Evaluating Upper-Extremity (Dys)function Using Inertial Measurement Unit Technology and its Applications to Resource-Constrained Settings

Alisa Aguirre
Health Equity Institute
San Francisco State University
San Francisco, USA
aaguirr1@mail.sfsu.edu

Abstract—The rural population of the world has grown slowly since 1950 and is expected to reach 3.1 billion by 2050. Individuals who reside in rural areas experience more functional limitations caused by chronic conditions (e.g., stroke, cerebral palsy) and have reduced access to healthcare services that do urban residents with similar conditions. Applications of robotics along with information and communications technology (ICT) have been proposed as a cost-effective solution to increase access and quality of care in rural areas. In this paper, I introduce the use of an inertial measurement unit (IMU) for the assessment of upper-extremity impairments, and present data demonstrating the reliability and validity of this device to capture the kinematics during activities of daily living (ADL) in 50 physically and neurologically healthy adults. This innovation solution integrates multiple technologies to provide rehabilitation services without in-person clinician-to-patient encounters, and could support care services for individuals who reside in geographically remote areas.

Keywords—low-income; rural; stroke; upper extremity function; inertial sensors; motor assessment

I. INTRODUCTION

A. Stroke

Every year, there are over 17 million first-time stroke incidences worldwide [1], with approximately 795,000 people in the United States experiencing a new or recurrent stroke each year [2]. Of the nearly 4 million people who have survived a stroke incidence, approximately 65% will suffer from permanent sensorimotor deficits (i.e., hemiparesis, spasticity, impaired motor coordination, sensory impairments) [3]. Sensorimotor impairments have significantly negative impact on the ability to perform activities of daily living (ADL, e.g., reaching and grasping, object manipulation) which can affect their level of independence and quality of life.

Thorough, consistent, and accurate assessment of motor function is an important component of the rehabilitation process. When assessing upper extremity function, clinicians are generally limited to the use of measurements that can be made during an office visit (e.g., physical exams, range of motion tests, self-report assessments). Conventional post-stroke motor assessments use an observational methodology in which the patient performs an ADL and performance is scored on an ordinal scale. For example, the Action Research Arm Test (ARAT) scores performance on various ADL using a 4-point scale, where 0 indicates no movement, 1 indicates task is partially performed, 2 indicates task is completed by takes abnormally long, and 3 indicates task is completed with normal performance (Lyle, 1981). Although the ARAT, and other post-stroke motor assessments are widely accepted and have high test-retest and interrater reliability, they rely on ordinal scoring and thus are insensitive to subtle differences of deficit and changes over the rehabilitation lifespan.

B. Healthcare Disparities for Rural Areas

Quality stroke assessment and rehabilitation is influenced by geographical location [4-6] and race/ethnicity [7-9]. In rural areas, there is limited technology at local hospitals, fewer stroke specialists and outpatient rehabilitation facilities, and clinicians often do not adhere to standard clinical guidelines for stroke management [10,11]. As such, individuals who reside in rural areas often have worse stroke outcomes than those who live in urban areas [4-6], unless they are able to travel to metropolitan area for specialized treatment. In 2014, there were just over 46 million people residing in rural areas (approximately 15% of U.S. residents) [12]. In addition, there exist racial/ethnic differences in stroke type, incidence, severity, and mortality in the U.S. population [10-12]. Thus, the large number of these residents who are people of color (37% African American, 33% Native, and 27% Hispanic) [13], are less likely to report good overall health status, health care access and quality. Given the healthcare utilization barriers that these individuals face, it is likely that this population would greatly benefit from the development of a low-cost assessment device with tele-rehabilitation capabilities.

C. Stroke Rehabilitation Technology

Recent technological advances have motivated researchers to examine whether rehabilitation robots [14,15] and commercially available motion capture systems [16-18] can accurately measure post-stroke motor function. This corpus of work demonstrate that rehabilitation robots and motion capture system are capable of detailed kinematic analysis during upper-limb movement. While there are certainly advantages to
rehabilitation robotics and motion capture systems, their use in a clinical setting is hampered by their high cost, large space requirements, lengthy training and setup time, and technical knowledge requirements. Thus, while effective assessment tools, they are not a realistic option for many clinical settings.

An alternative option is to record movement kinematics during ADL using low-cost MEMS (microelectromechanical systems) accelerometers. Compared to commercial motion capture systems, accelerometers are relatively inexpensive, provide continuous analysis and have the added benefit of being portable. There has been a resurgence in their use due to technological advances, and accelerometer-based technologies (ABTs) are widely used as an activity monitoring technique in clinical and non-clinical rehabilitation settings [19]. In comparison, relatively few studies have examined whether accelerometers have the potential to accurately measure upper extremity movements [20-23]. Despite the small number of studies conducted, there is empirical evidence that accelerometers can accurately capture kinematic characteristics indicative of motor impairment. For example, Knorr and colleagues [21] had eight stroke patients perform tasks derived from two common motor function assessments while wearing accelerometers placed on the hand, forearm, and upper-arm. The authors reported that kinematic characteristics indicative of motor impairment were captured using the accelerometers, although kinematic data collected from the hand were more highly correlated with clinical scores than data collected from the forearm and upper-arm.

Taken together, these results indicate that accelerometers could be used to address the aforementioned issues in post-stroke rehabilitation in rural populations. However, prior studies have not examined standard kinematic variables used in human neuromotor control research to provide specific quantitative information about movement strategies and quality. Furthermore, to date very few studies have examined the accuracy (validity) and consistency (reliability) of the accelerometer compared to an optoelectric motion capture system (i.e., the gold standard). Establishing the accuracy (validity) and consistency (reliability) of the proposed device is an important first step in the development of clinical accelerometer applications.

As such, the aim of the present paper is to evaluate the reliability and validity of an inertial measurement unit (IMU) for upper-extremity motor assessment. The IMU system yielded similar test-retest reliability values compared to the Vicon system, and could accurately measure movement time and mean acceleration variability. Taken together, results of the present study provide preliminary evidence that tri-axial IMU’s are reliable tools for investigating upper limb movement kinematics, and thus could be used to evaluate motor (dys)function in rural or socioeconomically disadvantaged areas.

II. METHODS

A. Participants

Fifty participants from the San Francisco State University campus and personal contact (mean age = 38.2, SD = 14.9, 23 men, 26 women, 1 did not specify) participated in the present study. Based on administration of the Revised Edinburgh Handedness Inventory [24] which ranks handedness in a battery of common tasks on a scale ranging from −1 (strongly left-handed) to 1 (strongly right-handed), 4 participants were left handed (mean = -53.75, SD = 12.88) and 46 were right handed (mean = 93.61, SD = 9.73). The study was approved by the San Francisco State University Institutional Review Board committee.

B. Apparatus

Kinematic data was recorded using an optical motion capture system (Bonita 10, VICON Motion Capture Systems, Oxford, U.K.), consisting of eight Bonita cameras that had a temporal and spatial resolution of 200 Hz and 1mm, respectively. Three retro reflective markers (14 mm diameter) were placed on an armband positioned on the thickest part of the forearm (Figure 1a).

A single inertial measurement unit (IMU), Memsense W2 IMU, Memsense, Rapid City, U.S.A.) was strapped immediately distal to the armband (Fig. 1a,b). The IMU consists of a 3D accelerometer (± 6 g range), a 3D gyroscope (± 500 °/s range) and a 3D magnetometer (± 10 Gauss range). IMU data was collected using the Memsense Inertial Insight program and transmitted to the host computer via Bluetooth™. Each trial was video recorded using a Bonita 720c HD digital video camera that was placed above the apparatus which provided a bird’s eye view of the participant. The digital video camera was used to identify any movement errors.

C. ADL Measures

Kinematic analysis was evaluated through performance of a total of four ADL tasks. The task list was comprised of three tasks from the Action Research Arm Test (ARAT) [25] and one task from the Frenchay Arm Test (FAT) [26].

1) ARAT 75: Taken from the Grasp subtest of the ARAT, participants were instructed to “grasp the 7.5cm block located on the tablet, lift it up, and place it on the top of the shelf.”

Fig. 1. Experimental setup depicting (a) device and marker placement on participant’s right arm, (b) Inertial Measurement Unit [IMU].
2) **Pour Water:** Taken from the Grip subtest of the ARAT, participants were instructed to “pour the water from this cup to the other empty cup.”

3) **Thin Tube:** Taken from the Grip subtest of the ARAT, participants were instructed to “grasp this tube and place it onto the peg.”

4) **Drink Glass:** Taken from the FAT, participants were instructed to “pick up this glass half full of water and drink some water and then put it back on the table.”

**D. Procedure**

After reading and filling out the written informed consent and handedness inventory forms, retroreflective markers were placed on the appropriate anatomical landmarks. Participants sat in a chair in front of a table placed 15cm away. Prior to each task, the participant placed the tested hand on the starting position in a pronated orientation and objects were placed on the table. Upon the verbal signal from the experimenter, the participant performed the specified action, and then returned the hand to the starting position. For all trials, participants were instructed to perform the movements at a comfortable speed (i.e., akin to performing similar activities in their home environment). Trials performed in a non-instructed manner (moving prior to verbal start command, performing the task incorrectly) were repeated immediately.

To be consistent with post-stroke clinical assessments, the order of tasks within each ADL motor assessment (i.e., ARAT, FAT) was kept constant. The order in which the assessments was performed, and the hand used to perform each task, was randomized and counterbalanced between participants. Participants performed each ADL task twice with each hand (dominant, non-dominant), yielding a total of 16 trials. The entire experiment lasted approximately 15 minutes.

**E. Data Processing**

The 3D coordinates of the retro reflective markers collected using the Vicon system were reconstructed and labeled for each individual trial in Vicon Nexus software (v2.2). Any missing data (less than 10 frames) were interpolated using a cubic spline, and exported as a csv file. The exported data was processed in Matlab using a custom program (The Mathworks, Version R2013a, Natick, MA). Data was filtered at a 5 Hz cut-off using a second order Butterworth filter. Movement velocity was calculated by differentiating 3D position data using a first order central difference technique. For each trial, only the time period between when the hand left the starting position (movement onset) to the time the hand returned to the starting position (movement offset) was further analyzed. Movement onset was determined as the instant when the velocity trace exceeded 1% of the first peak-acceleration. In addition, movement offset was determined as the moment when the acceleration trace dropped below 1% of the final peak-acceleration. At this point, acceleration parameters were calculated from the cropped time series.

**F. Parameters**

**Movement Time (MT)** is defined as the amount of time (ms) taken to complete the task (time period from when the hand left the start position to when it was placed back on the table). This parameter is considered a measure of general movement associated with most motor behaviors [15-18], as well as stroke severity and recovery [27,28].

**Mean Acceleration (MA)** reflects the intensity of the performed movement, and is defined as follows:

\[
MA = \frac{1}{T} \sum_{t=0}^{T} MA(t),
\]

with \(MA(t) = \sqrt{a_x(t)^2 + a_y(t)^2 + a_z(t)^2}\)

over the complete task, where \(a_x(t)\), \(a_y(t)\), \(a_z(t)\) refers to acceleration of the x-, y-, and z-axis (respectively) at time \(t\).

**Mean acceleration: variability (MAV)** indicates the spread of accelerations that were measured while performing the task, and is calculated as the variation of acceleration. Clinically, it is used to measure the variability of movements.

**Maximal jerk (MJ)** was used to capture the smoothness of movement.

### III. RESULTS

Pearson’s correlation coefficients were used to examine the linear relationship between devices (Vicon vs Memsense), separately for each task and parameter. Values range between -1 and +1, where 1 indicates a total positive correlation, -1 indicates a total negative correlation, and 0 indicates that there is no correlation between the two variables. The strength of the relationship was determined using Evans [29] empirical classifications, in which a correlation coefficient lower than 0.20 is very weak, 0.20-0.39 is weak, 0.40 to 0.59 is moderate, 0.60 to 0.79 is strong, and 0.80 to 1.0 is very strong.

Correlation coefficients between Vicon and Memsense data for each task and parameter are reported in Table 1 and scatterplots of the correlations for each parameter for ARAT 75 are depicted in Fig. 2a-d. With regards to movement time, there was a strong positive correlation between devices for all four parameters (mean \(r = .913\), all \(p’s < 0.01\). For **mean acceleration variability**, there was a strong positive correlation for ARAT 75 and Pour Water tasks (\(r’s = .785\) and .642, respectively), and moderately strong correlations for the Drink
Glass (r = .568), and Thin Tube tasks (r = .568, and .517, respectively), all p’s < 0.01. Correlations between devices for maximum jerk differed between ADL tasks. Although there was a strong positive correlation for the ARAT 75 task (r = .666, p < 0.001) and a moderately strong correlation for the Drink Glass task (r = .394, p < 0.01), the correlation between devices for the Pour Water and Thin Tube tasks were weak (r = .322, respectively), p’s < 0.01. The correlation of mean acceleration values between devices varied between tasks (range = .001 - .540). There was a moderately strong positive correlation between devices for the ARAT 75 task (r = .540), a weak correlation for the Drink Glass task (r = .385), and very weak correlations for the Thin Tube and Pour Water tasks (.201 and .001, respectively).

**Table 1. Correlation Coefficients Between Devices for the Four ADL Tasks and Kinematic Parameters.**

<table>
<thead>
<tr>
<th></th>
<th>Movement Time</th>
<th>Mean Acceleration</th>
<th>Mean Acceleration Variability</th>
<th>Maximum Jerk</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARAT 75</td>
<td>.872**</td>
<td>.540**</td>
<td>.785**</td>
<td>.666**</td>
</tr>
<tr>
<td>Pour Water</td>
<td>.922**</td>
<td>.001</td>
<td>.642**</td>
<td>.394**</td>
</tr>
<tr>
<td>Thin Tube</td>
<td>.895**</td>
<td>.201**</td>
<td>.517**</td>
<td>.322**</td>
</tr>
<tr>
<td>Drink Glass</td>
<td>.963**</td>
<td>.385**</td>
<td>.568**</td>
<td>.432**</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level

Test-retest reliability for the Vicon motion capture system (black bars) and Memsense IMU (grey bars) for the four parameters and tasks is depicted in Fig. 3. Test-retest reliability was conducted separately for each task, device, and parameter using Pearson’s Product Moment Correlation coefficient. Test-retest reliability values above .8 are indicative of good reliability, values between .7 and .8 indicate acceptable reliability, and values between .6 and .7 indicate questionable reliability [30].

With regards to movement time, good test-retest reliability was observed for the ARAT 75 (Vicon = .877, Memsense = .821) and Pour Water (Vicon = .886, Memsense = .853) task, regardless of device. Test-retest reliability for the Drink Glass task was acceptable (Vicon = .747, Memsense = .763). For the Thin Tube task, test-retest reliability for Vicon was acceptable (.745), whereas test-retest reliability for Memsense was questionable (.651). As can be seen in Fig. 3(top left), test-retest reliability was similar for both devices, indicating that both devices consistently measure movement time values at two different time points. For mean acceleration, good test-retest reliability was seen for the ARAT 75 (Vicon = .905, Memsense = .859), Pour Water (Vicon = .854, Memsense = .902), and Drink Glass (Vicon = .841, Memsense = .849) tasks. For the Thin Tube task, there was good test-retest reliability for Vicon (.842) and acceptable (.793) for Memsense. Test-retest reliability was similar for both devices, indicating that both devices consistently measure mean acceleration values at two different time points.

Fig. 2. Scatterplots depicting the correlations of ARAT 75 between Vicon (x axis) and Memsense data (y axis) for movement time (a), mean acceleration (b), mean acceleration variability (c), and maximum jerk (d).
Bar graphs of correlations between trials. Bar graphs depicting the correlations between Vicon and Memsense data of trial 1 (black) and trial 2 (grey) for movement time (top left), mean acceleration (top right), mean acceleration variability (bottom left), and maximum jerk (bottom right).

For **mean acceleration**: variability, there was good test-retest reliability for the Pour Water (Vicon = .848, Memsense = .822) task, regardless of device. For the ARAT 75, Thin Tube, and Drink Glass tasks, there was good test-retest reliability for Vicon (.867, .868, .812, respectively) and acceptable test-retest reliability for Memsense (.794, .739, .732, respectively). Lastly, for **maximum jerk**, test-retest reliability was questionable for Memsense during the Drink Glass (.630) task whereas, poor to unacceptable test-retest reliability was observed for both devices for all other tasks (range = .340-.592) and is depicted in Fig. 3(bottom right).

**IV. DISCUSSION**

In the present study, the ability of a single IMU (Memsense) to capture the kinematics during activities of daily living (ADL) in 50 physically and neurologically healthy adults was examined, and compared to data obtained by a commercially available motion capture system (Vicon).

Before a technological device can be utilized in a clinical setting to evaluate motor (dys)function clinicians must be confident that the device measures relevant metrics the same way each time it is used. Test-retest reliability is a method of estimating a device's reliability at two or more time points. Overall, we observed good to acceptable test-retest reliability for three of the four kinematic parameters (movement time, mean acceleration, mean acceleration variability) when upper extremity kinematics were captured using the Vicon motion capture system. These results are not surprising given that motion capture systems are the gold standard among research scientists. The IMU system yielded similar test-retest reliability values compared to the Vicon system, indicating that they are reliable tools for investigating upper limb movement kinematics. An interesting finding of the present study was that maximum jerk test-retest values ranged between .334 and .630 across all four tasks, and were lower for the motion capture compared to the IMU system. This is congruent with prior research [28] in which test-retest reliability values for healthy participants and stroke patients were lower for movement smoothness parameters (i.e., normalized speed peaks, logarithm of dimensionless jerk, spectral arc length) compared to the indices mean velocity and mean acceleration. The finding that maximum jerk test-retest reliability values were not acceptable for both devices, indicates that caution should be employed when using movement smoothness metrics to evaluate upper extremity function.

Correlations were also conducted to evaluate how well the data collected by the Memsense IMU corresponds to the gold standard (i.e., Vicon motion capture system). Overall, there was a strong positive correlation between devices for movement time, between moderately strong and strong for mean acceleration variability, between weak and strong for maximum jerk, and between very weak and strong for mean acceleration. These results indicate that the Memsense IMU can accurately measure movement time and mean acceleration variability (to a lesser degree), and thus would be a useful, low-cost assessment tool to evaluate motor dysfunction in rural or socioeconomically disadvantaged areas.

Results of the current study are promising, and I will continue this line of research by examining which kinematic parameters are most appropriate for IMU-based motor evaluation, and utilize different methodological approaches to filter and analyze the data. With respect to the latter, sensor fusion approaches (stochastic Extended Kalman Filter, complementary non-linear observer) will be applied to the raw accelerometer and gyroscope data, which will likely improve the accuracy of the IMU measurements. In addition, data will be collected from stroke patients and will include all ADL tasks within the FAT and ARAT assessments. In doing so, this
will allow me to examine the concurrent validity of the IMU data against common clinical assessments.

In contrast to the lower extremity [29-31], only a few studies have examined the reliability and validity of low-cost motor assessment tools during upper extremity ADL [19,32]. This is unfortunate given the number of individuals with stroke and other neurological disorders that affect upper extremity function. Moving forward, efforts will be focused on developing a working prototype that leverages tele-monitoring and tele-consultation in rural areas within the U.S.

REFERENCES