Multiple Sclerosis and Related Disorders 70 (2023) 104521

Contents lists available at ScienceDirect

Multiple Sclerosis and Related Disorders

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

Original article

Reliability, validity and clinical usability of a robotic assessment of finger proprioception in persons with multiple sclerosis

Monika Zbytniewska-Mégret [a,*], Christoph M. Kanzler [a,c], Joke Raats [b,d], Cigdem Yilmazer [b,d], Peter Feys [b,d], Roger Gassert [a,c], Olivier Lambercy [a,c,#], Ilse Lamers [b,d,e,#]

[a] Rehabilitation Engineering Laboratory, Institute of Robotics and Intelligent Systems, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland
[b] REVAL Rehabilitation Research Center, Faculty of Rehabilitation Sciences, Hasselt University, Hasselt, Belgium
[c] Future Health Technologies, Singapore-ETH Centre, Campus for Research Excellence And Technological Enterprise (CREATE), Singapore
[d] Universitair MS Centrum UMSC Hasselt, Pelt, Belgium
[e] Noorderhart Rehabilitation and MS Centre, Pelt, Belgium

ARTICLE INFO

Keywords:
Multiple sclerosis
Assessment
Rehabilitation
Somatosensory function
Proprioception

ABSTRACT

Background: Multiple sclerosis often leads to proprioceptive impairments of the hand. However, it is challenging to objectively assess such deficits using clinical methods, thereby also impeding accurate tracking of disease progression and hence the application of personalized rehabilitation approaches.
Objective: We aimed to evaluate test-retest reliability, validity, and clinical usability of a novel robotic assessment of hand proprioceptive impairments in persons with multiple sclerosis (pwMS).
Methods: The assessment was implemented in an existing one-degree of freedom end-effector robot (ETH MIKE) acting on the index finger metacarpophalangeal joint. It was performed by 45 pwMS and 59 neurologically intact controls. Additionally, clinical assessments of somatosensation, somatosensory evoked potentials and usability scores were collected in a subset of pwMS.
Results: The test-retest reliability of robotic task metrics in pwMS was good (ICC=0.69–0.87). The task could identify individuals with impaired proprioception, as indicated by the significant difference between pwMS and controls, as well as a high impairment classification agreement with a clinical measure of proprioception (85.00–86.67%). Proprioceptive impairments were not correlated with other modalities of somatosensation. The usability of the assessment system was satisfactory (System Usability Scale ≥73.10).
Conclusion: The proposed assessment is a promising alternative to commonly used clinical methods and will likely contribute to a better understanding of proprioceptive impairments in pwMS.

1. Introduction

Hand somatosensory impairments are common and are one of the earliest symptoms of Multiple Sclerosis (MS) (Wallin et al., 2019; Bertoni et al., 2015; Kister et al., 2013). amongst somatosensory modalities, proprioception is of particular interest, since it is crucial for the generation of coordinated movements and hence for the hand use in many activities of daily living (ADLs) (Miall et al., 2018). However, assessing proprioception is challenging, as there is a lack of sensitive assessments. Commonly used clinical assessments of proprioception are human-administered (Stolk-Hornsveld et al., 2006). While their execution is simple and rapid, they show poor interrater reliability, have an ordinal scale and are subjective (Lincoln et al., 1991), making it challenging to detect subtle changes in impairment severity over time. As an alternative assessment approach, neurophysiology measurements, namely Somatosensory Evoked Potentials (SSEPs), can be applied. This assessment is advantageous in its interval scale and reliability (Brown et al., 2017). Increased SSEPs latency is common in MS due to demyelination within the central fibres of the dorsal column and has been shown to coincide with sensory symptoms (Walsh, 2005). However, the use of SSEPs in regular clinical practice has been questioned, since their recording is time consuming, labour intensive and requires trained personnel (Aminoff, 1984). Therefore, novel assessment approaches are needed, which could allow to quantitatively measure proprioception in

* Corresponding author.
E-mail address: relab.publications@hest.ethz.ch (M. Zbytniewska-Mégret).
# These two authors contributed equally to this work

https://doi.org/10.1016/j.msard.2023.104521
Received 3 January 2022; Received in revised form 31 December 2022; Accepted 13 January 2023
Available online 14 January 2023

a clinically meaningful and applicable way. A recent approach is to use robotics for the assessment of proprioception (Rinderknecht et al., 2018; Ingemanson et al., 2019; Zbytniewska et al., 2021). In such assessment paradigm, instead of an examiner, it is a robotic device that provides a precise stimulus (e.g., displacement of the limb of the tested subject) and objectively measures a resulting response, thereby offering the possibility to sensitively quantify proprioceptive ability. Most of the existing robotic platforms capable of assessing hand proprioception have only been evaluated with stroke patients or healthy individuals (Rinderknecht et al., 2018; Ingemanson et al., 2019; Zbytniewska et al., 2021). It is unclear whether they are also applicable to persons with Multiple Sclerosis (pwMS), while there is a need to evaluate clinimetric properties of newly proposed outcome measures in target populations (Shirota et al., 2019; Schwarz et al., 2019). amongst properties of importance, reliability and measurement error are essential to understand the capability of an assessment metric to capture incremental progress over time (distinguish real improvement from measurement noise) (Lexell and Downham, 2005). Discriminant and concurrent validity determine how accurate a novel metric is at capturing impairment (Kanzler et al., 2020). Usability is necessary to ensure users are engaged and therefore do their best at the assessment (ISO 9241–11 2018).

The objective of this work was to evaluate test-retest reliability, validity and clinical usability of a robotic assessment of hand proprioception, based on a passive position matching task, in pwMS. The proposed robotic assessment was implemented on the ETH MIKE robot, a one degree-of-freedom platform focusing on the index finger metacarpophalangeal (MCP) joint (Zbytniewska et al., 2023). We hypothesized, that the proposed robotic metrics are reliable, given the objectivity of their scale and that they are capable of discriminating pwMS according to their hand proprioceptive impairment. This work aspires to contribute to the field of neurorehabilitation by providing an objective, sensitive and usable assessment of proprioception, which could deepen the understanding of sensory deficits in pwMS and aid in personalizing therapies.

2. Methods

2.1. Participants

Participants with MS were recruited in the Noorderhart Rehabilitation and MS Centre, Pelt, Belgium. The inclusion criteria were older than eighteen years and diagnosis with MS (according to the McDonald criteria (Hartung et al., 2019)). Participants were excluded if they had a relapse or relapse-related treatment(s) within the last three months, a complete paralyses of both upper limbs, were not able to detect any passive movements of the hand and fingers, were not able to place the hand into the robot without discomfort or pain, had marked or severe intention tremor (Fahn's tremor rating scale on finger-to-nose $>$ 3 (Hooper et al., 1998)), had marked or severe spasticity or stiffness in the finger flexors, elbow flexors or shoulder adductors (Modified Ashworth Scale $>$ 3 (Gregson et al., 1999)), had other medical conditions which can influence the function of the hand (e.g. pain, oedema, orthopaedic impairments) and/or had severe cognitive or visual impairments interfering with testing and training. Neurologically-intact control subjects were recruited in Hasselt, Belgium and in Zurich, Switzerland. Exclusion criteria for control subjects were any history of neurological, orthopaedic or rheumatologic disease affecting wrist or hand function.

2.2. Primary outcome measure

The robotic assessments were performed using the ETH MIKE (Motor Impairment and kinesthetic Evaluation), a one degree of freedom end-effector robot (Fig. 1) (Zbytniewska et al., 2023). The device can provide well-controlled displacements at the index finger MCP joint, as well as measure its torque, velocity and position. While performing the robotic assessment, the participants were seated in front of the device, one hand grasping a 3D-printed handle, and with the index finger attached to a finger interface using Velcro straps. A tablet computer with a Graphical User Interface (GUI) was placed above the hand, so that the

Fig. 1. The ETH MIKE device and its graphical user interface. This one degree of freedom end-effector robot can provide well-controlled displacement to the index finger, which is crucial for an objective and sensitive proprioception assessment. In the gauge position matching task protocol, the participant's finger is passively moved by the robot from a starting position (0° angle at the MCP joint, 30° from the middle of device's workspace) to another, random position. Then, the participant needs to indicate the perceived finger position on the tablet screen with a virtual gauge indicator, using the non-tested hand. When this was not possible because of impairments of the non-tested hand, the experimenter moved the indicator based on participant's oral feedback. This was repeated for 21 different positions (integer values [10–30°] in flexion from the starting position).

vision of the index finger was constrained. In order to evaluate proprioceptive impairments, the gauge position matching task was used, as previously described in detail (Zbytniewska et al., 2021; Rinderknecht et al., 2016). Briefly, the finger was passively moved by the robot from a starting position to a different position in the flexion direction. The subject was prompted to indicate with the other hand, on the tablet screen placed directly above the tested hand, the perceived finger position. Within one experimental session this was repeated for 21 different positions, ranging from 10° to 30° in flexion from the starting position (0° angle at the MCP joint, 30° from the middle of device's workspace). The outcome measures consist of the constant error (CE = average error), absolute error (AE = average absolute error), variable error (VE = standard deviation of errors) and total variability (E = root mean square errors), all expressed in degrees. An 'error' refers to the difference between the reported position and the presented position.

2.3. Secondary outcome measures

Secondary outcome measures consisted of clinical assessments of somatosensation: the Erasmus MC modification of the Nottingham Sensory Assessment (EmNSA, the proprioception subscale was of particular interest) (Stolk-Hornsveld et al., 2006), Semmes-Weinstein Monofilaments (SWM) (Tracey et al., 2012), Rydel Seiffer Tuning Fork (RSTF) (Panosyan et al., 2016) and Somatosensory Evoked Potentials (SSEPs) obtained from electrical stimulation at the median nerve of the wrist (Walsh, 2005). We then analysed the cortical latency and amplitude of the SSEPs signal (N20). The shortest latency and the greatest amplitude out of three trials were used for statistical analysis. We used the N20 latency of 20.0 ms as the abnormality threshold (Chiappa and Ropper, 1982). Additionally, as measures of hand dexterity, the Box & Block Test (BBT) (Mathiowetz et al., 1985) and the Nine Hole Peg Test (NHPT) (Feys et al., 2017) were used. The clinical usability of the ETH MIKE system was evaluated with pwMS using the System Usability Scale (SUS) (Brooke, 1996).

2.4. Experimental protocol

Experiments with pwMS were conducted on two days within a time span of maximum one week. On the first day (test), three sessions of the robotic assessment were performed consecutively (with a short break after each session). Additionally, demographic information was collected on the first day (age, gender, handedness, EDSS-Expanded Disability Status Scale (Kurtzke, 1983), as shown in Table 1). Clinical assessments were also performed on the first day. On the second test day (retest), only the robotic assessment was repeated (all three sessions), as well as the System Usability Scale was collected. SSEPs were extracted from medical records if data were recently collected (maximum 1 month before or after the first test day). Control subjects performed only one experimental session with the robotic assessment. For pwMS and control subjects, in each robotic assessment session both hands were tested, one side at a time.

Table 1
Participants' demographics and clinical characteristics.

	pwMS	Control
n	43	59
Age	48.60 ± 12.46	62.56 ± 12.28
Gender	29 F, 14 M	28 F, 31 M
Handedness	34 R, 5 L, 4 A	55 R, 3 L, 1 A
EDSS	4.21 ± 2.10	–
Clinical test	**pwMS Right**	**pwMS Left**
NHPT [s]	22.70 ± 8.25	24.37 ± 8.14
BBT	47.43 ± 11.76	48.50 ± 12.13
EmNSA total	36.13 ± 5.00	37.20 ± 3.96
EmNSA prop.	7.87 ± 0.43	7.83 ± 0.37
SWM finger	2.60 ± 1.02	2.57 ± 1.02
RSTF index	7.53 ± 0.76	7.47 ± 0.88

Legend: F-female, M-male, L-left, R-right, A-ambidextrous, EDSS-Expanded Disability Status Scale. Handedness was evaluated using the Edinburgh Handedness Inventory. NHPT-Nine Hole Peg Test, BBT-Box & Block Test, EmNSA-Erasmus MC modification Nottingham Sensory Assessment, EmNSA prop. - proprioception subscale of EmNSA, SWM-Semmes Weinstein Monofilaments, RSTF-Rydel Seiffer Tuning Fork.

2.5. Statistical analysis

Intraclass correlation coefficient ICC(A,k) was used to calculate absolute agreement between test and retest, based on a two-way analysis of variance, taking into account all assessment sessions (i.e. 3 sessions on test and 3 sessions on retest) (Koo and Li, 2016). ICC values above 0.7 were considered acceptable (Prinsen et al., 2018). Further, smallest real difference (SRD) and SRD% (% with respect to the range across all sessions) were calculated. Desired SRD% is below 30.3% (Kanzler et al., 2020). To quantify the learning effect (LE), a difference between two sessions within one day, as well as between two days (mean across 3 sessions on test and retest), normalized with respect to the range, were calculated. Learning effect outside of the range of [−6.35 and 6.35] has previously been defined as undesired (Kanzler et al., 2020). For validity analysis, only results of robotic assessments from the first session on day 1 were taken into account, in order to best represent a clinical use scenario. To evaluate discriminant validity, robotic assessment results of pwMS were compared to control subjects. This was performed using Kruskal-Wallis test and the Area Under the Curve (AUC) of the Receiver Operating Characteristic (Kanzler et al., 2020). The AUC method defines a rate of classification of each subject into the two groups (pwMS/control). Desired AUC is above 0.7[16]. Moreover, the percentage of pwMS with a score worse than the 95th percentile of control subjects was calculated. Concurrent validity was evaluated by comparing the robotic assessment of proprioception to clinical measures of somatosensation and hand dexterity as well as to SSEPs, using Spearman correlation. P-values were Bonferroni corrected (10 correlations for each metric). The correlation strength was defined as: $0.4< \rho <0.69$ moderate, $\rho >0.7$ strong (Schober et al., 2018). For all statistical analysis left- and right-hand measurements were pulled together, due to no significant difference between hands in both pwMS and controls.

3. Results

3.1. Feasibility and clinical usability

In total 73 pwMS were contacted for study recruitment purposes based on their known clinical records and expected compliance to the inclusion/exclusion criteria. From these, 2 were excluded due to a recent relapse, 26 were not willing to participate in the study due to various reasons not related to the exclusion criteria (no time, no interest, did not feel well, lived too far, didn't react on the second phone call to make an appointment). Fourty-five pwMS agreed to participate in the study, 43 completed all measurements on both hands with the ETH MIKE (Table 1). Out of the 45 recruited pwMS, 30 completed clinical assessments on both hands, while SSEPs were collected for 19 individuals (both hands). In total 59 control subjects performed the robotic assessments on both hands, across two study locations. None of the controls had to be excluded.

Overall, the robotic assessment was found feasible in pwMS given the high protocol completion rate. Moreover, a single measurement session was fast to perform, it took on average 3.60±0.87 min (excluding setup time and instructions). The whole protocol on a single day (including task repetition, setup, instructions and breaks) took approx. 1 hour. The SUS score for the robotic system was equal to 73.10±20.14 (N = 29) on the first day and it was equal to 75.09±19.67 (N = 27) on the retest.

3.2. Test-retest reliability

Reliability was good for the four robotic task metrics (Table 2, Fig. 2). ICC was above 0.7 for 3/4 metrics, AE was just below the threshold (0.69). SRD% was below 30.3% for all metrics (scores ranging from 12.03 to 28.12). Learning effect was negligible within a single day, but it was above the threshold between days for AE and E.

3.3. Discriminant and concurrent validity

It was possible to discriminate between control subjects and pwMS for 3 out of 4 robotic metrics (all but CE), as indicated by AUC above 0.7 and a significant difference between pwMS and controls ($p < 0.001$) - Table 3. Generally, this population was not strongly impaired according to both robotic metrics (13.33–36.67% of pwMS impaired on left or right hand) and the clinical measure of proprioception (20.0%). Those subjects that were classified as impaired on EmNSA proprioception showed poorer performance in the gauge position matching task than controls and than pwMS that scored within norm on EmNSA proprioception. There was a significant difference between these three groups (Fig. 3a). Moreover, there was a high level of classification agreement between robotic metrics and EmNSA proprioception (85.00–88.33%) - Fig. 3b.

Robotic metrics were not correlated with clinical assessments describing other modalities of somatosensory function than proprioception, or hand dexterity (Table 4). There was a moderate significant correlation of robotic metrics with EmNSA proprioception subscale ($\rho =$ 0.40, −0.42, −0.53, $p < 0.05$ for AE, E and VE), however 80% of the scores were in the ceiling of the clinical scale. Moreover, no significant correlations were found between the gauge position matching task outcomes and neurophysiological measures of somatosensation (SSEPs latency and amplitude). More subjects were classified as impaired according to SSEPs latency (63.16% - Table 3) than to the robotic assessment in at least one hand for that same sample (15.79–36.84%).

4. Discussion

The goal of this work was to evaluate clinimetric properties of a novel robotic assessment of proprioception in pwMS. This paper showed that the proposed method is reliable, valid and clinically usable in pwMS, and therefore suggests that it is suitable to be implemented in clinical practice to regularly monitor proprioceptive deficits. The key novelty of the ETH MIKE gauge position matching task is that it can objectively and sensitively quantify hand proprioceptive deficits by focusing on the MCP joint of the index finger, which reduces platform's complexity and increases its clinical usability.

Test-retest reliability of the robotic assessment in pwMS was generally satisfactory for all four metrics and the achieved result is in line with literature considering technology-aided assessments (ICC 0.7–0.9) (Schwarz et al., 2019). However, in another study performed on the same device with stroke subjects, ICC of AE was higher (0.90 on the more affected side (Zbytniewska et al., 2021)). One aspect contributing to higher ICC in that study was higher inter-subject variability and factor severity, given a larger range of impairments in the studied population (BBT 20.90±20.16 in stroke vs 47.43±11.76 in pwMS). Further, we found that although within-day learning/fatigue effects were minimal, the learning effect between test and retest was above the threshold for two metrics (AE & E) in pwMS. It might be that through repeated practice of the task, a learning process occurred, which got consolidated during a few days break between test and retest. Therefore, for future study protocols with pwMS it would be recommended to include one day for familiarization with the system.

Overall, the robotic task could identify individuals with proprioceptive impairment. Up to 36.67% of pwMS in this study had proprioceptive deficits, which is comparable to previous findings. Another study that used an alternative robotic assessment of proximal joints of the upper limb revealed similar prevalence - 9/41, 22% of pwMS were impaired in proprioception (Simmatis et al., 2020). However, in our study pwMS were more severely affected (EDSS 4.21±2.10 vs 2.5 ± 2.5 in Simmatis et al. (Simmatis et al., 2020)), which might explain the higher prevalence in our study. Further, the impairment classification agreement with EmNSA proprioception was high (up to 86.67%). In fact, more subjects were classified as impaired according to the ETH MIKE robotic metrics. That is an expected result given higher sensitivity of the robotic assessment method. Indeed, the robotic assessment does not suffer from any ceiling effect and its scale has a higher resolution, hence subtle deficiencies can be detected.

The proposed robotic assessment is specific to measuring proprioception, since we found no significant correlations with clinical measures of other modalities of somatosensation (e.g., perception of vibration with tuning fork or tactile sensitivity with monofilaments). The correlation of the robotic scores was found significant only with the EmNSA proprioception subscale. However, that scale was strongly affected by the ceiling effect (80% of pwMS reached the maximum score), therefore results of this correlation analysis should be treated with caution and it's more appropriate to use classification agreements to compare these two scales (Fig. 3b). The lack of stronger association between the position matching task and clinical assessments of somatosensation could also be explained by the involvement of high-level processing in the robotic task, adding a cognitive confound on top of the measure of proprioception. The task requires subjects to integrate visual information with proprioceptive feedback to match finger's position with a virtual gauge on a tablet computer screen, while most of the other clinical assessments exclude vision. Further, an explanation for the dissociation between BBT/NHPT and the position matching task can come from the large influence of the motor capabilities in the outcome of the former, while the robotic task is purely passive. Moreover, proprioceptive deficits can be compensated with vision in tests such as BBT/NHPT.

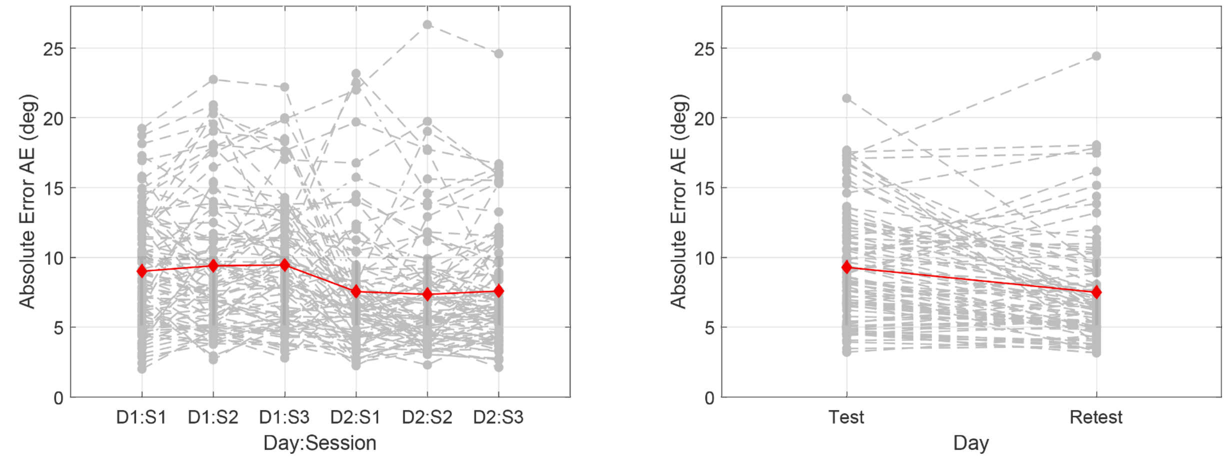
We found that more subjects had abnormal SSEPs latency than impaired proprioception as measured by the robotic task (63.16% vs max. 36.84%). This result is in agreement with literature, as it has been shown that upper limb SSEPs abnormalities occur in about half of pwMS who have no sensory symptoms (Chiappa and Ropper, 1982), and the overall incidence of SSEPs abnormalities has been reported to be up to 80% (Walsh, 2005). Indeed, SSEPs can capture demyelination occurring within the central fibres of the dorsal column or in the brain, which is not necessarily linked to somatosensory symptoms (Aminoff, 1984). Hence SSEPs can be seen as a measure describing the overall integrity of the sensory system, rather than a specific somatosensory deficit. Therefore, behavioural measures, such as the proposed robotic task, and neurophysiology complement each other and potentially need to be used together to provide a full picture of MS disease progression.

The robotic system was found clinically usable, as the average SUS score of 73–75 is above the previously defined usability threshold of 68 (Lewis and Sauro, 2018). This result is comparable to another study

Table 2
Test-retest reliability of the gauge position matching task in pwMS.

	AE	CE	VE	E
ICC (CI)	0.69 (0.59–0.76)	0.78 (0.71–0.82)	0.87 (0.83–0.90)	0.73 (0.64–0.79)
SRD (deg)	6.94	10.83	3.11	6.78
SRD (%)	28.12	23.60	12.03	25.14
LE within	2.20	2.35	0.89	1.72
LE betw.	−7.26	1.99	−1.64	−6.85

Legend: $N = 86$ (both hands for 43 pwMS). AE-Absolute Error, CE-Constant Error, VE-Variable Error, E-Total Variability, ICC-intraclass correlation coefficient (A,k), CI-confidence interval, SRD-smallest real difference, LE within-learning effect within one measurement day (session 3 - session 1), LE between-learning effect between days (mean across 3 sessions on test and retest).



(a) Test-retest reliability across all measurement sessions **(b)** Test-retest reliability across two measurement days

Fig. 2. Test-retest reliability of the gauge position matching task absolute error (AE), as an example. Similar results were obtained for the three other proprioception assessment metrics. The grey points represent one person with MS and the red points represent the mean across all subjects for each measurement session. Almost a straight red line can be seen within one day indicating high reliability. Abbreviations: D-day, S-session. Test is the mean across all 3 sessions on day 1 and Retest is the mean across all 3 sessions on day 2.

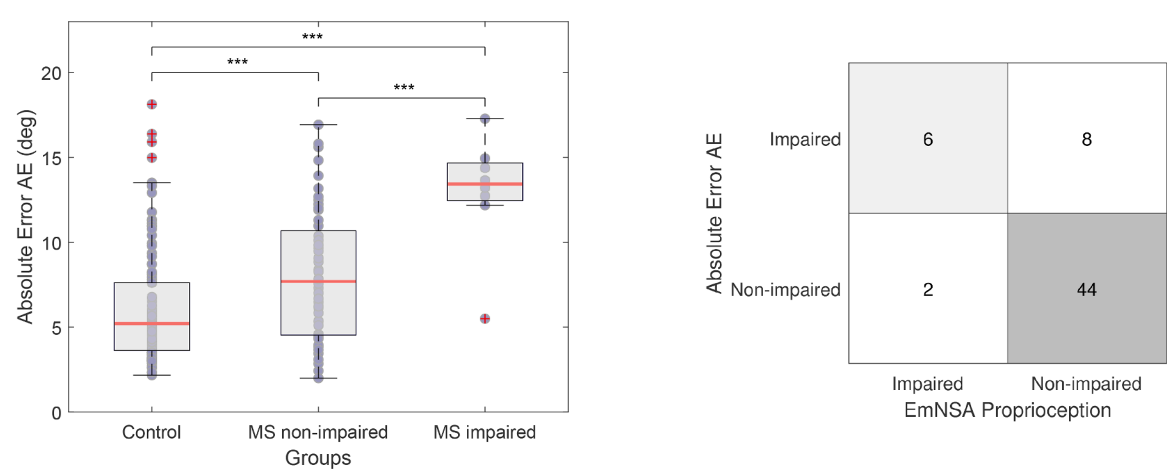
Table 3
Discriminant validity of robotic metrics in pwMS.

	AUC	% impaired	% agreement
AE	0.73	33.33	85.00
CE	0.43	13.33	88.33
VE	0.78	33.33	86.67
E	0.75	36.67	86.67
EmNSA	–	20.00	100.00
SSEPs lat.	–	63.16	-

Legend: $N = 118$ control subjects, $N = 60$ robotic task and EmNSA, $N = 38$ SSEP lat. For each measure, for each subject two data points were considered, corresponding to left and right hand. AUC- Area Under the Curve,%impaired- subjects with left or right hand impaired,% agreement- classification agreement with EmNSA proprioception, EmNSA-Erasmus MC modification Nottingham Sensory Assessment (proprioception), SSEPs lat. - Somatosensory Evoked Potentials latency.

evaluating technology-based training system in pwMS (73.75–77.50) (Knippenberg et al., 2021). The SUS score increased on retest, which means that familiarization might be needed until participants feel comfortable performing robotic assessments.

Some limitations of this study need to be acknowledged. SSEPs were not specifically conducted for the purpose of this study, hence also the exact timing between the robotic measurement and when SSEPs were collected was not matching, which limits their comparability. Further, the control group was on average older than pwMS group, while it has been shown that proprioceptive acuity might decrease with age (Rinderknecht et al., 2017). Therefore, it could be that the impairment threshold is higher than it would have been in an age-matched control group, leading to a lower number of pwMS being classified as impaired according to the robotic proprioception assessment. Finally, the robotic method assesses the index finger only, and it is not yet clear to what extent those results generalize to the whole hand or upper limb. Nevertheless, the index finger MCP joint is relevant in many ADLs and evaluating only one degree of freedom simplifies the robotic technology, increasing its clinical applicability (Zbytniewska et al., 2021; Zbytniewska et al., 2023).



(a) Discriminant validity **(b)** Impairment classification matrix

Fig. 3. Validity of the gauge position matching task absolute error (AE), as an example. Figure a) shows AE for three groups control subjects ($N = 118$), pwMS classified as impaired ($N = 52$) and as non-impaired ($N = 8$) according to EmNSA proprioception. Figure b) depicts an impairment classification matrix - number of pwMS classified as impaired / non-impaired according to AE / EmNSA proprioception. For each subject, both hands were considered together in the figures.

Table 4
Concurrent validity of the robotic proprioception assessment metrics.

	AE	CE	VE	E
NHPT	0.02	−0.14	0.03	0.04
BBT	−0.25	0.07	−0.25	−0.27
EmNSA	−0.24	0.02	−0.33	−0.25
EmNSA prop.	−0.40*	−0.07	−0.53**	−0.42**
SWM thumb	0.26	−0.28	0.19	0.29
SWM finger	0.13	−0.22	0.13	0.15
RSTF ulnar	0.03	0.25	−0.07	0.00
RSTF index	−0.01	0.11	−0.14	−0.02
SSEPs lat.	−0.02	−0.04	0.09	0.00
SSEPs amp.	−0.14	0.27	−0.17	−0.14

Legend: $N = 60$ for clinical assessments, $N = 38$ for SSEPs. NHPT-Nine Hole Peg Test, BBT-Box & Block Test, EmNSAErasmus MC modification Nottingham Sensory Assessment (total), EmNSA prop.-proprioception subscale of EmNSA, SWM-Semmes Weinstein Monofilaments, RSTF-Rydel Seiffer Tuning Fork, SSEPs-Somatosensory Evoked Potentials, lat.-latency, amp.-amplitude. Statistical significance: p-val. < 0.05:*, p-val. < 0.01:**.

5 Conclusions

The proposed robot-assisted assessment is reliable, valid and clinically usable in pwMS. Due to its satisfying reliability, the task can be utilized in the future for regular monitoring of proprioceptive impairments, e.g., in response to targeted therapies. The proposed assessment is specific to index finger proprioception and it is not correlated with other modalities of somatosensation. Due to its high sensitivity, it can spot subtle proprioceptive deficits, previously undetectable by conventional methods. Overall, the presented assessment is a promising complement to commonly used clinical methods and will likely contribute to a better understanding of proprioceptive impairments in pwMS, which could positively influence future choices of therapies.

Data availability

The data presented in this manuscript are available upon reasonable request and under consideration of the ethical regulations.

Research ethics and patient consent

This study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki. All subjects gave written informed consent before participating in the experiment. This study was approved by the committee of medical ethics CME2017/748 of the University of Hasselt and the Noorderhart Rehabilitation and MS Centre (Belgium) and by the ETH Ethics Committee EK 2019-N-108 (Switzerland).

CRediT authorship contribution statement

Monika Zbytniewska-Mégret: Methodology, Formal analysis, Software, Visualization, Writing – original draft. **Christoph M. Kanzler:** Methodology, Formal analysis, Validation, Writing – review & editing. **Joke Raats:** Methodology, Data curation, Project administration, Investigation, Writing – review & editing. **Cigdem Yilmazer:** Methodology, Data curation, Project administration, Investigation, Writing – review & editing. **Peter Feys:** Conceptualization, Resources, Supervision, Writing – review & editing. **Roger Gassert:** Conceptualization, Methodology, Resources, Funding acquisition, Supervision. **Olivier Lambercy:** Conceptualization, Methodology, Resources, Supervision, Funding acquisition, Validation, Writing – review & editing. **Ilse Lamers:** Conceptualization, Methodology, Data curation, Investigation, Project administration, Supervision, Validation, Writing – review & editing.

Declaration of Competing Interests

The Authors declare that there is no conflict of interest.

Acknowledgments

The authors would like to thank master students of Hasselt University involved in the data collection Lore Schildermans, Suzanne van Kooij, Laura Verwaest and Jasmien Hooybergs. The research was partially conducted at the Future Health Technologies programme which was established collaboratively between ETH Zurich and the National Research Foundation Singapore.

Funding

This work was supported by the Swiss National Science Foundation, project 320030L_170163 and by the ETH Zurich Foundation in collaboration with Hocoma AG. Ilse Lamers has received teaching honoraria from Sanofi Genzyme Europe. This research was supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

References

Aminoff, M.J., 1984. The clinical role of somatosensory evoked potential studies: a critical appraisal. Muscle Nerve 7 (5), 345–354. https://doi.org/10.1002/mus.880070502.

Bertoni, R., Lamers, I., Chen, C.C., et al., 2015. Unilateral and bilateral upper limb dysfunction at body functions, activity and participation levels in people with multiple sclerosis. Mult. Scler. J. 21 (12), 1566–1574. https://doi.org/10.1177/1352458514567553.

Brooke, J., 1996. SUS: a 'quick and dirty' usability scale. *Usability Evaluation in Industry*. July. CRC Press, pp. 207–212 eBook ISBN: 9780429157011.

Brown, K., Lohse, K., Mayer, I., et al., 2017. The reliability of commonly used electrophysiology measures. Brain Stimul. 10 (6), 1102–1111. https://doi.org/10.1016/j.brs.2017.07.011.

Chiappa, K.H., Ropper, A.H., 1982. Evoked potentials in clinical medicine. N. Engl. J. Med. 306 (20), 1205–1211. https://doi.org/10.1056/NEJM198205133061904.

Feys, P., Lamers, I., Francis, G., et al., 2017. The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis. Multi. Scler. J. 23 (5), 711–720. https://doi.org/10.1177/1352458517690824.

Gregson, J.M., Leathley, M., Moore, A., et al., 1999. Reliability of the tone assessment scale and the modified ashworth scale as clinical tools for assessing poststroke spasticity. Arch. Phys. Med. Rehabil. 80 (9), 1013–1016. https://doi.org/10.1016/S0003-9993(99)90053-9.

Hartung, H.P., Graf, J., Aktas, O., et al., 2019. Diagnosis of multiple sclerosis: revisions of the McDonald criteria 2017 – continuity and change. Curr. Opin. Neurol. 32 (3), 327–337. https://doi.org/10.1097/WCO.0000000000000699.

Hooper, J., Taylor, R., Pentland, B., et al., 1998. Rater reliability of Fahn's tremor rating scale in patients with multiple sclerosis. Arch. Phys. Med. Rehabil. 79 (9), 1076–1079. https://doi.org/10.1016/s0003-9993(98)90174-5.

Ingemanson, M.L., Rowe, J.R., Chan, V., et al., 2019. Neural correlates of passive position finger sense after stroke. Neurorehabil. Neural Repair 33 (9), 740–750. https://doi.org/10.1177/1545968319862556.

ISO 9241-11, 2018. (en), Ergonomics of human-system interaction — Part 11: usability: definitions and concepts. Standard. Int. Org. Standardiz., Geneva, CH (March 2018). https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en.

Kanzler, C.M., Rinderknecht, M.D., Schwarz, A., et al., 2020. A datadriven framework for selecting and validating digital health metrics: use-case in neurological sensorimotor impairments. npj Digit. Med. 3 (1), 80. https://doi.org/10.1038/s41746-020-0286-7.

Kister, I., Bacon, T.E., Chamot, E., et al., 2013. Natural history of multiple sclerosis symptoms. Int. J. MS Care 15 (3), 146–156. https://doi.org/10.7224/1537-2073.2012-053.

Knippenberg, E., Lamers, I., Timmermans, A., et al., 2021. Motivation, usability, and credibility of an intelligent activity-based client-centred training system to improve functional performance in neurological rehabilitation: an exploratory cohort study. Int. J. Environ. Res. Public Health 2021 18 (14), 7641. https://doi.org/10.3390/ijerph18147641.

Koo, T.K., Li, M.Y, 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J. Chiropr. Med. 15 (2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012.

Kurtzke, J.F., 1983. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 33, 1444. https://doi.org/10.1212/WNL.33.11.1444, 1444.

Lewis, J.R., Sauro, J., 2018. Item benchmarks for the system usability scale. J Usabil. Stud. 13, 158–167. https://dl.acm.org/doi/abs/10.5555/3294033.3294037.

Lexell, J.E., Downham, D.Y., 2005. How to assess the reliability of measurements in rehabilitation. Am. J. Phys. Med. Rehabil. 84 (9), 719–723. https://doi.org/10.1097/01.phm.0000176452.17771.20.

Lincoln, N., Crow, J., Jackson, J., et al., 1991. The unreliability of sensory assessments. Clin. Rehabil. 5 (4), 273–282. https://doi.org/10.1177/026921559100500403.

Mathiowetz, V., Volland, G., Kashman, N., et al., 1985. Adult norms for the box and block test of manual dexterity. Am. J. Occup. Ther. 39 (6), 386–391. https://doi.org/10.5014/ajot.39.6.386.

Miall, R.C., Kitchen, N.M., Nam, S.H., et al., 2018. Proprioceptive loss and the perception, control and learning of arm movements in humans: evidence from sensory neuronopathy. Exp. Brain Res. 236 (8), 2137–2155. https://doi.org/10.1007/s00221-018-5289-0.

Panosyan, F.B., Mountain, J.M., Reilly, M.M., et al., 2016. Rydel-Seiffer fork revisited: beyond a simple case of black and white. NeurologyNeurology 87 (7), 738–740. https://doi.org/10.1212/WNL.0000000000002991.

Prinsen, C.A.C., Mokkink, L.B., Bouter, L.M., et al., 2018. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual. Life Res. 27 (5), 1147–1157. https://doi.org/10.1007/s11136-018-1798-3.

Rinderknecht, M.D., Popp, W.L., Lambercy, O., et al., 2016. Reliable and rapid robotic assessment of wrist proprioception using a gauge position matching paradigm. Front. Hum. Neurosci. 10, 316. https://doi.org/10.3389/fnhum.2016.00316.

Rinderknecht, M.D., Lambercy, O., Raible, V., et al., 2017. Age-based model for metacarpophalangeal joint proprioception in elderly. Clin. Interv. Aging 12, 635–643. https://doi.org/10.2147/CIA.S129601.

Rinderknecht, M.D., Lambercy, O., Raible, V., et al., 2018. Reliability, validity, and clinical feasibility of a rapid and objective assessment of post-stroke deficits in hand proprioception. J. Neuroeng. Rehabil. 15 (1), 47. https://doi.org/10.1186/s12984-018-0387-6.

Schober, P., Boer, C., Schwarte, L.A., 2018. Correlation Coefficients. Anesthes. Analges. 126 (5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864.

Schwarz, A., Kanzler, C.M., Lambercy, O., et al., 2019. Systematic review on kinematic assessments of upper limb movements after stroke. Stroke 50 (3), 718–727. https://doi.org/10.1161/STROKEAHA.118.023531.

Shirota, C., Balasubramanian, S., Melendez-Calderon, A., 2019. Technology-aided assessments of sensorimotor function: current use, barriers and future directions in the view of different stakeholders. J. Neuroeng. Rehabil. 16 (1), 53. https://doi.org/10.1186/s12984-019-0519-7.

Simmatis, L.E., Jin, A.Y., Taylor, S.W., et al., 2020. The feasibility of assessing cognitive and motor function in multiple sclerosis patients using robotics. Mult. Scler. J. Exp. Transl. Clin. 6 (4), 205521732096494 https://doi.org/10.1177/2055217320964940.

Stolk-Hornsveld, F., Crow, J.L., Hendriks, E.P., et al., 2006. The Erasmus MC modifications to the (revised) Nottingham Sensory Assessment: a reliable somatosensory assessment measure for patients with intracranial disorders. Clin. Rehabil. 20 (2), 160–172. https://doi.org/10.1191/0269215506cr932oa.

Tracey, E.H., Greene, A.J., Doty, R.L., 2012. Optimizing reliability and sensitivity of Semmes–Weinstein monofilaments for establishing point tactile thresholds. Physiol. Behav. 105 (4), 982–986. https://doi.org/10.1016/j.physbeh.2011.11.002.

Wallin, M.T., Culpepper, W.J., Nichols, E., et al., 2019. Global, regional, and national burden of multiple sclerosis 1990–2016: a systematic analysis for the global burden of disease study 2016. Lancet Neurol. 18 (3), 269–285. https://doi.org/10.1016/S1474-4422(18)30443-5.

Walsh, P., 2005. The clinical role of evoked potentials. J. Neurol., Neurosurg. Psychiatry 76 (suppl 2), ii16–ii22. https://doi.org/10.1136/jnnp.2005.068130.

Zbytniewska, M., Kanzler, C.M., Jordan, L., et al., 2021. Reliable and valid robot-assisted assessments of hand proprioceptive, motor and sensorimotor impairments after stroke. J. Neuroeng. Rehabil. 2021 18:1 18 (1), 1–20. https://doi.org/10.1186/s12984-021-00904-5.

Zbytniewska, M., Rinderknecht, M.D., Lambercy, O., et al., 2023. Design and characterization of a robotic device for the assessment of hand proprioceptive, motor, and sensorimotor impairments. In: 2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR). IEEE, pp. 441–446. https://doi.org/10.1109/ICORR.2019.8779507. ISBN 978-1-7281-2755-2.