



MASTER THESIS

Test-retest reliability of an IMU sensor-based method for measuring quality metrics, during reach to grasp movement in healthy adults.

Ilana Davita Lakerveld

Biomedical Signals and Systems (BSS)

EXAMINATION COMMITTEE

Dr. Ir. B.J.F. Van Beijnum

Dr. G.H.R. Regterschot

M. Van Mierlo MSc

November 2021

UNIVERSITY OF TWENTE.

Test-retest reliability of an IMU sensor-based method for measuring quality metrics, during the reach to grasp movement in healthy adults.

November 18, 2021

Abstract

BACKGROUND Inertial Measurement Units(IMUs) are compact, reusable, and cheap. Therefore, it is a logical option for home revalidation for Cerebro Vascular Accident(CVA) patients. Unfortunately, there is not enough evidence for their validity, responsiveness, and reliability [1]. In this paper, the test-retest reliability of an IMU will be investigated for measuring different quality metrics during the reach to grasp movement in healthy adults.

RESEARCH QUESTION “What is the test-retest reliability of an IMU sensor-based method for measuring quality metrics, during the reach to grasp movement in healthy adults?”

METHODS Ten healthy adults, of which six females, participated in this experiment. They performed five 2D reach to grasp tasks, and one functional drinking task, each fifteen times, with five IMUs. They repeated these tasks three times with donning and doffing of the sensor system in between. To get the orientation in the earthframe, the madgwick filter will be used [2]. The beginning of a movement is defined when the angular velocity of the hand IMU was ≥ 0.05 rad/s and the endpoint is defined as were the angular velocity of the hand IMU ≤ 0.05 rad/s. For each metric the Kruskal Wallis test, intraclass correlation coefficient (ICC), and the standard error measurement (SEM) were calculated.

RESULTS The reach to grasp movements had a good to excellent relative reliability for all the different metrics: Movement time (range ICC:0.73-0.92), smoothness (range ICC: 0.55-0.93), trunk displacement (range ICC: 0.62-0.97), mean velocity (range ICC: 0.63-0.9), max velocity (range ICC: 0.52-0.93) and time to maximum velocity (range ICC: 0.52-0.93). The reach to grasp movements had an excellent absolute reliability for the metrics: Movement time (range SEM: 5.4-9.2%), and smoothness (range SEM: 2.2-2.6 %). The reach to grasp movements had an excellent to poor absolute reliability for the metrics: Mean velocity (range SEM: 6.4-12.6%), maximum velocity (range SEM: 6.6-17.2 %), and time to maximum velocity (range SEM: 6.1-19 %). The reach to grasp movements had a poor absolute reliability for the metric trunk displacement (range SEM: 13-40.9 %). The functional task had an excellent relative reliability for the metric maximum velocity (ICC: 0.96), a good relative reliability for the metrics: Movement time (ICC 0.53), smoothness (ICC: 0.65), and trunk displacement (ICC: 0.6), and a poor relative reliability for the metric: Mean velocity. The absolute reliability were excellent for the metrics: Smoothness (SEM: 8.3%), mean velocity (SEM 10%) and maximum velocity (SEM 3.1%). The absolute reliability was poor for the metrics: Movement time (SEM: 19.6%) and trunk displacement (SEM: 47.4%). The results for velocity profiles had a high correlation coefficient (range r: 0.95-0.99).

CONCLUSION This study evaluates the test-retest reliability of IMU design with associated metrics of the upper extremity(UE) by healthy subjects. This research indicates good to excellent test-retest reliability for reach to grasp movements by the metrics movement time, smoothness, velocity profiles, mean velocity and maximum velocity.

Keywords: *Inertial Measurement Unit (IMU), test-retest reliability, upper extremity(UE)*

Introduction

Cerebro Vascular Accident (CVA, also known as stroke) is a worldwide problem. In the United States and Europe, around 1.3 million first-ever strokes are approximately expected each year [3]. With age as one of the most important factors, this number is expected

to increase strongly due to the aging of the population [3]. Furthermore, stroke is the main cause of long-term disability [4]. Therefore, it has a high impact on the patients and society.

At hospital admission 85% of all patients with stroke have an upper extremity (UE) deficit [5] and 50 to 70 % of the patients with stroke suffer from deficits

of the UE in daily life [6]. Deficits of the UE are muscle weakness and contracture, changes in muscle tone, joint laxity, and impaired motor control. This induces disabilities in frequently used activities like reaching, picking up, and holding an object [7].

Nowadays rehabilitation at home is based on self-training. Unfortunately, there are some weaknesses with self-training. Firstly, if the training is not appropriately adapted to the needs of the patient, the patient may not benefit from the training. It is important to train specific motor skills, as they are represented in memory in a highly specialized way [8, 9]. Secondly, when there are home exercises, the physician cannot give feedback or adjust the exercise, and it is demotivating for the patient when exercises are too hard or too easy [9]. Furthermore, if an exercise is too easy a patient does not learn from the exercise [9], and lastly, rehabilitation depends highly on the motivation of the patients [10].

This is worrisome because the effects of intensive and prolonged arm use (both via exercises and/or in daily life) on recovery can be large [11]. Patients who do not use their arm frequently have an increased risk of learned non-use [9, 12].

Currently, the intensiveness and high-cost are the major barriers to give high-quality (home) rehabilitation [13]. This is because a patient's movement has to be almost continuously monitored, so that if needed they can be corrected [14]. With a still-growing number of CVA patients, expensive care is not preferable. Because of increasing opportunities in telemedicine, there could be an alternative that will increase the recovery with lower cost.

Some research has already been done on how to monitor these activities. The major groups of systems that can monitor activities are optical marker systems, electromagnetic position tracking systems, markerless optical systems and Inertial Measurement Units (IMUs) [15]. Markerless optical systems are not useful for home rehabilitation because they do not provide accurate triplanar measurements in real time [15]. A disadvantage of using electromagnetic position tracking is the possible disturbance of the environmental magnetic field. [15, 16]. Therefore, the two more logical options to use are IMUs and optical marker systems.

Before looking further into IMU and optical marker systems first some information about the outcome metrics will be given. Independent of which system is used, metrics are used to say something about the movements. These are calculated based on the movement data. Examples of these metrics are movement time, peak velocity, smoothness, trunk displacement.

When using a system for home rehabilitation, metrics will be calculated to get insight into the movement self or into the quality of the movement of the patient. Since IMUs give acceleration they are good

in acceleration metrics while camera-based systems provide position data and are therefore better in position-based metrics. It is possible to calculate position from acceleration but this is subject to drift.

An optical-based system works great when using in a laboratory setting but is less feasible to implement in every patient's home. IMUs are more compact, reusable, and cheaper. Therefore, it is a logical option for home revalidation. However, in guidelines made by Kwakkel et al. [1], it is stated that there is not enough evidence for their validity, responsiveness, and reliability. In that study it is recommended to use it only in combination with a camera-based system. [1]

As stated, using only IMUs could be an interesting option for home rehabilitation for CVA patients. However, there needs to be more evidence about their reliability, validity and responsiveness. In this paper the focus will be on the test-retest reliability. Test-retest reliability tests the correlation between two (or more) measurements at different times, on the same subject(s) [17]. Test-retest reliability can be calculated within or between measurement sessions. This research focuses on the test-retest reliability between measurement sessions. The task should be specific because motor skills are represented in memory in a highly specialized way [9]. It should be a movement that can safely be done in a home environment and is often used in daily life. Reach to grasp is a movement that is commonly used in daily life, therefore, learning this has a good impact on the quality of life [11]. Thus, the task will be the reach to grasp movement. Furthermore, grasping is a more complex movement than only reaching tasks. Before investigating the test-retest reliability for CVA patients, the reliability for healthy adults needs to be examined. To test the test-retest reliability an experiment with healthy adults performing several reach to grasp task and a functional task will be done. This data will be used to calculate some metrics. From this metrics the relative and absolute reliability will be calculated.

This leads to the following question: "What is the test-retest reliability of an IMU sensor-based method for measuring quality metrics, to during the reach to grasp movement in healthy adults?"

Methods

Measurement configuration

The IMUs used are from Xsens (Enschede, the Netherlands) and have a sample frequency of 50 Hz. The placement of the five IMUs (Xsens IMU MTw Awinda (MTw-38A70G20)) is shown in figure 1. One sensor on the body of the sternum (height: between second and third rib), one on the shoulder (on the acromion), one in the middle of the lateral side of the upper arm, one in the middle of the dorsal side of the

lower arm and one on the dorsal side of the hand (on the third metacarpal bone).

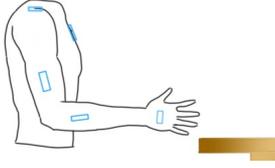


Figure 1: Placement of the 5 IMUs

Participants

Ten healthy adults, of which six female, participated in this experiment. The age ranged between 20 and 27 years (mean 23.3, SD \pm 2.45 years). Their height ranged between 165-199 cm (mean 178, SD, \pm 10.25 cm). Inclusion criteria were: age between 18-65, the right hand is dominant, and understanding Dutch and/or English. Exclusion criteria were: having the left hand as the dominant hand, having a neural or muscular condition or currently having trauma on the upper extremity.

Measurement protocol

The first step was explaining the experiment to the participant. After that, the sensors were placed as described in the section of the measurement configuration. There are two types of movements: reach to grasp and a functional task. The reach to grasp movement was reaching to a cup. There were five different positions where the cup was placed on the table, as shown in figure 2. The functional task was a 3D movement where besides reaching for a glass to pick it up, it was picked up and taken a sip out of it. Below a description is given on how the begin position, movement, and end position were defined.

Start position 2D movement

During the measurements, the subject was seated in a chair (back against the chair's backrest) by the table, with the feet flat on the floor. The arm was raised to a level that allows shoulder and elbow movement in a horizontal plane. The start position of the arm was at the trunk midline at 2/3rd of the subjects' arm length, marked with the s in figure 2. The elbow flexed \pm 90 degrees. The shoulder at \pm 70 degrees elevation and adduction. There were five different places the object was placed, namely, 20 cm before, after, to the left and the right of the starting position, and the far-right on 90% of the arm length of the subject. The object was a hard-plastic cup of 9-10 cm in height and 6-7 cm in diameter. These places are marked as 1 to 5 in figure 2.

Movement

When the researcher gave a sign to start, the subject moved the hand from starting position to an end

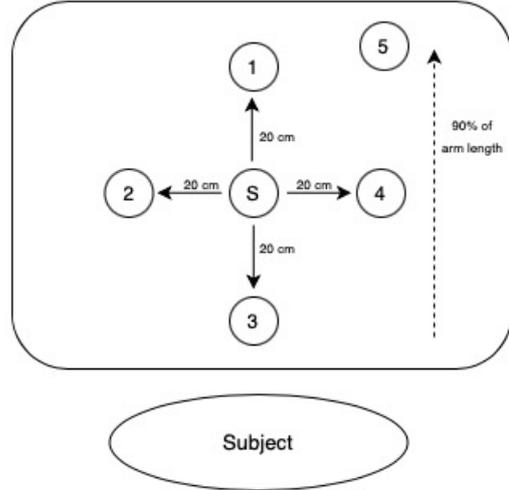


Figure 2: Schematic upper view of placement of the object. The rectangle is the table. The positions 1 to 5 are the end positions of the reach to task movement, so the positions where the cup is placed. S is the start position.

position. The movement was done at user preferred speed.

End position

When the object was reached and the person held the object, the person held still in this position for 1-2 seconds. After that, the hand released the cup and returned to the starting position. The hand was there at rest for 5- 10 seconds before it started reaching to the object again.

Besides these five 2D movements one 3D movement was made. This is picking up a glass to drink.

Movement	Measurement 1	M. 2	M. 3
1	15	15	15
2	15	15	15
3	15	15	15
4	15	15	15
5	15	15	15
Functional task	15	15	15

Table 1: Illustration of amount of trials for each movement and each measurement.

Start position 3D movement

The subject was seated in a chair back against the chair's backrest, having the hips and knees at 90 degrees. The elbow should have been 90 degrees and the hand pronated, resting on the edge of the table, where the wrist was aligned with the edge of the table. The shoulder should not have been elevated. The hard-plastic cup (same as in 2D movement) was

filled with approximately 100 mL water and placed in the midline of the body, 30 cm from the table edge.

Movement

The subject grasped the cup, brought it to the face, and took a sip. The subject had to keep the cup for about 3 seconds at the mouth. After that, the cup was returned to the initial position and the hand returned to the end position.

End position

The end position is the same as the initial position. After this movement, there were 10 seconds rest before starting the functional task again.

These six movements (five reach to grasp movements and one functional task movement) were repeated fifteen times. This was done in a predefined order, where first 15 times position one is reached then 15 times position two, then positions three, then four, then five, and last 15 times the functional task will be done. Before the 15 repetitions of each movement the hands were raised. The next step was to save the data and remove all the sensors from the subject. The placement and movements were repeated 2 times, so these steps were performed three times by every subject. Table 1 visualises the amount of trails in every movement and measurement session for one subject.

Data (pre)processing

Data was captured by Xsens MVN studio software. All the preprocessing steps are shown in figure 3. Preprocessing of the data and the selection of begin and endpoint were done in Matlab (The Mathworks, Inc; 2020b). Intregation of the data to velocity and position, calculation of the metrics and all statistical analyses were done in Python (Anaconda, Spider 3.3.6)

First the acceleration data was filtered by a lowpass butter filter with a cut off frequency of 10 Hz. To get the orientation in the earth frame the madgwick filter is used. [2]. The input of the madgwick filter is the accelerometer data and the gyroscope data as can be seen in 3. The convention made by this algorithm is that the gravitation of the earth frame is on the vertical axis [2]. The beta value of the madgwick filter was set to 0.1.

Because all the metrics are calculated with the norm of the data, there was no need to do a sensor to segment calibration.

Begin and endpoints detection

First the points between every reach to grasp movement were found manually using the raw accelerometer data by looking for the point where subjects raised their hand. Finding the begin and end points for every trial (one reach to grasp movement) in those movements was not done manually but using

detection script. This script used the norm of the gyroscope data.

Reach to grasp movement detection

First the reach to grasp movement and back to starting point were detected using the peaks in the norm of the gyroscope data. The movement of interest, was the reach to grasp movement, so only the first of the two peaks was used. From the peak point backward, there was looked for the begin point. This is where the angular velocity of the sensor on the hand is ≤ 0.05 rad/s. From the point forward there is also looked were the angular velocity of the sensor on the hand is ≤ 0.05 rad/s that is the endpoint.

Functional task movement detection

This worked quite similar as in the reach to grasp movement. The major difference is that from the two peaks that were defined the first one was used to find the starting point backward and the second one was used for finding the endpoint forwards. These begin and endpoint were stored so that the metric calculations are done only at these parts.

Velocity and position calculation.

For some of the metrics, the acceleration data needed to be changed to velocity or position. These were calculated by integrating the acceleration data. This was done for every trial separately so that the integration was done on small amounts of time. Before integrating the data, the gravitation of the vertical axis was removed first. After that, the data were integrated in all three axes to get the velocity data in all three axes. This velocity has drift due to integration. It is known that the velocity at the begin point $v(0)$ and the velocity at the end point $v(\text{end})$ are zero. If the velocity at the endpoint is higher than zero, this is due to drift. Based on this fact the drift can be identified and removed. A linear line made using the begin and end point of the velocity, is the drift of this trial. This drift was removed by deducting the linear line from the velocity data. The velocity without the drift was integrated in all three axes to get the position data in all three axes.

Metric computation

In figure 4 a schematic overview is shown, which displays which input is used for each metric. The acceleration data is the preprocessed as described above. The metric calculations were calculated for every trail of every movement using the preprocessed data.

Movement time

Movement time is defined as the time to reach the end goal and it is a well established metric [7]. Movement time is calculated as end time minus begin time.

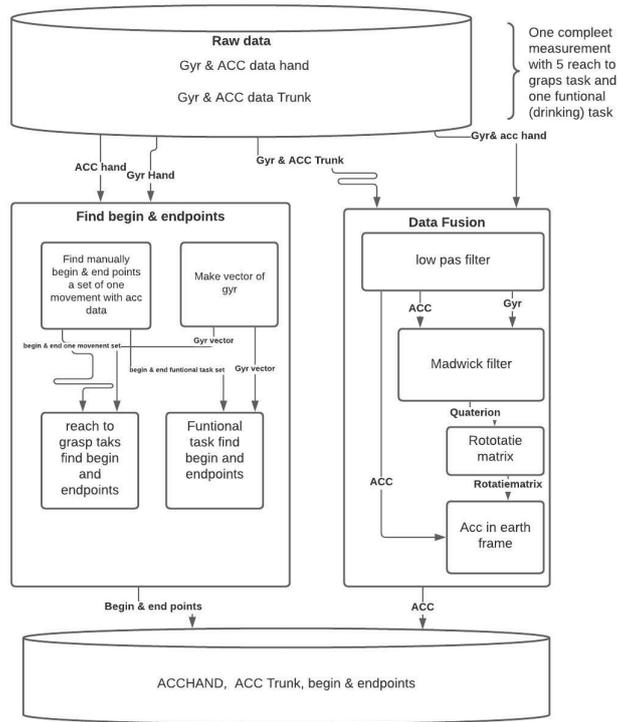


Figure 3: Schematic overview of the preprocessing steps and begin- and endpoints detection

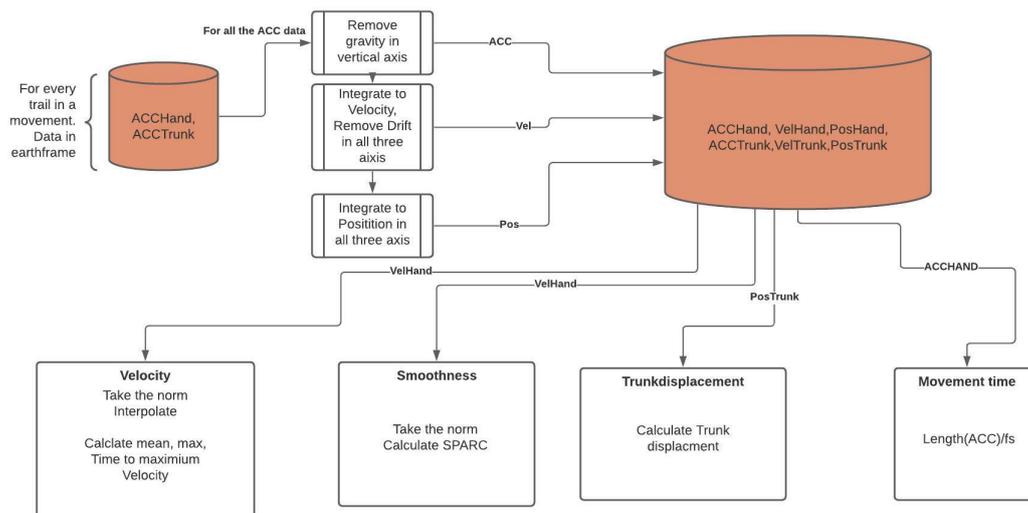


Figure 4: Schematic overview of different metric calculations and there input

This was based on the accelerometer data from the hand.

Smoothness

Stroke survivors have less smooth and more segmented movements. There are different ways to calculate smoothness. Mohamed Refai et al [18] concluded that correlation metric (CM) and Spectral Arc Length adapted (SPARC) are the best ways to define smoothness in ‘reach to point tasks’ and SPARC, as defined by Balasubramanian et al [19], gives also good results for reach to grasp tasks [18]. Therefore the SPARC will be used as smoothness metric.

As input the norm of the velocity is used.

$$SPARC = - \int_0^{\omega_c} \left[\left(\frac{1}{\omega_c} \right)^2 + \left(\frac{d\hat{V}(\omega)}{d\omega} \right)^2 \right]^{\frac{1}{2}} d\omega \quad (1)$$

$$\hat{V}(\omega) = \frac{V(\omega)}{V(0)} \quad (2)$$

$$\omega_c = \min\{\omega_c^{max}, \min\{\omega, \hat{V}(r) < \bar{V} \forall r > \omega\}\} \quad (3)$$

(ω) is the Fourier magnitude spectrum $v(t)$ and $\hat{V}(\omega)$ is the normalized magnitude spectrum, normalised with respect to the magnitude of the DC component $V(0)$. \bar{V} is a given threshold that is upperbounded by ω_c^{max} . [19]. SPARC is in dimensionless units. The smoothness value for reach to grasp movements for healthy subjects SPARC is -1.6. [19]

Trunk displacement

Individuals with stroke demonstrated greater trunk displacement during reaching [20]. This is used as a compensation mechanism in order to reach a goal or to create a larger range of motion (ROM)[21][7]. Trunk displacement will be calculated using the position data of the trunk. The displacement is mostly affected in the sagittal plane but Repnik et al [7] found better results using all three directions. Therefore, the trunk movement is calculated using all three directions.

$$Trunkdisplacement = \max(\sqrt{a})$$

$$a = (pos_x - pos_x(0))^2 + (pos_y - pos_y(0))^2 + (pos_z - pos_z(0))^2 \quad (4)$$

Velocity

The velocity is calculated with the norm of the velocity data of the hand IMU. The time axis of this data was transformed to a percentage of movement completion because not every trail has the same duration. From this velocity profile the mean and maximum velocity and the time to peak velocity are calculated. The mean velocity is the mean velocity of the velocity profile. The maximum velocity is were the velocity profile has the highest value. The last velocity based metric is the time to maximum velocity. This metric

looks at the place on where the maximum velocity value is in the duration of the movement.

Statistics

The statistics were calculated over each metric separately. A Kruskal Wallis test (>2 groups) checked the intrasubject reliability between the three measurements. This statistic was used because this can be done on non-normally distributed data, but also on normally distributed data. This was done with a confidence interval of 95 % (alpha = 0.05). Also, intraclass correlation coefficients (ICCs) were calculated (two way model type consistency, average unit). An ICC ≥ 0.75 is excellent relative reliability, an ICC between 0.4 and 0.75 means a fair to good relative reliability, an ICC <0.4 means bad relative reliability [22]. With the ICC also the standard error measurement (SEM) was calculated. The SEM is estimated as;

$$SEM = SD * \sqrt{(1 - ICC)} \quad (5)$$

Where the SD is based on all scores from all subjects. A SEM smaller than 10% of the mean is indicated as an excellent absolute reliability [23]. Furthermore, for the velocity profiles, the correlation coefficient between the three measurement sessions were calculated.

Results

From the ten participants described in the methods, one participant was left out because two out of three measurements were not usable due to technical problems (e.g. missing data in one or all of the sensors). From three other participants, only two of the three measurements were included because the third measurement had technical problems. Some trials were left out because the movement was done incorrectly.

Movement Time

The data of the metric movement time is normally distributed for the reach to grasp movements and non normally distributed for the functional task, as can be seen in appendix A. In figure 5, for every 2D movement three boxplots are shown, one for each measurement session. In figure 6 the boxplot for the functional task is shown with three boxes, one for each measurement session.

In table 2 the P-value for the Kruskal Wallis test, the ICC, and the SEM are shown. As can be seen in table 2 for all the different movements the three measurement sessions do not significantly differ from one another (alpha = 0.05). The ICC of movement 1, 2, 4 and 5 is larger than 0.75, which means that it has an excellent relative reliability. Movement 3 has

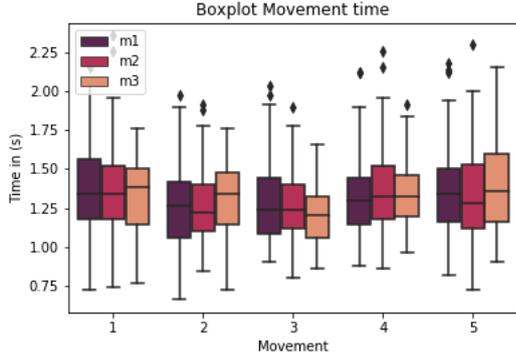


Figure 5: Boxplot of movement time of reach to grasp tasks. On the horizontal axis the different movement and on the vertical axis the time in seconds.

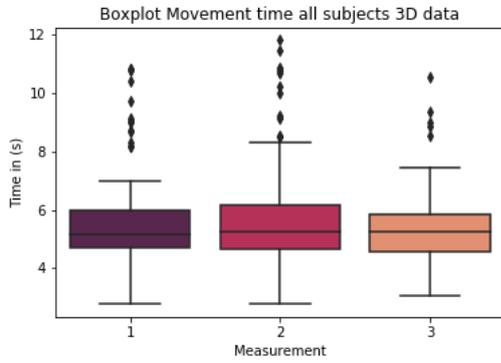


Figure 6: Boxplot of movement time of functional task. On the horizontal axis the different measurements and on the vertical axis the time in seconds.

an ICC of 0.73, which indicates a fair to good relative reliability. The functional task has an ICC of 0.53, so the relative reliability is good. All the reach to grasp movement have a SEM below 10% which indicates an excellent absolute reliability. The functional task has a SEM above 10%, which indicates a poor absolute reliability.

Smoothness

The data of the metric smoothness was skewed distributed, for both the reach to grasp movements as for the functional task, as can be seen in appendix

Movement	P value	ICC	SEM
1	0.83	0.79	0.12 (9.2%)
2	0.14	0.92	0.07 (5.4 %)
3	0.21	0.73	0.11 (9.1%)
4	0.42	0.82	0.10 (7.6%)
5	0.17	0.84	0.11 (8.3 %)
Functional task	0.84	0.53	1.10 (19.6%)

Table 2: Movement time

A. In figure 7, for every 2D movement three boxplots are shown, one for each measurement session. In figure 8 the boxplot for the functional task is shown with three boxes, one for each measurement session. In table 3 the P-value for the Kruskal Wallis test, the ICC, and the SEM are shown. For all the different movements, except for movement 1, the three measurement sessions do not significantly differ from one another ($\alpha = 0.05$), as can be seen in table 3.

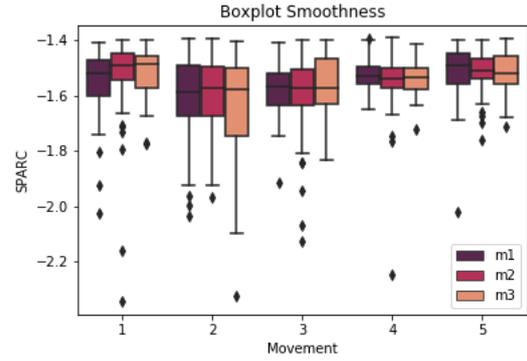


Figure 7: Boxplot of smoothness of reach to grasp tasks. On the horizontal axis the different movement and on the vertical axis the SPARC.

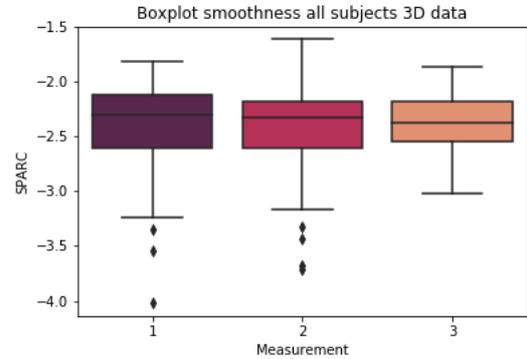


Figure 8: Boxplot of smoothness of functional task. On the horizontal axis the different measurements and on the vertical axis the SPARC.

The ICC of movements 1, 2, 3 and 5 is larger than 0.75, which means that it has excellent relative reliability, movement 4 and the functional task have an ICC between 0.4 and 0.75, which indicates a fair to good relative reliability. All movements have an SEM below 10% which indicates an excellent absolute reliability.

Trunk displacement

The data of the metric trunk displacement was skewed distributed both, for the reach to grasp movements as for the functional task, as can be seen in appendix A. In figure 9, for every 2D movement three boxplots are shown, one for each measurement session. In figure 10

Movement	P value	ICC	SEM
1	0.002	0.86	0.04 (2.6 %)
2	0.61	0.93	0.04 (2.5%)
3	0.79	0.88	0.03 (2.2%)
4	0.45	0.55	0.05 (3.1%)
5	0.56	0.79	0.03 (2.2%)
Functional task	0.47	0.65	0.2 (8.3%)

Table 3: Smoothness

Movement	P value	ICC	SEM
1	0.0029	0.90	0.004(19.2%)
2	0.014	0.62	0.004(40.9%)
3	0.0423	0.73	0.003(34.1%)
4	0.004	0.68	0.004(38.8%)
5	0.75	0.97	0.002(13.0%)
Functional task	0.016	0.6	0.032(47.4%)

Table 4: Trunk displacement

the boxplot for the functional task is shown with three boxes, one for each measurement session.

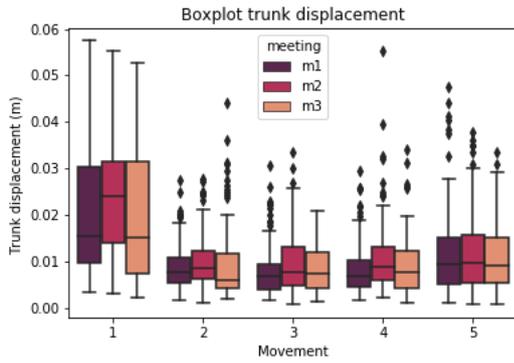


Figure 9: Boxplot of trunk displacement of reach to grasp tasks. On the horizontal axis the different movement and on the vertical axis the trunk displacement in meters.

In table 4 the P-value for the Kruskal Wallis test, the ICC, and the SEM are shown. For all the different movements, except for movement 5, the three measurement sessions do significantly differ from one another ($\alpha = 0.05$), as can be seen in table 4. The ICC of movements 1 and 5 is larger than 0.75, which means that it has an excellent relative reliability. The ICC of movements 2, 3, 4, and the functional task are between 0.4 and 0.75, which indicates a fair to good relative reliability. Both the functional task and all the reach to grasp movements have a SEM above 10%, which indicates a poor absolute reliability.

Velocity

Velocity profile

In figure 11 the mean velocity profile for movement 1 is shown. Not every movement has the same duration,

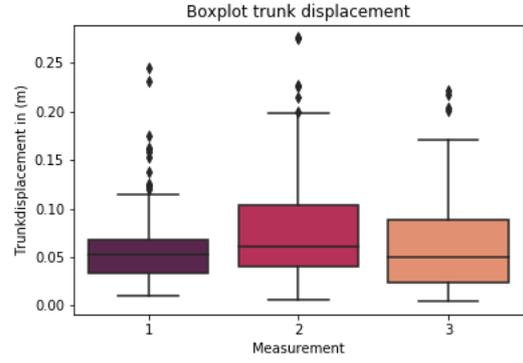


Figure 10: Boxplot of trunk displacement of functional task. On the horizontal axis the different measurements and on the vertical axis the trunk displacement in meters.

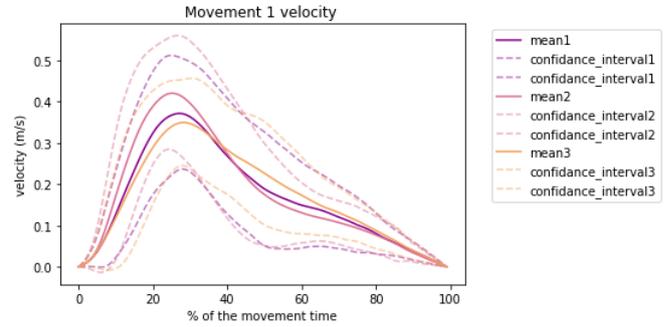


Figure 11: Velocity over time movement one

therefore the data has been interpolated to 100% of the movement time. In addition to the mean value, the 95%-confidence level is shown.

The correlation between measurements one and two is 0.98, between one and three is 0.99, and between two and three is 0.95. In figure 12 the mean velocity for movement 2 over every part of the movement is shown. In addition to the mean value, the 95%-confidence level is shown. The correlation between measurements one and two is 0.99, between one and three is 0.99, and between two and three is 0.98.

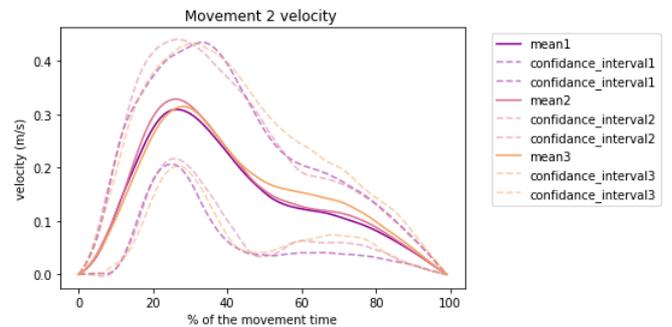


Figure 12: Velocity over time movement two

In figure 13 the mean velocity for movement 3 over

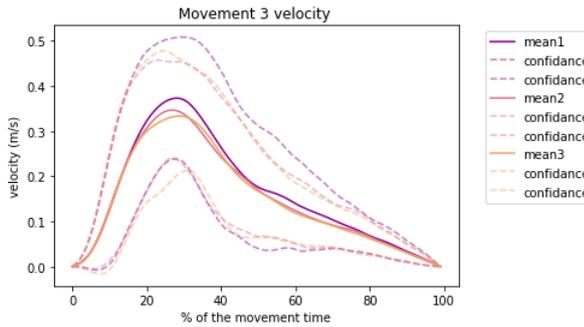


Figure 13: Velocity over time movement three

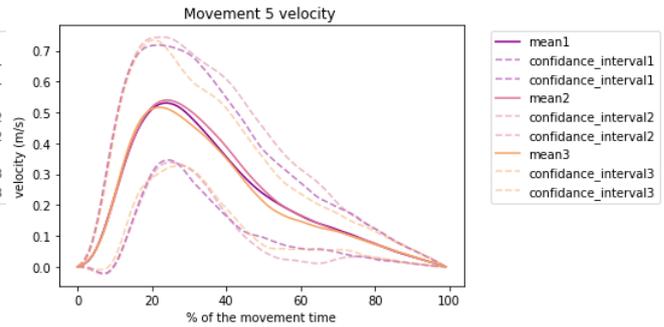


Figure 15: Velocity over time movement five

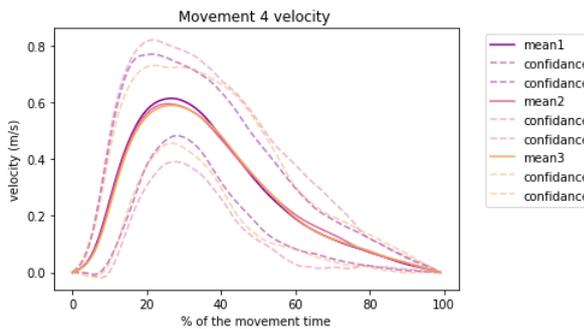


Figure 14: Velocity over time movement four

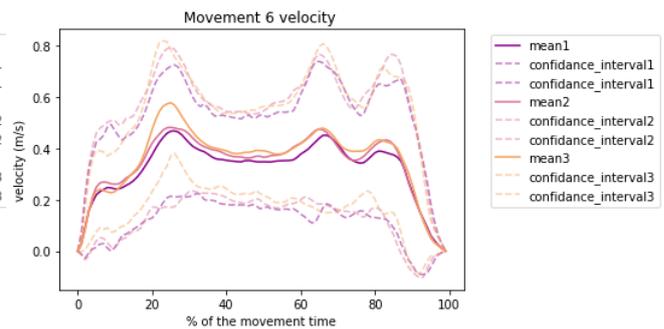


Figure 16: Velocity over time movement Functional task

every part of the movement is shown. In addition to the mean value, the 95%-confidence level is shown. The correlation between measurements one and two is 0.99, between one and three is 0.99, and between two and three is 0.99.

In figure 14 the mean velocity for movement 4 over every part of the movement is shown. In addition to the mean value, the 95%-confidence level is shown. The correlation between measurements one and two is 0.99, between one and three is 0.99, and between two and three is 0.99.

In figure 15 the mean velocity for movement 5 over every part of the movement is shown. In addition to the mean value, the 95%-confidence level is shown. The correlation between measurements one and two is 0.99, between one and three is 0.99, and between two and three is 0.99.

In figure 16 the mean velocity for the functional task over every part of the movement is shown. In addition to the the mean value, the 95%-confidence level is also shown. The correlation between measurements one and two is 0.99, between one and three is 0.98, and between two and three is 0.98.

Mean velocity

The data of the metric mean velocity was normally distributed, for both the reach to grasp movements as for the functional task, as can be seen in appendix A. In figure 17, for every 2D movement three boxplots are shown, one for each measurement session. In figure 18 the boxplot for the functional task is shown

with three boxes, one for each measurement session. In table 5 the P-value for the Kruskal Wallis test the ICC, and the SEM are shown. For reach to grasp movements 2 and 4, the three measurement sessions do not significantly differ from one another ($\alpha = 0.05$), as can be seen in table 5. For the other movements the three measurement sessions are significantly different.

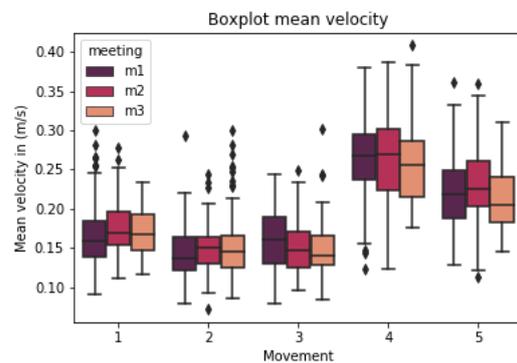


Figure 17: Boxplot of mean Velocity of reach to grasp tasks. On the horizontal axis the different movement and on the vertical axis the velocity in m/s.

The ICC of movements 1, 2, 3 and 4 is larger than 0.75, which means that it has an excellent relative reliability. The ICC of movements 5 and the functional task are between 0.4 and 0.75, which

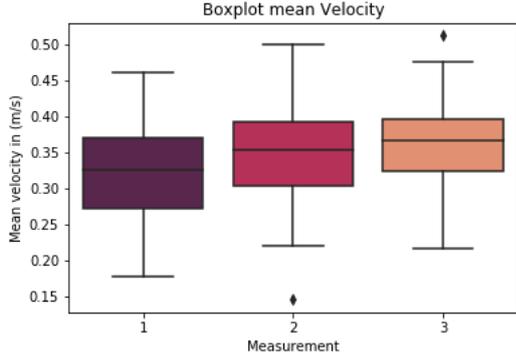


Figure 18: Boxplot of mean Velocity of functional task. On the horizontal axis the different measurements and on the vertical axis the velocity in m/s.

Movement	P value	ICC	SEM
1	0.03	0.90	0.01(6.4%)
2	0.06	0.85	0.01(8.7%)
3	0.005	0.80	0.02(10%)
4	0.42	0.85	0.02(7.73%)
5	0.02	0.63	0.03 (12.6%)
Functional task	0.0002	0.29	0.03(10%)

Table 5: Mean velocity

indicate a fair to good relative reliability. Movements 1, 2 and 4 have a SEM below 10%, movements 3 and the functional task have a SEM of 10%, and movement 5 has a SEM of 12.6%, which indicates a poor absolute reliability.

Maximum velocity

The data of the metric maximum velocity was normally distributed, for both reach to task movements as for the functional task, as can be seen in appendix A. In figure 19 the boxplot for the functional task is shown with three boxes, one for each measurement session. In figure 20, for every 2D movement three boxplots are shown, one for each measurement session.

In table 6 the P-value for the Kruskal Wallis test, the ICC, and the SEM are shown. For all the different movements, except for movement 4, the three measurement sessions do significantly differ from one another ($\alpha = 0.05$), as can be seen in table 6. The ICC of movements 2, 3, 4 and the functional task are larger than 0.75, which means that it has an excellent relative reliability. The ICCs of movements 1 and 5 are between 0.4 and 0.75, which indicates a fair to good relative reliability. Movements 2, 3, 4 and the functional tasks have a SEM below 10% and movements 1 and 5 have a SEM above 10%, which indicates a poor absolute reliability.

Time to max velocity

The data of the metric time to max velocity was

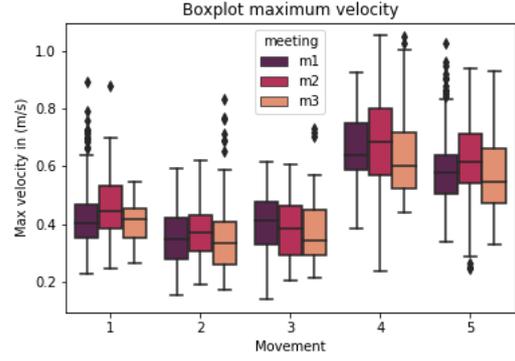


Figure 19: Boxplot of max Velocity of reach to grasp tasks. On the horizontal axis the different movement and on the vertical axis the velocity in m/s.

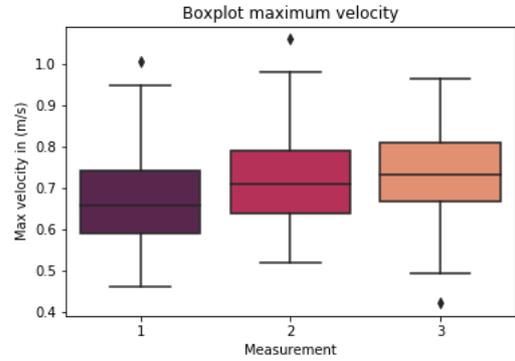


Figure 20: Boxplot of max Velocity of functional task. On the horizontal axis the different measurements and on the vertical axis the velocity in m/s

normally distributed as can be seen in appendix A. In figure 21 are for every 2D movement three boxplots shown, one for each measurement session.

In table 7 the P-value for the Kruskal Wallis test, the ICC, and the SEM are shown. For the functional task the calculation is not made. For all the different movements, except for movement 1, the three measurement sessions do not significantly differ from one another ($\alpha = 0.05$), as can be seen in table 7. The ICCs of movements 3, 4 and 5 are larger than 0.75, which means that it has an excellent relative reliability. The ICCs of movement 1 and 2 are between 0.4 and 0.75, which indicates a fair to good relative

Movement	P value	ICC	SEM
1	8.44e-5	0.52	0.07(17.2%)
2	0.04	0.90	0.03(9.2%)
3	0.004	0.93	0.03(7.1%)
4	0.09	0.92	0.04(6.6%)
5	0.009	0.71	0.07(12.4%)
Functional task	2.37e-5	0.96	0.02(3.1%)

Table 6: Max velocity

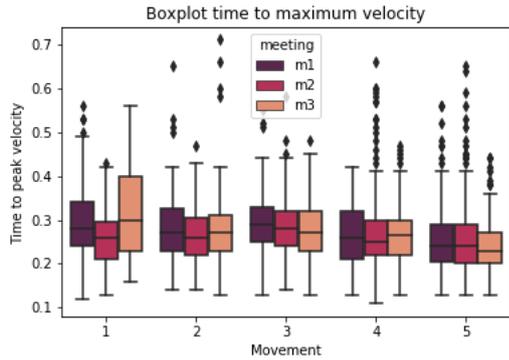


Figure 21: Boxplot time to max Velocity of reach to grasp tasks. On the horizontal axis the different movement and on the vertical axis decimal were on the completion the maximal velocity of the movement is.

Movement	P value	ICC	SEM
1	1.71e-5	0.69	0.05(16,8%)
2	0.22	0.57	0.05(19,0%)
3	0.26	0.94	0.02(6,1%)
4	0.99	0.87	0.03(11,5%)
5	0.11	0.87	0.03(12,2%)
Functional task	-	-	-

Table 7: Time to max velocity

reliability. Movement 3 has a SEM below 10%. All the other movements have a SEM above 10%.

Discussion

The aim of this study was to determine the test-retest reliability of an IMU sensor-based method for measuring quality metrics, during the reach to grasp movement in healthy adults. Therefore, the reliability of all the different metrics were calculated. These reliability outcomes should give insight which metrics are reliable and gain evidence of the reliability. Kwakkel et al [1] stated that these are needed for IMU measurement system before it can be used as independent system for UE (stroke) rehabilitation. For such use, the test-retest reliability should be excellent.

Discussion on results

In figure 22 are the results summarized in colors. The color green is excellent results, yellow good/moderate result and red poor results. As can be seen in figure 22 the five reach to grasp movement have better outcome than the functional task. Furthermore, of the reach to grasp movements movement 5 was the least with especially lower results by the velocity metrics. Moreover, trunk displacement was the metrics with the least results, followed by the metric

time to max velocity. The metric that had the best results was smoothness.

The study shows that velocity profiles has the best results, the correlation coefficient range was 0.95-0.99. The metric movement time showed excellent results for the reach to grasp movements. With all the SEM's below 10% and all reach to grasp movements had an ICC above 0.75 except movement three that had an ICC that is close to 0.75, namely 0.73. The results of the functional task were not as good as the results of the reach to grasp movements. The ICC was good but not excellent and the functional task had a poor SEM. The metric smoothness had also good results with both the reach to grasp movement and the functional task SEM's below 10%. Four out of five reach to grasp movements have an excellent ICC (above 0.75) and one reach to grasp movement a good ICC (between 0.4-0.75). The functional task had also an ICC between 0.4-0.75. The metric trunk displacement had for the reach to grasp movements good to excellent ICC. For the functional task the ICC was good. The SEMs of all the reach to grasp movements and for the functional task were above poor. The metric mean velocity has for the reach to grasp movements mostly excellent ICCs and one ICC that was good. The corresponding SEM varied around the 10%. The ICC for the functional task is good and the SEM was on the turning point of being excepted as excellent. The metric max velocity has ICC for reach to grasp movements between good and excellent. The SEM for reach to grasp movement varied above and below 10%. For the functional task the ICC and the SEM were excellent. The metric time to max velocity was only used for the reach to grasp movements. The ICC was between good and excellent. For one movement the SEM was excellent. For the other the movements the SEM was poor.

Movement time

For the metric movement time, almost all of the reach to grasp movement showed excellent ICC results. The only movement that did not have excellent but good result was only 0.02 point away from having an excellent result. Furthermore, all the SEMs were also excellent. Also the different reach to grasp movements are in the same duration range, which was as expected because the first four movements are all the same distance.

The functional task has a good and not an excellent ICC and a poor SEM. This could be due to the complexity of the functional task. This task consists of reaching to the cup, bringing the cup to the mouth, taking a sip and bringing the cup back to starting position. That there is some fluctuation in duration could be due to more fluctation in duration of the different sub-tasks of which the functional task consists.

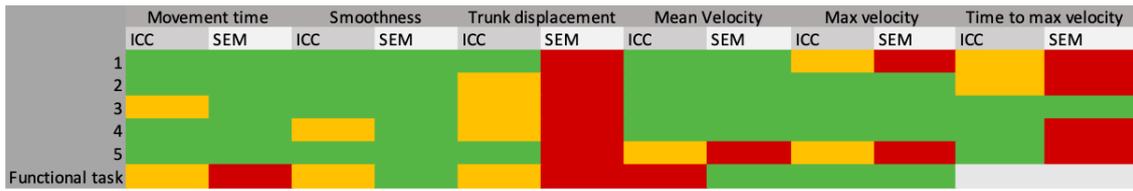


Figure 22: Schematic view on outcome results were green is excellent, yellow is good/moderate and red is poor outcome.

In appendix B the difference between subjects for the different reach to grasp movements are shown. As can be seen there are some differences between subjects. Some subjects had a shorter duration than others. However, the difference between measurements sessions of one subject are on average small, indicating that the the duration of the movements in different measurements are close.

In the research of Patterson et al [24], movement time was calculated with a motion capture system (VICON). This was done for reach to grasp movement. Averaging the outcome of the five reach to grasp movements, the relative reliability was about the same. The absolute reliability of the research of Patterson et al. was a little bit better.

Movement time is a metric that is easy to be calculated and the outcome is easy to be interpreted. The metric itself does depend on how consistent the begin and endpoints are taken.

Smoothness

The metric smoothness has excellent results with high ICC and the SEM are low meaning an excellent relative and absolute reliability. The exception is movement 4 with a good but not excellent relative reliability. In appendix B the difference between subjects for the different reach to grasp movements are shown. When comparing movement 4 with the other reach to grasp movement, movement 4 is not worse than the other movements. Moreover the SPARC values look even closer to eachother than for example, movement 3. The values of movement one until five are around -1.6, which is the normal value for healthy subjects in reach to grasp movements [19]. The results for the functional task are excellent, meaning this could be an interesting metric for a functional task. It would be interesting to have a normal healthy value for a the drinking task.

There are some outliers in the reach to grasp movement. There was something during the experiment that could explain these outliers. During the experiment the subject had to wait for a sign of the researcher to begin the reaching task. This had to be done quite some times. There were a view cases that a subject starting moving before the researcher gave the sign to move. When that happened the subject

had to continue the movement. Most of the times the subject hesitated before continuing the movement. This hesitating in a movement leads to a less smooth movement pattern. This could explain the outliers. During the experiment it has not been marked which movement these were, and therefore they could not be removed in advance.

The calculation of SPARC is based on the pathlength of the data in the frequency domain. This is done with a cut-off frequency of 20 Hz. In the preprocessing the data was filtered with a low pass filter of 10 Hz, removing frequencies above 10 Hz. This makes the pathlength from 10 - 20 Hz as small as possible (a straight line). This could have lead to an overestimation of the results for the SPARC, making movement more smooth.

Trunk displacement

For the trunk displacement, the SEM is very high, so there is a poor absolute reliability. The ICC is high, so the relative reliability is high. The values are really small; therefore, a small difference can have great impact on the absolute reliability. This could explain the poor absolute reliability. Futhermore, the fact that the movements are of really small values, it is important to explore what actually is measured. Due to intergrational problems position data does have some noise. The question raises whether the measured displacement is in fact displacement or that this metric mostly compares just the noise.

The small values could be due to the fact that for healthy subjects the placement of the cup was in a range that they can reach it without using their trunk. For movement 5, this placement was on 90% of arm length, and movement one, placed on the 20 cm above the start position, was the most far away for the subject. These movements have also the lowest SEM and the highest ICC, but still the SEMs are too high. It could be interesting to see, in healthy subjects, if the metric trunk displacement has a better reliability when healthy subject need to move their trunk.

Patterson et al [24] also calculated trunk displacement for reach to grasp movements, and showed about the same relative reliability but much lower absolute

reliability. this difference is probably due to the fact that Patterson et al did their research with an optical system (VICON) which is position based so does not have the problems due to integration.

Velocity

The correlation coefficients of the velocity profile were very high. This could be biased since this is calculated by taking the mean of the measurements and not the measurements themselves. These mean correlation gave high results, but this is calculated over average different subjects it is interesting to see how the individual trials and individuals measurement are correlated. The velocity pattern from one measurement could deviate from this mean and can therefore result in a lower correlation coefficient. To ensure the correlation values are indeed high, also some correlations of single measurements are calculated. In the appendix C, is a table where for movement one, each measurement of one subject is compared to all the other measurements (both the other measurements of that subject as all the measurements of the other subjects). In general these individual correlation coefficient gave also high results. Furthermore, in the appendix C are also figures visualizing the velocity profile of different trials in one measurement of one person. This also shows that the separate trials are more or less the same as the averaged one.

It's interesting to notice that all the following metrics (mean velocity, max velocity and time to max velocity) are all metrics depending on the velocity profile. Moreover, there are some metrics that are not considered in the research, like skewness, that also depend on this velocity profile. When visually inspecting the velocity profiles, once can get an indication on all of these metrics at one, because this are characteristics of the velocity profile, making these plots very interesting. Clinicians look at all these factors like mean, max, and time to max velocity in the revalidation of their patients [25]. Based on both the good results of the velocity profile, and the fact that this has all the information of the other metric within this plots, these velocity profiles would be very interesting to look at during revalidation.

For both mean velocity as for max velocity the difference between subjects were high, as can be seen in appendix B. Most logically, if the mean and maximum velocity are high, then the duration should also be shorter, because the velocity is higher but the length is not longer, indicating that the time should be shorter. When inspecting the results this relationship can be seen. For example movement one subject five has high mean and max velocity and low movement time.

As described in section smoothness the data was filtered with a low pass filter. This could also had have influence on the velocity results. When the acceleration data is lowpass filtered the higher

frequencies are left out. When integration to velocity the faster acceleration changes are left out, giving smoother velocities. This could have influenced the velocity profiles.

In a study by Colombo et al [26], the ICC for the mean velocity was 0.95 for healthy persons. This was measured with a robotic device, and this difference could explain the difference in ICC. Robotic devices measure position and the differentiated that to velocity. IMUs measure acceleration and integrate that to velocity. However, in contrast to differentiation, integrating is subjected to drift. In the research of Patterson et al [24], the maximum velocity was also calculated for reach to grasp movements. For comparison, the ICC and SEM of the five reach to grasp movement were averaged. The relative and absolute reliability is better then that of Patterson et al.

Discussion on the methods

It is important to highlight that for the Kruskal Wallis test (and statistics in general), the hypothesis is not automatically true when one cannot reject the zero hypothesis. It could be true however, it could also mean that there is not enough "evidence" to reject the hypotheses. It is not possible to change the zero hypotheses, but it is important to be critical on the data. With a large data set, it is more likely that the zero hypothesis is true.

ICC confidence intervals are quite large (between 0.1 – 0.8 is not uncommon), this is due to the small number of subjects (six subjects). Initially, this should be ten subjects, however, due to the already described problems only six of the subjects had three useful measurements sessions. Another option is to do the ICC only with two measurements sessions. The number of subjects is larger since there are three additional subjects with two measurements but lose one "rater", because only two measurement sessions are used instead of three measurement sessions. The impact of this on the confidence interval was small, therefore it is chosen to analyze the six subjects who have three complete measurements sessions.

It is interesting to know if the differences are due to donning and doffing of the system or due to differences between subjects, difference within subjects or due to other factors. The influence of the subject variability on the ICC is described by Bruton et al [27]. There they explain "So, if a variance between subjects is sufficiently high, then the reliability will inevitable appear to be high" [27]. So, this high ICC does not always indicate an excellent relative reliability, and that the donning and doffing of a system has no effect on the outcome, but can also be that the high results are due to the variability between the subjects.

Discussion on data preparation

It is interesting to look what the effects of the data preparation could have been on the results. There has already been described how the low pass filter has influence on the metric smoothness. Another influence is in the extraction of the begin and endpoint on the results. It is important to notice that the begin- and endpoints have a big influence on the results. The begin- and endpoint detection influences different metrics. That a different detection leads to a different movement time is quite obvious. But this is not the only metric the detection has an impact on. A different begin or end detection could also affect the metric smoothness. Specifically, where the end point is defined as endpoint could in the grasping phase be interesting. Small correction to actually holding the cup could be of influence calculating the smoothness. Furthermore, a too early detection could lead to a lower mean velocity and a too early end detection to a higher mean velocity. Having an earlier endpoint had also consequences for the time to peak velocity. Having impact on all these metrics make important to part in the process. In this research, the script generates begin and endpoints of the movement. This was done by finding points where the angular velocity is almost zero. The main advantage of this is that it works really fast and can be used to give fast results, which is a must if it is used in the future for real time feedback. But unfortunately, it is not perfect. When assessing this script, sample-wise, there was around 10% of the begin- and endpoints wrongly assessed. This looks like a lot, but most of them were repetition errors which means more than one error in the same measurement session. This is something that can be improved. Some of these wrong assessment lead to outliers in metrics. Another option was to choose the begin and endpoint manually. The main disadvantage is that is a tremendously large job. For a small data set as in this research, more than 5000 points have to be manually chosen. It is human to make mistakes and by doing such painstaking like this, it is very likely that there will also be some mistakes. To make it more accurate, and to overcome any researcher bias, at least one extra person needs to choose these points manually and the found points should be compared. The time-consuming argument and potential of the methods make that it will be recommended to generate it from the script but make some improvements on the script. This would have a positive effect on the result. Another option could be to have generated begin and endpoint through non-IMU data. This could be done by the pressure sensor on the begin point and the end object. A benefit for this method could be that is insensitive to the noise in the IMU data. A disadvantage would be that this system and the IMU should have some kind of calibration so that time data can be aligned. So it would only be recommended when the other suggested improvements do not give

the desired outcome.

Discussion on equipment

Besides the methodical improvements that can be made and discussion on the results, a brief reflection on the measurement system and the measurements will also be given. During the experiments, not all measurements went well. The measurements were done with two different IMU systems and a total of 9 individual IMUs. During the experiment, it happened that one of those nine IMUs lost connection (mostly from the XSens system). In that case, part of the measurement had to be redone. For the experiment itself, this was time-consuming and annoying but doable. However, it is not preferable to have unstable measurement objects when applying this system on patients. Furthermore, even though the problem was known (out of some test cases), and being extra careful during the experiments, there still had to be some measurements left out due to problems (large errors in data or loss of signal) in the data. Three measurements were left out due to large errors in data or loss of signal, which holds about 10% of the measurements. Fortunately, this is not a common problem of IMUs, because other people with more experience in working with IMUs did not recognize this problem. The most logical conclusion for these problems is therefore that the IMUs used in this experiment were often used and have battery problems or other maintenance issues. Making this not a problem for possible other research or further steps towards implementation, provided that not specifically these sensors are used. But if these problems do occur they should first be fixed before executing the experiments, to avoid problems during the measurements.

Limitation

One of the main limitations of the research is the small data set. With more data the confidence intervals of the ICC will be smaller and therefore more conclusions can be made. Another limitation is that there is no comparison with a reference system.

Validation

To investigate validity and reliability of the metrics based on the 2M IMU system, two sets of IMUs were placed on the arm. One sensor set was of XSens and the other one of 2M engineering. The sensors of 2M engineering were a prototype. Unfortunately, after the experiments, it came to light that the sampling frequency of the 2M IMU system was unreliable and changed during the experiment. Therefore, the validation study could not be done, and the data of the 2M engineering were not used in further analysis. Hence, there are no results available on the validity

of the 2M IMU sensors. In the present study only test-retest reliability with the XSens system was investigated.

Clinical goal

When doing this kind of experiment, it is interesting to keep the focus on the main goal, and how the current experiment contributes to reaching this goal. The reason to study the test-retest reliability in healthy subjects is because there is not enough evidence for the reliability of sensor-based arm kinematic metrics [1]. There is also a lack of evidence for the validity and responsiveness, but that was not addressed in this research. An IMU system could be an interesting option for stroke rehabilitation because IMUs are cheap, compact, portable, and reusable measurement devices. Results showing an excellent validity and test-retest reliability of an IMU-based method could support the next step: application of the technology in clinical settings. To use technology like IMU systems in clinical settings, only good results are not high enough. The test-retest reliability should be excellent in order to take steps toward clinical implementation. Fortunately, most of the metrics give these excellent results for reach to grasp movement. For the functional task the metrics did not have these excellent results meaning that there first needs to be some improvements made, which increase the relative and absolute reliability.

Recommendation

A recommendation would be, to improve the methods that take the begin- and endpoints out of the data. As described, this has impacted different metrics. One improvement would be, to get rid of the repetitions errors. This is because there is looked at a set of the movements which is the reach to grasp movement and the movement back. If large noise is seen as one of those movements, it could give repetitions errors. For future research it may be beneficial to include a large number of participants, to create a smaller confidence interval. In response to this research, some other research questions would be interesting to investigate. Firstly, as described at trunk displacement it would be interesting to know how the results are for movements where the trunk should be displaced a known distance. This could give more information on whether this could be used. Secondly, the functional task had good but no excellent results. Some specific research, with improvement specifically on this task, could make an improvement for this task. Thirdly, for the metrics that have excellent results a next step would be to test the test-retest reliability by CVA patient. Fourthly, comparing the results of the metrics between CVA patients and healthy subjects to see if can be discriminate between those metrics. Lastly, to use IMUs in a clinical setting not only the test-

retest reliability should be good but also the other clinimetric properties should have excellent results in order to make a next step towards implementation of IMUs for stroke rehabilitation. To investigate the other clinimetric parameters this IMU system with algorithms used in this research should be tested with a reference system.

Conclusion

This study evaluates the test-retest reliability of an IMU sensor-based method for measuring quality metrics, during the reach to grasp movement in healthy adults. This research indicates good to excellent relative test-retest reliability for reach to grasp movements for some of the metrics namely: movement time, smoothness, velocity profiles, mean velocity and maximum velocity. The metrics trunk displacement and time to max velocity had poor absolute reliability.

References

- [1] G. Kwakkel, E. Van Wegen, J. H. Burrige, C. J. Winstein, L. van Dokkum, M. Alt Murphy, M. F. Levin, and J. W. Krakauer, "Standardized measurement of quality of upper limb movement after stroke: Consensus-based core recommendations from the Second Stroke Recovery and Rehabilitation Roundtable.," *International journal of stroke : official journal of the International Stroke Society*, vol. 14, pp. 783–791, oct 2019.
- [2] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan, "Estimation of imu and marg orientation using a gradient descent algorithm," pp. 1–7, 2011.
- [3] A. Di Carlo, "Human and economic burden of stroke," *Age and Ageing*, vol. 38, no. 1, pp. 4–5, 2009.
- [4] J. Held, *Rehabilitation of stroke patients with sensor-based systems*. PhD thesis, University of Twente, Netherlands, 2 2019.
- [5] M. Hoonhorst, R. Nijland, J. Berg, C. Emmelot, B. Kollen, and G. Kwakkel, "How do fugl-meyer arm motor scores relate to dexterity according to the action research arm test at 6 months poststroke?," *Archives of physical medicine and rehabilitation*, vol. 96, 07 2015.
- [6] M. Michielsen, R. Selles, H. Stam, G. Ribbers, and J. Bussmann, "Quantifying nonuse in chronic stroke patients: A study into paretic, nonparetic, and bimanual upper-limb use in daily life," *Archives of physical medicine and rehabilitation*, vol. 93, 03 2012.
- [7] E. Repnik, U. Puh, N. Goljar, M. Munih, and M. Mihelj, "Using inertial measurement units and electromyography to quantify movement during action research arm test execution," *Sensors (Switzerland)*, vol. 18, no. 9, 2018.

- [8] D. W. F. Etienne Burdet and T. E. Milner, *HumanRobotics*. 55 Hayward Street, Cambridge: MIT Press book, 2003.
- [9] J. Kleim and T. Jones, "Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage.," *Journal of speech, language, and hearing research : JSLHR*, vol. 51 1, pp. S225–39, 2008.
- [10] P. Langhorne, J. Bernhardt, and G. Kwakkel, "Stroke rehabilitation," *The Lancet*, vol. 377, no. 9778, pp. 1693–1702, 2011.
- [11] J. J. Daly, J. P. McCabe, J. Holcomb, M. Monkiewicz, J. Gansen, and S. Pundik, "Long-Dose Intensive Therapy Is Necessary for Strong, Clinically Significant, Upper Limb Functional Gains and Retained Gains in Severe/Moderate Chronic Stroke," *Neurorehabilitation and Neural Repair*, vol. 33, no. 7, pp. 523–537, 2019.
- [12] E. Taub, G. Uswatte, V. Mark, and D. Morris, "The learned nonuse phenomenon: Implications for rehabilitation," *Europa medicophysica*, vol. 42, pp. 241–56, 10 2006.
- [13] S. H. Chae, Y. Kim, K.-S. Lee, and H.-S. Park, "Development and clinical evaluation of a web-based upper limb home rehabilitation system using a smartwatch and machine learning model for chronic stroke survivors: Prospective comparative study," *JMIR Mhealth Uhealth*, vol. 8, p. e17216, Jul 2020.
- [14] H. Costin, S.-I. Bejinariu, B. Dimitriu, and B. Anton-Prisacariu, "Assessment of rehabilitation degree in case of stroke by using image analysis of range of motion," in *EPE 2014 - Proceedings of the 2014 International Conference and Exposition on Electrical and Power Engineering*, pp. 557–560, 2014.
- [15] A. RajKumar, F. Vulpi, S. R. Bethi, H. K. Wazir, P. Raghavan, and V. Kapila, "Wearable inertial sensors for range of motion assessment," *IEEE Sensors Journal*, vol. 20, no. 7, pp. 3777–3787, 2020.
- [16] T. Ungi, A. Lasso, and G. Fichtinger, "Tracked ultrasound in navigated spine interventions," *Lecture Notes in Computational Vision and Biomechanics*, vol. 18, pp. 469–494, 01 2015.
- [17] d. i. M. v. D. prof. dr. LM. Bouter and prof.dr.ir G.A. Zielhuis, *Epidemiologisch onderzoek Opzet en interpretatie*. Vohn stafleu van loghum, 5,herziende druk ed., 2005.
- [18] M. I. M. Refai, B. L. S. Mique Saes, J. van Kordelaar, C. G. M. Meskers, G. K. Erwin E. H. van Wegen, and B.-J. F. van Beijnum, "Smoothness metrics for reaching performance after stroke 1. which one to choose?," *Submitted*.
- [19] S. Balasubramanian, A. Melendez-Calderon, A. Roby-Brami, and E. Burdet, "On the analysis of movement smoothness," *Journal of NeuroEngineering and Rehabilitation*, vol. 12, no. 1, p. 112, 2015.
- [20] K. C. Collins, N. C. Kennedy, A. Clark, and V. M. Pomeroy, "Kinematic Components of the Reach-to-Target Movement After Stroke for Focused Rehabilitation Interventions: Systematic Review and Meta-Analysis.," *Frontiers in neurology*, vol. 9, p. 472, 2018.
- [21] M. C. Cirstea and M. F. Levin, "Compensatory strategies for reaching in stroke," *Brain*, vol. 123, pp. 940–953, 05 2000.
- [22] G. R. H. Regterschot, W. Zhang, H. Baldus, M. Stevens, and W. Zijlstra, "Test-retest reliability of sensor-based sit-to-stand measures in young and older adults.," *Gait & posture*, vol. 40, no. 1, pp. 220–224, 2014.
- [23] W. J. P. J. J.P., "Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM," *Journal of Strength and Conditioning Research*, vol. 19, no. 1, pp. 231–240, 2005.
- [24] T. S. Patterson, M. D. Bishop, T. E. McGuirk, A. Sethi, and L. G. Richards, "Reliability of Upper Extremity Kinematics While Performing Different Tasks in Individuals With Stroke," *Journal of Motor Behavior*, vol. 43, no. 2, pp. 121–130, 2011.
- [25] a multidisplinair team with between people from UTwente TU Delft Erasmus MC and A. UMC, "focus group armcoach4stroke project,"
- [26] R. Colombo, I. Cusmano, I. Sterpi, A. Mazzone, C. Delconte, and F. Pisano *IEEE Transactions on Neural Systems and Rehabilitation Engineering*.
- [27] A. Bruton, J. H. Conway, and S. T. Holgate, "Reliability: What is it, and how is it measured?," *Physiotherapy*, vol. 86, no. 2, pp. 94–99, 2000.

Appendix A

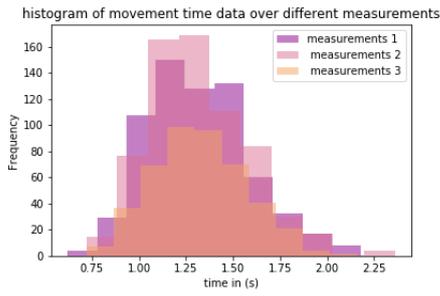


Figure 23: Histogram of movement time data over different measurements

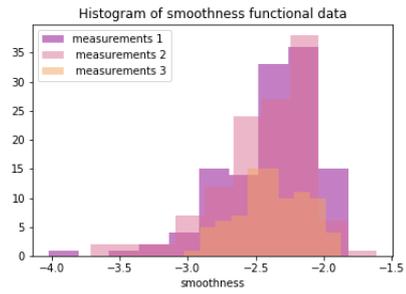


Figure 26: Histogram of smoothness functional time different measurements

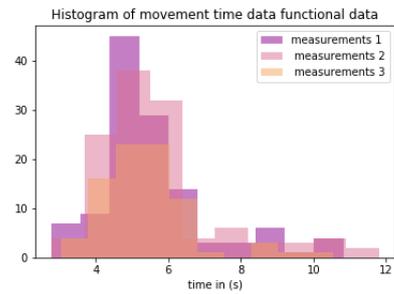


Figure 24: Histogram of movement time functional time different measurements

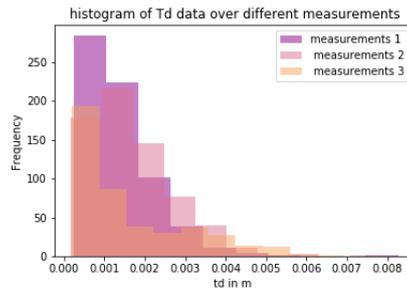


Figure 27: Histogram of trunk displacement data over different measurements

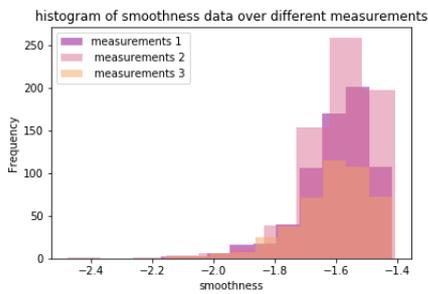


Figure 25: Histogram of smoothness data over different measurements

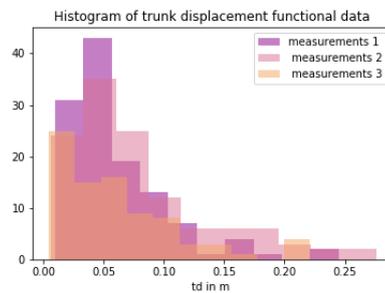


Figure 28: Histogram of trunk displacement functional time different measurements

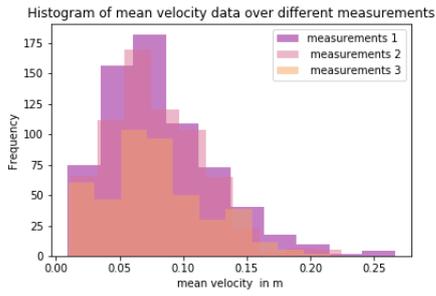


Figure 29: Histogram of mean velocity data over different measurements

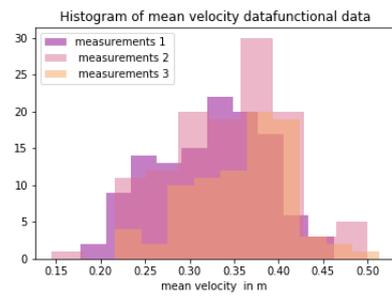


Figure 30: Histogram of mean velocity functional time different measurements

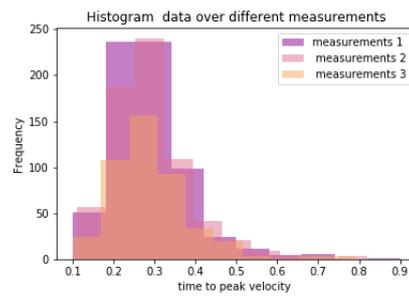


Figure 33: Histogram of time to peak velocity data over different measurements

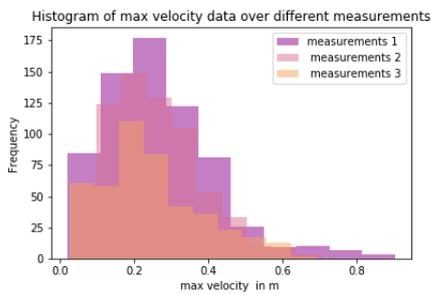


Figure 31: Histogram of max velocity data over different measurements

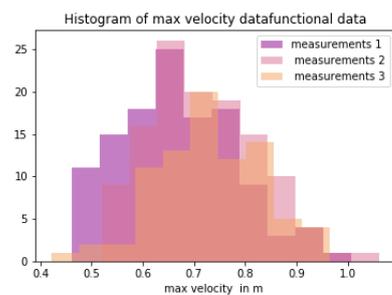


Figure 32: Histogram of max velocity functional time different measurements

Appendix B

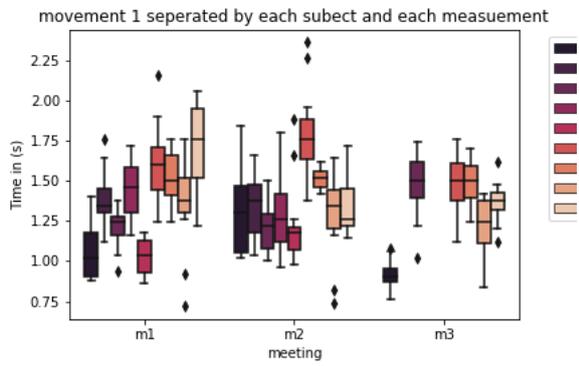


Figure 34: Movement one, Data seperated for each subject and each measurement for metric movement time

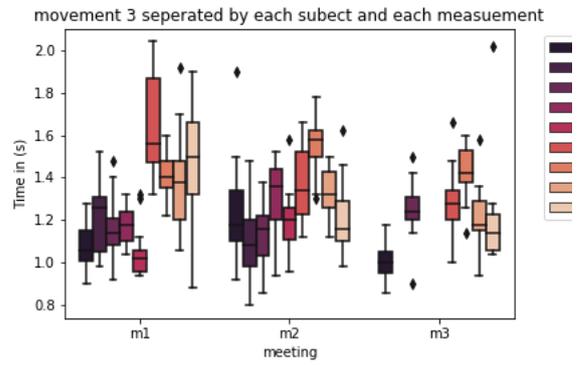


Figure 36: Movement three, Data seperated for each subject and each measurement for metric movement time

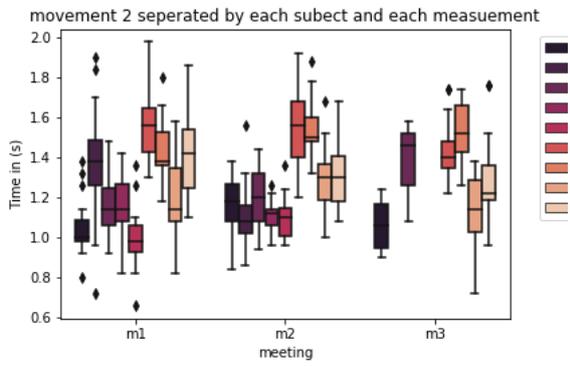


Figure 35: Movement two, Data seperated for each subject and each measurement for metric movement time

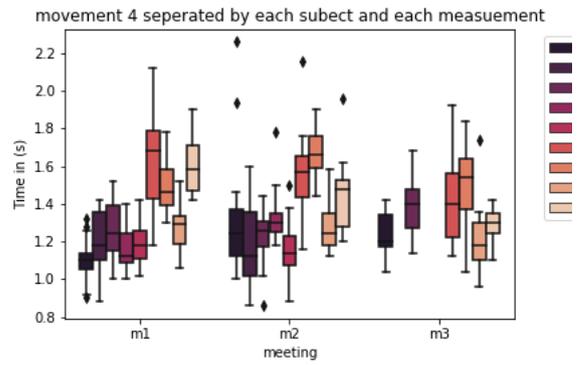


Figure 37: Movement four, Data seperated for each subject and each measurement for metric movement time

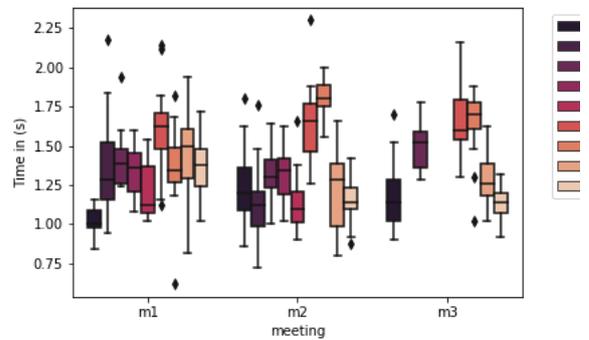


Figure 38: Movement five, Data seperated for each subject and each measurement for metric movement time

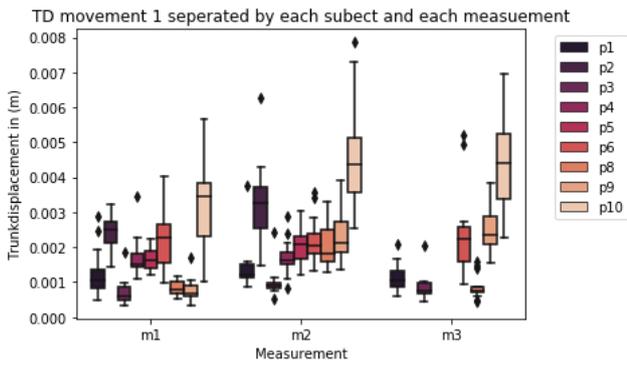


Figure 39: Movement one, Data separated for each subject and each measurement for metric trunk displacement

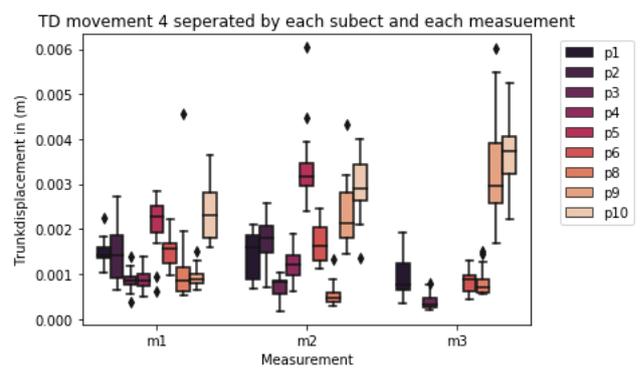


Figure 42: Movement four, Data separated for each subject and each measurement for metric trunk displacement

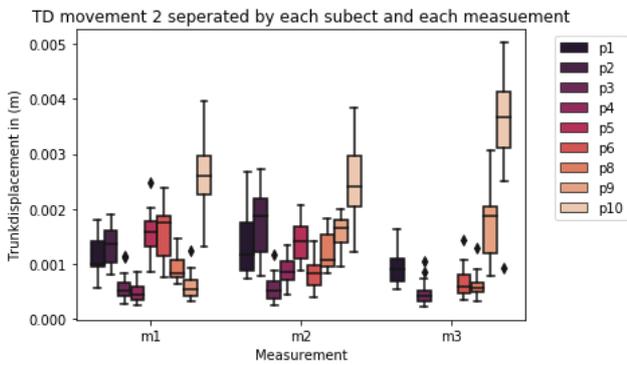


Figure 40: Movement two, Data separated for each subject and each measurement for metric trunk displacement

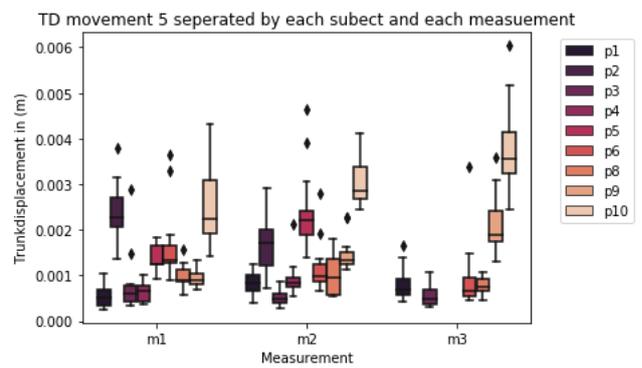


Figure 43: Movement five, Data separated for each subject and each measurement for metric trunk displacement

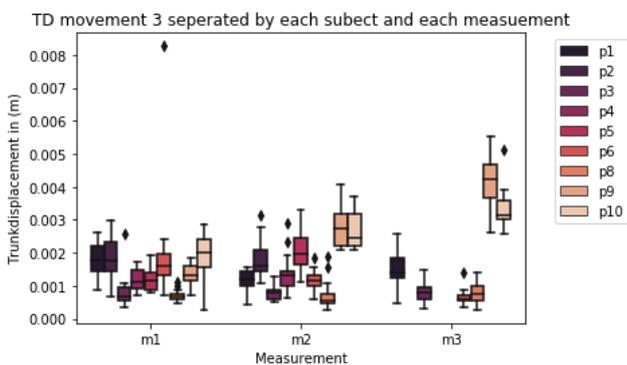


Figure 41: Movement three, Data separated for each subject and each measurement for metric trunk displacement

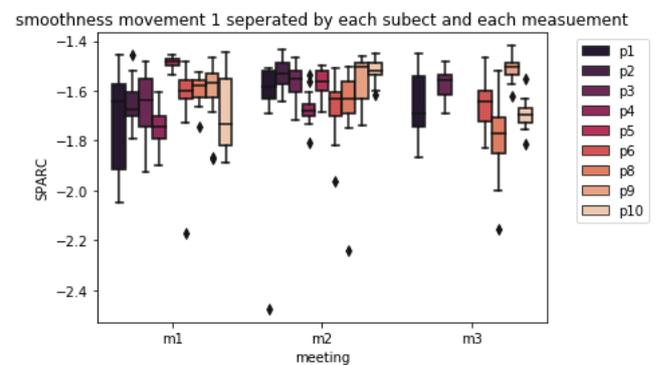


Figure 44: Movement one, Data separated for each subject and each measurement for metric smoothness

smoothness movement 2 separated by each subect and each measurement

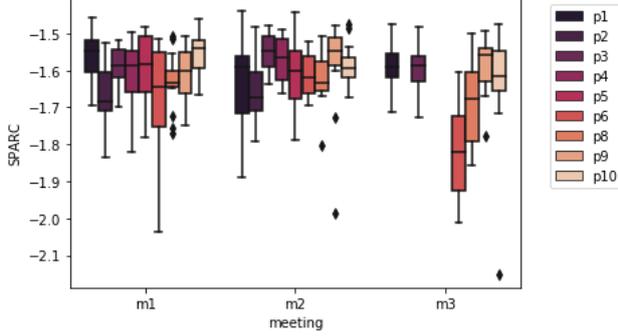


Figure 45: Movement two, Data separated for each subject and each measurement for metric smoothness

smoothness movement 5 separated by each subect and each measurement

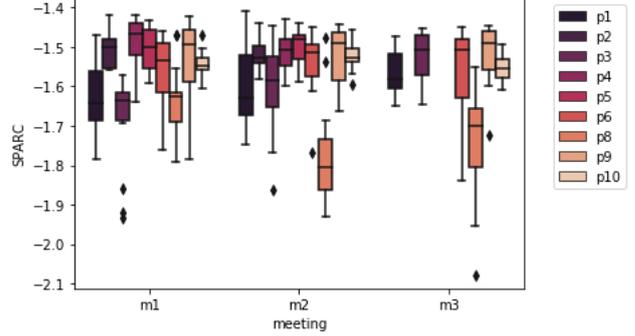


Figure 48: Movement five, Data separated for each subject and each measurement for metric smoothness

smoothness movement 3 separated by each subect and each measurement

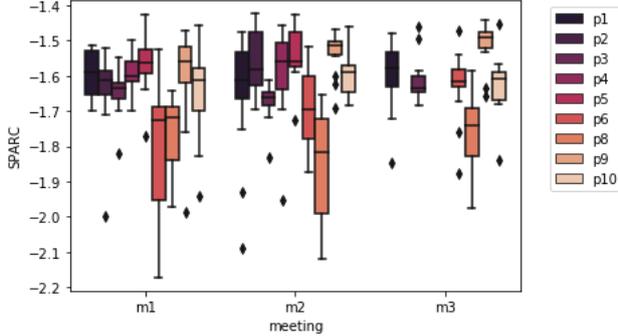


Figure 46: Movement three, Data separated for each subject and each measurement for metric smoothness

Mean velocity movement 1 separated by each subect and each measurement

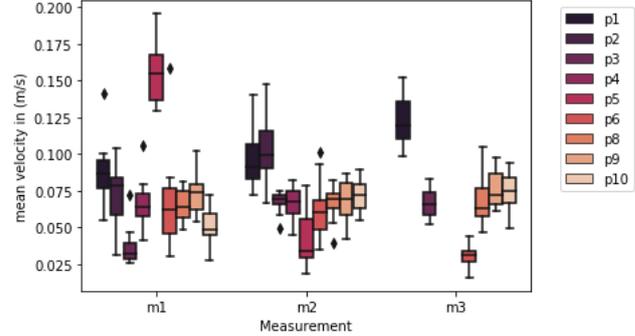


Figure 49: Movement one, Data separated for each subject and each measurement for metric mean velocity

smoothness movement 4 separated by each subect and each measurement

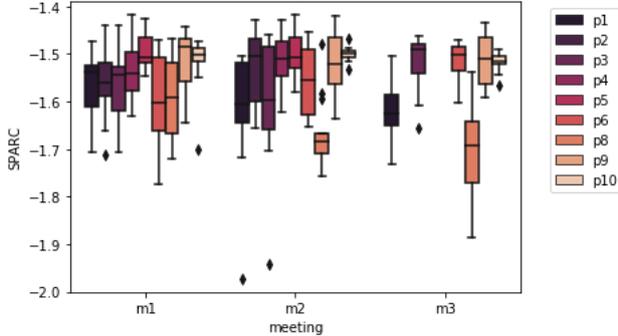


Figure 47: Movement four, Data separated for each subject and each measurement for metric smoothness

Mean velocity movement 2 separated by each subect and each measurement

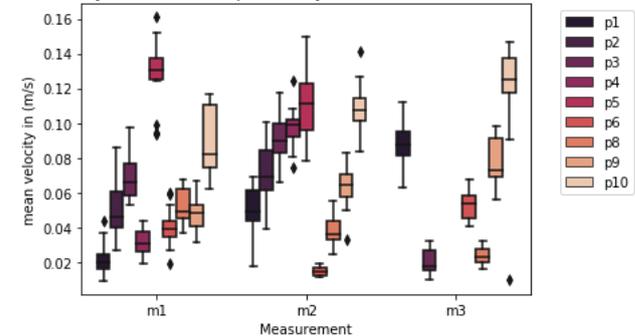


Figure 50: Movement two, Data separated for each subject and each measurement for metric mean velocity

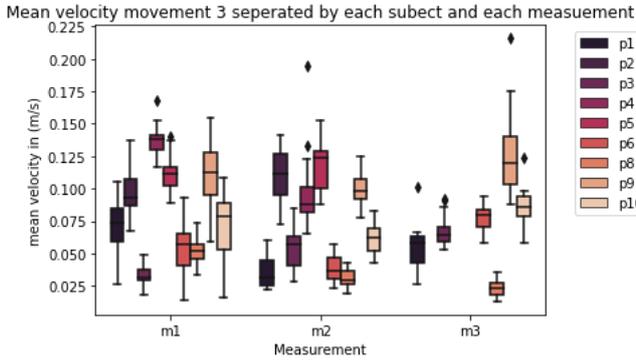


Figure 51: Movement three, Data separated for each subject and each measurement for metric mean velocity

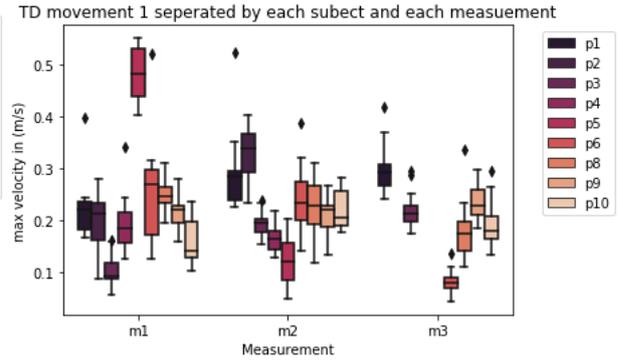


Figure 54: Movement one, Data separated for each subject and each measurement for metric max velocity

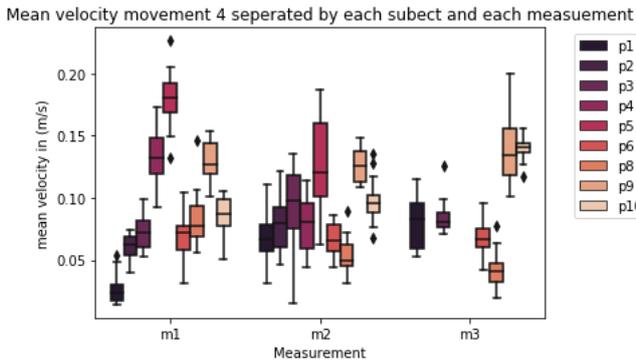


Figure 52: Movement four, Data separated for each subject and each measurement for metric mean velocity

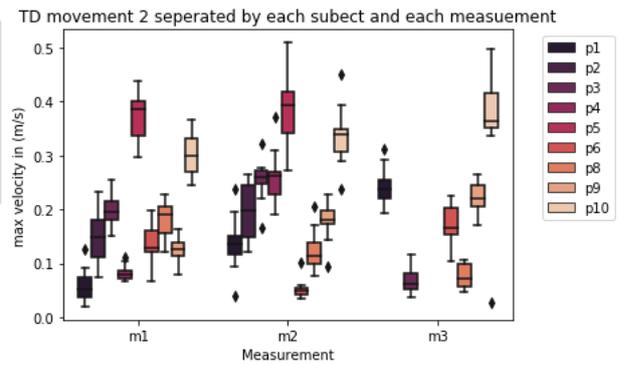


Figure 55: Movement two, Data separated for each subject and each measurement for metric max velocity

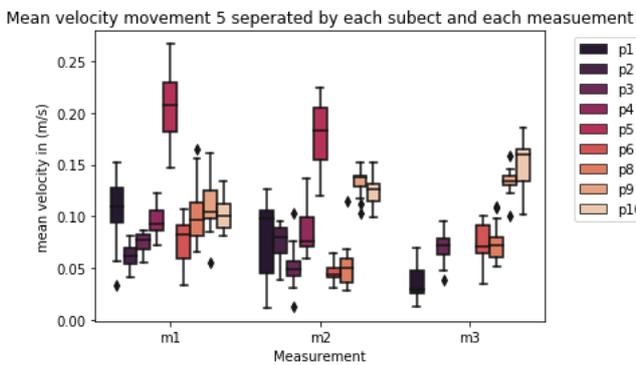


Figure 53: Movement five, Data separated for each subject and each measurement for metric mean velocity

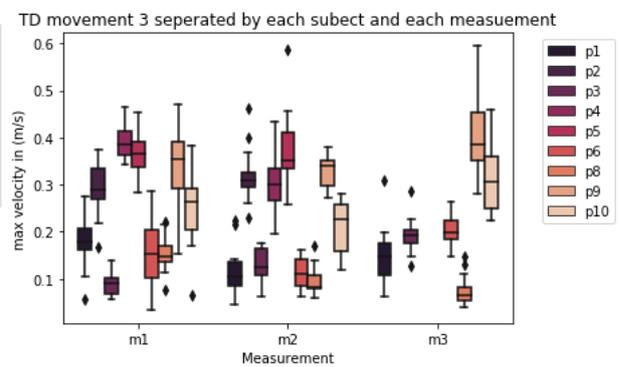


Figure 56: Movement three, Data separated for each subject and each measurement for metric max velocity

maximum velocity movement 4 separated by each subject and each measurement

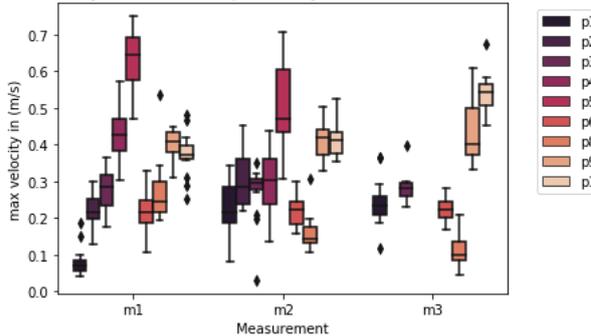


Figure 57: Movement four, Data separated for each subject and each measurement for metric max velocity

MTime to peak velocity movement 2 separated by each subject and each measurement

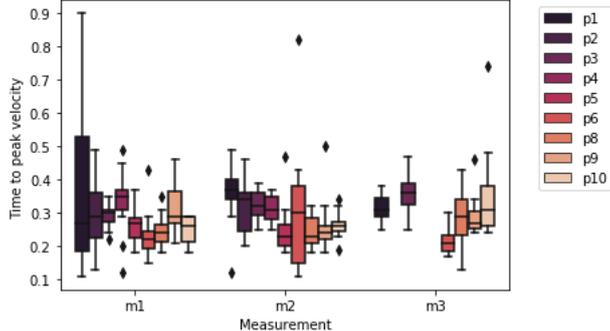


Figure 60: Movement two, Data separated for each subject and each measurement for metric time to max velocity

TD movement 5 separated by each subject and each measurement

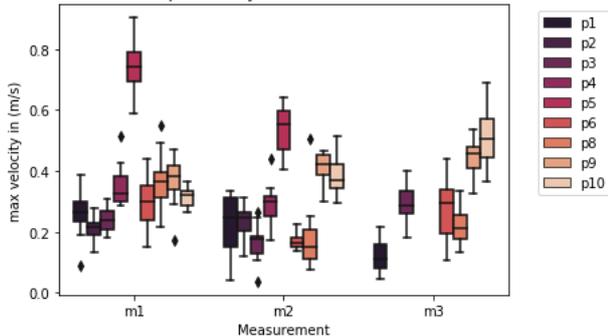


Figure 58: Movement five, Data separated for each subject and each measurement for metric max velocity

Time to peak velocity movement 3 separated by each subject and each measurement

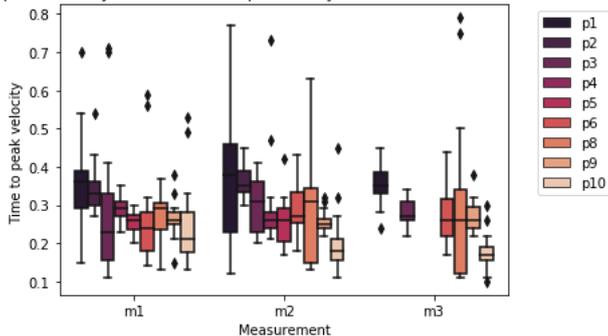


Figure 61: Movement three, Data separated for each subject and each measurement for metric time to max velocity

MTime to peak velocity movement 1 separated by each subject and each measurement

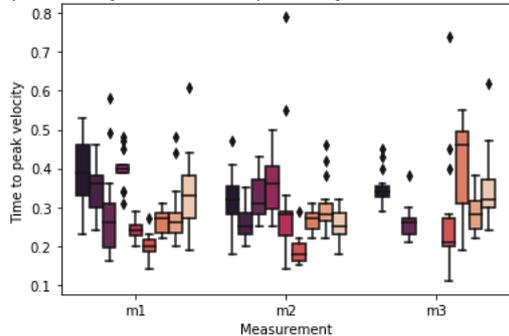


Figure 59: Movement one, Data separated for each subject and each measurement for metric time to max velocity

Time to peak velocity movement 4 separated by each subject and each measurement

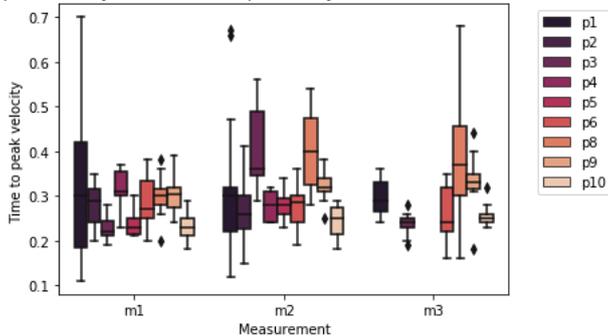


Figure 62: Movement four, Data separated for each subject and each measurement for metric time to max velocity

Time to peak velocity movement 5 separated by each subject and each measurement

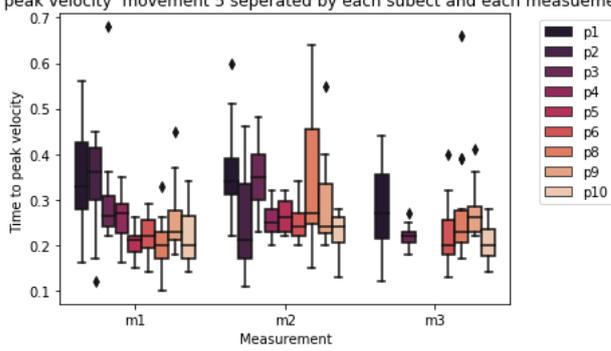


Figure 63: Movement five, Data separated for each subject and each measurement for metric time to max velocity

Appendix C

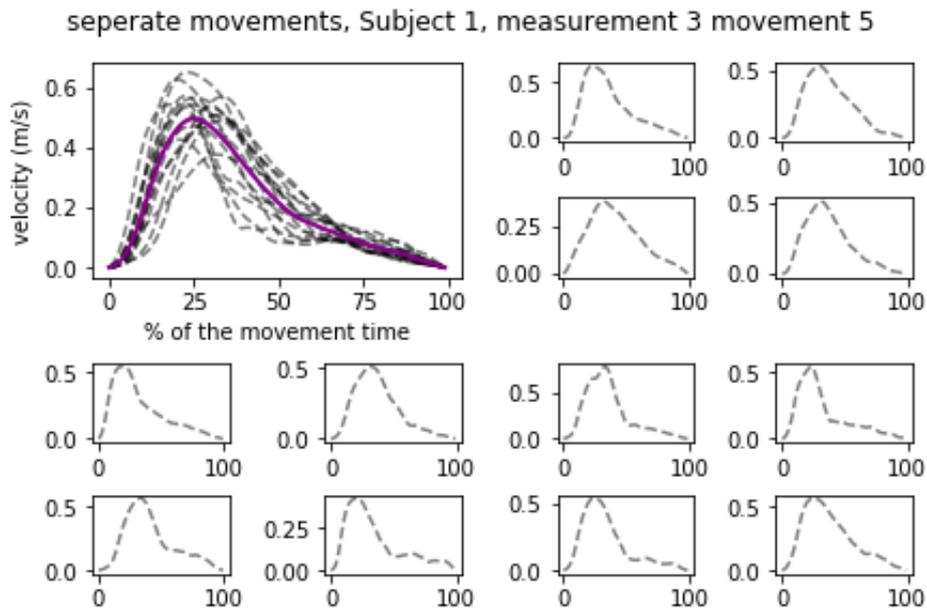


Figure 64: Separate Trails of movement 5 and their mean visualised.

separate movements, Subject 1, measurement 1 movement 6

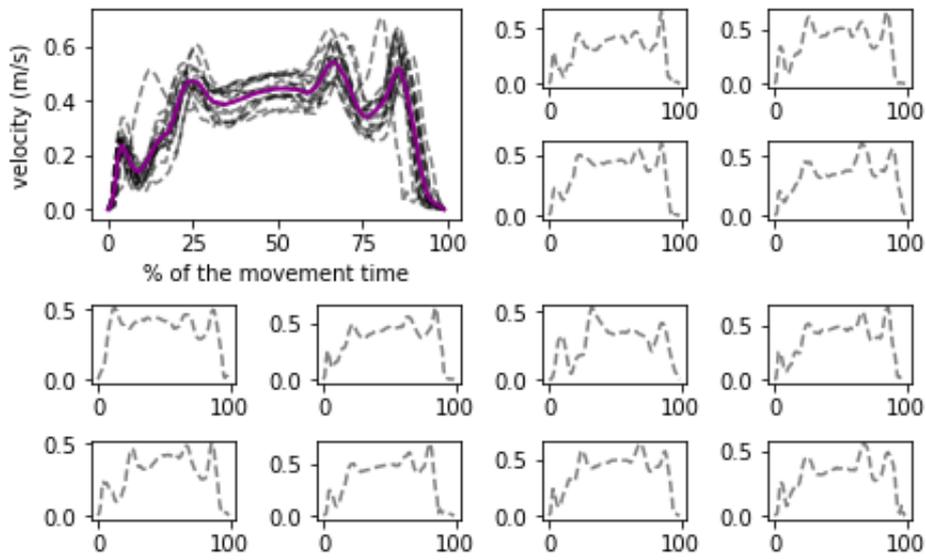


Figure 65: Separate Trails of functional task and their mean visualised.

	P1M	P1M	P1M	P1M	P2M	P2M	P3M	P3M	P3M	P4M	P4M	P5M	P5M	P6M	P6M	P8M	P8M	P9M	P9M	P10M	P10M	P10M	P10M		
	1	2	3	1	1	2	1	2	3	1	2	1	2	1	2	1	2	1	2	1	2	3			
P1M1	0.98	0.99	0.95	0.99	0.95	0.99	0.98	0.96	0.95	0.98	0.98	0.82	0.83	0.75	0.74	0.49	0.95	0.86	0.83	0.95	0.85	0.94	0.83	0.88	0.98
P1M2	0.98	1	0.99	0.93	0.98	0.98	0.98	0.99	0.95	0.98	0.97	0.89	0.80	0.76	0.72	0.51	0.94	0.86	0.84	0.95	0.82	0.93	0.78	0.85	0.99
P1M3	0.99	0.99	1	0.95	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.83	0.83	0.73	0.73	0.47	0.95	0.86	0.82	0.94	0.84	0.94	0.83	0.89	0.98
P2M1	0.95	0.93	0.95	1	0.97	0.94	0.97	0.97	0.97	0.97	0.97	0.98	0.95	0.85	0.87	0.62	0.98	0.95	0.92	0.97	0.95	0.97	0.95	0.96	0.95
P2M2	0.99	0.98	0.99	0.97	1	0.97	0.99	0.95	0.98	0.98	0.98	0.93	0.85	0.78	0.76	0.51	0.95	0.88	0.86	0.95	0.86	0.94	0.85	0.89	0.98
P3M1	0.98	0.98	0.99	0.94	0.97	1	0.99	0.97	0.99	0.99	0.99	0.94	0.86	0.74	0.76	0.53	0.97	0.88	0.84	0.96	0.87	0.96	0.84	0.92	0.98
P3M2	0.96	0.98	0.99	0.97	0.99	0.99	1	0.98	0.99	0.99	0.99	0.96	0.90	0.80	0.82	0.58	0.98	0.92	0.88	0.98	0.91	0.98	0.89	0.93	0.99
P3M3	0.95	0.95	0.95	0.97	0.95	0.97	0.95	0.97	0.98	0.97	0.98	0.92	0.86	0.88	0.88	0.68	0.99	0.96	0.94	0.99	0.95	0.99	0.91	0.94	0.98
P4M1	0.98	0.98	0.98	0.97	0.98	0.99	0.99	0.99	0.97	1	0.99	0.96	0.89	0.82	0.82	0.62	0.99	0.93	0.90	0.99	0.91	0.98	0.88	0.93	0.99
P4M2	0.98	0.97	0.99	0.97	0.98	0.99	0.99	0.99	0.97	0.99	1	0.96	0.90	0.75	0.79	0.53	0.97	0.89	0.84	0.96	0.90	0.97	0.89	0.94	0.97
P5M1	0.92	0.89	0.93	0.98	0.93	0.98	0.93	0.94	0.96	0.96	0.96	1	0.98	0.81	0.89	0.64	0.98	0.95	0.90	0.96	0.97	0.98	0.97	0.99	0.93
P5M2	0.83	0.80	0.83	0.95	0.85	0.86	0.90	0.92	0.89	0.90	0.90	0.98	1	0.84	0.94	0.72	0.94	0.95	0.90	0.91	0.99	0.94	0.99	0.85	0.85
P6M1	0.75	0.76	0.73	0.85	0.78	0.74	0.80	0.86	0.82	0.75	0.81	0.84	1	0.94	0.91	0.87	0.94	0.95	0.97	0.88	0.89	0.85	0.82	0.79	0.81
P6M2	0.74	0.72	0.73	0.87	0.76	0.76	0.76	0.82	0.88	0.82	0.79	0.89	0.94	0.94	1	0.90	0.90	0.97	0.94	0.88	0.96	0.89	0.92	0.89	0.80
P6M3	0.49	0.51	0.47	0.62	0.51	0.53	0.58	0.68	0.62	0.53	0.64	0.72	0.91	0.90	1	0.70	0.83	0.84	0.70	0.79	0.69	0.67	0.64	0.59	0.80
P8M1	0.95	0.94	0.95	0.98	0.95	0.98	0.95	0.97	0.98	0.99	0.99	0.97	0.98	0.94	0.90	0.70	1	0.97	0.93	0.99	0.96	0.99	0.92	0.96	0.97
P8M2	0.86	0.86	0.86	0.95	0.88	0.88	0.92	0.96	0.93	0.89	0.95	0.95	0.95	0.97	0.83	0.97	1	0.98	0.96	0.98	0.96	0.94	0.94	0.90	0.90
P8M3	0.83	0.84	0.84	0.92	0.86	0.84	0.88	0.94	0.90	0.84	0.90	0.90	0.97	0.94	0.84	0.93	0.98	1	0.95	0.95	0.93	0.89	0.89	0.88	0.89
P9M1	0.95	0.95	0.94	0.97	0.95	0.96	0.98	0.99	0.99	0.99	0.99	0.96	0.91	0.88	0.88	0.70	0.99	0.96	0.95	1	0.94	0.99	0.90	0.93	0.98
P9M2	0.85	0.82	0.84	0.95	0.86	0.87	0.91	0.95	0.91	0.90	0.90	0.87	0.99	0.89	0.96	0.79	0.96	0.98	0.95	0.94	1	0.96	0.98	0.97	0.88
P9M3	0.94	0.93	0.94	0.97	0.94	0.97	0.98	0.98	0.99	0.98	0.97	0.98	0.94	0.85	0.89	0.69	0.99	0.96	0.93	0.99	0.96	1	0.92	0.96	0.97
P10M1	0.83	0.78	0.83	0.95	0.85	0.84	0.89	0.91	0.88	0.89	0.89	0.97	0.99	0.82	0.92	0.67	0.92	0.94	0.89	0.90	0.98	0.92	1	0.98	0.84
P10M2	0.88	0.85	0.89	0.96	0.89	0.92	0.93	0.94	0.93	0.94	0.93	0.99	0.99	0.79	0.89	0.64	0.96	0.94	0.88	0.93	0.97	0.96	0.98	1	0.89
P10M3	0.98	0.99	0.98	0.95	0.98	0.98	0.99	0.98	0.99	0.97	0.97	0.93	0.85	0.81	0.80	0.59	0.97	0.90	0.89	0.98	0.88	0.97	0.84	0.89	1

Figure 66: Correlation coefficients movement 1 calculated for each separate measurement

Appendix D

Samenvatting Nederlands

ACHTERGROND Internal Measurement Units (IMUs) zijn compacte, herbruikbare en goedkope optie voor thuis revalidatie voor Cerebro Vasculair Accident (CVA) -patiënten. Helaas is er niet genoeg bewijs voor hun validiteit (Geldigheid), responsiviteit, en betrouwbaarheid om het IMU-systeem als losstaande optie te gebruiken. Kwakkel 2019 In dit artikel wordt de test-hertest betrouwbaarheid van een IMU onderzocht voor het meten van verschillende kwaliteitsmetrieken tijdens de reikbeweging bij gezonde volwassenen.

VRAAGSTELLING “Wat is de test-hertestbetrouwbaarheid van een IMU-sensorgebaseerde methode voor het meten van kwaliteitsmetrieken tijdens de reikbeweging bij gezonde volwassenen?”

METHODE Tien gezonde volwassenen, waarvan zes vrouwen, namen deel aan dit experiment. Ze voerden vijf 2D reikbewegingen en één functionele drinktaak uit, elk vijftien keer, met vijf IMU's op de arm. Ze herhaalden deze taken drie keer met tussendoor aan- en uittrekken van het sensorsysteem. Om de oriëntatie naar het “earth frame” te krijgen, is het Madwick filter gebruikt [2]. Het begin van een beweging wordt gedefinieerd als de snelheid $\geq 0,05$ rad/s is en het eindpunt is gedefinieerd als de snelheid $\leq 0,05$ rad/s. Voor elke metriek worden de Kruskal Wallis-test, de intraklas-correlatiecoëfficiënt (ICC) en de standaardfoutmeting (SEM) berekend.

RESULTATEN De relatieve betrouwbaarheid van de metriek bewegingstijd was voor de reikbeweging goed tot uitstekend (bereik ICC: 0,73-0,92) en goed voor de functionele taak (ICC 0,53). De absolute betrouwbaarheid was voor de reikbeweging uitstekend (bereik SEM: 5,4-9,2%), maar voor de functionele taak slecht (SEM: 19,6%). De relatieve betrouwbaarheid van de metriek vloeiendheid van de beweging (smoothness) was goed tot uitstekend voor de reikbeweging (bereik ICC: 0,55-0,93), en goed voor de functionele taak (ICC: 0,65). De absolute betrouwbaarheid was voor zowel voor de reikbeweging (bereik SEM: 2,2-2,6 %) als de functionele taak (SEM: 8,3 %) uitstekend. De relatieve betrouwbaarheid van de metriek rompverplaatsing was voor de reikbeweging goed tot uitstekend (bereik ICC: 0,62-0,97), en goed voor de functionele taak (ICC: 0,6). De absolute betrouwbaarheid was voor zowel de reikbeweging (bereik SEM: 13-40,9%) als voor de functionele taak (SEM: 47,4%) slecht. De resultaten voor snelheidsprofielen hadden een hoge correlatiecoëfficiënt (bereik r : 0,95-0,99). De relatieve betrouwbaarheid van de metriek gemiddelde snelheid was voor de reikbeweging goed tot uitstekend (bereik ICC: 0,63-0,9) en voor de functionele taak slecht (ICC 0,29). De absolute betrouwbaarheid was voor de reikbeweging uitstekend tot slecht (bereik SEM: 6,4-12,6%) en voor de functionele taak uitstekend (SEM 10 %). De relatieve betrouwbaarheid van de metriek maximale snelheid voor de reikbeweging was goed tot uitstekend (bereik ICC: 0,52-0,93), en voor de functionele taak uitstekend (ICC: 0,96). De absolute betrouwbaarheid was voor de reikbeweging uitstekend tot slecht (bereik SEM: 6,6-17,2 %) en voor de functionele taak uitstekend (SEM 3,1 %) De relatieve betrouwbaarheid van de metriek tijd tot maximale snelheid voor de reikbeweging was goed tot uitstekend (bereik ICC: 0,57-0,87). De absolute betrouwbaarheid was voor de reikbeweging uitstekend tot slecht (bereik SEM: 6,1-19 %).

CONCLUSIE Deze studie evalueert de test-hertest betrouwbaarheid van een IMU-design met bijbehorende metrieken van de bovenste extremititeit (UE) bij gezonde proefpersonen. Dit onderzoek wijst op een goede tot uitstekende test-hertestbetrouwbaarheid voor reikbewegingen door de metrieken bewegingstijd, vloeiendheid van de beweging, gemiddelde snelheid en maximale snelheid.