Gait & Posture 71 (2019) 253–260

Contents lists available at ScienceDirect

Gait & Posture

journal homepage: www.elsevier.com/locate/gaitpost

Full length article

Reliability of upper limb movement quality metrics during everyday tasks

Susannah M. Engdahl[a], Deanna H. Gates[a,b,*]

[a] Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI, USA
[b] School of Kinesiology, University of Michigan, Ann Arbor, MI, USA

ARTICLE INFO

Keywords:
Upper limb
Activities of daily living
Movement quality
Smoothness
Test-retest reliability

ABSTRACT

Background: Quantitative assessments of an individual's functional status commonly involve the use of movement quality metrics.
Research question: The purpose of this work was to quantify the reliability of movement quality metrics in healthy adults during a variety of unconstrained activities of daily living (ADLs).
Methods: Nineteen participants performed six ADLs (lifting a laundry basket, applying deodorant, turning a doorknob, placing a pill in a pillbox, placing a pushpin in a bulletin board, and drinking water from a glass) during two separate sessions. The ADLs were divided into reaching and object manipulation phases. Movement quality for each phase was assessed using three measures of smoothness (log dimensionless jerk, spectral arc length, and number of submovements) and one measure of straightness (index of curvature). Within- and between-session reliability was quantified using intraclass correlation coefficients (ICCs) and minimum detectable changes in measured units and as a percentage of their mean value (MDC%).
Results: Reliability was generally lower within-session than between-session and for object manipulation tasks compared to reaching tasks. The ICCs exceeded 0.75 for 5% of the within-session metrics and 73% of the between-session metrics. The average MDC% was 35% for the within-session metrics and 20% for the between-session metrics. Reliability was similar for most metrics when averaged across the tasks, but the number of submovements consistently indicated much lower reliability.
Significance: Unconstrained ADLs can reliably be used to assess movement quality in functional settings that mimic real-world challenges. However, the specific movement quality metrics used in the assessment should be chosen carefully since some metrics perform dissimilarly when applied to the same data. In particular, it may be advisable to use the number of submovements in combination with other metrics, if it is to be used at all.

1. Introduction

Movement smoothness is a fundamental property of skilled, well-coordinated motor behavior [1]. It likely reflects the nervous system's attempt to minimize movement error [2] and energy costs [3], as well as intrinsic mechanical filtering properties of muscle [4]. In healthy individuals, unconstrained point-to-point movements are approximately straight with symmetric velocity profiles [5] and highly blended submovements that overlap in time [6]. These characteristics are altered or absent in movements made by individuals with upper limb impairments, such as those caused by stroke [7], Parkinsonism [8], cerebral palsy [9], and upper limb loss [10]. Consequently, smoothness is a useful tool for assessing performance deficits and improvement following therapeutic interventions.

Smoothness can be quantified using a variety of metrics that focus on different aspects of the movement. Some metrics relate to movement speed, such as the number of peaks in the speed profile, normalized mean speed, and the mean arrest period ratio (proportion of movement time during which the speed exceeds a given percentage of the peak speed) [7]. Other metrics are based on jerk (rate of change of acceleration), including integrated squared jerk and root mean squared jerk (see [1] for additional examples). Several metrics quantify the complexity of the Fourier magnitude spectrum [11,12] since smooth movements contain primarily low frequency components, while unsmooth movements contain additional high frequency components. Most of these metrics do not meet basic utility requirements—that is, they are not dimensionless, robust to signal noise, or monotonically responsive and sensitive to changes within a range covering the physiological spectrum of healthy and pathological movements [11,12]. Recent comparative work suggests that the natural logarithm of dimensionless integrated squared jerk and arc length of the Fourier magnitude spectrum may be well-suited for meeting these constraints

* Corresponding author at: 401 Washtenaw Ave, Ann Arbor, MI, 48109-2214, USA.
E-mail address: gatesd@umich.edu (D.H. Gates).

https://doi.org/10.1016/j.gaitpost.2019.04.023
Received 27 August 2018; Received in revised form 15 March 2019; Accepted 23 April 2019
0966-6362/

[11,12], although other metrics remain prevalent in the literature.

When using any of these metrics to assess movement quality, it is important to understand their test-retest reliability. This information facilitates appropriate interpretation of the data so that meaningful results can be distinguished from experimental errors or natural variability in subject performance [13]. Reliability has already been quantified in patient and healthy populations for a variety of movement quality metrics, including index of curvature [14–16], number of peaks in the speed profile [15,16], normalized jerk [17,18], and the median logarithm of instantaneous curvature [19]. However, these studies have several drawbacks that limit their generalizability. Most studies focused on targeted reaching movements, which may not fully represent the complexity of movements performed in daily life. Two studies [16,19] included simulated activities of daily living (ADLs), which have been shown to involve different movement patterns than functional ADLs [20]. It is important to understand how tasks are accomplished in unconstrained environments that mimic real-world settings, which may present additional challenges that are not involved in targeted reaching movements or simulated ADLs.

Additionally, all studies except [14] and [15] reported reliability in relative terms using Pearson correlations and intraclass correlation coefficients (ICCs). Neither of these metrics define the expected noise in the data in terms of the units in which the original measurement was made, as the standard error of measurement (SEM) and minimum detectable change (MDC) do. Given the expected magnitude of a movement quality metric for a normative population, deviations from normative movement can easily be identified using one of these absolute indicators of reliability. As such, the primary purposes of this study were to 1) establish the range of normative movement quality metrics for a variety functional ADLs, and 2) establish the test-retest reliability of these metrics in healthy adults. A secondary purpose was to explore the effects of filtering on the magnitude and reliability of these metrics.

2. Methods

2.1. Subjects

Nineteen participants (9 male, age: 22 ± 4 years, height: 1.72 ± 0.10 m, weight: 71.4 ± 14.2 kg) provided written informed consent for this institutionally approved study. Exclusion criteria included a history of serious musculoskeletal, cardiovascular, neurological, respiratory, or visual problems. Additional data from the same participants is reported in [21].

2.2. Experimental protocol

Participants performed a series of six ADLs (Table 1; also described in [21]) during two identical data collection sessions. The interval between sessions was at least one day (mean: 12 ± 10 days). During each session, the ADLs were repeated 10 times at a comfortable pace. Since we expected participants to be familiar with these common daily tasks, we did not provide practice time or instructions on how to complete the tasks. However, we did require that each ADL repetition began from a consistent initial posture. For standing tasks, the arms were relaxed at the sides. For seated tasks, the arms rested flat on the table at shoulder width. Object position was based on participant anthropometry (Table 1) and the table was aligned to the bottom of the rib cage while seated.

Reflective markers on the right wrist (radial and ulnar styloid processes) and hand (3rd and 5th metacarpal heads and base of the 3rd metacarpal) were tracked at 120 Hz using a 19-camera motion capture system (Motion Analysis Corporation, Santa Rosa, CA).

2.3. Data analysis

Marker position data were filtered in Visual 3D (C-Motion, Germantown, MA) using a fourth-order low-pass Butterworth filter. All analyses were performed on three separate data sets where data were filtered with either a 6, 10, or 20 Hz cutoff frequency. Analyses focused on the wrist joint center for all tasks except for turning the doorknob, which focused on the center of gravity of the right hand (calculated in Visual 3D). The wrist joint center was defined as the midpoint between the styloid markers. First, second, and third derivatives were calculated as the vector sum of derivatives of the wrist joint center or center of gravity position across all three planes of movement (Fig. 1). Low-pass filters were applied after each differentiation [22].

Each ADL was divided into independent movement phases that were classified as “reaching” or “object manipulation” tasks. A total of 11 reaching and manipulation tasks were included for analysis (Table 1). The outcome measures for each reaching task included smoothness and straightness, while outcome measures for each manipulation task included only smoothness (Fig. 2). Smoothness was measured as log dimensionless jerk (LDJ), spectral arc length (SPARC) and number of submovements (nSUB). LDJ was the natural logarithm of the dimensionless integrated squared jerk:

$$LDJ := \ln\left(\frac{(t_2-t_1)^5}{A^2}\int_{t_1}^{t_2}\dddot{x}(t)^2dt\right),$$

where $x(t)$ was the position of the endpoint (wrist joint center or center of gravity), A was the arc length, t_1 was the movement start time, and t_2 was the movement end time. Because jerk is the rate of change of acceleration, increased LDJ magnitude reflects decreased smoothness (due to more rapid changes in acceleration). SPARC was the arc length of the Fourier magnitude spectrum of the speed profile [12]:

$$SPARC: = \int_0^{\omega_c}\sqrt{\left(\frac{1}{\omega_c}\right)^2+\left(\frac{d\hat{V}(\omega)}{d\omega}\right)^2}d\omega;$$

$$\hat{V}(\omega) := \frac{V(\omega)}{V(0)};$$

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

where $\hat{V}(\omega)$ was the Fourier magnitude spectrum normalized to the DC magnitude ($V(0)$) and ω_c was an adaptively selected cutoff frequency. This cutoff frequency was the smaller of two values: 1) a fixed frequency (ω_c^{max}) chosen to cover the anticipated spectrum of the movement analyzed and 2) the frequency at which the magnitude for all greater frequencies is below a certain threshold ($\bar{V}_{threshold}$). According to recommendations from [12], $\bar{V}_{threshold} = 0.05$ and $\omega_c^{max} = 20\,Hz$. The Fourier magnitude spectrum for a smooth movement will be a smooth function of frequency [12]. The shape of the frequency spectrum for an unsmooth movement will be more complex, such that the arc length of the spectrum is increased. Here, SPARC was defined as positive to reflect this relationship (i.e., an increase in magnitude corresponds to a decrease in smoothness). Submovements were defined as local maxima in the velocity profile, where $\frac{dv}{dt} = 0$ and $\frac{d^2v}{dt^2} < 0$. Straightness was defined using index of curvature (IOC), or the arc length of the position trajectory divided by the length of a straight line between the initial and final position. An IOC of one indicates a perfectly straight movement.

2.4. Statistics

Within-session reliability metrics were calculated based on all 10 repetitions from the first session using data filtered with a cut-off frequency of 10 Hz. Between-session reliability metrics were calculated using the averages of all 10 repetitions from each session. Within- and between-session ICCs were calculated using (2,1) and (2,k) models for absolute agreement, respectively. We assessed heterogeneity of the data by checking the significance ($p < 0.05$) of the between-subjects variance from the two-way ANOVA. In order for the ICC to be valid, there

Table 1
Each ADL was segmented into reaching and manipulation tasks based on position or velocity of the wrist.

ADL	Procedure	Task Name	Task Type	Segmentation Definition
BASKET	Participants stood facing the table with a 5 lb. laundry basket placed on the ground in front of it. They lifted the basket and placed it on the table.	Basket_Reach	Reaching	START: Velocity of WJC exceeds 5 cm/s STOP: WJC reaches global minimum in vertical direction
		Basket_Transport	Reaching	START: WJC reaches global minimum in vertical direction STOP: WJC reaches global maximum in anteroposterior direction
DEO	Participants stood 75% of arm's length from a stick of deodorant placed on the table along the midline of the body. They lifted the deodorant with the right hand, removed the cap with the left hand and simulated three swipes on the left axilla.	Deo_Reach	Reaching	START: Velocity of WJC exceeds 5 cm/s STOP: WJC reaches global maximum in anteroposterior direction
DOOR	Participants stood 75% of arm's length from a doorknob placed at waist height along the midline of the body. They turned the knob clockwise until its latch was fully retracted.	Door_Reach	Reaching	START: Velocity of WJC exceeds 5 cm/s STOP: Mediolateral velocity of RSP reaches first local minimum
		Door_Turn	Manipulation	START: Mediolateral velocity of RSP reaches first local minimum STOP: Mediolateral velocity of RSP reaches last local minimum
PILL	Participants held a pill in the right hand while seated at the table. They placed the pill in a pillbox located at the midline of the body and in between the hands.	Pill_Reach	Reaching	START: Velocity of WJC exceeds 5 cm/s STOP: Velocity of WJC reaches global minimum
PIN	Participants stood at 75% of arm's length from a corkboard with a 1 inch diameter circle drawn at eye level along the midline of the body. They placed a pushpin inside of the circle.	Pin_Reach	Reaching	START: Velocity of WJC exceeds 5 cm/s STOP: Anteroposterior velocity of WJC returns below 5 cm/s
		Pin_Place	Manipulation	START: Anteroposterior velocity of WJC returns below 5 cm/s STOP: Anteroposterior velocity of WJC exceeds 5 cm/s
WATER	Participants sat at the table with a solid plastic cup containing approximately 150 mL of water placed at the midline of the body and in between the hands. They lifted the cup and took a sip.	Water_Reach	Reaching	START: Velocity of WJC exceeds 5 cm/s STOP: Mediolateral velocity of WJC reaches first local minimum
		Water_Transport	Reaching	START: Vertical velocity of WJC reaches last minimum before peak vertical velocity STOP: Anteroposterior velocity of WJC returns below 5 cm/s
		Water_Drink	Manipulation	START: Anteroposterior velocity of WJC returns below 5 cm/s STOP: Anteroposterior velocity of WJC exceeds 5 cm/s

WJC = wrist joint center; RSP = radial styloid process.

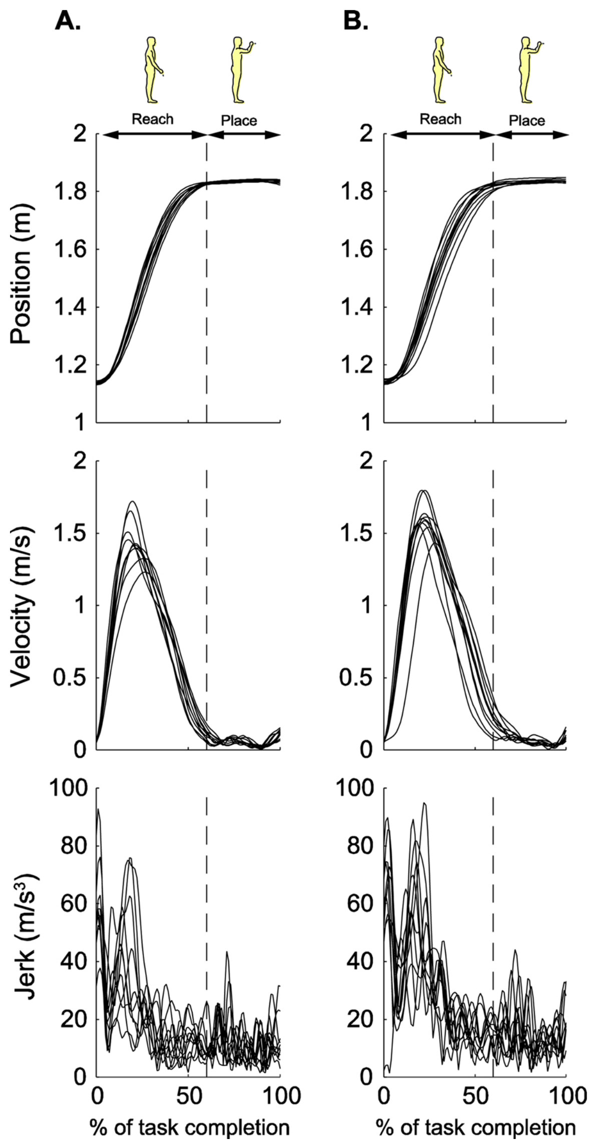

Fig. 1. Position, velocity, and jerk trajectories for ten trials are shown for one representative subject performing the Pin task during session 1 (**A**) and session 2 (**B**). The movement is divided into reaching and placing phases (shown by the vertical dashed line) following the conventions described in Table 1. The trials are normalized to 100% of task completion for visual clarity, although data was not normalized prior to calculating the movement quality metrics.

must be significant between-subject variance [23]. Eight invalid ICCs (4 within-session, 4 between-session) were excluded (Appendix 1, in Supplementary information). Six of the invalid ICCs were calculated for nSUB. Following recommendations in [23], we considered ICCs > 0.90 to be clinically reliable, ICCs > 0.75 to indicate good reliability in other contexts, and ICCs < 0.75 to indicate poor to moderate reliability. Within- and between-session MDCs were calculated according to $MDC_{95} = SEM * 1.96 * \sqrt{2}$, where 1.96 corresponds to a 95% confidence interval and SEM is the square root of the mean square error term from the two-way ANOVA [24]. Because the movement quality metrics are all quantified on a different scale, MDCs are also presented as a percentage of mean values from the first session (MDC%) to facilitate comparison between metrics. All statistical analyses were performed using SPSS 24 (IBM, Armonk, NY).

3. Results

3.1. Normative movement quality metrics

In general, movement quality was higher for reaching tasks (LDJ: 7.65 ± 0.96; SPARC: 1.50 ± 0.07; nSUB: 1.38 ± 0.55) compared to manipulation tasks (LDJ: 9.65 ± 0.41; SPARC: 3.14 ± 0.50; nSUB: 2.78 ± 0.34) (Table 2). Within the reaching tasks, the basket transport task had the poorest movement quality across all four metrics.

3.2. Within- and between-session reliability metrics

Reliability was generally poorer within-session than between-session (Figs. 3 and 4). Only 5% of all valid within-session ICCs were greater than 0.75 (mean: 0.53), compared to 73% of all valid between-session ICCs (mean: 0.80) (Fig. 2). Similarly, the mean MDC% across all metrics (Fig. 4) was higher within-session (35%) than between-session (20%).

These patterns are influenced by the fact that the reliability for nSUB was consistently worse than for the other three metrics. If nSUB is excluded, the discrepancy between the mean within- and between-session MDC% becomes much smaller (within-session: 16%; between-session: 11%). However, the discrepancy between the within- and between-session ICCs does not change (within-session mean: 0.55; between-session mean: 0.76). This is likely because only seven of the 11 within-session ICCs for nSUB were valid in the first place (Appendix 1, in Supplementary information).

3.3. Reliability metrics across tasks

Reliability for LDJ, IOC, and SPARC was similar when averaged across tasks. The mean within-session ICC for these three metrics was 0.55 (LDJ: 0.55, IOC: 0.61, SPARC: 0.49) and the mean within-session MDC% was 16% (LDJ: 19%, IOC: 12%, SPARC: 18%). The mean between-session ICC was 0.82 (LDJ: 0.80, IOC: 0.83, SPARC: 0.81) and the mean between-session MDC% was 11% (LDJ: 13%, IOC: 8%, SPARC: 11%).

The three object manipulation tasks (turning a doorknob, drinking water, and placing a pushpin) had the largest MDC% when averaged across LDJ, IOC and SPARC. This was true both within-session and between-session. Reaching to the doorknob and reaching with the pin had the smallest MDC% when averaged across LDJ, IOC and SPARC for both within-session and between-session. Interestingly, turning the doorknob also had the largest ICC across tasks when averaged across LDJ, IOC and SPARC for both within-session and between-session. There were no consistent patterns in the ICCs for the remaining 10 tasks.

All ICCs, SEMs, and MDCs are included as supplementary material (Appendix 1).

3.4. Effects of filtering

The choice of cutoff frequency affected the magnitude of the movement quality metrics (Appendix 2, in Supplementary information). Higher cutoff frequencies resulted in higher magnitudes (less smooth movements), primarily for LDJ and nSUB. While MDC and MDC% also increased, ICCs were minimally affected by cutoff frequency.

4. Discussion

This work established normative movement quality metrics for unconstrained ADLs performed by healthy adults. The manipulation tasks had poorer movement quality than the reaching tasks, possibly

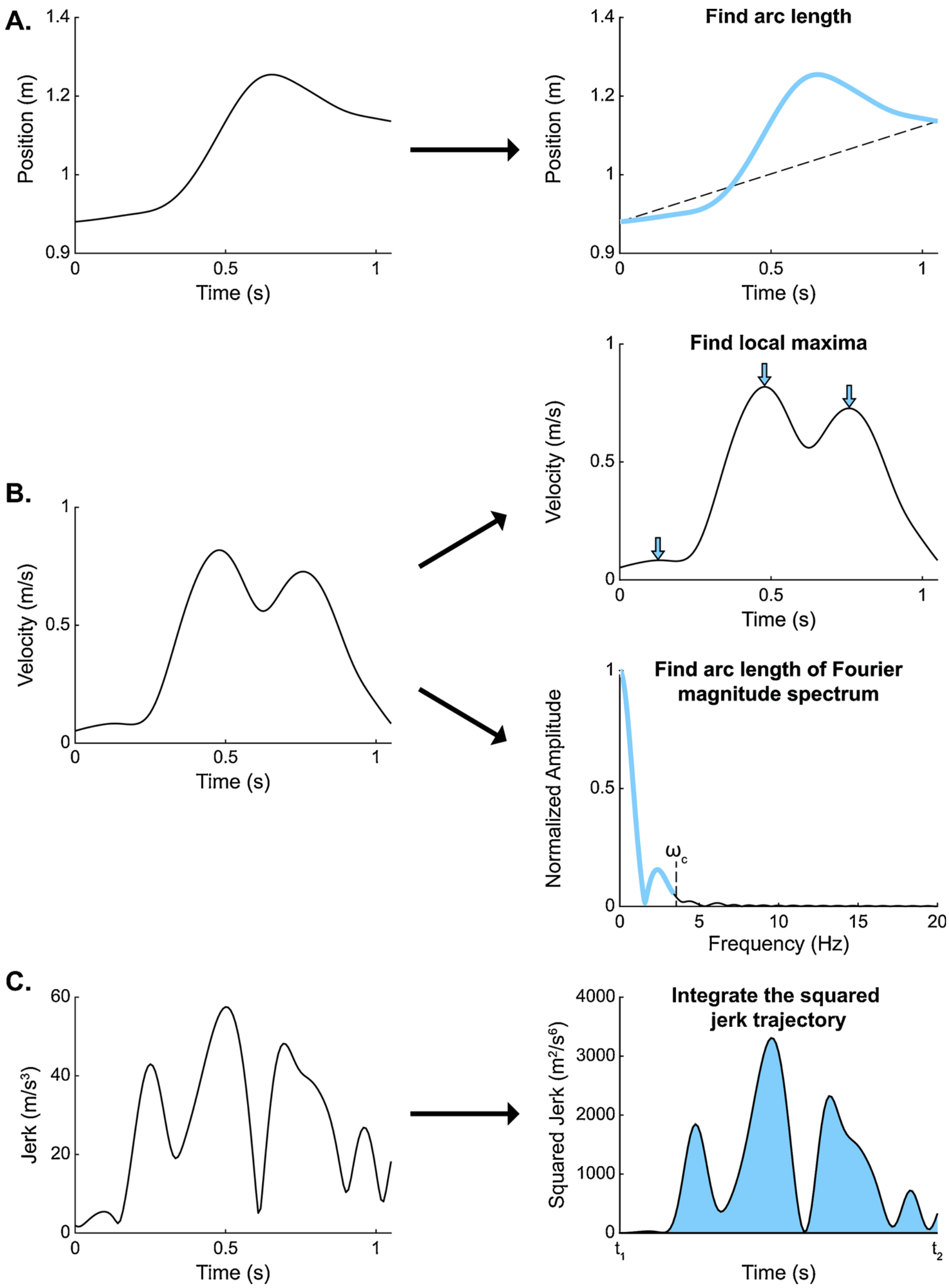

Fig. 2. Movement quality was quantified using four metrics of smoothness and straightness. Derivations of each metric are illustrated using a representative trial from the Deo_Reach task. **A)** The index of curvature (IOC) is a ratio of the arc length of the position trajectory (blue line) to the length of a straight line connecting the movement start and end points (dashed line). **B)** The number of submovements (nSUB) and the spectral arc length (SPARC) are determined using the speed profile. nSUB is the number of local maxima in the speed profile (blue arrows). SPARC is the arc length of the normalized Fourier magnitude spectrum (blue line) between 0 Hz and an adaptively selected cutoff frequency (ω_c). **C)** The log dimensionless jerk (LDJ) is based on the squared jerk trajectory, which is integrated (shaded area) over the duration of the movement (t_2 - t_1). This integral value is normalized based on the duration and arc length (blue line in A) of the movement (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Table 2
Mean (SD) movement quality metrics for sessions 1 and 2.

		LDJ		IOC		SPARC		nSUB	
		Session 1	Session 2	Session 1	Session 2	Session 1	Session 2	Session 1	Session 2
Reaching	Basket_Reach	7.63 (0.45)	7.48 (0.36)	1.07 (0.04)	1.06 (0.03)	1.47 (0.04)	1.46 (0.05)	1.16 (0.18)	1.09 (0.11)
	Basket_Transport	10.02 (0.47)	9.84 (0.36)	1.56 (0.07)	1.54 (0.07)	1.65 (0.07)	1.65 (0.07)	2.85 (0.54)	2.64 (0.49)
	Deo_Reach	7.06 (0.48)	6.80 (0.34)	1.08 (0.07)	1.05 (0.02)	1.48 (0.05)	1.46 (0.01)	1.07 (0.25)	1.02 (0.04)
	Door_Reach	7.15 (0.45)	7.05 (0.38)	1.06 (0.03)	1.06 (0.03)	1.44 (0.02)	1.45 (0.02)	1.21 (0.26)	1.23 (0.25)
	Pill_Reach	6.86 (0.56)	6.76 (0.57)	1.09 (0.04)	1.10 (0.05)	1.46 (0.03)	1.47 (0.03)	1.04 (0.09)	1.03 (0.08)
	Pin_Reach	7.87 (0.45)	7.86 (0.33)	1.04 (0.02)	1.03 (0.01)	1.46 (0.01)	1.45 (0.02)	1.06 (0.10)	1.15 (0.18)
	Water_Reach	7.42 (0.65)	7.46 (0.58)	1.28 (0.20)	1.25 (0.14)	1.60 (0.12)	1.57 (0.10)	1.49 (0.44)	1.43 (0.33)
	Water_Transport	7.53 (0.46)	7.55 (0.49)	1.02 (0.01)	1.02 (0.01)	1.50 (0.06)	1.49 (0.04)	1.31 (0.19)	1.29 (0.26)
Manipulation	Door_Turn	10.20 (0.93)	9.94 (0.92)	n/a	n/a	2.89 (0.76)	2.84 (0.51)	3.27 (0.96)	3.08 (0.97)
	Pin_Place	9.69 (1.49)	8.99 (1.73)	n/a	n/a	2.82 (0.50)	2.74 (0.43)	2.84 (1.30)	2.48 (1.36)
	Water_Drink	9.62 (1.30)	9.47 (1.49)	n/a	n/a	3.83 (0.75)	3.73 (0.62)	2.57 (1.48)	2.45 (1.68)

LDJ = log dimensionless jerk; IOC = index of curvature; SPARC = spectral arc length; nSUB = number of submovements.

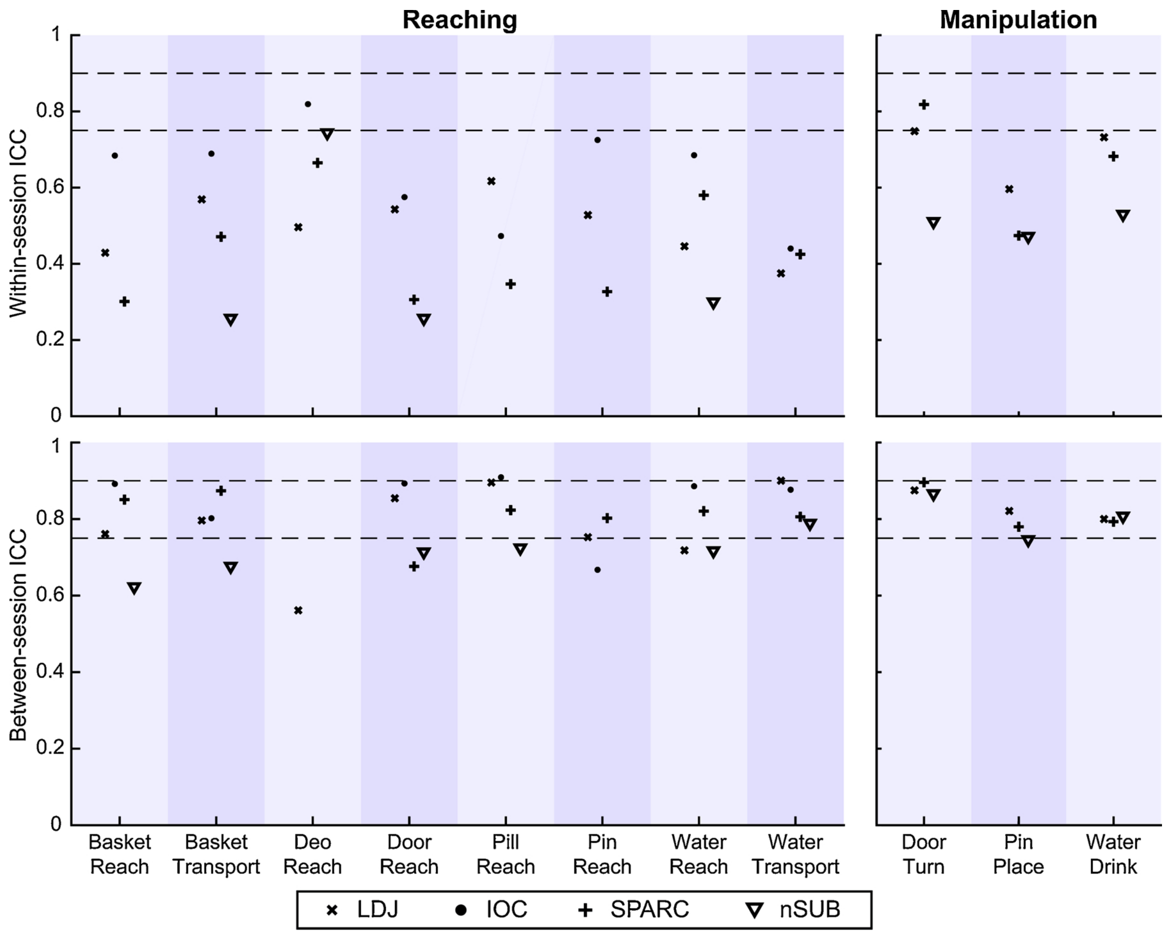

Fig. 3. Within-session (top) and between-session (bottom) intraclass correlation coefficients (ICCs) are shown for each task. Invalid ICCs are not presented. Dashed lines indicate suggested thresholds for interpretation (ICCs > 0.90 indicate clinical reliability, ICCs > 0.75 indicate good reliability in other contexts, and ICCs < 0.75 indicate poor to moderate reliability).

reflecting the fact that multiple small corrections were needed to accomplish the task goal. Within the reaching tasks, the basket transport task had the poorest movement quality. This task was also the most physically demanding task, requiring to participants to lift a weighted laundry basket from the floor and place it on a table. The weight of the basket combined with the difficulty of navigating it around the table to avoid collision may have contributed to the poorer movement quality. The remaining reaching tasks either did not involve moving an object or involved a smaller, lighter object than was easier to transport (i.e., the plastic cup in the water drinking task). Despite the poorer movement quality, the basket task may still be valuable to use in a functional assessment given its unique task requirements.

Additionally, this work quantified the reliability of movement quality metrics in healthy adults performing unconstrained ADLs. Within-session reliability is often better than between-session reliability because differences in experimental set-up between sessions can introduce unintended variation in task performance. Although we attempted to keep our set-up consistent by standardizing the participants' initial posture and distance from the objects, this variability cannot be completely eliminated. In this study, within-session reliability was lower than between-session, suggesting that participants were not consistent in how they performed the tasks. However, this explanation contrasts with our previous work [21], which explored the reliability of peak upper limb and trunk joint angles in healthy adults during ADLs. Several of the tasks included in [21] were also included in this study (applying deodorant, turning a doorknob, and placing a pushpin). Since the within-session peak joint angle reliability for these tasks was generally high, it seems unlikely that there were major differences in participants' movement patterns between repetitions of the tasks. The low within-session reliability for the movement quality metrics might instead be attributed to how they were calculated. Since the peak joint angles were calculated directly from positional data at a single point in time, they characterized movement patterns primarily in a global sense. In contrast, the movement quality metrics were calculated using continuous positional data (IOC) or derivatives of continuous positional data (LDJ, SPARC, and nSUB). As such, they may be more reflective of small changes in performance than peak joint angles. They may also be more affected by signal noise and filtering characteristics, especially for the metrics that are based on derivatives. Indeed, increasing the filter cutoff frequency resulted in increased magnitude for the movement quality metrics (Appendix 2, in Supplementary information). Of the three metrics calculated using differentiated data, filtering had the greatest impact on nSUB and LDJ. This is consistent with other work showing that SPARC is robust to changes in signal-to-noise ratio and filtering characteristics, especially in comparison to LDJ [12].

Nonetheless, within-session comparisons are less common than between-session comparisons. Most performance assessments involve making comparisons between different time points (e.g., before and after a therapeutic intervention), rather than within a single testing session. Between-session reliability is more relevant than within-session reliability in these situations. In fact, only one previous study of movement quality reliability even included within-session reliability [19]. For most tasks, the between-session ICC exceeded the threshold for good reliability (ICC > 0.75; defined by [23]) and the MDC% was low. This suggests that the tasks would be an effective basis for identifying changes in movement quality across multiple testing sessions.

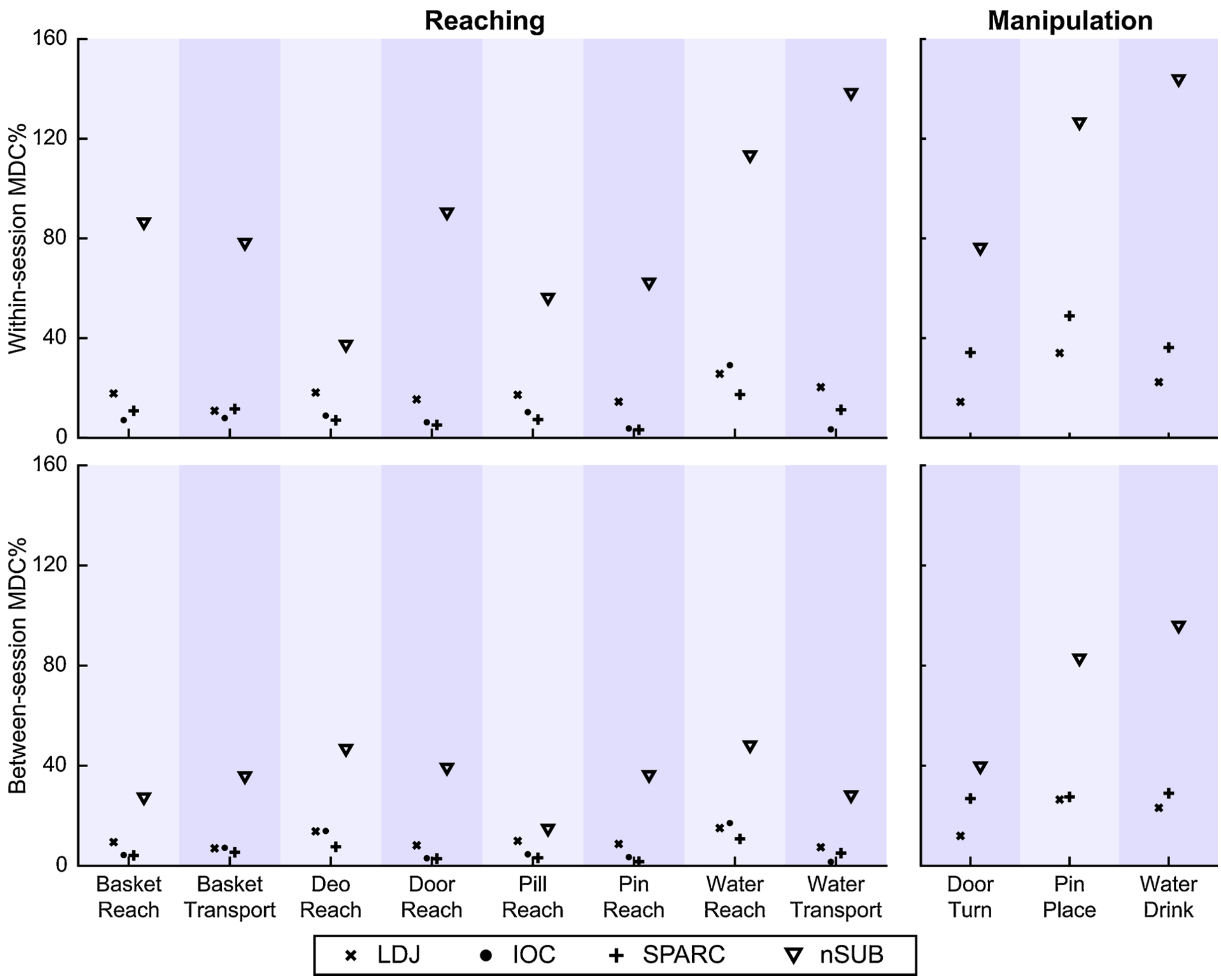

Fig. 4. Within-session (top) and between-session (bottom) minimum detectable change values for each task are presented as a percentage of the mean value from session 1 (MDC%).

The reliability metrics presented here are comparable to those reported previously for different tasks. Between-session reliability for the IOC during reach to grasp and reach to target tasks has been reported for healthy adults (ICC range: 0.64–0.88; MDC range: 0.00–0.08) [14], individuals with stroke (ICC ranges: 0.92–0.93 [14] and 0.08–0.95 [15]; MDC range: 0.08–0.19 [14]; MDC% range: 7.4%–28.9% [15]), and children with cerebral palsy (ICC range: 0.59–0.81) [16]. Between-session reliability for nSUB has also been reported for individuals with stroke (ICC range: 0.43–0.84; MDC% range: 24.4%–67.6%) [15] and children with cerebral palsy (ICC range: 0.82–0.91) [16]. These similarities are encouraging, as they suggest that tasks do not need to be tightly constrained in order for individuals to perform them consistently. Unconstrained, functional ADLs can still be used in assessments of movement quality.

Reliability tended to be slightly worse for manipulation tasks than reaching tasks, as indicated by the comparatively larger MDC and MDC%. In contrast, the ICCs were similar across all tasks, which might be attributed to shortcomings in the definition of the ICC. ICC magnitude depends on between-subjects variability, such that high between-subject variability in a dataset can artificially increase the ICC even if trial-to-trial variability is high [24,25]. Although the elevated MDC and MDC% values suggest the trial-to-trial variability was high (i.e., poor reliability), this may not be reflected in the ICCs if the between-subject variability was also high. Since the object manipulation tasks involved making multiple small adjustments to accomplish a goal, it is possible that larger movement variability was involved. Indeed, the within-subject standard deviations tended to be higher for the manipulation tasks (mean: 0.65) than the reaching tasks (mean: 0.18) (Appendix 3. in Supplementary information). However, the comparatively lower reliability for manipulation tasks does not necessarily invalidate the use of those tasks in performance assessments. Manipulation tasks challenge different aspects of movement than reaching tasks, so they may be helpful in developing a more comprehensive assessment of functionality.

Of the four movement quality metrics, nSUB tended to show the worst reliability when applied to any task. This is likely due to several problems with how the metric is calculated. Despite the intuitive nature of this metric, it can be difficult to accurately identify peaks [11]. Even when peaks are identified accurately, the number of peaks may not truly reflect the number of underlying submovements. The summation of multiple submovements can obscure peaks or create spurious peaks that do not correspond to any of the component submovements, and this issue may be further exacerbated by signal noise [26] and signal filtering characteristics. Aggressive filtering may conceal true peaks, while minimal filtering may not remove any spurious peaks (see Appendix 2, in Supplementary information for the effect of filter cutoff frequency on the magnitude of nSUB). Furthermore, this metric is insensitive to changes in the periods of arrest between submovements [1,11], which is problematic since the temporal separation of submovements is strongly linked to overall movement smoothness. Additionally, the ordinal scale of this metric means that for movements with few submovements, a small change in the number of peaks reflects a disproportionately large change in smoothness [11]. This issue is clearly seen in the elevated MDC% values for nSUB (range:

15%–144%), even though the range of MDC values is only 0.16–3.71 (Appendix 1, in Supplementary information). Ultimately, metrics that rely on the entire continuum of data (such as LDJ and SPARC) are more likely to respond consistently and accurately to changes in movement smoothness than metrics like nSUB that focus on only isolated features of the data [1,11].

This study established the reliability of movement quality metrics for healthy adults, expanding on prior work by including a wider variety of movement quality metrics and functional ADLs. Additionally, we provided normative data for smoothness metrics during ADLs, which can be used as a reference for future studies. These values are provided for three filter cut-off frequencies to facilitate comparison with other studies that used different methodology. Additional work is needed to quantify the reliability of these measures in specific patient populations and their sensitivity to detect changes over time.

Conflicts of interest

None.

Acknowledgements

This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs through the Orthotics and Prosthetics Outcomes Research Program under Award No. W81XWH-16-1-06548. The opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the Department of Defense.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.gaitpost.2019.04.023.

References

[1] N. Hogan, D. Sternad, Sensitivity of smoothness measures to movement duration, amplitude, and arrests, J. Mot. Behav. 41 (2009) 529–534.
[2] C.M. Harris, D.M. Wolpert, Signal-dependent noise determines motor planning, Nature 394 (1998) 780–784.
[3] J. Nishii, Y. Taniai, Evaluation of trajectory planning models for arm-reaching movements based on energy cost, Neural Comput. 21 (2009) 2634–2647.
[4] A.M. Krylow, W.Z. Rymer, Role of intrinsic muscle properties in producing smooth movements, IEEE Trans. Biomed. Eng. 44 (1997) 165–176.
[5] T. Flash, N. Hogan, The coordination of arm movements: an experimentally confirmed mathematical model, J. Neurosci. 5 (1985) 1688–1703.
[6] B. Rohrer, S. Fasoli, H.I. Krebs, B. Volpe, W.R. Frontera, J. Stein, et al., Submovements grow larger, fewer, and more blended during stroke recovery, Motor Control 8 (2004) 472–483.
[7] B. Rohrer, S. Fasoli, H.I. Krebs, R. Hughes, B. Volpe, W.R. Frontera, et al., Movement smoothness changes during stroke recovery, J. Neurosci. 22 (2002) 8297–8304.
[8] J.R. Tresilian, G.E. Stelmach, C.H. Adler, Stability of reach-to-grasp movement patterns in Parkinson's disease, Brain 120 (1997) 2093–2111.
[9] J.C. Van Der Heide, J.M. Fock, B. Otten, E. Stremmelaar, M. Hadders-Algra, Kinematic characteristics of reaching movements in preterm children with cerebral palsy, Pediatr. Res. 57 (2005) 883.
[10] J. Cowley, L. Resnik, J. Wilken, L. Smurr Walters, D. Gates, Movement quality of conventional prostheses and the DEKA Arm during everyday tasks, Prosthet. Orthot. Int. 41 (2017) 33–40.
[11] S. Balasubramanian, A. Melendez-Calderon, E. Burdet, A robust and sensitive metric for quantifying movement smoothness, IEEE Trans. Biomed. Eng. 59 (2012) 2126–2136.
[12] S. Balasubramanian, A. Melendez-Calderon, A. Roby-Brami, E. Burdet, On the analysis of movement smoothness, J. Neuroeng. Rehabil. 12 (2015) 112.
[13] M.H. Schwartz, J.P. Trost, R.A. Wervey, Measurement and management of errors in quantitative gait data, Gait Posture 20 (2004) 196–203.
[14] T.S. Patterson, M.D. Bishop, T.E. McGuirk, A. Sethi, L.G. Richards, Reliability of upper extremity kinematics while performing different tasks in individuals with stroke, J. Mot. Behav. 43 (2011) 121–130.
[15] J.M. Wagner, J.A. Rhodes, C. Patten, Reproducibility and minimal detectable change of three-dimensional kinematic analysis of reaching tasks in people with hemiparesis after stroke, Phys. Ther. 88 (2008) 652–663.
[16] S. Schneiberg, P. McKinley, E. Gisel, H. Sveistrup, M.F. Levin, Reliability of kinematic measures of functional reaching in children with cerebral palsy, Dev. Med. Child Neurol. 52 (2010) e167–e173.
[17] M. Caimmi, S. Carda, C. Giovanzana, E.S. Maini, A.M. Sabatini, N. Smania, et al., Using kinematic analysis to evaluate constraint-induced movement therapy in chronic stroke patients, Neurorehabil. Neural Repair 22 (2008) 31–39.
[18] M. Caimmi, E. Guanziroli, M. Malosio, N. Pedrocchi, F. Vicentini, L.M. Tosatti, et al., Normative data for an instrumental assessment of the upper-limb functionality, Biomed Res. Int. 35 (2015) 36.
[19] R. Osu, K. Ota, T. Fujiwara, Y. Otaka, M. Kawato, M. Liu, Quantifying the quality of hand movement in stroke patients through three-dimensional curvature, J. Neuroeng. Rehabil. 8 (2011) 62.
[20] S.A.F. Taylor, A.E. Kedgley, A. Humphries, A.F. Shaheen, Simulated activities of daily living do not replicate functional upper limb movement or reduce movement variability, J. Biomech. (76) (2018) 119–128.
[21] S.M. Engdahl, D.H. Gates, Reliability of upper limb and trunk joint angles in healthy adults during activities of daily living, Gait Posture 60 (2018) 41–47.
[22] L.H. Salmond, A.D. Davidson, S.K. Charles, Proximal-distal differences in movement smoothness reflect differences in biomechanics, J. Neurophysiol. 117 (2017) 1239–1257.
[23] L.G. Portney, M.P. Watkins, Foundations of Clinical Research: Applications to Clinical Practice, 3rd ed, Prentice Hall, 2009.
[24] J.P. Weir, Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM, J. Strength Cond. Res. 19 (2005) 231–240.
[25] S. Vaz, T. Falkmer, A.E. Passmore, R. Parsons, P. Andreou, The case for using the repeatability coefficient when calculating test–retest reliability, PLoS One 8 (2013) e73990.
[26] B. Rohrer, N. Hogan, Avoiding spurious submovement decompositions: a globally optimal algorithm, Biol. Cybern. 89 (2003) 190–199.