Published in final edited form as:

IEEE Trans Neural Syst Rehabil Eng. 2012 May ; 20(3): 361–370. doi:10.1109/TNSRE.2012.2195330.

Development of a Haptic Elbow Spasticity Simulator (HESS) for Improving Accuracy and Reliability of Clinical Assessment of Spasticity

Hyung-Soon Park [Member, IEEE], Jonghyun Kim, and Diane L. Damiano

National Institutes of Health, Clinical Center, Rehabilitation Medicine Department, Bethesda, MD 20892 USA (parkhs@cc.nih.gov)

Abstract

This paper presents the framework for developing a robotic system to improve accuracy and reliability of clinical assessment. Clinical assessment of spasticity tends to have poor reliability because of the nature of the in-person assessment. To improve accuracy and reliability of spasticity assessment, a haptic device, named the HESS (Haptic Elbow Spasticity Simulator) has been designed and constructed to recreate the clinical “feel” of elbow spasticity based on quantitative measurements. A mathematical model representing the spastic elbow joint was proposed based on clinical assessment using the Modified Ashworth Scale (MAS) and quantitative data (position, velocity, and torque) collected on subjects with elbow spasticity. Four haptic models (HMs) were created to represent the haptic feel of MAS 1, 1+, 2, and 3. The four HMs were assessed by experienced clinicians; three clinicians performed both in-person and haptic assessments, and had 100% agreement in MAS scores; and eight clinicians who were experienced with MAS assessed the four HMs without receiving any training prior to the test. Inter-rater reliability among the eight clinicians had substantial agreement ($\kappa = 0.626$). The eight clinicians also rated the level of realism (7.63 ± 0.92 out of 10) as compared to their experience with real patients.

Keywords

Elbow spasticity; haptic simulation; inter-rater reliability; modified Ashworth scale; spasticity assessment

I. Introduction

Spasticity is one of the most common and potentially disabling complications that affect individuals with neurological disorders. Accurate assessment of spasticity is important for designing optimal treatment plans, properly evaluating potential effects of treatment interventions or monitoring progression of recovery.

In clinical settings, spasticity is manually tested. Typically, clinician supports the proximal limb and then moves the patient's distal segment at one or more velocities while the patient is instructed to relax. The clinician judges the degree of spasticity based on resistance from patient's body. The Ashworth Scale, Modified Ashworth Scale (MAS), and Tardieu Scale are widely used clinical measures of spasticity; however, many studies report inconsistent

results on reliability of those clinical scales [1]–[5], demonstrating the need for improving the reliability by providing clinicians with structured training [6]–[10].

The major reasons for the inconsistent results are clinician and subject variability involved in spasticity testing. If different raters move at different speeds, the same patient may respond differently due to the velocity-dependent characteristics of the spasticity [11]–[14]. Patient factors affecting spasticity include the time of the day that they are tested, fatigue level, body posture and position [15], and level of anxiety or attention paid during the test. Raters can be directed to perform the test at similar speeds by using visual or auditory feedback [16]; however, it would still be difficult to control all of the patient factors.

Another reason for the poor reliability of spasticity assessment is likely to be the qualitative descriptions of these clinical scales. For example, in the MAS, terms such as “considerable (MAS 3), more marked (MAS 2), or slight increase (MAS 1 or 1+) in muscle tone” are used which are clearly qualitative, thus different scores can be assigned to the same feel of muscle tone if raters perceive it differently. For improving accuracy and reliability of spasticity assessment, it is important to let trainees feel spastic joints which would help them clearly distinguish among these levels while they are being taught the clinical scales; however, involving many patients for training is impractical because 1) the variability due to the patient factors mentioned above will still induce inconsistency; 2) it is difficult to continually recruit patients with diverse degrees of spasticity for this purpose; and 3) patients may experience fatigue or boredom with repetitive testing.

As a solution to these difficulties, other programmable robotic systems that simulate patients' responses have been developed as training tools. For elbow spasticity, the upper limb patient simulator reproduced stiffness or muscle tone using magnetorheological (MR) brake [17], [18]. Another haptic simulator implemented velocity dependence of reflex onset angle and variable stiffness/damping at different degrees of spasticity [19]. A leg robot was developed for demonstrating ankle clonus, a related symptom to ankle spasticity seen in neurological patients [20]. Another device simulates contracture in the hand for training hand stretching exercises [21]. If those haptic systems are to be used as training tools, their accuracy and reliability needs to be evaluated with respect to the existing clinical scales; however, there is a paucity of data on clinical evaluation of existing training devices.

This paper presents a framework to standardize clinical assessment of spasticity including development of a novel haptic device capable of simulating spasticity in the elbow joint, named the HESS (Haptic Elbow Spasticity Simulator), mathematical modeling of spastic elbow joints, and a paradigm for training clinicians in the MAS. A novel mathematical model representing various levels of elbow spasticity is proposed here based on clinical data collected from subjects with elbow spasticity. By using the proposed model, four different MAS scores are implemented on the HESS, and the accuracy and reliability are evaluated with experienced clinicians. A preliminary version of this work has been reported [22], [23], but this paper presents engineering details, more clinical data and complete haptic models enhanced from the previous work.

II. Mathematical Modeling of Elbow Spasticity

The purpose of developing a mathematical model is twofold: 1) to identify realistic quantitative parameters that have potential to be used as objective measures representing severity of spasticity and 2) to program these into the haptic device so that the device can be controlled to output accurate joint torques at given input conditions (joint velocity, and MAS score).

A. Clinical Data Collection

First, we accumulated a database of quantitative information (position, velocity, and resistance torque) each with corresponding MAS scores. If we are to build standardized mathematical models representing MAS instrument, it is necessary to collect quantitative data corresponding to some gold standard for MAS scoring; however, the MAS is known for poor inter-rater reliability [4], [24]–[28] and there is minimal information in the literature on quantitative data that are correlated to MAS scores. This led us to consider the quantitative data sets with which a majority of experienced clinicians rated the same MAS scores on as standard data sets. Therefore, our strategy of building standard mathematical models is 1) to build a database including quantitative data and corresponding MAS scores; 2) extract data sets that experienced clinicians have strong agreement on; 3) and use the selected data sets to build mathematical models.

Three clinicians examined nine children with cerebral palsy (CP). All guardians of the children (mean age: 13.3 ± 3.5) gave written informed consent approved by the National Institutes of Health IRB. In order to obtain quantitative data, we developed a manual spasticity evaluator (MSE) which can measure elbow joint angles by a digital encoder (0.09 deg of resolution), angular velocities (by numerical differentiation of position), and the force (or torque) exerted (or felt) by clinicians while they manipulate the subject's elbow joint [Fig. 1(a)]. Force was measured by a force transducer (MLP-100, Transducer Techniques Inc., Temecula, CA) placed between clinician's hand and subject's arm. Surface EMG sensors were attached to the skin over the biceps and triceps muscle bellies to record muscle activation during the test. Position, velocity, force, and EMG data were sampled at 1 kHz by using a NI-PCIE-6321 board with a custom Labview program (National Instruments Inc., Austin, TX). The subjects were seated upright with their shoulder positioned at 40° abduction and 40° flexion. After aligning the MSE device with the patient's elbow joint [Fig. 1(b)], the clinicians were asked to assess spasticity in a same manner as they would do in clinic except that they were asked to move the patient's forearm in a horizontal plane. Each performed both slow and fast extension of the elbow and determined the MAS score based on the written instructions (Appendix).

Among the nine subjects, data from four subjects were selected each representing MAS scores of 1, 1+, 2, and 3. The raters had at least 67% of agreement on those four subjects listed in Table I. MAS scores 0 and 4 are excluded because they are obvious from the instruction (Appendix): either no resistance felt (score 0) or no movement possible (score 4). Quantitative data including time courses of synchronized position, velocity, and resistance torque collected at each MAS scores were used to build mathematical models.

B. Modeling of Spastic Elbow Joint

There have been several previous attempts to build a model of spasticity, but the models reported are too simple to describe sophisticated responses [19]–[21], [29], purely descriptive [30], [31] or focused only on modeling the spastic catch [32], which is defined as a sudden appearance of increased muscle tone during the fast passive movement [33]. The modeling provided in this paper attempts to quantitatively describe the known characteristics of spasticity in the MAS instrument (Appendix) and the literature [29], [32], [35]. For an accurate and realistic spasticity model, we divided the spastic response during passive movement into three phases: pre-catch, catch, and post-catch (Fig. 2). Equations describing each phase were proposed based on the clinical data and the descriptions in the MAS instructions.

In the pre-catch phase, it is assumed that the resistance torque from the elbow joint can be regarded as a linear mass-spring-damper system

$$\tau_{\text{pre}} = m\ddot{\theta} = b\dot{\theta} + k\theta \quad (1)$$

where m denotes mass (inertia) of the forearm and hand; b damping coefficient; and k stiffness.

Previous studies applied (1) for representing nonspastic [34] or spastic elbow joints [29]. Moreover, the clinical data collected under slow stretch also matched with this equation. Since the catch may not occur during slow stretch, the whole data set collected under slow stretch can be considered as a pre-catch phase. For instance, in Fig. 3, the measured torque during the slow stretching assessment (Subject #3) and the reconstructed torque calculated from (1) and position (velocity, and acceleration) data match with small error (11.4% of average error).

The pre-catch phase ends with the occurrence of the catch at a specific joint angle which is called as catch angle or stretch reflex threshold [35]. Hence, determining the catch angle is essential in defining the end of the pre-catch phase. From the literature, stretching at a higher velocity typically evokes the catch sooner, and thus the catch angle is negatively proportional to the stretch velocity [35]. Moreover, the catch angle depends on the initial posture (or length of the muscle) [36]. Based on this information, the catch angle is determined as follows:

$$\theta_{\text{catch}} = L - \alpha\dot{\theta}_{\text{pre}} \quad (2)$$

where L denotes for the catch angle constant; α (> 0) the slope of the stretching velocity versus catch angle curve; and $\dot{\theta}_{\text{pre}}$ the average speed of stretch during the pre-catch phase.

In the catch phase, it was reported that an impulse-like function was a suitable form for modeling the sudden increase of muscle tone [32]. The magnitude of the impulse is greater for higher MAS scores, and is velocity-dependent. In addition, our data show that the ratio of residual torque to the peak torque is different for each individual and the ratio is greater for higher MAS scores. Considering all the above factors, the following equation describes the torque during the catch phase:

$$\begin{aligned} \tau_{\text{catch}} &= H\dot{\theta}_{\text{catch_init}}\delta(t) + \tau_{\text{pre_end}} \\ \text{with } \delta(t) &= \begin{cases} 1, & \text{if } t - t_{\text{catch_init}} < \Delta T_{\text{peak}} \\ Q(Q < 1), & \text{if } t - t_{\text{catch_init}} \geq \Delta T_{\text{peak}} \end{cases} \quad (3) \end{aligned}$$

where H denotes for catch torque constant that relates the stretching speed to the peak torque at catch; $\dot{\theta}_{\text{catch_init}}$ the stretching speed at the beginning of catch phase; $\tau_{\text{pre_end}}$ the torque at the end of pre-catch phase; Q (< 1) a constant scale factor that determines the amount of the residual torque after the peak torque (named as residual torque constant); $t_{\text{catch_int}}$ the time when the catch phase begins; and ΔT_{peak} the time duration that the impulsive peak torque at catch maintains (Fig. 4).

In addition to H and Q which determine the amount of peak and residual torque, the time duration of the catch phase needs to be determined which automatically defines the end of catch phase and the beginning of post-catch phase. From the clinical data collected, we observed that the time duration of the catch phase, the end of which is delineated by the rapid decrease in residual torque from its peak value that occurs either before the end of the movement or before a second smaller rise in resistance torque if the movement has not yet been completed, is inversely proportional to the stretch velocity which is modeled as follows:

$$\Delta T_c = \frac{D}{\bar{\omega}_c} \quad (4)$$

where D is named as the catch duration constant that determines time duration of the catch phase, and $\bar{\omega}_c$ the average stretching speed during the catch phase. Practically, (4) is implemented by calculating $\bar{\omega}_c(t)$ at every sampling time, and the catch phase ends when time spent for catch phase ($t - t_{\text{catch_init}}$) starts to become greater than $\Delta T_c(t)$. In the 12 trials performed at different stretching speeds for subject #3, $\Delta T_c \cdot \bar{\omega}_c$ remained at constant level which validates (4) (Fig. 5).

After the catch phase, assuming the limb movement has not yet been completed, there is a secondary increase in resistance torque which starts from the residual torque determined by Q in (3) from the catch phase. The rate of increase in the post-catch phase is slower than that in the catch phase. To our knowledge, no research has modeled this phenomenon; however, the secondary increase also is described in the MAS instruction. For example, the absence or presence of increased resistance after catch distinguishes MAS 1 and 1+, respectively (Appendix). From the data collected, we found that the secondary increase in resistance torque is position-dependent. Hence, the torque in the post-catch phase is described as follows:

$$\tau_{\text{post}} = k_{\text{post}}(\theta - \theta_{\text{post_i}}) + m\ddot{\theta} + b\dot{\theta} \quad (5)$$

where k_{post} denotes for the elbow stiffness in post-catch phase; $\theta_{\text{post_i}}$ the elbow joint angle at the beginning of the post-catch phase. Note that we do not need to determine the duration of the post-catch phase because post-catch phase ends if the speed of stretch is smaller than the speed threshold ($\dot{\theta}_m$), which occurs when the clinician stops the movement.

III. Design and Control of HESS

A. Design of HESS

HESS (Haptic Elbow Spasticity Simulator) has been designed to recreate the resistance that the clinicians would have felt during an in-person assessment [23]. The device consists of a mannequin forearm, brushless DC motor and controller (Barrett Technology Inc., Cambridge, MA), and a cable-driven speed-reducing mechanism (Fig. 6). The cable-driven mechanism consists of two stages of speed reduction, and the ratio at each stage is calculated by the ratio of diameters of input and output shafts.

Since the brushless DC motor has about ten times greater bandwidth as compared to the MR brake used in the other previously referenced patient simulator [18], [20], the device is able to implement fast responses such as catch in greater accuracy. The speed reducing transmission was implemented by using the cable-driven speed reducing mechanism [37]. It utilizes frictionless rolling contact of two adjacent pulleys driven by two steel cables that are pre-tensioned but do not stretch. Therefore, the speed reducing mechanism implements small friction and near zero-backlash which are desirable mechanical properties of haptic devices. The detailed mechanical specification of the HESS is provided in Table II.

B. Control of HESS

1) Implementation of the Spasticity Model—The mathematical model described by (1)–(5) is implemented in the control scheme illustrated in Fig. 7. In the pre-catch phase, the haptic device is controlled by (1) until the joint angle reaches the catch angle determined by

(2). Equation (2) yields unrealistic catch angles when $\dot{\theta}_{pre}$ is too small or too big. For practical implementation, we have modified the equation

$$\theta_{catch} = \begin{cases} \text{no catch,} & \text{if } \dot{\theta}_{pre} < v_L \\ L - \alpha \dot{\theta}_{pre}, & \text{if } v_L \leq \dot{\theta}_{pre} < v_H \\ \theta_L, & \text{if } v_H \leq \dot{\theta}_{pre} \end{cases} \quad (6)$$

θ_L represents the smallest feasible catch angle from the in-person assessment; whereas v_L and v_H denote the lower and the upper bounds of $\dot{\theta}_{pre}$ that the linear relationship in (2) holds. v_L and θ_L are determined from the data collected during the in-person assessment, $v_H = (L - \theta_L)/\alpha$ and from (2). From the in-person data collected from nine patients and the three clinicians, all measured $\dot{\theta}_{pre}$ were above $60^\circ/s$ and the smallest catch angle was 60° ; therefore v_L and θ_L was set to $60^\circ/s$ and 60° , respectively. These numbers, however, can vary depending on the data collected from the in-person assessment and should be adjusted accordingly. For example, if a clinician moved a patient's arm at a speed slower than $60^\circ/s$ and the catch was observed, v_L could be set to a smaller value.

If the joint angle becomes greater than θ_{catch} under a fast enough speed, the catch phase starts where the device is controlled by (3). The catch phase ends after ΔT_c which is calculated by (4), and post catch phase starts where the rater feels the secondary increase in torque determined by (5) throughout the remainder of ROM. All parameters in (1)–(5) are estimated from the position and torque data collected during the in-person assessment which are listed in Table III. Among the parameters, L , H , Q , and D had monotonic relationships with MAS scores which implies that those parameters can potentially be used for quantifying the degree of spasticity.

The control scheme was implemented under a real-time operating system, Xenomai, with 1 kHz sampling rate. The calculated torque commands from (1), (3), and (5) were sent to the motor amplifier (PUCK, Barrett Technology Inc., Cambridge, MA) under torque control mode (25 kHz sampling rate). Since the mannequin forearm has the average inertia and size of the subject's forearm, the inertial effect [$m \ddot{\theta}$ in (1) and (5)] was not programmed into the control commands. This avoids high-level noise due to numerical differentiation involved in the calculation of acceleration from the position measurement.

2) Compensation of Friction—Friction force often deteriorates the haptic feel and backdrivability is considered an important mechanical property of a haptic device. In order to verify the amount of friction torque that would be felt by raters, we have tested the backdrivability of HESS. The resistance torque felt by a rater was measured by a torque sensor (TRT-200, sensor stiffness 2300 Nm/rad, Transducer Technique Inc., Temecula CA) while the rater moved the mannequin arm and the motor command was set to zero. The maximum torque felt by the rater was greater than 0.5 Nm mainly due to the viscous and static friction [Fig. 8(a)]. In order to compensate for the friction torque, the viscous and the Coulomb friction coefficients of the following model were identified. Even though the motor controller controlled for current of the motor, there might be effect of uncompensated back EMF. The friction model also includes any effect due to uncompensated back EMF by the motor controller.

$$\tau_{ext} = I_{motor} \ddot{\theta} + B \dot{\theta} + \beta \text{sgn}(\dot{\theta}) \quad (7)$$

where τ_{ext} represents the torque measured during the backdrivability test, I_{motor} the inertia of the motor and the speed reduction mechanism, B the viscous damping coefficient, β and

the Coulomb friction coefficient. By using least square estimation [38], $I_{\text{motor}} = 0.0097$ (kg m²), $B = 0.168$ (Nm sec), and $\beta = 0.189$ (Nm). With viscous and static friction compensated, τ_{ext} measured during the backdrivability test was reduced below 0.2 Nm [Fig. 8(b)], which is small enough not to deteriorate the feel of the catch and residual resistance. In the real implementation, the output torque commands calculated from (1), (3), and (5) were updated after compensating the estimated friction torque in (7).

3) Safety Consideration—The flowchart in Fig. 7 assumes that raters will hold the device until the completion of the post-catch phase; however, the mannequin arm can slip out of their hands in the middle of the catch or post-catch phase if the resistance torque is too strong (e.g., MAS 3). When the device slips out of the operator's hand, it quickly accelerates opposite to the direction it has been moving. The loss of grasp was detected here by monitoring the velocity of the device. If a sudden change in velocity is detected, it is considered as loss of grip and the control mode is switched to the pre-catch phase to reduce resistance torque and thereby decrease the velocity.

C. Mechanical Behaviors of HESS

Mechanical behaviors of the four haptic models representing MAS 1, 1+, 2, and 3 are compared to those measured from in-person assessment (Fig. 9). Position was measured by the digital encoder installed along the motor shaft (Table II), and the force transducer (MLP-100, Transducer Techniques Inc., Temecula, CA) that is used for MSE device was installed between the rater's hand and the mannequin forearm to measure the torque felt by the raters. Since the raters cannot make same movement over trials, it is natural to see mismatching position and torque responses between the in-person and the haptic assessments. Note that the goal was to present mechanical behaviors that distinguish one MAS score from the others.

In the experimental results of MAS 1 [Fig. 9(a)], the peak torque is smaller (1 ~ 1.5 Nm) than the other MAS scores, and there is no post-catch phase so that the raters reach the joint limit within the catch phase. Therefore, the raters did not experience the secondary increase in resistance torque. However, in the other MAS scores [Fig. 9(b)–(d)], catch phases are brief and the resistance torque drops immediately after the catch followed by a secondary increase of resistance torque which is the post-catch phase. Higher MAS scores showed greater peak torques (related to H), smaller catch angles (related to L), shorter catch phase (related to D), and greater resistance torque at the end of post-catch phase. Note that the effect of parameter Q can be explained by (3). Parameter Q is described as the ratio of the decrease from peak torque to residual torque (local minimum after the catch) to the increase from torque at the end of pre-catch phase to the peak torque. Q is not seen directly from the graphs, but all four graphs show greater for Q higher MAS scores.

IV. Clinical Validation

A. Methods

The accuracy and reliability of the proposed HMs were validated by experienced clinicians. Eight experienced clinicians (six physical therapists and two physicians) including the three clinicians who collected clinical data shown in Table I participated in the validation process. All raters were experienced with using the MAS instrument from their previous institutions. They were provided with a written description of the scores but no additional training was provided during this study. In addition to the four models in Table III, 12 dummy models were added by using intermediate parameters of L , H , Q , and D to hide the identity of the four supposedly “true” models from the raters, similar to randomized “controls”.

Each rater came to the laboratory individually and was asked to manipulate HESS to give MAS scores to the HMs while the 16 models (including the four models in Table III and 12 dummy models) were presented in a randomized order. After reviewing MAS scoring instruction (Appendix) prior to the test, they moved the mannequin forearm as many times as they wanted; however, no training or instruction was provided regarding how much resistance corresponds to which MAS scores. They rated MAS scores based on their prior experience and the written MAS scoring instruction. After all trials, they scored level of realism [from 1 (worst) to 10 (best)] by comparing the haptic assessment with their prior experience with clinical assessments. For test–retest reliability after six months three of the eight clinicians tested the 16 HMs in a randomized order.

From the data collected, the accuracy of the four HMs was evaluated by using the MAS scores made by the three physical therapists who collected clinical data shown in Table I. The percent agreement between the MAS scores they made on in-person and haptic assessments were used to determine how closely the haptic models recreated the muscle tone of the four patients. The inter-rater reliability was evaluated from the data collected from all seven clinicians by using Fleiss's kappa statistic for multiple raters [39]. The test–retest reliability was evaluated from the data collected from three clinicians.

B. Results

First, the MAS scores that the three clinicians rated during the in-person and the haptic assessments had 100% agreement (Table IV). Table V shows the MAS scores rated on each haptic model (HM). Seven out of eight clinicians gave correct (or intended) scores to HM #1 (87.5% agreement), while six assigned correct (or intended) scores to HM #2 and #4 (75.0%). For HM #3, all eight gave correct scores (100%). Overall, mean agreement for the four HMs was $84.4 \pm 12.0\%$. The average stretching speed during pre-catch phase ($\dot{\theta}_{pre}$) was different for raters and trials (Table V).

Fleiss's kappa was obtained from Table IV to test inter-rater reliability. The kappa value ($\kappa = 0.626$) means that there was the substantial agreement [40] of MAS score among the eight raters. The inter-rater reliability of other 12 dummy models was poor ($\kappa = 0.141$). For test–retest reliability, each of the three raters had substantial ($\kappa = 0.771$) agreement on the four standard HMs while each of them had fair agreement ($\kappa = 0.232$) on the 12 dummy HMs. Lastly, from the questionnaire, the mean score regarding the level of realism of the device in comparison to a patient was 7.63 ± 0.92 out of a possible score of 10.

V. Discussion and Conclusion

This paper presents a framework for developing a robotic system which will help in standardizing the clinical assessment of spasticity. It involves 1) quantitative data collection from clinical assessment, 2) mathematical modeling of the clinical assessment, and 3) implementation and clinical evaluation of the model.

In order to improve the poor reliability of clinical spasticity assessment, there have been studies which attempted to quantify spasticity by correlating quantitative parameters to the clinical instruments [19], [20], [29]–[32], [35], [41]. It was reported that rate of change in resistance and the onset angle of stretch are closely related to the Ashworth Scale [41]. Most of the previous studies, however, modified the existing assessment protocol to obtain more consistent data [2], [29], [31], [35], [41] and/or tried to fit sophisticated phenomenon into single event [18]–[20].

The proposed mathematical model has advanced features which will ultimately enhance reliability of existing clinical assessment tools. First, it is a more precise and accurate model

such that the raters judged it to be realistic. By dividing a spastic response into three phases, the model could deal with sophisticated responses precisely and accurately. The score was notably high given the fact that the “dummy” simulations were also included. Second, the proposed model uses existing clinical assessment tools without any modification, while it can potentially improve reliability of them. Therefore, clinicians can still use the same clinical instruments that they have been using. Lastly, it accounts for human variability which is unavoidable in clinical practices. In a clinical assessment, clinicians control for the stretching speed which is variable throughout the ROM as well as across the raters. For example, clinicians tend to slow the stretching speed immediately after they feel the sudden increase of muscle tone and stretching speeds are different across raters as shown in Table V. The proposed model accounts for this variability and Table V show promising results that the same model received similar scores even though it was operated at different speeds.

Among the parameters used in the mathematical model, four parameters [L in (2), H and Q in (3), and D in (4)] seemed most closely related to the MAS scores. From Table II, one can see that higher MAS scores have smaller L and D , and larger H and Q . We were able to create 12 dummy trials in between the four gold standard models. All seven raters did not report any weird feeling (different from realistic muscle tone) during the experiment; however, the inter-rater reliability on those twelve trials was not as good as the four standard HMs which is understandable because those trials sit between two MAS scores would be harder to distinguish. This paper did not employ statistical methods for correlating the four parameters with the degree of spasticity. Future study will be carried out to see whether the parameters can be used as quantitative measures indicating the degree of spasticity when more rigor or quantification is desired.

After four haptic models (HMs) were implemented, our primary concern was how to evaluate accuracy and reliability of HMs objectively. Generally, accuracy is defined as “degree of closeness to the absolute value”; however, it was hard to find a gold standard data of MAS scores beyond the written instruction. Even expert clinicians might perceive the same instruction differently. Therefore, the standard models were assumed to be the data from four representative subjects that experienced clinicians rated the same (or similarly). More data should be collected in the future to propose more generalized standard models. For the additional data to be collected the modeling, implementation, and evaluation methods will likely remain the same as the framework presented in this paper.

After refining the gold standard, we could compare position/force data collected from in-person assessment and haptic assessment to evaluate accuracy; however, it was our thought here that the clinical outcome measures should be used to evaluate accuracy if we are to improve accuracy of existing clinical assessment. This led us to compare the MAS scores, the most widely used clinical measure.

The accuracy of HMs was evaluated in two ways: 1) by comparing MAS scores from in-person assessment and corresponding haptic assessment, and 2) by percent agreement to the intended MAS scores. The first method would indicate how close the HMs reconstructed responses of the corresponding patients, and the haptic assessment had 100% matching MAS scores (Table IV). The second method would show how close the intended MAS scores of HMs are to the MAS scores perceived by experienced clinicians. The result ($84.4 \pm 12\%$ agreement) and the survey results imply that the HMs implemented the haptic feeling that clinicians generally expect to feel.

As to the inter-rater reliability, there have been inconsistent results on inter-rater reliability of MAS score on elbow spasticity by using kappa statistics [2]–[4]. The agreements have been reported as poor to moderate ($\kappa = 0.16 \sim 0.42$) with four clinicians [4] and as moderate

($\kappa = 0.52$) with three clinicians [3]. In contrast, a study reported very good agreement ($\kappa = 0.868 \sim 0.892$) with two clinicians [2]. Note that all clinicians who participated in the previous studies had training session right before they tested which will significantly enhance reliability. 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. In our reliability test, substantial agreement ($\kappa = 0.626$) was obtained from participating eight raters even without training prior to the test. All raters were experienced with the MAS instrument prior to the test. They were not only completely blinded to the identity of four standard HMs but also confused by the 12 dummy HMs randomly placed in between the four HMs. It is noteworthy that without training and even with the impediment, the four HMs still had substantial agreement. This shows that most raters intuitively perceived the haptic feel from the four standard HMs as the intended feel described in MAS instruction, which implies that HESS will be promising as a training tool.

The HESS and haptic models can provide a novel clinical analysis and training strategy that provides consistent stimuli that should yield consistent responses. There are many studies that tested reliability of the MAS instrument; however, the existing results are affected by three sources of variability: 1) patient variability, 2) a qualitatively described instrument, and 3) rater variability. In order to analyze the rater variability, it is necessary to isolate the third source from the others. The haptic device removes the patient variability and the mathematical modeling quantifies the qualitatively described instrument which let us study rater variability alone. From our pilot results, the raters had substantial test-retest agreement for the four standard HMs ($\kappa = 0.771$), but fair agreement on the other 12 dummy HMs ($\kappa = 0.232$). This implies that there are cases where raters have greater variability, and those cases might not be described clearly by one MAS score. Designing appropriate training programs for these cases will be more efficient. For example, we plan to collect enough data to find cases that cause greater intra-and inter-rater variability, to determine which MAS scores to assign, and to train raters for those cases using HESS.

The framework presented in this paper can also be applied to simulate other types of mechanical behaviors that involve haptic interaction. Rigidity in Parkinsonism is one example. First, a haptic device resembling patients' body part needs to be designed. Mechanical specifications should be satisfied such as adequate joint ROM, minimal friction from the device, and sufficiently high torque/force levels so as to mimic the targeted mechanical behaviors, and realistic inertial and size properties of the device based on anthropometric data. Second, quantitative data (position, force/torque, EMG, and etc.) needs to be collected from patients along with the clinical judgments. Third, the mechanical behaviors need to be identified as programmable language (e.g., mathematical equations) based on the data and the instructions of the clinical assessment. Lastly, the target behaviors should be implemented in the device and validated in a clinical setup.

The level of realism rated by the clinicians was high enough that they thought the HMs mimicked the spastic elbow joint. It is understandable that the clinicians could not rate perfect level (score 10) because the HMs were not as variable as the patients. The consistency in and the precise knowledge of what the clinicians feel from the HMs are important factors for why the HESS can be used as a training tool. Hence, it provides a feasible training opportunity to clinicians for improving accuracy and reliability of clinical assessments.

Acknowledgments

The authors would like to thank all subjects and clinicians volunteered for the study. Specifically, the authors would like to thank Dr. L. Prosser and Dr. C. Zampieri-Gallagher for their valuable comments.

This work was supported in part by the intramural research program of the National Institutes of Health (protocol number 90-CC-0168) and in part by the Center for Neuroscience and Regenerative Medicine (G192HF).

Appendix

Modified Ashworth Scale (MAS) [42]

- 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.

Biographies

Hyung-Soon Park (M'05) received the Ph.D. degree in mechanical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, Korea, in 2004.

He is a staff scientist in the Rehabilitation Medicine Department, Clinical Center at the National Institutes of Health, Bethesda, MD. His current research interest focuses mainly on application of robotics and control technology on rehabilitation medicine, and biomechanics of human movement.

Jonghyun Kim (S'05) received the Ph.D. degree from the Department of Mechanical Engineering from Korea Advanced Institute of Science and Technology, Daejeon, Korea, in 2010.

Since 2010, he has been with the Rehabilitation Medicine Department, Clinical Center at the National Institutes of Health, Bethesda, MD, where he is currently a Visiting Fellow. His

active research interests include rehabilitation robotic systems, haptic display, and the design and control of bilateral teleoperation.

Diane L. Damiano received the Ph.D. degree in research methods/biomechanics from the University of Virginia, Charlottesville, in 1993.

She is Chief of the Functional and Applied Biomechanics Section in the Rehabilitation Medicine Department, Clinical Center at the National Institutes of Health, Bethesda, MD. Her area of expertise is in the investigation of both existing and novel activity-based rehabilitation approaches in children with cerebral palsy which has helped to revolutionize the treatment of these patients.

References

- [1]. Fleuren JF, et al. Stop using the Ashworth Scale for the assessment of spasticity. *J. Neurol. Neurosurg. Psychiatry*. Jan; 2010 81(no. 1):46–52. [PubMed: 19770162]
- [2]. Kaya T, et al. Inter-rater reliability of the Modified Ashworth Scale and modified Modified Ashworth Scale in assessing poststroke elbow flexor spasticity. *Int. J. Rehabil. Res.* Mar; 2011 34(no. 1):59–64. [PubMed: 20671560]
- [3]. Mehrholz J, et al. 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.* Jan; 2005 19(no. 1):63–72. [PubMed: 15704510]
- [4]. Mehrholz J, et al. Reliability of the modified Tardieu Scale and the modified Ashworth Scale in adult patients with severe brain injury: A comparison study. *Clin. Rehabil.* Oct; 2005 19(no. 7): 751–9. [PubMed: 16250194]
- [5]. Yam WK, Leung MS. Interrater reliability of modified Ashworth Scale and modified Tardieu Scale in children with spastic cerebral palsy. *J. Child Neurol.* Dec; 2006 21(no. 12):1031–5. [PubMed: 17156693]
- [6]. Gracies JM, et al. Reliability of the Tardieu Scale for assessing spasticity in children with cerebral palsy. *Arch. Phys. Med. Rehabil.* Mar; 2010 91(no. 3):421–8. [PubMed: 20298834]
- [7]. Haas BM, et al. The inter rater reliability of the original and of the modified Ashworth Scale for the assessment of spasticity in patients with spinal cord injury. *Spinal Cord.* Sep; 1996 34(no. 9): 560–4. [PubMed: 8883191]
- [8]. Klingels K, et al. Upper limb motor and sensory impairments in children with hemiplegic cerebral palsy. Can they be measured reliably? *Disabil. Rehabil.* 2010; 32(no. 5):409–16. [PubMed: 20095955]
- [9]. Pandyan AD, et al. A review of the properties and limitations of the Ashworth and modified Ashworth Scales as measures of spasticity. *Clin. Rehabil.* Oct; 1999 13(no. 5):373–83. [PubMed: 10498344]
- [10]. Scholtes VA, et al. Clinical assessment of spasticity in children with cerebral palsy: a critical review of available instruments. *Dev Med. Child Neurol.* Jan; 2006 48(no. 1):64–73. [PubMed: 16359597]
- [11]. Haugh AB, Pandyan AD, Johnson GR. A systematic review of the Tardieu Scale for the measurement of spasticity. *Disabil. Rehabil.* Aug; 2006 28(no. 15):899–907. [PubMed: 16861197]

- [12]. Lee HM, et al. Quantitative analysis of the velocity related pathophysiology of spasticity and rigidity in the elbow flexors. *J. Neurol. Neurosurg. Psychiatry*. May; 2002 72(no. 5):621–9. [PubMed: 11971049]
- [13]. Nielsen JB, Crone C, Hultborn H. The spinal pathophysiology of spasticity-from a basic science point of view. *Acta Physiologica*. Feb; 2007 189(no. 2):171–180. [PubMed: 17250567]
- [14]. Schmit BD, Rymer WZ. Identification of static and dynamic components of reflex sensitivity in spastic elbow flexors using a muscle activation model. *Ann. Biomed. Eng.* Apr; 2001 29(no. 4): 330–9. [PubMed: 11339330]
- [15]. Shumway-Cook, A.; Woollacott, MH. *Motor Control: Translating Research into Clinical Practice*. Lippincott Williams Wilkins; Philadelphia, PA: 2007. p. 115
- [16]. Wu YN, et al. Characterization of spasticity in cerebral palsy: Dependence of catch angle on velocity. *Dev. Med. Child Neurol.* Jun; 2010 52(no. 6):563–9. [PubMed: 20132137]
- [17]. Fujisawa, T., et al. *IEEE Int. Conf. Rehabil. Robot.* Noordwijk, The Netherlands: 2007. Basic research on the upper limb patient simulator; p. 48-51.
- [18]. Takhashi, Y., et al. *IEEE Int. Conf. Rehabil. Robot.* Zurich, Switzerland: 2011. Development of an upper limb patient simulator for physical therapy exercise; p. 1117-1120.
- [19]. Grow, DL., et al. *IEEE Symp. Haptic Interfaces Virtual Environ. Teleoperator Syst.* Reno, NV: Haptic simulation of elbow joint spasticity; p. 475-476.
- [20]. Kikuchi T, Oda K, Furusho J. Leg-robot for demonstration of spastic movements of brain-injured patients with compact magnetorheological fluid clutch. *Adv. Robot.* 2010; 24(no. 5–6):671–686.
- [21]. Mouri, T., et al. *IEEE/RSJ Int. Conf. Intell. Robots Syst.* San Diego, CA: Development of robot hand for therapist education/training on rehabilitation; p. 2295-2300.
- [22]. Kim, J.; Park, HS.; Damiano, DL. *IEEE Int. Conf. on Eng. Med. Biol. Soc.* Boston, MA: Accuracy and reliability of haptic spasticity assessment using HESS (Haptic Elbow Spasticity Simulator); p. 8527-8530.
- [23]. Park, H-S.; Kim, J.; Damiano, DL. *Proc. IEEE Int. Conf. Rehabil.* Zurich, Switzerland; 2011. Haptic recreation of elbow spasticity; p. 1-6.
- [24]. Blackburn M, Van Vliet P, Mockett SP. Reliability of measurements obtained with the modified Ashworth Scale in the lower extremities of people with stroke. *Phys. Ther.* Jan; 2002 82(no. 1): 25–34. [PubMed: 11784275]
- [25]. Clopton N, et al. Interrater and intrarater reliability of the modified Ashworth Scale in children with hypertonia. *Pediatr. Phys. Ther.* 2005; 17(no. 4):268–74. [PubMed: 16357682]
- [26]. Craven BC, Morris AR. Modified Ashworth Scale reliability for measurement of lower extremity spasticity among patients with SCI. *Spinal Cord*. Mar; 2010 48(no. 3):207–13. [PubMed: 19786977]
- [27]. Mutlu A, Livanelioglu A, Gunel MK. Reliability of Ashworth and Modified Ashworth Scales in children with spastic cerebral palsy. *BMC Musculoskelet Disord.* 2008; 9:44–44. [PubMed: 18402701]
- [28]. Tederko P, et al. Reliability of clinical spasticity measurements in patients with cervical spinal cord injury. *Ortop. Traumatol. Rehabil.* Sep-Oct; 2007 9(no. 5):467–83. [PubMed: 18026067]
- [29]. McCrea PH, Eng JJ, Hodgson AJ. Linear spring-damper model of the hypertonic elbow: Reliability and validity. *J. Neurosci. Methods*. Sep. 2003 128(no. 1–2):121–8. [PubMed: 12948555]
- [30]. Fee JW Jr, Foulds RA. Neuromuscular modeling of spasticity in cerebral palsy. *IEEE Trans. Neural Syst. Rehabil. Eng.* Mar; 2004 12(no. 1):55–64. [PubMed: 15068188]
- [31]. Koo TK, Mak AF. A neuromusculoskeletal model to simulate the constant angular velocity elbow extension test of spasticity. *Med. Eng. Phys.* Jan; 2006 28(no. 1):60–9. [PubMed: 15908257]
- [32]. Zhang LQ, et al. System identification of tendon reflex dynamics. *IEEE Trans. Rehabil. Eng.* Jun; 1999 7(no. 2):193–203. [PubMed: 10391590]
- [33]. Mayer NH. Clinicophysiology concepts of spasticity and motor dysfunction in adults with an upper motoneuron lesion. *Muscle Nerve*. 1997; 6:S1–13. [PubMed: 9826979]

- [34]. Bennett DJ, et al. Time-varying stiffness of human elbow joint during cyclic voluntary movement. *Exp Brain Res*. 1992; 88(no. 2):433–42. [PubMed: 1577114]
- [35]. Calota A, Feldman AG, Levin MF. Spasticity measurement based on tonic stretch reflex threshold in stroke using a portable device. *Clin. Neurophysiol.* Oct; 2008 119(no. 10):2329–37. [PubMed: 18762451]
- [36]. Sheean G. The pathophysiology of spasticity. *Eur. J. Neurol.* May; 2002 9(no. Suppl. 1):3–9. [PubMed: 11918643]
- [37]. Townsend, WT. Ph.D. dissertation, Mechan. Eng. Dept. MIT; Cambridge: 1988. The effect of transmission design on force-controlled manipulator performance.
- [38]. Ljung, L. *System Identification-Theory For the User*. 2nd ed.. Upper Saddle River: PTR Prentice Hall; 1999.
- [39]. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol. Bull.* 1971; 76(no. 5):378–382.
- [40]. Landis JR, Koch GG. Measurement of observer agreement for categorical data. *Biometrics*. 1977; 33(no. 1):159–174. [PubMed: 843571]
- [41]. Damiano DL, et al. What does the Ashworth Scale really measure and are instrumented measures more valid and precise? *Dev. Med. Child Neurol.* Feb; 2002 44(no. 2):112–8. [PubMed: 11848107]
- [42]. Bohannon RW, Smith MB. Interrater reliability of a modified Ashworth Scale of muscle spasticity. *Phys. Ther.* Feb; 1987 67(no. 2):206–7. [PubMed: 3809245]

Fig. 1. Prototype MSE. (a) Design of MSE. (b) MSE attached at elbow.

Fig. 2.
Typical experimental data showing three phases of elbow spasticity: (i) Pre-catch, (ii) Catch, and (iii) Post-catch.

Fig. 3.
Comparison between measured force during slow movement and estimated force using (1).

Fig. 4.
Schematic model of catch phase.

Fig. 5.

Clinical data supporting (4): $\Delta T_c \cdot \bar{\omega}_c$ from 12 trials at different stretching speeds (all data collected from subject #3).

Fig. 6.
Haptic Elbow Spasticity Simulator (HESS).

Fig. 7.
Control flowchart implemented in HESS.

Fig. 8. Backdrivability test. (a) Without friction compensation. (b) With friction compensation.

Fig. 9. Mechanical behaviors during in-person and haptic assessments: (i) Pre-catch phase (ii) Catch phase, and (iii) Post-catch phase.

TABLE I

MAS Scores of Four Selected Subjects.

Subject		#1	#2	#3	#4
MAS scores	Rater 1	1	1+	2	3
	Rater 2	1	1+	2	2
	Rater 3	Absent*	1+	2	3
MAS Scores used for the modeling		1	1+	2	3

TABLE II**Mechanical Specification of HESS**

Specifications	Value
Max. continuous torque	28Nm
Max. speed (at motor shaft)	300 deg/s
Position resolution at the elbow joint	0.005 deg
Gear ratio	18.8:1
Weight of forearm and hand	0.8 kg (forearm), 0.3 kg (hand)

TABLE III

Parameters of the Four HMs

	L	H	Q	D	Other constants
HM#1	130	1.4	0.15	60	
HM#2	120	2.0	0.3	50	$m = 0$, $b = 1.26$ (Nms/rad), and $k = 3.15$ (Nm/rad) in (1) and (5) $\alpha = 0.35$ (sec) in (2) $\Delta T_{peak} = 0.1$ (sec) in (3)
HM#3	115	2.8	0.6	30	$k_{post} = 5.8$ (Nm/rad) in (5) $\dot{\theta}_n = 20$ (deg/s) in Fig.6
HM#4	110	3.8	0.8	10	

Among all parameters, ΔT_{peak} and $\dot{\theta}_{th}$ are constants that can be freely chosen within a reasonable range, b and k was estimated from the trials at slow velocity by using least square parameter estimation. The stiffness in post-catch phase, k_{post} is estimated from the data around the joint limit. Q , H , and D can be measured directly from the torque data collected in the in-person assessment. L and α are parameters found in the linear relationship between the stretching velocity and the catch angle curve that are obtained from multiple trials at different velocities.

TABLE IV

Comparison of MAS Scores From In-Person and Haptic Assessment

MAS	Assessment	Rater 1 [*]	Rater 2 [*]	Rater 3 [*]
1	In-person (subject 1)	1	1	1
	Haptic (HM#1)	1	1	1
1+	In-person (subject 2)	1+	1+	Absent ^{**}
	Haptic (HM #2)	1+	1+	1+
2	In-person (subject 3)	2	2	2
	Haptic (HM #3)	2	2	2
3	In-person (subject 4)	3	2	3
	Haptic (HM #4)	3	2	3

* Rater 1,2, and 3 are same as the rater 1,2, and 3 in Table I, respectively.

** Rater 3 did not test with subject 2.

TABLE V

Rated MAS Scores and Stretching Speeds for Haptic Assessment

	HM#1 (MAS 1)		HM#2 (MAS 1+)		HM#3 (MAS 2)		HM#4 (MAS 3)	
	rated score	$\dot{\theta}_{pre}$	rated score	$\dot{\theta}_{pre}$	rated score	$\dot{\theta}_{pre}$	rated score	$\dot{\theta}_{pre}$
Rater 1 *	1	122	1+	124	2	108	3	68
Rater 2 *	1	100	1+	83	2	92	2	65
Rater 3 *	1	86	1+	70	2	88	3	64
Rater 4	1+	123	1+	131	2	110	3	72
Rater 5	1	80	1+	70	2	77	3	60
Rater 6	1	92	1	80	2	78	2	65
Rater 7	1	82	1+	86	2	73	3	61
Rater 8	1	89	1	93	2	80	3	60

(MAS #) denotes the intended MAS score; $\dot{\theta}_{pre}$ (deg/sec) denotes for average stretching velocity during pre-catch phase; values in the table are collected from a typical trial among 3~7 trials that each rater performed.

* Rater 1, 2, and 3 are same as the rater 1, 2, and 3 in Table I and IV.