	90.63	97.92	0.778

Bias: average of (automatic score − expert score); $|\Delta| < n$ **points**: percentage of automatic scores this far from expert scores; **correlation**: Pearson's linear correlation coefficient.

Since the magnetometer assumes a homogeneous magnetic field, distortions by metal objects in the lab could reduce robustness of the algorithm for dynamic exercises. For static exercises, the differences between accelerometer and magnetometer results were hardly noticeable. Since the assessment was done on a previously established database, the magnetometer sensors could not be accurately calibrated. Instead, accelerometer data was searched for close to neutral moments (pitch/roll close to zero), and the correct offset to the magnetic field components was calculated during these periods.

4. Discussion and outlook

We reported on the development of GymSkill in an iterative process, a system for individualized assessment and feedback on physical exercising through sensor data. GymSkill is a representative example for future mobile interactive applications that incorporate activity recognition, processing and advanced reasoning and interact with physical devices such as exercising equipment. We proved the acceptance of such mobile assistance systems and their potential to motivate users to regular activity in a five day study. Our automated skill assessment proved to be accurate in an evaluation against expert assessments.

In the future, the system could be extended in various ways. More exercise types with more degrees of freedom (circular movements) or multiple boards (individual left/right foot movements) can be integrated. Coupling with devices like heart rate monitors using e.g. ANT+ further would increase the sensed database and allow for further, more detailed physical and physiological assessments. GymSkill could be integrated with health platforms and interact as remote component of a home infrastructure in relation with other health and fitness applications. The smartphone has the advantage of immediate feedback and time- and location-independent training. With our second iteration, we addressed the limitations in feedback presentation on the small phone display by textual aspect-based feedback. Other user interface types than the presentation on the mobile will be investigated, such as interaction with public displays. A TV over e.g. DLNA can enable more sophisticated feedback, such as PCBA renderings, on larger screens, or foster the social aspect of training by integrating multiple users. We will also conduct user studies with focus on acceptance of different feedback presentations and usability aspects of mobile physical exercise support. Our future work will also have to investigate if this detailed feedback, potentially combined with social network platforms, can actually sustain the long-term motivation of the user.

Acknowledgments

References

[1] R. Ryan, C. Frederick, D. Lepes, N. Rubio, K. Sheldon, Intrinsic motivation and exercise adherence, International Journal of Sport Psychology 28 (1997) 335–354.
[2] R.W. Jeffery, R.R. Wing, C. Thorson, L.R. Burton, Use of personal trainers and financial incentives to increase exercise in a behavioral weight-loss program, Journal of Consulting and Clinical Psychology 66 (5) (1998) 777–783.
[3] M. Kranz, A. Schmidt, P. Holleis, Embedded interaction: interacting with the Internet of things, IEEE Internet Computing 14 (2) (2010) 46–53. http://dx.doi.org/10.1109/MIC.2009.141.
[4] L. Bao, S. Intille, Activity recognition from user-annotated acceleration data, in: A. Ferscha, F. Mattern (Eds.), Pervasive Computing, in: Lecture Notes in Computer Science, vol. 3001, Springer, Berlin, Heidelberg, 2004, pp. 1–17. http://dx.doi.org/10.1007/978-3-540-24646-6_1.
[5] J. Lester, T. Choudhury, G. Borriello, A practical approach to recognizing physical activities, Pervasive Computing (2006) 1–16.
[6] M. Ermes, J. Parkka, J. Mantyjarvi, I. Korhonen, Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions, Information Technology in Biomedicine, IEEE Transactions on 12 (1) (2008) 20–26.
[7] A. Mannini, A. Sabatini, Machine learning methods for classifying human physical activity from on-body accelerometers, Sensors 10 (2) (2010) 1154–1175.
[8] S. da Costa Ribeiro, M. Kleinsteuber, A. Möller, M. Kranz, A compressive sensing scheme of frequency sparse signals for mobile and wearable platforms, in: R. Moreno-Díaz, F. Pichler, A. Quesada-Arencibia (Eds.), Computer Aided Systems Theory — EUROCAST 2011, in: Lecture Notes in Computer Science, vol. 6928, Springer, Berlin, Heidelberg, 2012, pp. 510–518. http://dx.doi.org/10.1007/978-3-642-27579-1_66.

[9] N. Keijsers, M. Horstink, S. Gielen, Ambulatory motor assessment in Parkinson's disease, Movement Disorders 21 (1) (2006) 34–44.

[10] S. Moore, H. MacDougall, J. Gracies, H. Cohen, W. Ondo, Long-term monitoring of gait in parkinson's disease, Gait & Posture 26 (2) (2007) 200–207.

[11] F. Michahelles, B. Schiele, Sensing and monitoring professional skiers, IEEE Pervasive Computing (2005) 40–46.

[12] A. Ahmadi, D. Rowlands, D. James, Towards a wearable device for skill assessment and skill acquisition of a tennis player during the first serve, Sports Technology 2 (3–4) (2009) 129–136.

[13] K.-H. Chang, M.Y. Chen, J. Canny, Tracking free-weight exercises, in: Proceedings of the 9th International Conference on Ubiquitous Computing, UbiComp '07, Springer-Verlag, Berlin, Heidelberg, 2007, pp. 19–37. URL http://dl.acm.org/citation.cfm?id=1771592.1771594.

[14] M. Kranz, P. Holleis, W. Spiessl, A. Schmidt, The therapy top measurement and visualization system — an example for the advancements in existing sports equipments, International Journal of Computer Science in Sport 5 (2) (2006) 76–80.

[15] A. Schmidt, P. Holleis, M. Kranz, Sensor virrig-a balance cushion as controller, 2004.

[16] J. Harding, J. Small, D. James, Feature extraction of performance variables in elite half-pipe snowboarding using body mounted inertial sensors, Progress in Biomedical Optics and Imaging 8 (2007) 50.

[17] S. Daukantas, V. Marozas, A. Lukosevicius, Inertial sensor for objective evaluation of swimmer performance, in: Electronics Conference, 2008. BEC 2008. 11th International Biennial Baltic, IEEE, 2008, pp. 321–324.

[18] C. Strohrmann, H. Harms, G. Tröster, S. Hensler, R. Müller, Out of the lab and into the woods: kinematic analysis in running using wearable sensors, in: Proceedings of the 13th International Conference on Ubiquitous Computing, ACM, 2011, pp. 119–122.

[19] R.C. King, L. Atallah, B.P.L. Lo, G.-Z. Yang, Development of a wireless sensor glove for surgical skills assessment, Information Technology in Biomedicine, IEEE Transactions on 13 (5) (2009) 673–679.

[20] Y. Enokibori, K. Mase, A method to evaluate metal filing skill level with wearable hybrid sensor, in: Proceedings of the 3rd Augmented Human International Conference, ACM, 2012, p. 6.

[21] S. Consolvo, D. McDonald, T. Toscos, M. Chen, J. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, L. LeGrand, R. Libby, et al., Activity sensing in the wild: a field trial of ubifit garden, in: Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems, ACM, 2008, pp. 1797–1806.

[22] T. Denning, A. Andrew, R. Chaudhri, C. Hartung, J. Lester, G. Borriello, G. Duncan, Balance: towards a usable pervasive wellness application with accurate activity inference, in: Proceedings of the 10th Workshop on Mobile Computing Systems and Applications, ACM, 2009, p. 5.

[23] N. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, A. Campbell, A survey of mobile phone sensing, Communications Magazine, IEEE 48 (9) (2010) 140–150. http://dx.doi.org/10.1109/MCOM.2010.5560598.

[24] A. Möller, J. Scherr, L. Roalter, S. Diewald, N. Hammerla, T. Plötz, P. Olivier, M. Kranz, Gymskill: mobile exercise skill assessment to support personal health and fitness, in: Ninth International Conference on Pervasive Computing, Pervasive 2011, Video Proceedings, San Francisco, CA, USA, 2011.

[25] A. Möller, L. Roalter, S. Diewald, J. Scherr, M. Kranz, N. Hammerla, P. Olivier, T. Plötz, Gymskill: a personal trainer for physical exercises, in: 2012 IEEE International Conference on Pervasive Computing and Communications, PerCom, IEEE, Lugano, Switzerland, 2012, pp. 213–220.

[26] T. McGuine, J. Keene, The effect of a balance training program on the risk of ankle sprains in high school athletes, The American Journal of Sports Medicine 34 (7) (2006) 1103.

[27] G. Waddington, R. Adams, The effect of a 5-week wobble-board exercise intervention on ability to discriminate different degrees of ankle inversion, barefoot and wearing shoes: a study in healthy elderly, Journal of the American Geriatrics Society 52 (4) (2004) 573–576.

[28] T. Ishigaki, T. Higuchi, K. Watanabe, Spectrum classification for early fault diagnosis of the LP gas pressure regulator based on the Kullback–Leibler kernel, in: Machine Learning for Signal Processing, 2006. Proceedings of the 2006 16th IEEE Signal Processing Society Workshop on, 2006, pp. 453–458. http://dx.doi.org/10.1109/MLSP.2006.275593.

[29] L.R. Olsen, et al., Multiscale spectral analysis for detecting short and long range change points in time series, Computational Statistics & Data Analysis 52 (2008) 3310–3330.

[30] D. Cox, D. Oakes, Analysis of survival data, in: Monographs on Statistics and Applied Probability, Chapman and Hall, 1984.

[31] G. Broll, S. Siorpaes, E. Rukzio, M. Paolucci, J. Hamard, M. Wagner, A. Schmidt, Supporting mobile service usage through physical mobile interaction, in: Pervasive Computing and Communications, 2007. PerCom'07. Fifth Annual IEEE International Conference on, IEEE, 2007, pp. 262–271.

[32] A. Geven, P. Strassl, B. Ferro, M. Tscheligi, H. Schwab, Experiencing real-world interaction: results from a NFC user experience field trial, in: Proceedings of the 9th International Conference on Human Computer Interaction with Mobile Devices and Services, ACM, 2007, pp. 234–237.