

Accuracy and Reliability of Haptic Spasticity Assessment Using HESS (Haptic Elbow Spasticity Simulator)

Jonghyun Kim, PhD, Hyung-Soon Park*, PhD, *Member, IEEE* and Diane L. Damiano, PhD, PT

Abstract— Clinical assessment of spasticity tends to be subjective because of the nature of the in-person assessment; severity of spasticity is judged based on the muscle tone felt by a clinician during manual manipulation of a patient's limb. As an attempt to standardize the clinical assessment of spasticity, we developed HESS (Haptic Elbow Spasticity Simulator), a programmable robotic system that can provide accurate and consistent haptic responses of spasticity and thus can be used as a training tool for clinicians. The aim of this study is to evaluate the accuracy and reliability of the recreated haptic responses. Based on clinical data collected from children with cerebral palsy, four levels of elbow spasticity (1, 1+, 2, and 3 in the Modified Ashworth Scale [MAS]) were recreated by HESS. Seven experienced clinicians manipulated HESS to score the recreated haptic responses. The accuracy of the recreation was assessed by the percent agreement between intended and determined MAS scores. The inter-rater reliability among the clinicians was analyzed by using Fleiss's kappa. In addition, the level of realism with the recreation was evaluated by a questionnaire on "how realistic" this felt in a qualitative way. The percent agreement was high ($85.7 \pm 11.7\%$), and for inter-rater reliability, there was substantial agreement ($\kappa=0.646$) among the seven clinicians. The level of realism was 7.71 ± 0.95 out of 10. These results show that the haptic recreation of spasticity by HESS has the potential to be used as a training tool for standardizing and enhancing reliability of clinical assessment.

I. INTRODUCTION

SPASTICITY is a heightened, velocity-dependent stretch reflex commonly found post brain injury such as stroke, traumatic brain injury, and cerebral palsy (CP) [1]. Clinical assessment of spasticity is important in determining treatment options and monitoring progression of rehabilitation.

The severity of spasticity is typically assessed by clinicians who manually manipulate the patient's limb at different speeds. By the nature of the in-person assessment, the result of the assessment tends to be subjective; several studies have reported insufficient agreement in the clinical scales such as Ashworth Scale [2], and Modified Ashworth Scale (MAS) [3-5]. The inter-rater agreement may be enhanced by improving the quality and amount of training that clinicians

receive [6]; however, having enough clinical training with real patients is challenging because 1) the need to recruit many patients with diverse severity of spasticity; 2) patient responses may vary within a session or across sessions and examiners; and 3) multiple repetitions may be fatiguing to patients.

As a solution to these challenges, programmable robotic systems have been proposed as training tools [7-11]. If an accurate and reliable haptic recreation implemented on a haptic device displays realistic responses from the patients, the robotic tool can be an appropriate substitute removing the need to having real patients during the training.

There have been a few studies that developed robotic devices for a similar training purpose [7-10]. For elbow spasticity, the upper limb patient simulator [7] and haptic simulator [9] were developed. A leg-robot was developed for displaying ankle clonus, a symptom of ankle spasticity [10]. Another device simulates contracture in the hand for training hand stretching [8]. If those haptic systems are to be applied for training clinicians, their accuracy and reliability need to be verified; however, there is a paucity of data evaluating these systems.

Recently, we have developed the Haptic Elbow Spasticity Simulator (HESS), a robotic system that recreates responses from the spastic elbow joint based on novel mathematical modeling that stems from clinical data collected [11].

This paper aims to evaluate the accuracy and reliability of haptic recreations implemented on HESS with a computer model of spasticity. From the clinical data collected from children with CP with elbow spasticity, we have modeled and implemented four levels (1, 1+, 2, and 3) of the MAS, the most widely used clinical measure for assessing spasticity [12]. The four models were programmed into HESS and seven experienced clinicians assessed the models by manipulating HESS. The percent agreement, inter-rater reliability, and level of realism were tested to assess the feasibility of HESS as a training tool.

II. HAPTIC ELBOW SPASTICITY SIMULATOR (HESS)

HESS, which consists of a haptic device and a control scheme, was developed to provide an accurate and readily available training opportunity for clinicians. A forearm manipulated by a brushless DC motor (Barrett Technology Inc., Cambridge MA) constituted the haptic device (Fig. 1). The mannequin forearm and hand were designed based on anthropometric data [13]. For the programmable recreation of elbow spasticity, a model-based control scheme was proposed

Manuscript received April 15, 2011. This work was supported in part by Intramural Research Program of the NIH, Clinical Center, protocol number 90-CC-0168.

Jonghyun Kim is with the Rehabilitation Medicine Department, Clinical Center, National Institutes of Health, Bethesda, MD 20892 USA.

Hyung-Soon Park is with the Rehabilitation Medicine Department, Clinical Center, National Institutes of Health, Bethesda, MD 20892 USA. (corresponding author to provide phone: 301-451-7533; fax: 301-451-7536; e-mail: parkhs@cc.nih.gov)

Diane L. Damiano is with the Rehabilitation Medicine Department, Clinical Center, National Institutes of Health, Bethesda, MD 20892 USA.

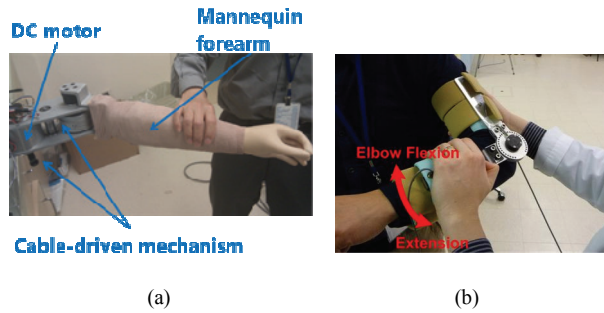


Figure 1. (a) Haptic Elbow Spastic Simulator (HESS) and (b) In-person assessment using a manual spasticity evaluator [11]

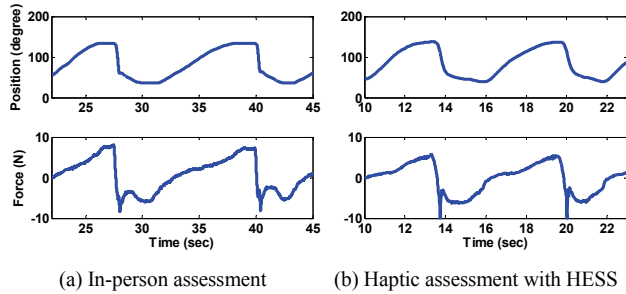


Figure 2. Comparison of position and force profiles measured [11]

by using the experimental data (position, velocity and force) collected from children with CP during the in-person clinical assessment of spasticity [11]. In the clinical assessment, a prototype manual spasticity evaluator equipped with a digital encoder and a force sensor was aligned with the subject's elbow joint. A clinician held the subject's upper arm and forearm while the subject was asked to relax, and moved quickly the forearm 2 to 5 times within the elbow range of motion (ROM) until the clinician determined MAS scores (Fig. 1b). Position, velocity and force data were sampled at 1 kHz by using NI-PCIe-6321 board with a Labview program (National Instruments., Austin, TX).

In our control scheme, elbow spasticity was implemented by dividing its movement into three phases: pre-catch, catch, and post-catch [11]. Note that the catch, defined as a sudden appearance of increased resistance during the fast passive movement [14], is a typical symptom of spasticity. We found that the following four parameters of the control scheme are closely correlated to the severity of spasticity recreated by using HESS [11]:

- 1) L is a time constant that determines how early the catch occurs; when L is smaller, the catch occurs earlier under same stretching speed; for faster speed, the catch occurs earlier for the same L . L works in between the pre-catch phase and the catch phase.
- 2) H is a torque constant that determines the peak torque at catch; when H is larger, the peak torque at catch is larger under same stretching speed; for faster speed, the peak torque at catch is larger for the same H . H works in the catch phase.
- 3) Q is a torque constant that determines the amount of the residual torque after peak torque at catch; when Q is larger,

the residual torque after the peak torque is larger. Q works in the catch phase.

- 4) D is a time constant that relates the average stretching speed to the time duration of the catch phase; when D is smaller, the catch phase ends earlier under same stretching speed; for faster speed, the catch phase ends earlier for the same D . D works in between the catch phase and the post-catch phase.

By using the control scheme with the programmable four parameters (L , H , Q , and D), HESS provides a haptic recreation (HR) for various levels of severity [11]. In Fig. 2, position and force measured during the haptic assessment with HESS were compared with those during the in-person assessment. In addition to the similar position and force profiles (Fig. 2), clinicians reported that the two assessments felt similar [11].

III. METHODS

During the clinical assessments mentioned in Section II, the clinicians determined MAS scores based on the muscle tone they felt. Six children with CP (mean age: 12.5 ± 4.1) were enrolled, and MAS scores were rated by two experienced physical therapists (PT). All guardians of the children gave written informed consent approved by the National Institutes of Health (NIH) IRB. The MAS scores rated are summarized in Table I. Among the subjects' scores, we chose three that the two raters reached agreement on to build representative (or standardized) HR #1, #2, and #3 in Table I.

The HRs for three MAS scores (1, 1+, and 2) were implemented by using different sets of parameters (L , H , Q , and D), respectively. In addition, although the two raters did not have an agreement, we used the subject #2's data (clinician #1 rated MAS 3) to implement HR #4 for MAS 3. The four sets represent different MAS scores (Table II). Note that MAS 0 and 4 were not considered in this study because these levels are obvious to distinguish from others.

TABLE I. MODIFIED ASHWORTH SCLAE OBTAINED FROM IN-PERSON ASSESSMENT

	Subject #1	#2	#3	#4	#5	#6
Rater 1	1	3	1+	2	1+	1+
Rater 2	1	2	1+	2	1+	1
choice	HR #1	HR #4	HR #2	HR #3	HR #5	HR #6

TABLE II. FOUR SETS OF PARAMETERS REPRESENTING MAS 1, 1+, 2, AND 3

	L	H	Q	D
HR #1 (for MAS 1)	2500	1.4	0.15	60
HR #2 (for MAS 1+)	1500	2.0	0.3	50
HR #3 (for MAS 2)	1000	2.8	0.6	30
HR #4 (for MAS 3)	300	3.8	0.8	10

In order to evaluate the four representative HRs, seven experienced clinicians (2 PTs who performed the assessment with the subjects in Table I, and five additional clinicians who are experienced with MAS) participated in an experiment on haptic assessment using HESS. Each clinician manipulated HESS 3 to 7 times, similar to the in-person assessment, and determined MAS scores based on the feel from the HRs. Prior to the experiment, all clinicians read the written MAS scoring instructions in the Appendix. In addition to the HRs in Table II, twelve arbitrary HRs in between the four HRs were added so that the four representative HRs can randomly appear. Thus, clinicians scored sixteen trials each. After all trials, the clinicians scored level of realism (from 1 (worst) to 10 (best)) by comparing the haptic assessment with their prior experience with in-person assessment.

Through the experiment above, the MAS scores rated by the seven clinicians were analyzed to evaluate accuracy and reliability. The accuracy was evaluated using two methods. First, we compared the MAS scores rated by the two clinicians who performed both in-person and haptic assessments. The MAS scores from in-person and haptic assessments were compared. Second, for all seven clinicians, the percent agreement was calculated between intended MAS scores (Table I) and the determined MAS scores from the haptic assessment.

Since this paper aims to see whether the HRs have potential for standardizing the spasticity assessment, we focused on the inter-rater reliability rather than the intra-rater reliability. The inter-rater reliability was analyzed by using a well-known Fleiss's kappa statistics for multi-rater [15].

IV. RESULTS

1) Accuracy: In-person assessment vs. Haptic assessment

First, Table III shows the MAS scores that the two clinicians rated during the in-person assessment and the MAS scores they rated for the corresponding HRs. The MAS scores rated from the haptic assessment had 100% agreement with the in-person assessment.

Second, Table IV shows the MAS scores rated on each HRs. Six out of seven clinicians assigned correct (or intended) scores to HR #1 and #2 (85.7% accuracy), while five assigned correct (intended) scores to HR #3 (71.4%). For HR #3, all seven gave correct scores (100%). Overall, mean agreement for the four HRs was 85.7±11.7%. Even though the clinicians were not trained in using the device, the agreement shows the high accuracy of clinical responses with HESS.

2) Reliability: Fleiss's kappa

For inter-rater reliability, Fleiss's kappa was obtained from Table IV. The kappa value ($\kappa=0.646$) means the substantial agreement [16] of MAS score among the seven raters.

3) Level of realism: Questionnaire

Lastly, from the questionnaire, the mean score about the level of realism was 7.71±0.95 out of maximum of 10. This score shows that the feel perceived by clinicians was similar to the muscle tone they experienced in in-person assessment.

TABLE III. COMPARISON OF MAS SCORES RATED IN-PERSON AND HAPTIC ASSESSMENT

Intended MAS	Assessment	Rater #1*	Rater #2*
1	In-person (subject #1)	1	1
	Haptic (HR #1)	1	1
1+	In-person (subject #3)	1+	1+
	Haptic (HR #2)	1+	1+
2	In-person (subject #4)	2	2
	Haptic (HR #3)	2	2
3	In-person (subject #2)	3	2
	Haptic (HR #4)	3	2

* Rater #1 and #2 are same to rater #1 and #2 in Table I, respectively.

TABLE IV. MAS SCORES DETERMINED BY CLINICIANS IN EXPERIMENT

	HR #1 (MAS 1)	HR #2 (MAS 1+)	HR #3 (MAS 2)	HR #4 (MAS 3)
Rater #1*	1	1+	2	3
#2*	1	1+	2	2
#3	1	1+	2	3
#4	1+	1+	2	3
#5	1	1+	2	3
#6	1	1	2	2
#7	1	1+	2	3

* Rater #1 and #2 are same to rater #1 and #2 in Table I, respectively. (MAS #) denotes the intended MAS score

V. DISCUSSION AND CONCLUSION

After four HRs were implemented, our concern was how we could evaluate these HRs objectively. In order to use haptic assessment in training, the requirements of the HRs are twofold; 1) the implemented HRs should accurately mimic the real elbow joint from which the parameters for implementing HR are calculated; 2) the MAS scores made on the four HRs should be reliably scored across experienced clinicians. Although Fig. 2 showed that the position and force responses from the haptic assessment were close to the ones from the in-person assessment, the practical outcome from clinical assessment is the clinical scale, not the position/force profiles. This led us to compare the MAS scores, the most common clinical measure [12].

Generally, the accuracy is defined as 'degree of closeness to the absolute value'; however, it was hard to find a gold standard of MAS scores other than the written instruction. Clinicians might perceive the same instruction differently. For example, slight, more marked, or considerable increase in muscle tone (see Appendix) is not clear for us to implement it in HESS. Therefore, we assumed the gold standard as the four position/force responses that the clinicians had same (or similar) score on. In Section IV, the first accuracy test with the two clinicians showed 100% agreement between the haptic assessment and the in-person assessment, proving that the four HRs closely mimicked the respective in-person elbow joints assessments from which the parameters are calculated. In addition, most of the raters gave correct scores to the HRs (Table IV) which implies that the four HRs can represent the four MAS scores, respectively.

There have been inconsistent results on inter-rater reliability of MAS score on elbow spasticity by using kappa

statistics [3, 4, 17]. The agreements were reported as poor to moderate ($\kappa=0.16\sim0.42$) with four clinicians [4] and as moderate ($\kappa=0.52$) with three clinicians [3]. In contrast, a study reported very good agreement ($\kappa=0.868\sim0.892$) with two clinicians [17]. Note that all clinicians who participated in the existing studies had training session right before they tested [3, 4, 17]. The existing results on reliability seem to be controversial; they, however, show that the better agreement (the higher κ) was obtained from a smaller number of raters. Since substantial agreement ($\kappa=0.646$) was obtained from participating seven raters even without having training sessions, the result shows that HESS is promising as a training tool.

From Table II, one can see that higher MAS scores have smaller L and D , and larger H and Q . Due to these quantitative parameters, we were able to create twelve dummy trials in between the four HRs. It is noteworthy that the clinicians did not report any weird feeling (different from realistic muscle tone) during the experiment; however, the reliability on those twelve trials was not as good as the four HRs. This is understandable because those trials sit between two MAS scores which are harder to distinguish.

In contrast to the HR #1, #2 and #3, the set of parameters for HR #4, intended to represent MAS 3, was not from the clinical data that the two clinicians had agreement. Owing to this limitation, rater #2 did not agree with rater #1 on MAS scores of HR #4 (Table III), resulting in the least percent agreement. This also proves that clinicians may perceive the same instruction differently; rater #1 perceived HR #4 as 'considerable increase in muscle tone' whereas rater #2 perceived it as 'more marked increase' (see Appendix).

The level of realism rated by the clinicians was high enough that they thought the HRs mimicked the spastic elbow joint. It is understandable that the clinicians could not rate perfect level (10) because the HRs were more consistent than the patients. The consistency and the fact that we know what the clinicians feel from the HRs are important factors that HESS can be used as a training tool.

It was verified that HESS is able to implement accurate and reliable (objective) HRs for elbow spasticity. Hence, it will provide a more feasible training opportunity to clinicians for improving accuracy and reliability of clinical assessments.

APPENDIX

Modified Ashworth Scale (MAS) [18]

- 0 No increase in tone
- 1 Slight increase in muscle tone, manifested by a catch and release or minimal resistance at the end of the ROM when the affected part is moved in flexion or extension
- 1+ Slight increase in muscle tone, manifested by a catch, followed by minimal resistance throughout the remainder (less than half) of the ROM
- 2 More marked increase in muscle tone through most of the ROM, but affected part easily moved
- 3 Considerable increase in muscle tone, passive movement difficult

4 Affected part rigid in flexion or extension

ACKNOWLEDGMENT

The authors thank all subjects and clinicians volunteered for the study. This research was supported by Intramural Research Program of the NIH, Clinical Center, protocol number 90-CC-0168.

REFERENCES

- [1] J. B. Nielsen, C. Crone, and H. Hultborn, "The spinal pathophysiology of spasticity - from a basic science point of view," *Acta Physiologica*, vol. 189, pp. 171-180, Feb 2007.
- [2] K. S. Sunnerhagen, "Stop using the Ashworth scale for the assessment of spasticity," *Journal of Neurology Neurosurgery and Psychiatry*, vol. 81, pp. 2-2, Jan 10 2010.
- [3] J. Mehrholz, Y. Major, D. Meissner, S. Sandi-Gahun, R. Koch, and M. Pohl, "The influence of contractures and variation in measurement stretching velocity on the reliability of the Modified Ashworth Scale in patients with severe brain injury," *Clin Rehabil*, vol. 19, pp. 63-72, Jan 2005.
- [4] J. Mehrholz, K. Wagner, D. Meissner, K. Grundmann, C. Zange, R. Koch, and M. Pohl, "Reliability of the Modified Tardieu Scale and the Modified Ashworth Scale in adult patients with severe brain injury: a comparison study," *Clin Rehabil*, vol. 19, pp. 751-9, Oct 2005.
- [5] W. K. Yam and M. S. Leung, "Interrater reliability of Modified Ashworth Scale and Modified Tardieu Scale in children with spastic cerebral palsy," *J Child Neurol*, vol. 21, pp. 1031-5, Dec 2006.
- [6] K. Klingels, P. De Cock, G. Molenaers, K. Desloovere, C. Huenaerts, E. Jaspers, and H. Feys, "Upper limb motor and sensory impairments in children with hemiplegic cerebral palsy. Can they be measured reliably?," *Disabil Rehabil*, vol. 32, pp. 409-16, 2010.
- [7] T. Fujisawa, M. Takagi, Y. Takahashi, and K. Inoue, "Basic research on the upper limb patient simulator," in *IEEE Int. Conf. on Rehabilitation Robotics (ICORR)*, 2007, pp. 48-51.
- [8] T. Mouri, H. Kawasaki, Y. Nishimoto, T. Aoki, and Y. Ishigure, "Development of robot hand for therapist education/training on rehabilitation," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2007, pp. 2295-2300.
- [9] D. I. Grow, M. Wu, M. J. Locastro, S. K. Arora, A. J. Bastian, and A. M. Okamura, "Haptic simulation of elbow joint spasticity," in *IEEE Symp. on Haptic Interfaces for Virtual Environments and Teleoperator Systems*, 2008, pp. 475-476.
- [10] T. Kikuchi, K. Oda, and J. Furusho, "Leg-Robot for Demonstration of Spastic Movements of Brain-Injured Patients with Compact Magnetorheological Fluid Clutch," *Advanced Robotics*, vol. 24, pp. 671-686, 2010.
- [11] H.-S. Park, J. Kim, and D. L. Damiano, "Haptic Recreation of Elbow Spasticity," in *IEEE Int. Conf. on Rehabilitation Robotics (ICORR)*, 2011 (Accepted).
- [12] V. A. Scholtes, J. G. Becher, A. Beelen, and G. J. Lankhorst, "Clinical assessment of spasticity in children with cerebral palsy: a critical review of available instruments," *Dev Med Child Neurol*, vol. 48, pp. 64-73, Jan 2006.
- [13] D. A. Winter, *Biomechanics and Motor Control of Human Movement*, Fourth ed.: John Wiley & Sons, Inc., 2009.
- [14] N. H. Mayer, "Clinicophysiological concepts of spasticity and motor dysfunction in adults with an upper motoneuron lesion," *Muscle Nerve Suppl*, vol. 6, pp. S1-13, 1997.
- [15] J. L. Fleiss, "Measuring Nominal Scale Agreement among Many Raters," *Psychological Bulletin*, vol. 76, pp. 378-382, 1971.
- [16] J. R. Landis and G. G. Koch, "Measurement of Observer Agreement for Categorical Data," *Biometrics*, vol. 33, pp. 159-174, 1977.
- [17] T. Kaya, A. G. Karatepe, R. Gunaydin, A. Koc, and U. Altundal Ercan, "Inter-rater reliability of the Modified Ashworth Scale and modified Modified Ashworth Scale in assessing poststroke elbow flexor spasticity," *Int J Rehabil Res*, vol. 34, pp. 59-64, Mar 2011.
- [18] R. W. Bohannon and M. B. Smith, "Interrater reliability of a modified Ashworth scale of muscle spasticity," *Physical Therapy*, vol. 67, pp. 206-7, Feb 1987.