To What Extent Can Arm-Hand Skill Performance—of Both Healthy Adults and Children—Be Recorded Reliably Using Multiple Bodily Worn Sensor Devices?

7 AUTHORS, INCLUDING:

Ryanne J M Lemmens
Maastricht University
8 PUBLICATIONS **32** CITATIONS

SEE PROFILE

Annick Timmermans
Hasselt University
34 PUBLICATIONS **296** CITATIONS

SEE PROFILE

Rob Johannes Smeets
Maastricht University
123 PUBLICATIONS **1,757** CITATIONS

SEE PROFILE

Yvonne J M Janssen-Potten
Adelante Zorggroep
29 PUBLICATIONS **327** CITATIONS

SEE PROFILE

To what Extent can Arm Hand Skill Performance – of Both Healthy Adults and Children – be Recorded Reliably using Multiple Body Worn Sensor Devices?

Ryanne J M Lemmens [1*], Henk A M Seelen [1, 2], Annick A A Timmermans [3], Marlous L A P Schnackers [2], Annet Eerden[2], Rob J E M Smeets [1,2], Yvonne J M Janssen-Potten [1,2].

Abstract—Neurological patients often encounter arm-hand problems in daily life. Body worn sensors may be used to assess actual performance by quantifying specific movement patterns associated with specific activities. However, signal reliability during activities of daily living should be determined first. Aim is to determine to what extent standardized arm-hand skill performance of both healthy adults and healthy children can be recorded reliably using a combination of multiple sensor devices. Thirty adults (aged>50 years) and thirty-two children (aged between 6-18 years) performed the activities drinking, eating and combing 5 times in a standardized setting. Sensor devices, each containing a triaxial accelerometer, gyroscope and magnetometer were attached to the arms, hands and trunk of the participants. Within-subject and between-subject reliability of the signal patterns amongst skill repetitions was determined by calculating Intraclass-Correlation-Coefficients (ICCs). Median reliability was good to very good for all activities performed (both within and between subjects). Regarding within-subject reliability (instruction-condition), median ICCs ranged between 0.76-0.90 and 0.68-0.92 for the adults and children respectively. For between-subject reliability (instruction-condition), median ICCs ranged 0.75-0.86 and 0.61-0.90 for the adults and children respectively. It can be concluded that the abovementioned sensor system can reliably record activities of daily living in a standardized setting.

Index Terms—Activities of Daily living, Outcome assessment Reliability, Reproducibility of results, Sensor, Upper extremity.

I. INTRODUCTION

In daily life, the upper extremities are continuously used in the performance of activities of daily living (ADL), including self-care, work, household and leisure. Patients with a neurological disorder such as stroke or cerebral palsy, often suffer from hemiparesis, spasticity and coordination disorders, leading to loss of arm-hand function and consequently loss of arm-hand performance [1-5]. This limits the execution of ADL, affecting their daily life extremely, and resulting in greater dependency, restricted social participation [6] and decreased quality of life [7].

In the field of rehabilitation medicine, measuring arm-hand use is very important for both clinical practice and research. In clinical practice it can, for example, provide insight in the progress the patient makes regarding arm-hand use during rehabilitation. Furthermore, information about arm-hand performance can be used to tailor treatment more efficiently to the specific needs and goals of the individual patient. This information may also be used to more objectively predict therapy outcome for (individual) patients. In research, assessment of arm-hand performance can be important when determining the effectiveness of (new) therapies or when determining the natural course of a disease regarding arm-hand problems. In addition to the available clinical and observational tests measuring a patients capacity or perceived performance [8], objectively measuring actual arm-hand use in daily life is essential, and can provide additional information about the functioning of patients and problems patients encounter in their home situation. Until now, instruments to determine actual performance are lacking [8]. One of the reasons for this is the fact that the upper extremity is a

neurophysiologically and kinematically complex system consisting of the shoulder, upper arm, forearm, wrist and hand, with many degrees of freedom, a large range of motion and movements requiring fine motor coordination. Also, in contrast to the lower limb, analysis of the upper extremity is far more complicated, due to several factors [9]. For instance, upper extremity movements are mostly non-cyclic, which increases complexity and makes it more difficult and less obvious to apply a time normalisation based on the movement cycle [9].

Body-worn sensors can be used to register movements. Many of such sensors are suitable to collect data in a daily life situation, with minimal inconvenience to the participants [10]. There is a wide range of sensors available, for example, accelerometers, gyroscopes, magnetometers, goniometers and pressure transducers. The choice of sensors to be used depends on the research questions to be answered. Rehabilitation treatment is focussed on the patient returning to his daily environment, daily pursuits and societal roles in an optimal way given his residual impairments. Our ultimate aim is to develop a measure gauging actual performance, which is capable of a) identifying arm-hand skill performance in uncontrolled daily life situations and b) measuring both amount and quality of performance. To identify activities and to determine the amount of arm-hand use as well as the quality of performance, sensors that can quantify specific movement patterns associated with specific activities should be used. A combination of accelerometers, gyroscopes and magnetometers, measuring acceleration, angular velocity and orientation towards the earth magnetic field respectively, is expected to be very useful for this purpose [10]. Signals of these 3 different sensors may each show a specific pattern related to the movement executed, i.e. the performance of an activity. The combination of all these movement patterns (=multi-array signal patterns) may characterize a specific activity and may be used to recognize that activity in a registration containing multiple different activities (e.g. during a recording of daily life pursuits of a patient). In addition it is possible to kinematically decompose activities into multiple, consecutive sub phases or sub tasks [11]. Similarly to the complete activity, every sub task may be characterized by specific signal patterns related to that movement. The sub task sequences may also be used to recognize specific activities in a registration containing a sequence of multiple activities.

Before a sensor system can be further developed and used in clinical practice to determine actual performance, it must be tested as to its psychometric properties, one of which is reliability. The technical reliability of the sensors (i.e. the 'Freescale MMA7361 accelerometer' (Freescale, Muenchen, Germany), the 'invensense 500 series MEMS gyroscope' (InvenSense Inc, San Jose, U.S.A.) and the 'Honeywell HMC5843 magnetoneter' (Honeywell International Inc, Plymouth, U.S.A.) has been reported to be good [12-14]. However, to our knowledge, it has not yet been investigated whether a combination of multiple of the aforementioned sensors may reliably record ADL, given the biological variation being present while in daily life. Reliability refers to the reproducibility of measurements and, in this study, is defined as 'the degree to which the measurement is free from measurement error' [15-17]. Reliability will be described using the Intraclass Correlation Coefficient (ICC) parameter. To determine reliability, multi-array signal patterns recorded with multiple sensors during repeatedly executed activities will be compared. As a first step, the reliability of the new system will be determined in healthy individuals during the execution of activities of daily living in standardized conditions. Several factors such as complexity and variability amongst performances of different activities might influence the extent to what these activities can be reliably recorded. Naylor and Briggs defined complexity of activities as the number of parts or components and the amount of information processing demands that characterize an activity [18]. More complex activities have more component parts and require more information processing than less complex activities [18]. Despite standardization, differences in the performance of activities will always be present. Without external movement guidance, it is nearly impossible to execute a movement twice using the exact same movement pattern. In this study, the differences in movement patterns exist within individuals as well as between individuals, and will be referred to in this study as 'variability'.

It is assumed that performance of different upper extremity activities can reliably be recorded / detected and that registrations of sub tasks may be more reliable relative to registrations of the complete task, since the former will be less complex and their duration will be shorter, leading to less variability among repetitions. The aim of this study is to determine to what extent, in standardized conditions, arm-hand skill performance of both healthy adults and healthy children can be recorded reliably using a combination of multiple sensor devices, each containing a triaxial accelerometer, triaxial gyroscope and triaxial magnetometer.

II. METOD

This study is reported using the guideline for reporting reliability and agreement studies [19].

A. Study design and participants

This cross-sectional study consisted of two parts, performed consecutively. Part I included 30 healthy adults aged 50 years and older (i.e. 14 women and 16 men, mean age was 58.0 ± 5.1 years). Participants were included between February 2012 and August 2012. Part II focussed on healthy children aged between 6 and 18 years. Thirty-two children were included between February 2013 and August 2013. Children were divided into two age groups: 16 children aged between 6-11 years old (9 girls, 7 boys, mean age 8.5 ± 1.7 years) and 16 children aged between 12 and 18 years old (8 girls, 8 boys, mean age 14.6 ± 1.5 years).

The sample size was based on a number of practical constraints, among others the number of persons (and parents where applicable, i.e. part II) consenting to participation within the time frame available. For both parts, the exclusion criterion was: motor problems with arm, hand or shoulder, which interfere with the performance of activities of daily living. Participants were recruited by advertisement among staff of Adelante Rehabilitation Centre and their family. This study was reviewed and approved by the Medical Ethical Committee of Adelante Rehabilitation Centre (Hoensbroek,

The Netherlands) and the Medical Ethical Committee of Maastricht UMC+ (Maastricht, The Netherlands, NL42965.068.12). Prior to the start of the measurements, written informed consent was obtained from the participants, i.e. from the healthy adults and children aged 12 years and older. From all children informed consent was obtained from both parents. Upon entry in the study, information concerning age, hand dominance and gender was obtained.

B. Measurements / experimental set-up

The adults in part I performed the activities 'drinking from a cup', 'eating with knife and fork' and 'combing hair'. The children in part II performed the activities 'drinking from a cup', 'eating with knife and fork', 'combing hair' and (additionally) 'opening a zipper'. These activities were chosen from a gross list of activities adult stroke patients and children with cerebral palsy reported wishing to train and improve on, established in earlier research [20, 21].

The measurements consisted of multiple trials (see figure 1 for an overview), each trial consisting of five repetitions of An activity performed at a self-selected speed with short resting intervals (i.e. about fifteen seconds) in between. If one of the trials failed (for instance because the activity was not performed as instructed) an additional trial was performed to end with five correctly performed repetitions. For the activities drinking, eating and combing, subjects were sitting comfortably on a chair behind a table, both height-adapted to the length of the participant to create a starting position in which their back was straight, their feet on the floor, their knees and hips flexed 90 degrees, their elbows flexed 90 degrees and the ulnar side of their hands touching the tabletop. After skill performance, the initial position was assumed. For the skill 'opening a zipper', subjects were standing straight with their arms along their body. All skill utensils (cup, knife, fork, plate with play-dough 'food', hair brush) were positioned on the tabletop in a standard way.
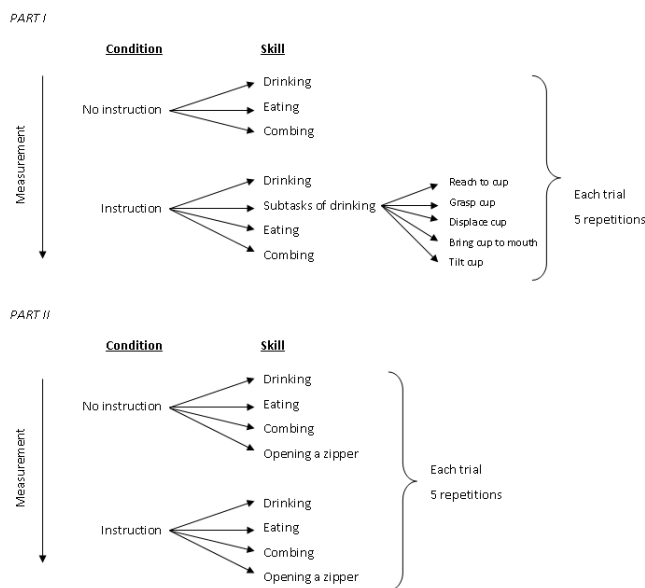

Fig. 1. Overview of the trials as measured chronologically.

All activities were performed in two conditions, i.e. a 'no-instruction condition (NI-condition) and an 'instruction condition' (I-condition). First, the subjects performed all activities five times the NI-condition, to simulate their daily habitual performance. Subsequently, all activities were performed five times in the I-condition with standardized verbal instruction, to homogenize the performance of the activity as much as possible. The standardized instructions given are presented in the appendix.

To determine to what extent *sub tasks* of skills can be recorded reliably, the activity 'drinking' was used as an example and decomposed in sub tasks based on a clinically used and ecologically valid skill decomposition technique [11]. Sub tasks were: 1) reach to the cup, 2) grasp the cup, 3) displace the cup, 4) bring cup to mouth, 5) tilt the cup, 6) put the cup back on the table, and 7) return to starting position. After a short verbal instruction, the adults performed each sub task 5 times. Measurements of sub tasks were only done for the activity 'drinking' performed by the adults since this will give an indication about the extent to what sub tasks can be recorded reliably. For conciseness of this paper, the evaluation of sub task signal matrices for the task 'drinking' as performed by healthy adults will be reported as proof-of-principle.

C. Sensor devices

In this study a combination of several sensor devices, each containing a triaxial accelerometer, triaxial gyroscope and triaxial magnetometer was used (SHIMMER Research, Dublin, Ireland). For part I, four of these lightweight devices (i.e. 27 gram) were attached to the dominant arm and hand and to the chest of the participant in a standardized manner as shown in figure 2a. One device was attached on the dorsal side of the hand (parallel to the. *metacarpophalangeal joints*) using a custom-made glove, one device on the dorsal side of the wrist (directly proximal to the *proc. styloideus ulnae*) using a custom-made fabric strap with Velcro, one device on the upper arm (halfway between the *epicondylus lateralis* and the *tuberculum majus,* parallel to the longitudinal axis of the *humerus)* using a custom-made fabric strap with Velcro, and one device on the chest (located on the *manubrium sterni*) using hypoallergenic tape. For part II, involving healthy children, devices were, in addition to the dominant arm-hand and chest, also attached to the non-dominant arm-hand (figure 2b). The devices on the hands were attached with hypoallergenic tape. Differences regarding the use of the sensor devices between part I and part II encompass the placement of the devices on the hand and the number of devices used. After having performed part I of this study it became clear that measuring both arms and hands would enhance insight into specific aspects of bimanual task performance. Therefore, in our ensuing work, bilateral recordings were used. Information about movement patters of the non-dominant arm-hand are also very valuable, especially when measuring patients, since their affected arm-hand is mostly use as assisting (non-dominant) arm-hand. Furthermore, we have chosen to attach the devices on the hand with hypoallergenic tape since the hands of the children are much smaller compared to the hands of the adults and therefore the glove used in part I would probably hamper skill execution in children.
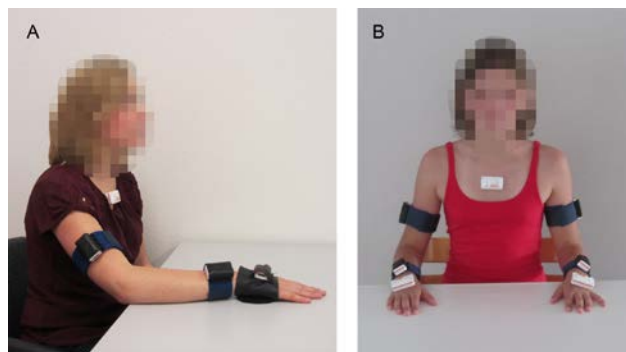
Fig. 2. Placement of the devices on the participant in a standardized manner. a) part I; for the measurements with healthy adults, b) part II; for the measurements with healthy children.

The wireless devices were connected to a computer via Bluetooth. Data were synchronously sampled and saved using the Multi-Shimmer Sync application version 1.1.0 (Shimmer, Dublin, Ireland) at a frequency of 256 Hz for part I and 128 Hz for part II. These sample frequencies differed to accommodate maximum data throughput (# devices * 9 channels * sample frequency * 2 byte) of the recording system.

D. Data analysis

Data processing

All data analyses were done off-line using purpose-designed Matlab software (MathWorks Inc., Natick, MA). Raw data of each trial consisted of 36 signals (four devices, each gauging three types of signals in three directions) for part I, and 63 signals for part II (seven devices, each gauging three types of signals in three directions). Data pre-processing included zero time-phase, low-pass filterin1g (4th order Butterworth filter, cut-off frequency: 1.28 Hz as determined in prior experiments) of all signals to obtain the low frequency signal content associated with the three-dimensional spatial displacements / kinematics of the skill performed. Additionally, a 3D resultant vector was calculated by applying Pythagoras' rule to each time point of the x-, y-, and z-components of the signals of each sensor. This was done for the accelerometer signals, the gyroscope signals and the magnetometer signals separately. For part I, a total of 12 signals remained (4 devices * 3 types of signals * 1 vector representing the three axes), and a total of 21 signals for part II (7 devices * 3 types of signals * 1 vector representing the three axes). Each trial consisted of 5 repetitions of an activity, saved in 1 data file. For part I, all subjects were right handed and performed the activities with their dominant right arm-hand. Only movements of this arm-hand were registered. For part II, movements of both arm-hands were recorded. This group consisted of right-handed as well as left-handed participants. In order to be able to compare movement patterns of the similar movement across participants, e.g. movement of the hand manipulating the knife (which can be either the right hand or the left hand), the arm-hand sensor data of the left-handed participants were mirrored regarding the ab-adduction component and rotation component, in order to compare these data to those of the right-handed participants.

Identification of repetitions

The identification of the time borders of the repetitions, i.e. the start and endpoints of each of the five attempts to perform an activity, was based on the gyroscope signals. The gyroscope signals were used for this because these signals demonstrate the most distinctive pattern between repetitions and resting periods (i.e. between activity-related movement and non-movement periods between repetitions). The 3D-resultant gyroscope signals of all devices were summed into one (artificial) resultant signal (X). This summation was performed to include the signal content of all gyroscope signals of *all* devices of one person, since it is not known beforehand which device (e.g. chest, upper arm, wrist or hand) will be moved first at the start of each activity execution and which device will be moved last at the end of each activity execution. From the resultant signal X, a threshold was determined, demarking 'resting' signal from 'active' signal. This threshold was determined experimentally. When the use of the standard threshold resulted in finding time points not associated with the performance of the activity, in an iterative process, the threshold was increased until this problem was resolved. A minimum duration of the specific activity at hand was set a-priori, based on the average length of the activity performed, as was a minimum duration of the rest phase between repetitions of the skill (a-priori parameters (APPs)), both expressed in data points, thereby avoiding false-positive findings regarding the activities start and endpoints. Based on the start and endpoints identified, a collection of five data matrices, each containing a time delimited epoch of *all signals* from *all devices* (e.g. hand, wrist, upper arm, chest) and *all sensors* (i.e. accelerometer, gyroscope and magnetometer), each representing one activity execution, was obtained. These signal matrix epochs were of unequal length since no instructions about performance speed were given. A schematic overview of the procedure of identifying start and endpoint of each attempt is depicted in figure 3. In some cases, it was impossible to identify all repetitions using the described APPs. In these cases, APPS were adjusted manually and repetitions were identified using the same algorithm. Unsuccessful repetitions, caused by problems with the data transfer or wrong execution of the activities (i.e. not following the instruction), were discarded. The repetitions identified were time-normalized to 1000 data points using linear interpolation.
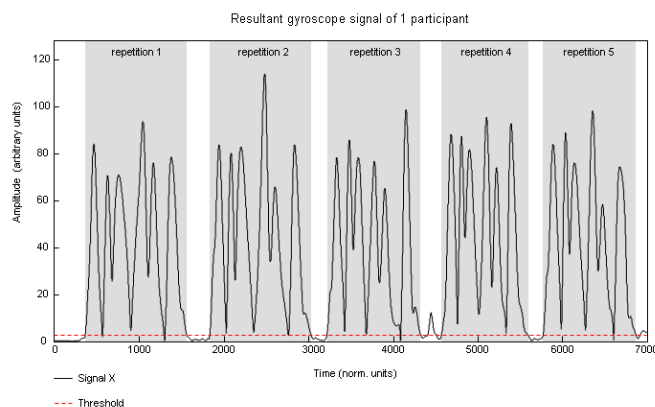
Fig. 3. Procedure of identifying start and endpoints of skill repetitions. The black line represents the summed gyroscope signal of multiple devices of one person. The dotted red line represents the threshold. With an algorithm making use of a threshold, minimal duration of a repetition and minimal duration between two repetitions, repetitions were identified.

Calculation of reliability

For each participant, for each activity, within-subject reliability was determined by calculating the Intraclass Correlation Coefficients (ICC's) between the time-normalized repetitions, i.e. the time epochs in which an activity was performed. This was done for all signals separately (i.e. 12 signals for part I and 21 signals for part II). Subsequently, for every participant, a mean ICC was calculated of the for each activity.

For the between-subject reliability, a mean signal of the repetitions was calculated for every participant, for each activity. This mean signal was used in the calculation of the ICC's across participants. This was done for every activity, for all signals separately (i.e. 12 signals for part I and 21 signals for part II).

The ICC's were classified based on the kappa statistic classification of Landis and Koch [22], shown in table 1.

Table 1: classification scheme for the ICC's based on Lanidis and Koch [22].

ICC		Classification
<	0.00	poor
0.00	0.20	slight
0.21	0.40	fair
0.41	0.60	moderate
0.61	0.80	good
0.81	1.00	very good

III. RESULTS

The mean number of repetitions per activity was 4.7 for the adults and 4.3 for the children. Reasons to discard repetitions were: participant did not follow the instructions during the performance of the activity, or just before or immediately after the performance of the activity the participant moved his arm-hand making it impossible to identify the repetitions without the extra movement. For the adults, 3.5% all repetitions were identified with manual adjustments of the a-priori parameters, compared to 11.5% of the repetitions of the children.

Figure 4 shows the within-subject reliability of the activities eating, combing and drinking, performed by healthy adults. Based on the classification of Landis and Koch [22], the median ICC's were good to very good for all activities . The activity drinking had a higher reliability compared to the activities eating and combing, whereas combing had a higher reliability compared to eating. Reliability was higher for the I-condition compared to the NI-condition, especially for the activities eating and combing.. Regarding the sub tasks of the activity drinking, the median ICC's were very good for all sub tasks .
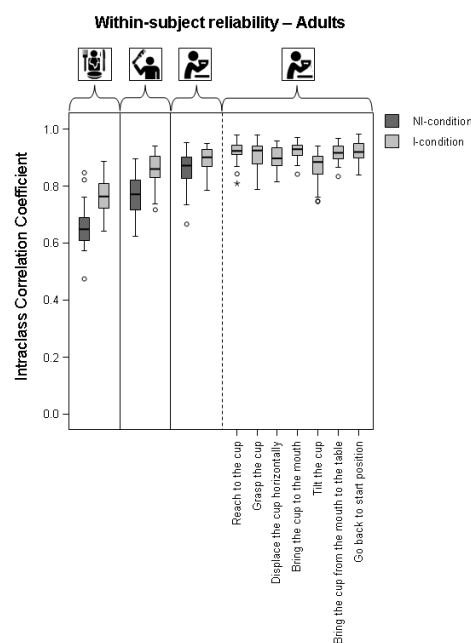

Fig. 4. Within-subject reliability for skills performed by the adults, expressed as intraclass correlation coefficients (ICC's), for the skills eating, combing and drinking. Dark grey bars: no-instruction condition; light grey bars: instruction condition.
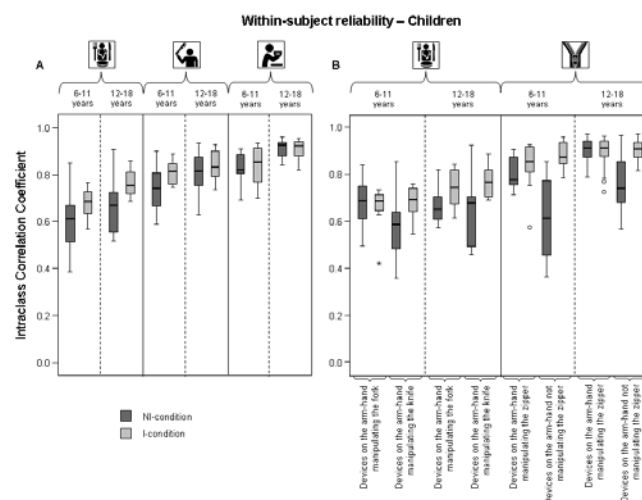

Fig. 5. Within-subject reliability for the skills performed by the children, expressed as intraclass correlation coefficients (ICC's), for the skills eating, combing and drinking. Dark grey bars: no-instruction condition; light grey bars: instruction condition. Panel A: boxplots representing the devices on the chest and on the arm-hand manipulating the knife, comb and cup; panel B: boxplots representing the devices of the two arm-hands separately.

In figure 5a, the within-subject reliability of the performance of the activities eating, combing and drinking by the children is shown. Similarly to the within-subject reliability of the activities performed by the adults, ICC's were calculated based on the signals of the devices on the chest and on the arm-hand manipulating the knife, comb and cup (i.e. using unilateral recordings). Reliability was good to very good for all three activities , and the performance in the I-condition had, in general, a higher reliability compared to the performance in the NI-condition. The within-subject reliability was higher for activities performed by the children aged

between 12-18 years compared to the results from the children aged between 6-11 years.

Figure 5b shows the within-subject reliability for the activities eating and opening a zipper, calculated for the two arm-hands separately (using bilateral recordings). For the activity eating, overall the reliability was good , with the exception for the arm-hand manipulating the knife in the NI-condition (moderate reliability). The I-condition resulted in a higher reliability compared to the NI-condition, except for the arm-hand manipulating the fork in the younger children. The performance of activities by the older children, in general, had a higher reliability compared to the younger children, except for the arm-hand manipulating the fork in the NI-condition. Regarding the activity opening a zipper, the main finding was the difference in reliability between the NI-condition and the I-condition for the arm-hand not manipulate the zipper, i.e. a good reliability for the NI-condition compared to a very good reliability for the I-condition.
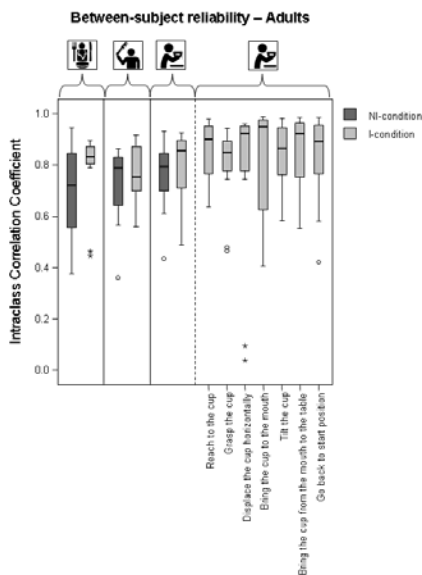

Fig. 6. Between-subject reliability for the skills performed by the adults, expressed as intraclass correlation coefficients (ICC's), for the skills eating, combing and drinking. Dark grey bars: no-instruction condition; light grey bars: instruction condition.

Between-subject reliability for the activities performed by the adults was good to very good for the activities eating, combing and drinking, is shown in figure 6. The median ICC was higher for the I-condition compared to the NI-condition for the activities eating and drinking. For all sub tasks of the activity drinking, the median ICC value was very good .
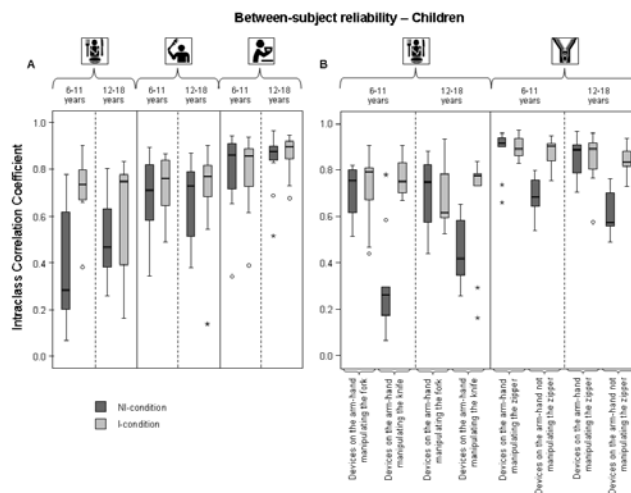

Fig. 7. Between-subject reliability for the skills performed by the children, expressed as intraclass correlation coefficients (ICC's), for the skills eating, combing and drinking. Dark grey bars: no-instruction condition; light grey bars: instruction condition. Panel A: boxplots representing the devices on the chest and on the arm-hand manipulating the knife, comb and cup; panel B: boxplots representing the devices of the two arm-hands separately.

Figure 7a shows the between-subject reliability for the activities eating, combing and drinking, performed by the children. Similarly to the between-subject reliability of the activities performed by the adults, ICC's were calculated based on the signals of the devices on the chest and on the arm-hand manipulating the knife, comb and cup (i.e. using unilateral recordings). For the activity eating, the median ICC for the performance in the NI-condition was fair for the youngest children and moderate for the oldest children, whereas the median ICC for the performance in the I-condition was good in both groups. For the activities combing and drinking, median ICC's were good and very good, respectively.

Figure 7b shows between-subject reliability for the activities eating and opening a zipper, for the two arm-hands separately (using bilateral recordings). For the activity eating, the median ICC was, in general, very good, except for the arm-hand manipulating the knife without instruction in both age groups. For the activity opening a zipper, the median ICC was overall very good, except for the signal registered with the devices on the arm-hand that did not manipulate the zipper when the activity was performed in the NI-condition. For the arm-hand manipulating the zipper, median ICC's were almost similar between the performance in the NI-condition and the I-condition.

IV. DISCUSSION

The aim of this study has been to determine to what extent arm-hand activity performance of both healthy adults and healthy children can be recorded reliably using a combination of multiple sensor devices, each containing a triaxial accelerometer, triaxial gyroscope and triaxial magnetometer.

Overall, the mean within-subject reliability was good to very good for all activities performed by the adults and the children as detected by a combination of multiple sensor devices. Between-subject reliability was good to very good for all activities performed by adults, measured with 4 sensor

devices. For the measurement with children, between-subject reliability was also good to very good for the sensor devices placed on the arm-hand manipulating the objects. However, more variance was found regarding the between-subject ICCs compared to the within-subject ICCs.

As to each activity separately, the within-subject reliability was higher for the activity drinking compared to the activity eating and combing in the adults and the children. This may be explained by activity complexity [23]. The activity drinking is less complex compared to eating since the latter is a bimanual activity demanding more cognitive effort and bimanual coordination, which are known to be more difficult compared to unimanual arm-hand use [18]. Furthermore, the activity eating takes longer, requiring more effort and information processing activities [23]. Complex activities consist of more component parts than less complex activities. The more component parts an activity consists of, the more variance is possible in the way an activity is executed, resulting in more variability in movement patterns, thereby decreasing the reliability rating. The higher reliability regarding the activity drinking compared to combing can be explained by the fact that the endpoint of the cup is relatively unchanging (i.e. the mouth) whereas the endpoint of the comb during the combing movements may vary, i.e. the orientation of the comb and the movement plane of the combing movement may vary between repetitions. This leads to more variability in movement patterns and thereby lower reliability.

Participants performed each activity in the 'no-instruction condition' (NI-condition) first, and subsequently in the 'instruction condition' (I-condition). Regarding the adults, the within-subject reliability for the activities eating and combing, was higher when performed in the I-condition. By giving instructions about how to perform the activity, the variability in execution of the activity was reduced, thereby increasing the reliability rating. For the activity drinking, the difference between performances in the two instruction conditions was minimal. This may be explained by the fact that there are not many ways to perform the activity drinking, i.e. this activity always includes reach, grasp, transport the cup to the mouth, take a sip and place the cup back on the table. Most people move along the shortest trajectory while performing the activity, therefore the variability in execution of the activity is rather small, even without instructions. Regarding the reliability of skill performance in the activities performed by the children, results similar to those of the healthy adults were found. But there was a difference between the two age groups of children, i.e. the difference in reliability of movement patterns in the two instruction conditions was smaller in the children aged between 12-18 years compared to the children aged between 6-12 years. This difference was especially large when the movement of the arm-hand performing the activity could be varied, i.e. when the arm-hand could follow different movement trajectories while executing the activity. For example, for the activity opening the zipper, the movement of the arm-hand not manipulating the zipper could make any (or no) movement while the subject was still able to complete the activity, whereas the movement of the arm-hand manipulating the zipper always included grasping the zipper, pulling the zipper down and releasing the zipper. The variability in possible movement trajectories, logically, resulted in a lower

reliability. When an instruction was given about which movement the arm-hand not manipulating the zipper should make, the reliability improved and hence the variability decreased. This was similar for the arm-hand manipulating the knife during the activity eating.

The results of the reliability of activities performed by children are in line with a reliability study of Speth et al. in which reliability of the Observational Skills Assessment Score (OSAS) was investigated [24]. In that study, test-retest reliability for the quality of use domain was higher in children aged between 12-16 years compared to children aged between 2.5 and 6 years old. These differences in reliability can be explained by differences in motor control and motor planning, changing with age. For example, Janssen et al. have shown a trend of motor planning improvement with increasing age in typically developing children [25], whereas Schneiberg et al. showed that movement profiles become smoother and more consistent with age [26]. According to the Fitts and Posner three stage model of motor learning [27], skill performance may become more consistent when subjects are in a more advanced stage of motor learning, which might explain the higher reliability in the older age groups.

To investigate the assumption that sub tasks will be recorded more reliably compared to those of a complete activity, sub tasks of the activity drinking were recorded in healthy adults and the reliability of both the complete activity drinking and sub tasks of drinking was determined. Reliability of the sub tasks was classified as very good. It was seen that the reliability of the sub phases was comparable or slightly higher compared to the reliability of the complete activity.

Between-subject reliability was good to very good, but had more spreading compared to within-subject reliability. This is reasonable since the variability between subjects is higher than within subjects. If someone is asked to perform an activity five times, it is very likely that the person performs the activity five times in a similar way, because everyone has a natural tendency to hold on to a certain movement pattern [28]. Whereas it is expected that different persons perform a similar activity differently. This is especially present in the activity eating performed by children. The reliability of the movement patterns of the performance in the NI-condition is much lower compared to the I-condition, indicating between subject differences in the way of executing the activities. When looking at the reliability of the arm-hand separately, it can be seen that this is caused by the arm-hand manipulating the knife. During the measurements it was already observed that the children manipulated the knife in different ways. Furthermore, many children, especially in the younger age group, did not use the knife during activity execution.

Some limitations of this study needs to be addressed. In the present study, reliability describes the resemblance in signal patterns between multiple similar movements. This reliability might be influenced by several factors. First of all, the recorded signals were low-passed filtered to retain only the low frequency signal content associated with the three-dimensional spatial displacement / kinematics of the activity performed. Filtering, however, transforms the original signal, reducing signal content. Low-pass filtering of the signals, in general, leads to improved ICC values across repetitions because the high frequency content is removed from the

signal. However, low-pass filtering of the data was essential to obtain the low frequency content of the movements of the upper extremity segments during activity execution. [29, 30]. Secondly, another factor that might have influenced the between-subject reliability is the placement of the sensors on the body, directly affecting the x, y and z vectors of the sensor signals. Although this happened in a very standardized manner, differences in sensor placement between subjects most certainly were present. In order to reduce this problem, for each sensor within each device, a 3D resultant signal was calculated. This (directionless) resultant signal is insensitive to small differences of positioning of the sensor on a body segment. Thirdly, no instructions were given regarding speed of performance. This resulted in variations in duration of the execution of the activities between repetitions and between persons. We have corrected for the differences in total activity duration by time-normalizing the signal per repetition. Using this time-normalisation across the entire time span of an activity, however, does not correct for small time-invariances (i.e. different time-lengths) of sub phases across activity repetitions. For example, the sub phase 'lifting the cup towards the mouth' may differ as to its time length between repetitions. This may have resulted in variations in the proportionate contribution and timing of the sub phases to the complete activity. These small variations in time were not normalized using the abovementioned method. This may have led to somewhat lower ICC values across activity repetitions (within subjects and between subjects). However, despite these time-normalization issues, the reliability was still found to be good to very good.

The measurements of the adults were performed first, after which some adjustments were made for the measurements with the children. For example, the measurements with the adults were performed with 4 devices (on the chest and dominant arm-hand), whereas the measurements with the children were performed with 7 devices (on the chest and both arms and hands). After having performed part I of this study it became clear that measuring both arms and hands would enhance insight into specific aspects of bimanual task performance. Information about movement patters of the non-dominant arm-hand are also very valuable, especially when measuring patients, since their affected arm-hand is mostly use as assisting (non-dominant) arm-hand. Since more devices were used during the measurements with children, the sample frequency was adjusted to accommodate the maximum throughput.

In conclusion, it can be stated that a combination of multiple sensors each consisting of a triaxial accelerometer, triaxial gyroscope and triaxial magnetometer can reliably record activities of daily living. The instrument as used in the current setting can be used to measure skilled arm-hand performance in healthy individuals reliably. In the present paper, the activities of daily living were performed in a standardized environment. The ultimate aim is to develop an instrument capable of identifying and assessing activities of daily living in the home situation of patients. The performance of activities in daily life may differ from those performed in a standardized environment. For instance, the participants' posture at the start and end of an activity may vary and/or the execution of activities may vary. However, despite this variability, it is possible to reliably recognize activities in both standardized situations and daily life situations [31], which is essential for future usability of the system in studying differences in skill performance within and between subjects/patients. Future research will focus on the investigation of the reliability of the system in recording activities performed by patients both during lab conditions and during their daily pursuits. Furthermore, in addition to the described activities, more activities should be investigated. The ultimate aim is to use this instrument to identify activities and determine the amount of use as well as the quality of performance of arm-hand related skills in patients during daily life.

REFERENCES

[1] M. Arner, A. C. Eliasson, S. Nicklasson, K. Sommerstein, and G. Hagglund, "Hand function in cerebral palsy. Report of 367 children in a population-based longitudinal health care program," *J Hand Surg Am,* vol. 33, pp. 1337-47, Oct 2008.
[2] J. G. Broeks, G. J. Lankhorst, K. Rumping, and A. J. Prevo, "The long-term outcome of arm function after stroke: results of a follow-up study," *Disabil Rehabil,* vol. 21, pp. 357-64, Aug 1999.
[3] E. Pagliano, E. Andreucci, R. Bono, C. Semorile, L. Brollo, and E. Fedrizzi, "Evolution of upper limb function in children with congenital hemiplegia," *Neurol Sci,* vol. 22, pp. 371-5, Oct 2001.
[4] S. Tyson and G. Turner, "Discharge and follow-up for people with stroke: what happens and why," *Clin Rehabil,* vol. 14, pp. 381-92, Aug 2000.
[5] M. van Eck, A. J. Dallmeijer, I. S. van Lith, J. M. Voorman, and J. Becher, "Manual ability and its relationship with daily activities in adolescents with cerebral palsy," *J Rehabil Med,* vol. 42, pp. 493-8, May.
[6] C. D. Wolfe, "The impact of stroke," *Br Med Bull,* vol. 56, pp. 275-86, 2000.
[7] D. S. Nichols-Larsen, P. C. Clark, A. Zeringue, A. Greenspan, and S. Blanton, "Factors influencing stroke survivors' quality of life during subacute recovery," *Stroke,* vol. 36, pp. 1480-4, Jul 2005.
[8] R. Lemmens, A. Timmermans, Y. Janssen-Potten, R. Smeets, and H. Seelen, "Valid and reliable instruments for arm-hand assessment at ICF activity level in persons with hemiplegia: a systematic review," *BMC Neurology,* vol. 12, pp. 1-17, 2012.
[9] G. Rau, C. Disselhorst-Klug, and R. Schmidt, "Movement biomechanics goes upwards: from the leg to the arm," *J Biomech,* vol. 33, pp. 1207-16, Oct 2000.
[10] S. J. Preece, J. Y. Goulermas, L. P. Kenney, D. Howard, K. Meijer, and R. Crompton, "Activity identification using body-mounted sensors--a review of classification techniques," *Physiol Meas,* vol. 30, pp. R1-33, Apr 2009.
[11] A. A. A. Timmermans, R. P. J. Geers, J. A. Franck, P. Dobbelsteijn, A. I. F. Spooren, H. Kingma, *et al.*, "T-TOAT: A method of task-oriented arm training for stroke patients suitable for implementation of exercises in rehabilitation technology," in *IEEE 11th International Conference on Rehabilitation Robotics* Kyoto International Conference Center, Japan., 2009, pp. 98-102.
[12] Freescale. (08-12-2014). *Accelerometer data sheet*. Available: http://cache.freescale.com/files/sensors/doc/data_sheet/MMA7361L.pdf
[13] InvenSense. (08-12-2014). *Gyroscope data sheet*. Available: http://invensense.com/mems/gyro/documents/PS-IXZ-2510A-00.pdf and http://invensense.com/mems/gyro/documents/PS-ISZ-2510A-00.pdf
[14] Honeywell. (08-12-2014). *Magnetometer data sheet*. Available: http://www51.honeywell.com/aero/common/documents/myaerospacecatalog-documents/Defense_Brochures-documents/HMC5843.pdf
[15] L. B. Mokkink, C. B. Terwee, D. L. Patrick, J. Alonso, P. W. Stratford, D. L. Knol, *et al.*, "The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes," *J Clin Epidemiol,* vol. 63, pp. 737-45, Jul 2010.

[16] H. C. W. De Vet, C. B. Terwee, L. B. Mokkink, and D. L. Knol, *Measurement in Medicine*. Cambridge: Cambridge University Press, 2011.

[17] J. E. Lexell and D. Y. Downham, "How to assess the reliability of measurements in rehabilitation," *Am J Phys Med Rehabil,* vol. 84, pp. 719-23, Sep 2005.

[18] R. A. Magill, "Motor Learning and Control. Concepts and Applications " in Eigth edition ed New York: The McGraw-Hill Companies, 2007, pp. 2-19.

[19] J. Kottner, L. Audige, S. Brorson, A. Donner, B. J. Gajewski, A. Hrobjartsson*, et al.*, "Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed," *Int J Nurs Stud,* vol. 48, pp. 661-71, Jun 2011.

[20] A. A. Timmermans, H. A. Seelen, R. D. Willmann, W. Bakx, B. de Ruyter, G. Lanfermann*, et al.*, "Arm and hand skills: training preferences after stroke," *Disabil Rehabil,* vol. 31, pp. 1344-52, 2009.

[21] R. J. M. Lemmens, Y. J. M. Janssen-potten, A. A. A. Timmermans, A. Defesche, R. J. E. M. Smeets, and H. A. M. Seelen, "Arm hand skilled performance in cerebral palsy: activity preferences and their movement components," *BMC Neurology,* vol. 14, 2014.

[22] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics,* vol. 33, pp. 159-74, Mar 1977.

[23] G. Wulf and C. H. Shea, "Principles derived from the study of simple skills do not generalize to complex skill learning," *Psychon Bull Rev,* vol. 9, pp. 185-211, Jun 2002.

[24] L. Speth, Y. Janssen-Potten, P. Leffers, E. Rameckers, A. Defesche, R. Geers*, et al.*, "Observational skills assessment score: reliability in measuring amount and quality of use of the affected hand in unilateral cerebral palsy," *BMC Neurol,* vol. 13, p. 152, 2013.

[25] L. Janssen and B. Steenbergen, "Typical and atypical (cerebral palsy) development of unimanual and bimanual grasp planning," *Res Dev Disabil,* vol. 32, pp. 963-71, May-Jun 2011.

[26] S. Schneiberg, H. Sveistrup, B. McFadyen, P. McKinley, and M. F. Levin, "The development of coordination for reach-to-grasp movements in children," *Exp Brain Res,* vol. 146, pp. 142-54, Sep 2002.

[27] P. M. Fitts and M. I. Posner, *Human Performance*. Oxford, England: Brooks and Cool, 1967.

[28] A. C. King, R. Ranganathan, and K. M. Newell, "Individual differences in the exploration of a redundant space-time motor task," *Neurosci Lett,* vol. 529, pp. 144-9, Nov 7 2012.

[29] J. Sinclair, P. J. Taylor, and S. J. Hobbs, "Digital filtering of three-dimensional lower extremity kinematics: an assessment," *J Hum Kinet,* vol. 39, pp. 25-36, Dec 18 2013.

[30] D. A. Winter, *Biomechanics and Motor Control of Human Movement*: John Wile & Sons Inc, 2009.

[31] R. J. M. Lemmens, Y. J. M. Janssen-Potten, A. A. A. Timmermans, R. J. E. M. Smeets, and H. A. M. Seelen, "Recognizing complex upper extremity skills using body worn sensors," *PLOS ONE,* accepted.