at – Automatisierungstechnik 2015; 63(4): 286–298 DE GRUYTER OLDENBOURG

Anwendungen

Robert Riener* and Domen Novak

Movement Onset Detection and Target Estimation for Robot-Aided Arm Training

Bestimmung des Bewegungsbeginns und Bewegungsziels bei der roboterunterstützten Armrehabilitation

Abstract: This paper presents a motion intention estimation algorithm that is based on the recordings of joint torques, joint positions, electromyography, eye tracking and contextual information. It is intended to be used to support a virtual-reality-based robotic arm rehabilitation training. The algorithm first detects the onset of a reaching motion using joint torques and electromyography. It then predicts the motion target using a combination of eye tracking and context, and activates robotic assistance toward the target. The algorithm was first validated offline with 12 healthy subjects, then in a real-time robot control setting with 3 healthy subjects. In offline crossvalidation, onset was detected using torques and electromyography 116 ms prior to detectable changes in joint positions. Furthermore, it was possible to successfully predict a majority of motion targets, with the accuracy increasing over the course of the motion. Results were slightly worse in online validation, but nonetheless show great potential for real-time use with stroke patients.

Keywords: Intention detection, physical human-robot interaction, rehabilitation robotics, sensor fusion.

Zusammenfassung: In dieser Arbeit stellen wir einen Algorithmus zur Abschätzung der Bewegungsintention vor. Der Algorithmus beruht auf Messungen von Gelenkmomenten, Gelenkwinkeln, elektromyographischen Muskelaktivitäten, Augenbewegungen und kontextbezogenen Informationen und wird für die roboterunterstützte Armrehabilitation zusammen mit Techniken der Virtuellen Realität eingesetzt. Zunächst wird der Beginn einer Streckbewegung mittels Messung von Gelenkmomenten und Muskelaktivitäten detektiert. Schliesslich wird das Bewegungsziel anhand einer Kombination von Augenzielbewegungsmessung und kontextbezogener Information prädiziert, um die Bewegung mittels Roboter in Richtung Bewegungsziel zu unterstützen. Der Algorithmus wurde zunächst offline an 12 gesunden Probanden und schliesslich in Echtzeit an 3 gesunden Probanden getestet. In einer offline Kreuzvalidierung auf der Basis der gemessenen Gelenkmomente und Muskelaktivitäten konnte der Bewegungsbeginn 116 ms vor einer messbaren Gelenkbewegung erkannt werden. Zudem konnte das Bewegungsziel für eine Mehrheit der Bewegungsziele korrekt vorhergesagt werden; die Genauigkeit nahm während der Bewegungsdurchführung zu. Die Ergebnisse verschlechterten sich zwar leicht in der Onlinevalidierung, sie erweisen sich jedoch für Echtzeitanwendungen mit Schlaganfallpatienten als sehr vielversprechend.

Schlüsselwörter: Intentionsdetektion, Mensch-Maschine Interaktion, Rehabilitationsrobotik, Sensorfusion.

DOI 10.1515/auto-2015-0004
Received November 18, 2014; accepted February 17, 2015

1 Introduction

Many robotic devices have been developed to assist, rehabilitate or replace the human upper limb. These devices may support a single part of the limb such as the hand [1] or elbow [2], but may also be robots capable of supporting the entire arm [3, 4] or robotic prostheses that replace the arm [5]. To provide appropriate support for the user, such devices require an adequate hardware design as well as intelligent control and decision-making systems.

One challenge for the design and application of assistive and rehabilitation robots is that their control systems need to recognize human intentions and voluntary capabilities. To provide optimal assistance, the robotic device should infer how the human wants to move the arm. Such motion intention estimation has been emphasized as a major challenge in, e.g., hand exoskeletons [1] and pow-

***Corresponding author: Robert Riener,** ETH Zurich, Zurich, e-mail: riener@hest.ethz.ch
Domen Novak: University of Wyoming, USA

ered hand prostheses [6], and numerous methods have been developed for it. Some motion intention estimation methods use only the robot's built-in sensors while others also utilize additional sensors such as surface electromyography (EMG) [7, 8] or eye tracking [9].

1.1 Intention estimation in rehabilitation robotics

While most intention estimation research has been done in the context of wearable assistive robotics and prosthetics [10], another possible application is neurorehabilitation of the upper extremities. There, robotic devices are used to support and guide the arm while performing exercises in virtual environments (VE). Robot-aided exercises are comparable to exercises with a therapist [4], and exercises in VE are potentially more effective than conventional therapy [11]. Intention estimation in rehabilitation robots can roughly be divided into detection of motion onset and prediction of the intended motion target.

1.1.1 Motion onset detection

Movements in motor rehabilitation should be self-initiated. Thus, a rehabilitation robot that aims to cooperatively interact with the patient should be able to detect the onset of an intended movement. A notable early implementation was a rehabilitation robot that began providing assistance when onset was detected from muscle activity recorded by EMG [12]. Recently, a study demonstrated that rehabilitation outcome can be significantly improved if robotic assistance is provided in response to motion onset detected via electroencephalography (EEG) [13]. Later studies have combined EEG with EMG to detect motion onset and activate a rehabilitation robot accordingly [14].

Many rehabilitation robots simply consider onset to be detected once the subject's limb velocity exceeds a threshold (e.g. the ARMin III [15]). However, both EMG and EEG can detect motion onset before the limb physically moves and can be used with severely impaired patients who cannot make large limb movements [12–14].

1.1.2 Prediction of intended motion target

The rehabilitation robot should not only detect when the patient wants to start moving, but should also determine what kind of motion he/she wants to perform so that appropriate robotic assistance can be provided. Without this knowledge, the robot can only force a predefined motion on the patient or work in a less supportive force-controlled way. On the other hand, if the robot can detect and assist any intended motion, the patient is free to do whatever he/she would like, feels more comfortable, and may be more motivated to exercise.

Assistive robots often use intention estimation to predict and enhance limb kinematics or dynamics [10, 16, 17]. State-of-the-art rehabilitation robots, on the other hand, generally impose an 'optimal' reference trajectory between the starting point and endpoint of the motion, and then correct deviations from this trajectory [18]. Thus, a rehabilitation robot's intention estimation algorithm needs to predict the target of the motion.

EEG is not a good candidate for target prediction, as its accuracy is relatively low, especially as the number of possible targets within a limited space increases [19]. EMG is more accurate, containing enough information to potentially reconstruct the entire motion trajectory [17]. A third promising candidate is eye tracking, as humans generally focus their gaze on the target before beginning the motion [20]. Frisoli et al. [9] recently demonstrated the first application of eye tracking in arm rehabilitation, using EEG to detect motion onset and then predicting the motion target among three possibilities using a head-mounted eye tracker.

1.2 Practical constraints of rehabilitation robotics

The most common target population for arm rehabilitation robots are stroke patients, who usually use the robots to perform reaching motions in VE [15]. In such a setup, the robot should be able to both detect motion onset and predict the motion target. However, since patients and therapists generally have only a limited amount of available time, the sensor setup should be as fast as possible so that it does not take time away from therapy.

EEG in particular has a very long setup time and is only practical for very severely impaired patients. EMG setup is less time-consuming, but still requires significant time due to placement and adjustment of electrodes on the body. Furthermore, factors such as abnormal muscle recruitment may limit the reliability of EMG-based intention estimation in stroke rehabilitation [21].

If we limit ourselves to stroke patients who use rehabilitation robots together with VEs, we can assume that users can initiate a motion at least to some degree; stroke patients who are completely paralysed rarely exercise with a VE. Therefore, a possible onset detection criterion would be that the limb has begun moving with a certain velocity or a certain interaction torque has been applied to the robot.

For motion target prediction, eye tracking represents the most promising practical choice. Since the VE is displayed on a fixed screen in front of the patient, the patient's gaze can be monitored with a remote (contactless) eye tracker rather than a head-mounted one. Such a remote eye tracker has a brief setup time and generalizes well to new users [19]. Furthermore, it is more comfortable for patients, who are usually older and do not like to wear additional equipment on the head.

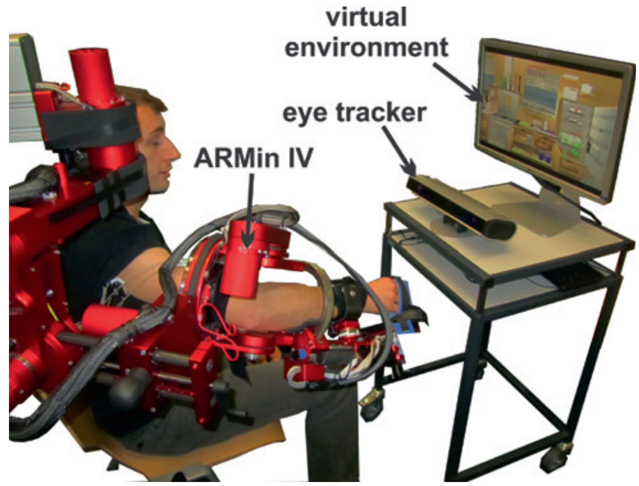

Figure 1: A person interacting with the ARMin IV robot, virtual environment displayed on screen, and eye tracker. Not visible is the EMG recording system.

1.3 Novelty and content of this study

This paper presents a technical implementation of an intention estimation system for arm rehabilitation using a robot and VE. Its novelties over the state of the art are that it combines both motion onset detection and target prediction in an arm rehabilitation exoskeleton, that it combines joint torques and EMG for onset detection, that it combines eye tracking with contextual information for target prediction, and that it presents an online evaluation of the system.

The Materials and Methods section is structured as follows: Sections 2.1 and 2.2 describe the setup: the hardware (2.1) and the VE (2.2). Section 2.3 describes the experimental protocol used to obtain data with which to train and test the intention estimation algorithm. Section 2.4 describes the intention estimation algorithm, which is then validated offline (Section 2.5) and online (Section 2.6).

2 Materials and methods

2.1 Hardware and signal pre-processing

Our system consists of four components: the ARMin IV arm rehabilitation robot, a screen, an eye tracker and EMG sensors (Figure 1). The ARMin IV is similar to the previous generation, ARMin III [15], in that it has an exoskeletal structure with seven actuated degrees of freedom, including a hand opening/closing module. The subject is connected to the robot with cuffs on the upper arm and forearm, and the hand is strapped to the hand module. The dimensions of the device are adjusted to each subject. Position sensors are built into each joint, allowing global end-effector position to be calculated using forward kinematics [15]. Furthermore, three 6-degree-of-freedom force/torque sensors (Mini45, ATI Industrial Automation) are placed in the two cuffs and the hand module. From the recorded interaction forces between the human and robot, joint torques are approximated via the Jacobian matrix of the current robot configuration. The sampling frequency for position and force sensors is 100 Hz.

A 22-inch widescreen monitor placed in front of the robot displays the VE (Section 2.2). Below and in front of the screen (Figure 1) is the SMI RED (SensoMotoric Instruments GmbH, Germany), a contactless remote eye tracker based on two infrared cameras. The included software uses the known position of the cameras as well as the measured position and orientation of the eyes to automatically calculate the gaze position on the screen at 60 Hz. In pretests, we found that the eye tracker's accuracy of gaze position on the screen was approximately ± 1 cm.

Finally, the subject's EMG signals are measured with a g.USBamp signal amplifier (g.tec Medical Engineering GmbH, Austria) and disposable dual Ag/AgCl electrodes (Noraxon Inc., USA). Electrodes are placed on the upper trapezius, anterior deltoid, posterior deltoid, medial deltoid and biceps brachii. These sites were selected based on previous intention estimation studies [14, 17, 19], the constraints of the ARMin's arm cuffs and pretests with reaching motions in the ARMin. Raw EMG is recorded at 1.2 kHz with an analog high-pass filter at 5 Hz followed by a digital bandpass filter (third-order Butterworth, 20–500 Hz bandpass). The filtered EMG is rectified and smoothed with a 50-ms moving average window.

2.2 Virtual environment

We developed a 'virtual kitchen' VE specifically for intention estimation studies [21]. It is displayed on the ARMin's monitor and consists of a frontal view of a kitchen (Fig. 2), with the ARMin's end-effector position shown as a red pointer. Moving the pointer from the far left to the far right corresponds to 90 cm of horizontal ARMin end-effector movement while moving it from the top to the bottom corresponds to 55 cm of vertical end-effector movement.

The VE is divided into six areas (Figure 2) that contain several objects:

– Area 1: shelf above stove (contains recipe book, salt, pepper, olive oil, frying pan and small saucepan),
– Area 2: spice jars and countertop (contain sugar, cardamom, cloves, ginger, cinnamon, nutmeg, mixing bowl, grater and chopping board),
– Area 3: shelves above fridge (contain medium and large saucepans),
– Area 4: stove and oven (contain stove and oven),
– Area 5: cupboard under counter (contains flour and spaghetti),
– Area 6: fridge (contains eggs, butter, salad rocket, tomatoes, mozzarella, milk, yoghurt, garlic and cucumbers).

A total of 30 objects in the VE can be touched (Figure 2), though several additional objects serve only as decoration. The six larger areas are considered as potential 'target areas' for purposes of motion intention estimation while the 30 touchable objects are considered as potential 'target objects'.

When the pointer is touching a movable object (e.g. milk) and the hand module is closed, the object 'sticks' to the pointer and can be moved around the VE while the hand module is closed. If the hand is opened and the object is released while it is touching another object, it interacts with that object:

– If released over a container (pot, pan, bowl), the object is placed inside it.
– If released over the grater, the object is replaced by a grated version of it. For instance, a cucumber is replaced by grated cucumber.
– If released over the chopping board, the object is replaced by a chopped version of it.

A recipe book is located above the stove. When grasped, it opens and takes up most of the screen. It displays the current recipe that the subject has to complete at a given point in the experiment protocol.

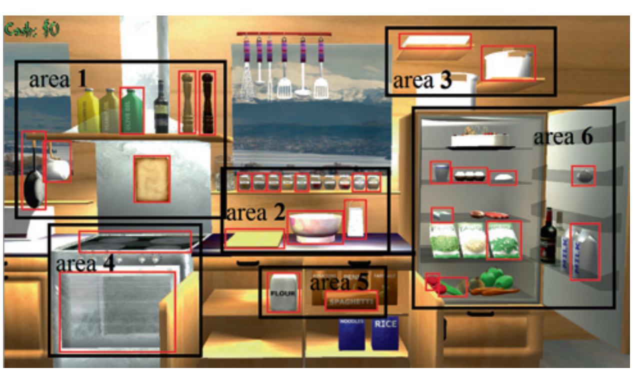

Figure 2: The virtual kitchen, with the six possible target areas marked by black frames and objects that can be touched marked by red frames.

2.3 Experimental protocol

Two experiments were conducted with a similar protocol. The goal of the first experiment was to obtain training data for the intention estimation algorithm and to evaluate the algorithm offline. Therefore, the robot remained passive. The goal of the second experiment was to show that intention estimation works online and can be used to activate the robot. Therefore, the robot actively provided assistance based on results of intention estimation.

2.3.1 Training data acquisition and offline crossvalidation

Twelve healthy subjects (9 males, 3 females) participated in this experiment. Their ages were between 25 and 35 years, mean age 29.3 years, standard deviation 3.1 years. They did not wear thick-rimmed glasses as this degraded eye tracker performance.

The experiment consisted of subjects completing six recipes in the VE while measurements were taken with the sensors. The recipes are: cooked spaghetti (place medium saucepan on stove, add salt and spaghetti), fried egg (place frying pan on stove, add butter, egg, salt and pepper), béchamel sauce (place small saucepan on stove, add butter, milk, flour and nutmeg), biscuits (place oven dish in oven, add butter, flour, sugar, cardamom, cinnamon, cloves and ginger), tzatziki (place yoghurt, grated cucumber, grated garlic, olive oil, salt, pepper into mixing bowl), and insalata caprese (place salad rocket, chopped tomato, chopped mozzarella, olive oil, salt, pepper into mixing bowl).

For each subject, the purpose and procedure of the experiment were first explained. The experimenter demonstrated motions with the ARMin and pointed out various objects in the VE. The subject then sat at the ARMin and was connected to the cuffs. He/she completed the cooked spaghetti and fried egg recipes as a practice run, with the experimenter providing verbal guidance. EMG electrodes were attached to the arm and trunk and the eye tracker was calibrated by having the subject look at five points on the screen in succession. The eye tracker calibration takes approximately 30 seconds, during which time the experimenter simultaneously checked the quality of the EMG recordings.

Following calibration, the subject completed all six recipes one after another in random order. The subject could consult the recipe as many times as he/she wished by moving to it and opening it. As we also wished to identify motion onset, the subject was asked to wait at least 2 s between consecutive motions. The robot did not actively guide the subject toward any target, but did compensate its own gravity and friction, allowing it to be moved by applying only small forces [15].

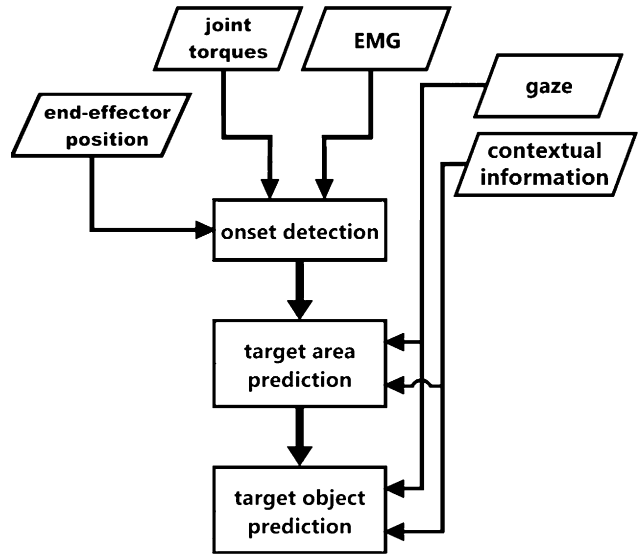

Figure 3: Structure of the intention estimation algorithm, with the input signals (end-effector positions, joint torques, EMG, gaze, contextual information), motion onset detection and target prediction.

2.3.2 Online validation

Three healthy male subjects participated in this experiment. Their ages were 25, 27, and 28. Subjects again completed the six recipes in random order while measurements were taken. The main difference was that the intention estimation algorithms (Section 2.4) had been trained using data from the previous experiment and were used online to detect motion onsets and predict the motion targets. Upon receiving the command from the intention estimation algorithm, the robot moved along a straight line to the predicted target.

2.4 Motion intention estimation strategy

The complete motion intention estimation algorithm has the basic structure as shown in Figure 3. Multiple signals (Section 2.4.1) are recorded and fed into three stages of the algorithm. The first stage detects the onset of a new motion (Section 2.4.2). Once onset has been detected, the motion target estimation algorithm (Section 2.4.3) first predicts the larger target area, then determines the specific target object that the subject is reaching for. Once the probability of the predicted target is sufficiently high, the algorithm also triggers robotic assistance toward the target (Section 2.4.4).

2.4.1 Signal acquisition and segmentation

The following signals were acquired from the hardware's built-in signal processing algorithms and from the VE:
- Robot end-effector position in Cartesian coordinates (3 signals for 3 dimensions);
- Human-robot interaction torques in joint space (6 signals for 6 robot joints);
- Filtered EMG of each muscle (5 signals for 5 muscles);
- Gaze position on the screen in screen coordinates (2 signals – horizontal and vertical coordinates);
- Is the subject holding an object in the VE? (1 binary signal – yes or no).

Based on whether the subject is holding an object, signals were divided into individual motion segments of two types:
- Carrying motions begin when the subject picks up an object and end when the subject releases the object by closing or opening the hand, respectively.
- Reaching motions begin when the subject has released an object and end when the subject picks up the next object.

The subject is not always moving during a motion segment; generally, he/she looks around and finds the target object before moving.

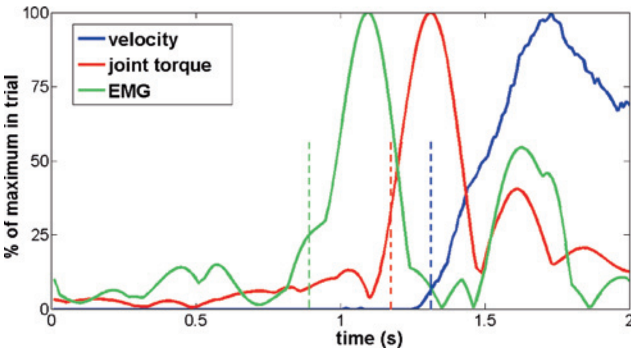

Figure 4: Examples of end-effector velocity, joint torque, and smoothed EMG before and during a reaching motion. Dashed vertical lines represent detected onsets.

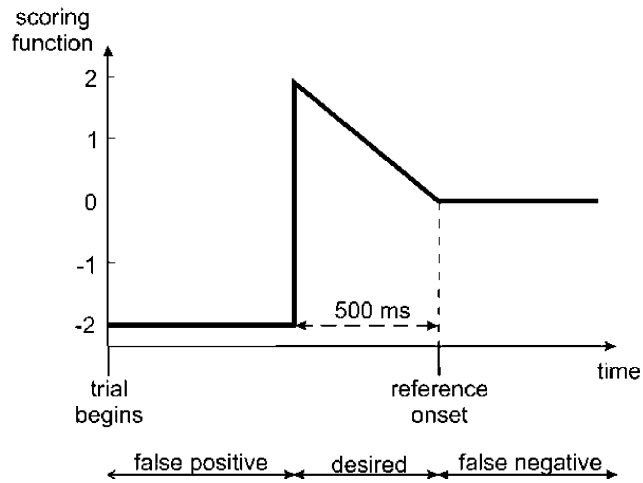

Figure 5: Scoring function used to evaluate a detected onset with respect to the reference onset. If onset is detected more than 500 ms before the reference onset, it is a false positive and penalized. If it is detected too late, it is a false negative, which is not penalized.

2.4.2 Motion onset detection

The intention estimation algorithm should first detect the beginning of a motion. In the previous version of the robot, ARMin III, this is done with a simple threshold on absolute end-effector velocity [15]. This threshold has been shown to be reliable and serves as a fall-back solution to motion onset detection. However, onset can be detectable earlier using EMG or the torque sensors of ARMin IV (Figure 4). The goal of our motion onset detection algorithm is, therefore, not simply to detect onset, but to detect it earlier than using a velocity threshold.

Our algorithm builds on the approach of Kirchner et al. [14], which we previously modified for detection of turn onset during gait [23]. The algorithm is meant specifically to detect onset earlier than a reference onset. This reference onset was defined using an end-effector velocity threshold that was set manually for each subject to be just slightly higher than the range of natural variability due to measurement uncertainty and arm shaking, as in ARMin III [15].

After defining the reference onset, we tried to detect onset using EMG and using joint torques independently from each other. As there are 5 EMG signals and 6 joint torque signals, these were first merged into a single signal for each modality as described in pseudocode:

at time T:
$EMG\_merged(T) = \max(\text{abs}(allEMGsignals(T)))$;
$torque\_merged(T) = \max(\text{abs}(alltorquesignals(T)))$;

Essentially, at any time during the motion, the merged EMG signal is the maximum of the absolute values of all 5 EMG signals at that time. Torque signals were merged in the same way.

Motion onset is then detected from the merged signal if the signal exceeds a certain threshold. The optimal threshold can be found by testing different thresholds on a training set of previously recorded trials. However, this requires a scoring function that evaluates the performance of a particular threshold.

The principle of the scoring function is as follows (see also Figure 5): Our additional sensor (EMG or torque) should detect onset earlier than the reference onset, but not too early (specifically not more than 500 ms earlier), as this would constitute a false positive. On the other hand, if our additional sensor does not detect an onset but the reference velocity sensor does, we can consider onset as detected since the reference has been shown to be reliable – to trigger the robot, we can still use the velocity threshold if other sensors fail. The criterion for false positives (more than 500 ms before reference onset) was previously used for EMG-based onset detection [14] and acceleration-based turn detection [23]. The values of ±2 (Figure 5) were chosen empirically during preliminary tests.

The scoring function was used in an optimization process that was run across all motion segments in the training data to calculate total score as a function of threshold. The optimal threshold was then selected as the one with the highest total score. We used the genetic algorithm in MATLAB 2011b's *optimtool* function for optimization. Optimal thresholds were determined separately for the EMG and torque signals.

After detecting onset from a single modality (torque or EMG), we also tested onset detection using both modalities together. In this case, onset was detected when the merged

EMG signal exceeds an EMG threshold AND the merged torque signal exceeds a torque threshold. The same scoring function and genetic algorithm were again used on training data to find optimal detection thresholds. Notably, threshold values are not the same as when detecting onset from a single modality, as the scoring function takes into account that onset is detected from a combination of both modalities.

2.4.3 Target prediction

Once onset is detected, we must predict the motion target among the many possible target objects (30 in our VE). Therefore, we combine gaze information with contextual information from the VE. We first make a rough prediction of the target area among the six areas in the VE (Figure 2), then predict the specific object inside the predicted target area. These predictions are calculated every 20 ms starting from the time of detected motion onset.

The advantage of first making a rough prediction is that it can be made relatively quickly, allowing the robot to rapidly begin assisting the motion. The specific prediction can be made later in the motion when more accurate data are available, fine-tuning the robotic assistance.

Target area prediction begins once a motion onset is detected. The prediction algorithm (Figure 6) first checks if the subject is currently holding an object in the VE. If the subject is holding an object, we assume that he/she will carry it to the stove, oven, mixing bowl, grater or chopping board. As all these objects are located in areas 2 and 4, the other areas can be excluded. On the other hand, if the subject is not holding an object, we assume that he/she will pick up an object. Area 4 contains no objects that can be picked up and can thus be excluded. The algorithm then also excludes areas that do not contain objects relevant to the current recipe, and finally predicts the target area as the one closest to the subject's gaze position.

Target object prediction selects the most probable target object within the predicted target area. The algorithm follows the same reasoning as with target area prediction. If the subject is holding an object, the prediction algorithm excludes target objects that could be picked up; on the other hand, if the subject is not holding an object, the algorithm excludes target objects that cannot be picked up (e.g. the stove and oven). The algorithm then checks the current recipe and additionally excludes objects that are not relevant to the current recipe. Finally, it calculates the distances between the current gaze position and all remaining possible target objects.

2.4.4 Controlling the robot

As described in Section 2.4.3, target predictions are made every 20 ms. However, this is not enough to simply make predictions; the robot must also assist the subject's motion. We control the ARMin IV based on the intention estimation as shown in Figure 7. Essentially, the robot is idle until motion onset has been detected, then starts predicting the target area and begins moving toward the target area once the same prediction has been made 3 times in a row. While it is moving, it continues predicting the target area, and stops if a different area is predicted. At the same time, it predicts the most probable target object within the target area and moves toward it. Once the robot comes sufficiently close to the currently predicted target object, it stops assisting the subject and waits for motion onset to be detected again. Alternatively, if the subject picks up or releases an object while the robot is moving, the robot also stops assisting the subject and waits for motion onset to be detected.

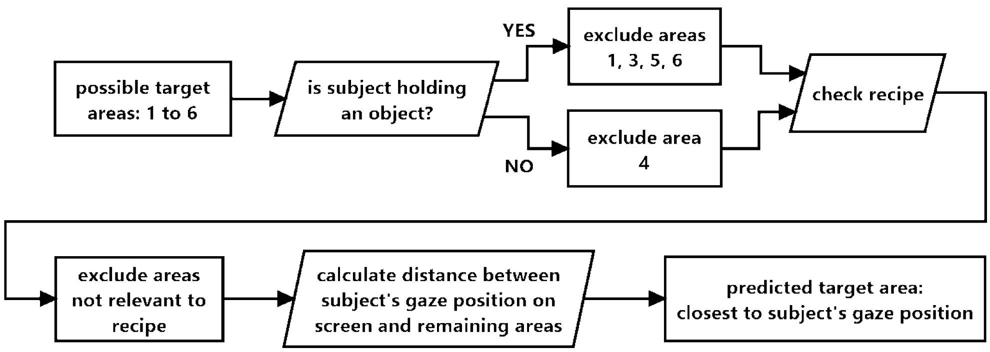

Figure 6: The target area prediction process, which combines contextual information with gaze position measurements.

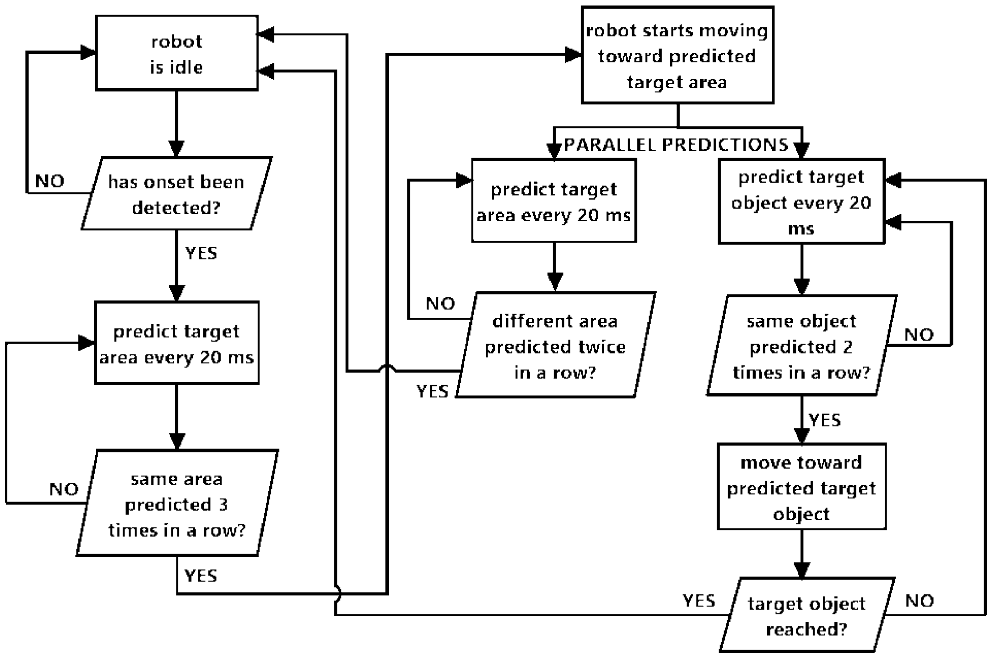

Figure 7: The robot's control process based on intention estimation (onset detection and target prediction).

2.5 Offline validation

In an offline validation mode, the data obtained from the 12 subjects was used to develop and test the onset detection and target prediction algorithms before they were implemented for real-time use. Therefore, for this mode, no commands are sent to the robot.

2.5.1 Onset detection

The onset detection algorithm is an example of supervised learning: it requires training data to learn the optimal onset detection threshold. This training data was obtained as described in Section 2.3.1 and also allows the accuracy of the algorithms to be evaluated offline; we must only make sure that the algorithms are not trained and tested on the same data. We thus use two types of crossvalidation: leave-motion-out crossvalidation and leave-subject-out crossvalidation. The same two types have been used in our previous work [19, 23].

Leave-motion-out crossvalidation can be considered as a subject-specific detection algorithm where the optimal threshold is tailored to each subject. The threshold is calculated based on all but one motion from that subject, then tested on the remaining motion. This is repeated as many times as there are motions for that subject. The process can be described in pseudocode as:

for *subject* = 1 to number of subjects
  for *motion* = 1 to number of motions for that subject
    trainingdata = all motions of *subject* except *motion*;
    define detection rules based on *trainingdata*;
    accuracy(*subject*,*motion*) = calculated by applying
      detection rules to *motion*
  end
end
overallaccuracy = mean(*accuracy*);

Leave-subject-out crossvalidation, on the other hand, can be considered as a general detection algorithm where the optimal threshold is calculated over all subjects. In this case, the optimal threshold is calculated based on all motions from all but one subject and tested on all motions from the remaining subject. The process can be described in pseudocode as:

for *subject* = 1 to number of subjects
  trainingdata = all motions of all subjects except *subject*;
  define detection rules based on *trainingdata*;
  accuracy(*subject*) = calculated by applying detection
    rules to *all* motions of *subject*
  end
end
overallaccuracy = mean(*accuracy*);

Both types of crossvalidation were performed separately for EMG, separately for joint torques, and for the third case of combining both data sources. The outputs of crossvalidation are the differences between time of detected onset and time of reference onset. False negatives do not occur, as onset can in the worst case always be detected with the reference velocity threshold. Therefore, there are two accuracy metrics for onset detection: *percentage of false positives* (when onset is detected more than 500 ms prior to the reference onset) and *mean detection improvement* (mean difference between the detected and reference onset).

2.5.2 Target prediction

Target prediction only uses gaze position and is not a supervised learning algorithm, so it does not require crossvalidation. However, as predictions are made every 20 ms, we wish to know how accurate the prediction is as a function of time. As the algorithm works on two levels (area and object prediction), we define two accuracy metrics: *percentage of correctly predicted areas* and *percentage of correctly predicted objects*. For offline evaluation, the *percentage of correctly predicted objects* assumes that the area was correctly predicted and the object only needs to be predicted among the possibilities in that area (not among all 30 possible objects). These metrics are calculated as a function of time, as the system makes predictions every 20 ms. The correct target of a motion is defined as the object eventually reached and manipulated by the subject.

2.6 Online validation

During online validation, motion onset was detected using joint torques, with the reference velocity threshold used as a ‘backup’ measure. The optimal detection threshold was calculated based on data from offline validation. Once onset was detected, targets were predicted using gaze position, and robotic assistance (Section 2.4.4) was provided based on the prediction results. Thus, online validation served as a way of demonstrating the performance of the intention estimation algorithm when it is used to control the robot in real time.

The provided assistance was simple: when a target was predicted, the robot end-effector moved toward it along a straight line using a simple proportional-integral controller whose inputs were the current position and desired target position in global coordinates. The force applied by the robot to the subject was moderately strong; while the subject could resist the robot’s guidance and move in a different direction, this required a significant force to be applied. The robot did not take the possibility of incorrect target prediction into account, and did not reduce its assistance if resistance from the subject was detected. Nonetheless, subjects were told that they should resist erroneous robotic assistance, and that they should also verbally state the object they are trying to reach if incorrect assistance is provided. This served as a way of verifying the correct target object.

3 Results

3.1 Offline crossvalidation

For onset detection, we excluded any motion segments where the subject does not wait at least 1.5 s after picking up or releasing an object. A total of 646 motions were included in onset detection (Section 3.1.1). For target prediction, we excluded any motion segments where the subject accidentally drops a carried object or changes the intended target in the middle of the segment. A total of 776 motions were included in area prediction and object prediction (Section 3.1.2).

3.1.1 Onset detection

The *mean detection improvement* and *percentage of false positives* are given for different inputs in Table 1.

3.1.2 Target area and object prediction

Results of both target area prediction and target object prediction as a function of time are shown graphically in Figure 8. They are shown for the first second of the motion; no motion is shorter than 1 s, and we did not wish to include time points where some motions have already been completed.

Table 1: Onset detection results in offline crossvalidation.

	Leave-motion-out crossvalidation		Leave-subject-out crossvalidation	
Input data	Improvement (ms)	False positives (%)	Improvement (ms)	False positives (%)
Joint torques	82.0	0.8	41.2	3.0
EMG	74.4	1.7	25.6	3.5
joint torques and EMG	115.8	0.0	40.7	0.0

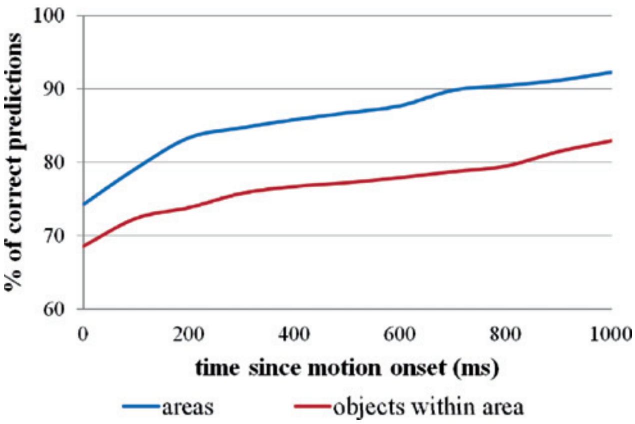

Figure 8: Percentages of correctly predicted target areas and objects within the area as a function of time since motion onset in offline validation.

3.2 Online validation

3.2.1 Onset detection

Motion onset was detected with the combination of joint torque and EMG recordings. As with offline crossvalidation, motion segments where the subject did not wait at least 1.5 s before starting the motion were not counted in the results. There were 153 valid motions for onset detection across the three subjects.

On average, onset was detected 35.4 ms prior to the reference onset, with 6 cases (3.9%) being false positives.

3.2.2 Target area prediction

As in offline crossvalidation, we excluded any motion segments where the subject either accidentally drops an object while carrying it or changes the intended target mid-trial. There were a total of 194 valid motion segments across all three subjects.

Robotic assistance was triggered as soon as the same target area was predicted 3 times in a row. Overall, when robotic assistance was first activated, the correct target area was predicted in 161 out of 194 (83.0%) motion segments. Robotic assistance was activated at $t = 40$ ms since motion onset (earliest possible time) in 38 trials, between $t = 60$ and 100 ms in 132 trials, between $t = 120$ and 160 ms in 10 trials, and later in 14 trials.

In all 33 motions where robotic assistance was initially triggered toward the wrong target area, the mistake was corrected automatically by the intention estimation algorithm. This happened within 100 ms of the original incorrect prediction in 12 of 33 motions, within 120–200 ms in 14 of 33 trials, within 220–300 ms in 5 motions, and later in 2 motions.

3.2.3 Target object prediction

Target object prediction began as soon as the robotic assistance was triggered. For the 161 motions where the correct target area was correctly predicted, the correct target object was simultaneously predicted in 102 motions (63.3%). This accuracy gradually increased through the motion, and the robot guided the subject to the correct target object in 126 motions (78.2%). In the other 35 motions, the subject had to apply increased force to the robot in order to overcome the erroneous guidance or wait for the robot to stop assisting the subject and then manually correct the motion.

4 Discussion

4.1 Onset detection

In leave-motion-out offline crossvalidation, both joint torques and EMG can reliably detect motion onset before it is visible in the end-effector velocity signal. We expected onset to be visible in EMG earlier than in torque measurements, though this was not evident in the results. In a qualitative examination of the signals, we saw that EMG is more 'noisy' and depends very strongly on the position of the arm, a problem that has been noted in previous studies [24]. It is, therefore, more prone to false positives than torque signals. Combining EMG and torque signals allows early onset detection with practically no false positives.

Among the torque signals, the most important information was obtained from the first two joints (horizontal and vertical shoulder); most other signals can be removed without greatly affecting onset detection accuracy. For EMG, the trapezius and all three deltoid signals contributed significantly to onset detection; only the biceps signal can be removed without affecting accuracy. Furthermore, though approximately 50 samples of motion onset were available per subject, not all are needed to find the optimal onset detection thresholds with the genetic algorithm; using 20 samples per subject yielded very similar results.

In leave-subject-out crossvalidation, results are worse for both modalities. This can be expected due factors such as variations of arm placement in the ARMin, variations of EMG electrode placements and differences of EMG patterns among subjects. Notably, EMG does not appear to have any advantage over joint torques with our setup. Indeed, while a 40-ms improvement in onset detection can be achieved without subject-specific algorithm training, further analysis is needed to determine whether this improvement is meaningful or whether the detection algorithm should be adapted to each specific subject. Additionally, we should consider different ways of calibrating and normalizing the EMG signals for each individual subject, as this may reduce inter-individual differences and allow better EMG-based onset detection.

4.2 Target prediction

Gaze tracking provided a reasonably accurate prediction of the target area, allowing the robot to guide the subject to the rough target area. While the algorithm does sometimes incorrectly identify the target area, these mistakes are usually corrected as the motion progresses. In offline validation, accuracy increases over time; this is in agreement with our previous work on gaze-based target prediction algorithms [19]. In real-time validation, we also observed that a subject who felt the robot moving in an undesired direction would refocus his gaze on the desired target, often leading to the robot correcting its estimate.

The algorithm more often misidentifies the desired target object within an area. This is not a critical mistake; as the target objects within an area are located relatively close together, the subject must only make a small correction (move the ARMin end-effector by less than 10 cm). However, the sensation can be unpleasant for the subject, as he/she must either resist the assistive action of the robot or wait for the robot to complete its action. Therefore, errors in intention estimation again limit the subject, reducing the amount of freedom. A possible additional feature would be to monitor interaction forces between the human and robot. When a high interaction force is detected, the human may be resisting the robot's action, and the intention estimation result should be re-evaluated.

4.3 The role of context

Contextual information (areas and objects involved in the current recipe) offered a high degree of reliable information that was usefully complemented with eye tracking. In a follow-up analysis using offline crossvalidation, we actually found that context alone can predict the target area in approximately 60% of cases and the target object in approximately 50% of cases. Furthermore, if contextual information is removed, the overall accuracy of area and object prediction decreases by 10–15%. This emphasizes that a motion intention estimation task can be simplified by exploiting task-related information and giving a higher weight to possibilities that make sense in a given situation.

The question then arises: to what degree are sensors needed and to what degree can the VE be designed so as to give the impression of freedom while relying on information about what the subject is likely to do. Our previous work already emphasized that intention estimation can be made easier by taking the subject's preferences into account [19]. However, we should not rely too heavily on context: ideally, subjects should feel that they have the freedom to do even unexpected or 'senseless' things in a virtual environment while still being intelligently supported by the robot.

4.4 Improving robot control

The assistive action of the robot was relatively simple, as it simply moved to the predicted target in a straight line, occasionally changing its direction when a different target was predicted. Clinical rehabilitation robots generally plan optimal trajectories to the target based on principles such as minimal jerk [18], and this would be a possible expansion of our control system. However, such trajectories are much easier to plan for only a small amount of possible starting points and targets, and it can be difficult to switch from one trajectory to another one in the middle of a motion.

4.5 Generalizability to stroke patients

A significant weakness of the proposed approach is that it was tested only with healthy subjects. Stroke patients are likely to exhibit different behavior, leading to different intention estimation results. EMG is particularly likely to be less effective with stroke patients, who exhibit pathological muscle recruitment patterns [21]. Eye movement would also be less useful in patients with hemispatial neglect, but eye-tracking-based user interfaces have previously been successfully used with stroke victims [25], and we believe that a majority of patients would be able to use our eye-tracking-based target prediction system. However, actual tests with patients are needed to determine the usability of the intention estimation. Additionally, even if intention estimation turns out to be feasible with the target population, it is necessary to determine whether it has concrete benefits for rehabilitation and increases patient motivation, exercise intensity, and finally improves therapy outcome.

5 Conclusion

We developed and tested an intention estimation algorithm for an arm rehabilitation exoskeleton that detects the onset of a reaching motion and predicts the intended target. Onset can be detected earlier using joint torques than with the position sensors commonly used in the ARMin, especially if the detection threshold is tuned to each subject. The intended target can be predicted based on eye tracking and context with a reasonable accuracy, allowing the rehabilitation robot to assist the subject in realtime.

As a next step, it will be necessary to link intention estimation more closely to the robotic support so that the robot could, for instance, detect when the subject is resisting the robot and re-evaluate its choice of assistance. Furthermore, the current assistive action simply moves to the target in a straight line; in the future, more natural trajectories such as minimum-jerk trajectories should be used instead. Finally, the system needs to be tested with stroke patients, and perhaps other patient groups, to determine not only technical performance, but whether it also increases patient motivation and/or exercise intensity during rehabilitation.

Acknowledgement: This work was supported by the Swiss National Science Foundation through the National Centre of Competence in Research Robotics (NCCR Robotics) and by the Clinical Research Priority Program KFSP "NeuroRehab" University of Zurich. The authors would like to thank Mark van Raai for programming assistance and Urs Keller for hardware support.

References

1. P. Heo, G. M. Gu, S. Lee, K. Rhee, and J. Kim, "Current hand exoskeleton technologies for rehabilitation and assistive engineering," *Int. J. Precis. Eng. Manuf.*, vol. 13, no. 5, pp. 807–824, May 2012.
2. N. Vitiello, T. Lenzi, S. Roccella, S. M. M. De Rossi, E. Cattin, F. Giovacchini, F. Vecchi, and M. C. Carrozza, "NEUROExos: A powered elbow exoskeleton for physical rehabilitation," *IEEE Trans. Robot.*, vol. 29, no. 1, pp. 220–235, Feb. 2013.
3. H. S. Lo and S. Q. Xie, "Exoskeleton robots for upper-limb rehabilitation: state of the art and future prospects," *Med. Eng. Phys.*, vol. 34, no. 3, pp. 261–268, Apr. 2012.
4. N. Norouzi-Gheidari, P. S. Archambault, and J. Fung, "Effects of robot-assisted therapy on stroke rehabilitation in upper limbs: systematic review and meta-analysis of the literature," J. Rehabil. Res. Dev., vol. 49, no. 4, pp. 479–496, 2012.
5. L. Resnik, S. L. Klinger, and K. Etter, "The DEKA arm: its features, functionaly and evolution during the Veterans Affairs study to optimize the DEKA arm," *Prosthet. Orthot. Int.*, in press.
6. R. G. E. Clement, K. E. Bugler, and C. W. Oliver, "Bionic prosthetic hands: a review of present technology and future aspirations," *Surg.*, vol. 9, no. 6, pp. 336–340, Dec. 2011.
7. C. Fleischer and G. Hommel, "A human-exoskeleton interface utilizing electromyography," *IEEE Trans. Robot.*, vol. 24, no. 4, pp. 872–882, Aug. 2008.
8. T. Lenzi, S. M. M. De Rossi, N. Vitiello, and M. C. Carrozza, "Intention-based EMG control for powered exoskeletons," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 8, pp. 2180–2190, Aug. 2012.
9. A. Frisoli, C. Loconsole, D. Leonardis, F. Banno, M. Barsotti, C. Chisari, and M. Bergamasco, "A new gaze-BCI-driven control of an upper limb exoskeleton for rehabilitation in real-world tasks," *IEEE Trans. Syst. Man, Cybern. – Part C Appl. Rev.*, vol. 42, no. 6, pp. 1169–1179, Nov. 2012.
10. D. Novak and R. Riener, "A survey of sensor fusion methods for wearable robotics," *Robot. Auton. Syst.*, in press, DOI 10.1016/j.robot.2014.08.012.
11. K. E. Laver, S. George, S. Thomas, J. E. Deutsch, and M. Crotty, "Virtual reality for stroke rehabilitation," Cochrane Database Syst. Rev., 2011, issue 9, art. no. CD 008349.
12. L. Dipietro, M. Ferraro, J. J. Palazzolo, H. I. Krebs, B. T. Volpe, and N. Hogan, "Customized interactive robotic treatment for stroke: EMG-triggered therapy," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 13, no. 3, pp. 325–334, Sep. 2005.
13. A. Ramos-Murguialday, D. Broetz, M. Rea, L. Läer, Ö. Yilmaz, F. L. Brasil et al., "Brain-machine interface in chronic stroke rehabilitation: a controlled study," *Ann. Neurol.*, vol. 74, no. 1, pp. 100–108, Jul. 2013.
14. E. A. Kirchner, M. Tabie and A. Seeland, "Multimodal movement prediction – towards an individual assistance of patients," *PLOS ONE*, vol. 9, no. 1, pp. e85060, Jan. 2014.
15. M. Guidali, A. Duschau-Wicke, S. Broggi, V. Klamroth-Marganska, T. Nef, and R. Riener, "A robotic system to train activities of daily living in a virtual environment," *Med. Biol. Eng. Comput.*, vol. 49, no. 10, pp. 1213–1223, Oct. 2011.
16. K. Kiguchi, T. Tanaka, and T. Fukuda, "Neuro-fuzzy control of a robotic exoskeleton with EMG signals," *IEEE Trans. Fuzzy Syst.*, vol. 12, no. 4, pp. 481–490, Aug. 2004.
17. E. Corbett, K. Körding, and E. Perreault, "Real-time evaluation of a noninvasive neuroprosthetic interface for control of reach," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 4, pp. 674–683, Jul. 2013.
18. L. Marchal-Crespo and D. J. Reinkensmeyer, "Review of control strategies for robotic movement training after neurologic injury," *J. Neuroeng. Rehabil.*, vol. 6, no. 20, Jan. 2009.
19. D. Novak, X. Omlin, R. Leins-Hess, and R. Riener, "Predicting targets of human reaching motions using different sensing technologies," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 9, pp. 2645–54, Sep. 2013.
20. R. S. Johansson, G. Westling, A. Bäckström, and J. R. Flanagan, "Eye-hand coordination in object manipulation," *J. Neurosci.*, vol. 21, pp. 6917–6932, Sep. 2001.

21. B. Cesqui, P. Tropea, S. Micera, and H. I. Krebs, "EMG-based pattern recognition approach in post stroke robot-aided rehabilitation: a feasibility study," *J. Neuroeng. Rehabil.*, vol. 10, no. 75, Jan. 2013.
22. D. Novak and R. Riener, "Enhancing patient freedom in rehabilitation robotics using gaze-based intention detection," in *Proc. IEEE Int. Conf. Rehabil. Robot.*, Seattle, WA, 2013.
23. D. Novak, M. Goršič, J. Podobnik, and M. Munih, "Toward real-time automated detection of turns during gait using wearable inertial measurement units," Sensors, vol. 14, pp. 18800–18822, Oct. 2014.
24. A. Fougner, E. Scheme, A. D. Chen, K. Englehart, and O. Stavdahl, "Resolving the limb position effect in myoelectric pattern recognition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 19, no. 6, pp. 644–651, Dec. 2011.
25. R. Bates, M. Donegan, H. O. Istance, J. P. Hansen, and K.-J. Räihä, "Introducing COGAIN: communication by gaze interaction," *Universal Access Inf. Soc.*, vol. 6., pp. 159–166, Aug. 2007.

Bionotes

Robert Riener
ETH Zurich, Zurich
riener@hest.ethz.ch

Robert Riener is full professor for Sensory-Motor Systems at the Department of Health Sciences and Technology, ETH Zurich, and professor of medicine at the University Clinic Balgrist, University of Zurich. Riener has published more than 400 peer-reviewed journal and conference articles, 20 books and book chapters and filed 21 patents. He has received more than 15 personal distinctions and awards including the Swiss Technology Award in 2006, the IEEE TNSRE Best Paper Award 2010, and the euRobotics Technology Transfer Awards 2011 and 2012. Riener's research focuses on the investigation of the sensory-motor actions in and interactions between humans and machines. This includes the study of human sensory-motor control and the design of novel user-cooperative robotic devices and virtual reality technologies. Main application areas are the fields of rehabilitation and sports.

Domen Novak
University of Wyoming, USA
dnovak1@uwyo.edu

Domen Novak received the M.Sc. and Ph.D. degrees in electrical engineering from the University of Ljubljana in 2008 and 2011, respectively. From 2012 to 2014, he was a postdoctoral researcher at the Sensory-Motor Systems Lab of ETH Zurich. Since September 2014, he is an assistant professor at the University of Wyoming, USA. His research interests include rehabilitation robotics, applied psychophysiology, machine learning, and patient motivation in virtual environments.