A new quantitative performance parameter for monitoring robotics rehabilitation treatment

Elisabetta Peri
NearLab, Politecnico di Milano
Milan, Italy
elisabetta.peri@polimi.it

Emilia Biffi
IRCCS E. Medea, Bioengineering Lab
Bosisio Parini (LC), Italy
emilia.biffi@bp.lnf.it

Cristina Maghini
IRCCS E. Medea
Bosisio Parini (LC), Italy
cristina.maghini@bp.lnf.it

Fernanda Servodio Iammarrone
IRCCS E. Medea,
Bosisio Parini (LC), Italy
servodiofernanda.iammarone@bp.lnf.it

Chiara Gagliardi
IRCCS E. Medea,
Bosisio Parini (LC), Italy
chiara.gagliardi@bp.lnf.it

Chiara Germiniasi
IRCCS E. Medea
Bosisio Parini (LC), Italy
chiara.germiniasi@bp.lnf.it

Alessandra Pedrocchi
NearLab, Politecnico di Milano
Milan, Italy
alessandra.pedrocchi@polimi.it

Anna Carla Turconi
IRCCS E. Medea,
Bosisio Parini (LC), Italy
annacarla.turconi@bp.lnf.it

Gianluigi Reni
IRCCS E. Medea, Bioengineering Lab
Bosisio Parini (LC), Italy
gianluigi.reni@bp.lnf.it

ABSTRACT

In clinical practice, the great potential of robots in extracting quantitative and meaningful data is not always exploited. The aim of the present paper is to propose a simple parameter that allows to follow the performance of subjects during upper limb robotic training with no additional effort for patients or clinicians. Armeo®Spring has been used to perform a training on 14 children affected by cerebral palsy (CP). Each session was evaluated with the new parameter allowing us to observe variations in subjects' performance over time. A preliminary validation of the parameter was performed by means of Melbourne clinical scale. Further studies are required to compare the results with kinematics data and other clinical scales.

Categories and Subject Descriptors

I.2.9 [**Artificial Intelligence**]: Robotics – *Commercial robots and applications.*

General Terms

Measurement, Experimentation, Standardization.

Keywords

Robotics rehabilitation, upper limb, Armeo®Spring Pediatric, cerebral palsy.

1. INTRODUCTION

In clinical practice, it is substantial to identify among the multiple variables involved in rehabilitation treatments which ones might have a larger impact on outcomes and influence recovery.

Moreover, such evaluations require the use of quantifiable, valid, and sensitive tools to guarantee reliable between-study comparisons and greatly improve the understanding of key treatment effects [1]. Unfortunately, many assessment methods are based on subjective impressions, which make it difficult to justify the effectiveness of therapy treatments. Further, sophisticated 3D posture analysis systems, which allow assessing posture and movements in a quantitative fashion, are not always available [2].

During the past few years, robot-assisted rehabilitation has become a very active area of research not only because rehabilitation robots can provide controlled, intensive task-specific training that is goal directed and cognitively engaging, but also because measures derived from robot data can contribute to the understanding of how different treatment variables (e.g., dosage, amount, and type of assistance provided) influence motor learning and recovery [3]. Previous studies have developed some ad-hoc assessment tools to extract outcome measures of patients' performance, such as position of a hand effector, ability in following a trajectory and measures of forces exchanged [4]. Nevertheless, only few robotic assessment methods have been compared and validated with clinical and functional scales [4]. Often, robots are provided with potentiometers that allow extracting trajectories and subjects' kinematics during the robotic training. In this direction, Merlo and colleagues proposed normative values of healthy subjects' upper limb functionality [5] during an exercise of assessment with Armeo®Spring. However, all the methods reported above require the development of ad-hoc technology and, in some cases, also additional time both for patients and clinicians to perform the rehabilitation assessments.

In this work we describe a simple parameter that can be easily derived from data saved by the robot and that gives an indication of subjects' performance. This assessment parameter combines information about time needed to finish an exercise, scores obtained during the exercise and level of difficulty. It can be used to follow the trend of a robot-aided treatment, to describe changes in performance before and after a rehabilitation and thus to investigate the effects of variations in the therapy on patients' motor and functional recovery. Here we used Armeo Spring

(Hocoma, AG) and we evaluated 8 exercises. However, the value of the extracted parameter does not depend on the robot employed and on the exercise performed.

2. MATERIAL AND METHODS

Fourteen inpatients, aged between eight and sixteen, affected by cerebral palsy performed a training with Armeo® Spring Pediatric device. Details about the subjects are reported in Table 1. All inpatients and their families gave voluntarily their consent to the clinical trial. The research protocol was approved by the medical ethics committee of IRCCS E. Medea.

2.1 Apparatus

Armeo®Spring is a five degree of freedom (3 in the shoulder, 1 in the elbow and 1 in the forearm) exoskeleton. It is not provided with robotic actuator but springs are used to guarantee passive arm weight support and guidance. The stiffness of the springs can be customized resulting in a different level of gravity support and patients' muscular involvement. This enables patients to achieve a larger range of motion within a 3D workspace with their own residual functionality. The Armeo exoskeleton can be adapted to the patient's morphology by changing the position and the length of the orthosis. A pressure-sensitive handgrip is also present providing grasp training. Through dedicated software, patients are engaged in exercises that aim at training functional and meaningful tasks (e.g. break eggs, clean a surface, etc.) involving different upper arm districts and joints.

2.2 Experimental design

The subjects underwent 4 weeks of training composed of 20 sessions lasting 30 minutes. During each session, subjects performed a pull of exercises, customized on each subject, with the supervision of a physiotherapist. The exercises were chosen in order to provide an engaging and gradual training increasing the difficulty level (very easy, easy, medium, difficult). During each training session, information about the exercise (e.g. difficulty level, working area, weight support), the score obtained by the subject and the time required to perform the exercise were automatically recorded by the system, with no additional effort for the physiotherapists.

Eight exercises were selected to evaluate subjects' performance over different joins and in different spaces (1D, 2D and 3D), accordingly to the indication of physiotherapists and clinicians. In particular, we evaluated a 1D exercise ("goalkeeper"), five exercises performed in a 2D space ("egg cracking", "fruit shopping", "stove cleaning", "moorhuhn" and "vertical catching") and two in a 3D space ("chase balloon" and "reveal panorama"). During the "goalkeeper" exercise the patient is supposed to use the prono-supination of the wrist to parry balls. As the difficulty level increases, more and quicker movements are required. The "egg cracking" and "fruit shopping" exercises require to reach an object with a cursor controlled by the end-point position in the 2D space (fruit or egg), moving it to a different portion of the space (cart or pot) involving flex-extension/abduction-adduction of the shoulder, flex-extension of the elbow and the grip-release of the

Table 1. Participants' details at baseline.

Etiology	CP (6 unilateral, 8 bilateral)
Age, years [*]	10,8 (2,9)
Gender, M/F	11/3
MACS (1/2/3) [**]	2/7/5

[*] median (interquartile range) is reported.
[**] MACS is Manual Ability Classification System.

hand, when the autogrip function is disabled. Higher difficulty levels require a higher number of more little objects to be moved. "Reveal panorama" involves the same body districts but require exploring the 3D space with a sponge that become littler as the difficulty level increases. The "stove cleaning" exercise requires moving a sponge in 2D space and involves abduction-adduction of the shoulder, flex-extension of the elbow and the grip-release of the hand, when the autogrip function is disabled. The difficulty level varies the dimension of the sponge. "Chase balloon" and "vertical catching" involve the same articular movement (flex-extension/abduction-adduction of the shoulder, flex-extension of the elbow) but are performed in the 2D and 3D space, respectively. Both exercises require reaching a target that appears on the screen. Higher difficulty level means higher number of targets to be reached to obtain the maximum score and, for the "vertical catching" exercise, bigger area covered by the targets. The last exercise considered is "Moorhuhn" that trains abduction-adduction of the shoulder, flex-extension of the elbow, prono-supination of the wrist and the grip-release of the hand, when the autogrip function is disabled. The task is to shoot some moving targets. Higher difficulty level means higher velocity of target movement and less time at patients' disposal.

In addition to the level of difficulty, the physiotherapists could modify other parameters to increase or decrease the difficulty of the exercises. In particular, if the grasping task was expected for the exercise, the autogrip could be enabled or disabled. The precision required in grasping could be modulated through thresholds. Other parameters such as the workspace and the weight support of each subject were not varied during our training.

All the components described above have been taken into account in a comprehensive performance parameter (p_i) computed as in Equation I for each ith exercise:

$$p_i = \frac{S_i / S_{i,TOT}}{T_i / T_{i,TOT}} D_i \qquad (I)$$

where S_i is the score obtained during the ith exercise, $S_{i,TOT}$ is the maximum score obtainable, T_i is the time required to complete the ith exercise, $T_{i,TOT}$ is the maximum time available and D_i is the difficulty coefficient that considers level of the exercise and variation in autogrip and control threshold for each subject during the training. If the ith exercise was not time or score limited, $S_{i,TOT}$ or $T_{i,TOT}$ have been considered equal to 1. In order to compare different exercises, p_i has been divided on the maximum performance achieved over time by the group of subjects ($p_{i,max}$) obtaining P_i.

In this work, we first used the performance parameter P_i to follow the training of subjects over time, considering every single session. Then, the median value of P_i within the first week (T_0), between the 12th and the 16th days ($T_{1/2}$) and within the fourth week (T_1) of training were computed for each subject. Finally, we computed the median values of P_i over the 14 subjects at the beginning, at the middle and at the end of the training. The median value of P_i for all the exercises has been considered as index of the overall performance (P).

We also evaluated the Melbourne Assessment of Unilateral Upper Limb Function [6], which is composed by some items (reach, grasp, drawing, release, manipulation, pointing, pronation/supination, hand to hand transfer, hand to mouth and down) that subjects have to perform. The overall score provides a reliable evaluation of upper limb functionality. Moreover, we computed the median values of the Melbourne Assessment at the

beginning and at the end of the training. Finally a preliminary validation of P has been proposed by comparing, for each subject, variation of P between T_0 and T_1 with variation of Melbourne Assessment referred to the maximum value obtained by the groups of patients in the same time points.

2.3 Statistical analysis

The effect of time on the performance was assessed by using the non-parametric Friedman test for paired samples. If values of P in T_0, $T_{1/2}$ and T_1 were found significantly different ($p<0.05$), a post-hoc analysis (Wilcoxon test) was performed comparing paired groups (T_0 vs $T_{1/2}$, T_0 vs T_1 and $T_{1/2}$ vs T_1).

3. RESULTS

3.1 During the training of each subject

Figure 1A reports an example of the training with the "vertical catching" exercise on two subjects, named S1 and S2 in the following. A clinical expert classified their functional impairment by the mean of Gross Motor Function Classification System (GMFCS) and of MACS as mild (S1 obtained GMFCS=I and MACS=1) and moderate (S2 obtained GMFCS=III and MACS=2). The performance parameter P was able to display the different results obtained by the two subjects: S1 maintained similar values of P over time (R=0.04) while S2 started from a lower value and showed a linear increasing trend (R=0.93). Oscillation of the results of single session may depend on motivation and effort of the patient at that time.

3.2 Evaluation at T_0, T_m, T_1

Table 2 reports medians and interquartile values of the performance parameter P_i, computed for each of the eight selected exercises, within the first week (T_0), between the 12th and the 16th days ($T_{1/2}$) and within the fourth week (T_1) of training over all the subjects. P varied significantly between T_0 and T_1 in all the exercises ($p<0.05$), but the trend is different in some cases. "Stove cleaning" and "reveal panorama" exercises did not significantly differ between $T_{1/2}$ and T_1 showing a plateau during the second half of the training. Differently, "buy fruits" exercise needs a longer training to vary as no significance was obtained between T_0 and $T_{1/2}$. Finally, other exercises significantly increased between T_0 and $T_{1/2}$ and $T_{1/2}$ and T_1 such as "vertical catching" as shown in Figure 1B. P shows the median and interquartile values of the

whole group of exercises for all the patients. As can be observed a statistically significant variation was obtained.

3.3 Preliminary validation of P

To validate P, the Melbourne scale has been used. This scale showed a significant improvement between T_0 and T_1 (84 vs. 87, p=0.002 as shown in Figure 2A), highlighting progresses in terms of upper limb functionality. These improvements were related to the increased performance observed with P (see Figure 2B). Moreover, the comparison between variation of Melbourne scale and P (ΔM vs. ΔP) for each subject shows a moderate positive correlation (R=0.4) and the root mean square error between data and linear fitting curve is adequate (RMSE=0.04). The slope of the linear fitting was 2.3.

4. DISCUSSION

The need of reliable, quantitative and repeatable evaluations of training effectiveness is an up-to-date theme in the clinical practice. In fact, in some cases clinical scales are operator-dependent and not sensitive enough to highlight changes felt by patients and their parents [4]. Some quantitative evaluation could be obtained by the use of ad-hoc technology (e.g. optoelectronic analysis of the kinematics, sensorized robots) but these methods are time and money consuming.

Here we propose a parameter P that takes into account the time needed to finish an exercise, the scores obtained during the exercise and the level of difficulty. This parameter is computed in a simple and quick way from data automatically acquired by the robot during the training, and it does not require extra expensive devices. An excel macro is available for other researchers if required.

Our data give preliminary evidence that an increase in the values of P corresponds to a functional improvement in terms of Melbourne scale, clinically validating the results obtained. In particular the slope equal to 2.3 may suggest that the parameter P is more sensitive to variations during training with respect to Melbourne scale. Anyway it should be considered that Melbourne scale is an outcome that gives a functional indication while the parameter P evaluates a trained task thus being influenced by learning process.

Table 2. P_i for the considered exercises over time (T_0, $T_{1/2}$ and T_1). N is the sample size on each exercise.

	T_0	$T_{1/2}$	T_1	p-value**		
				T_0vs$T_{1/2}$	T_0vsT_1	$T_{1/2}$vsT_1
$P_{Goalkeeper}$* N	0.25 (0.09) 12	0.33 (0.31) 12	0.66 (0.18) 10	**0.002**	**0.005**	**0.022**
$P_{Moorhuhun}$* N	0.32 (0.14) 13	0.45 (0.40) 13	0.72 (0.25) 11	**0.009**	**0.008**	**0.007**
$P_{Stove\ cleaning}$* N	0.25 (0.18) 12	0.53 (0.20) 12	0.54 (0.10) 8	**0.003**	**0.018**	0.889
$P_{Fruit\ shopping}$* N	0.45 (0.12) 11	0.43 (0.26) 11	0.62 (0.24) 9	0.248	**0.011**	**0.011**
$P_{Egg\ cracking}$* N	0.13 (0.19) 14	0.16 (0.14) 14	0.26 (0.12) 11	**0.022**	**0.033**	**0.003**
$P_{Reveal\ panorama}$* N	0.10 (0.08) 13	0.17 (0.15) 12	0.30 (0.47) 10	**0.012**	**0.005**	0.066
$P_{Chase\ balloon}$* N	0.26 (0.15) 13	0.47 (0.24) 12	0.68 (0.26) 10	**0.013**	**0.013**	**0.015**
$P_{Vertical\ catching}$* N	0.35 (0.11) 13	0.50 (0.15) 13	0.54 (0.14) 10	**0.002**	**0.005**	**0.021**
P N	0.27 (0.21) 101	0.43 (0.30) 99	0.55 (0.37) 79	**<0.001**	**<0.001**	**<0.001**

* median (interquartile range) is reported. **Bold values are statistically significant ($p<0,05$).
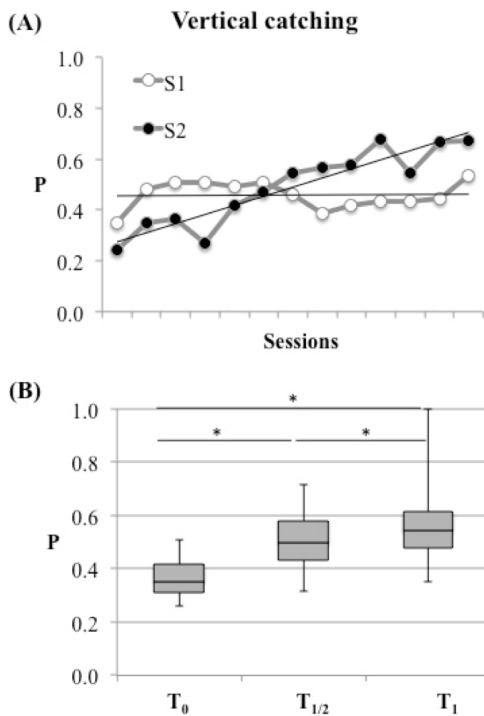
Figure 1. Results of P obtained during "vertical catching."

A: comparison between normalized performances over time of two subjects with different impairment at baseline (S1 mild and S2 severe impairment). The trendline is also shown.
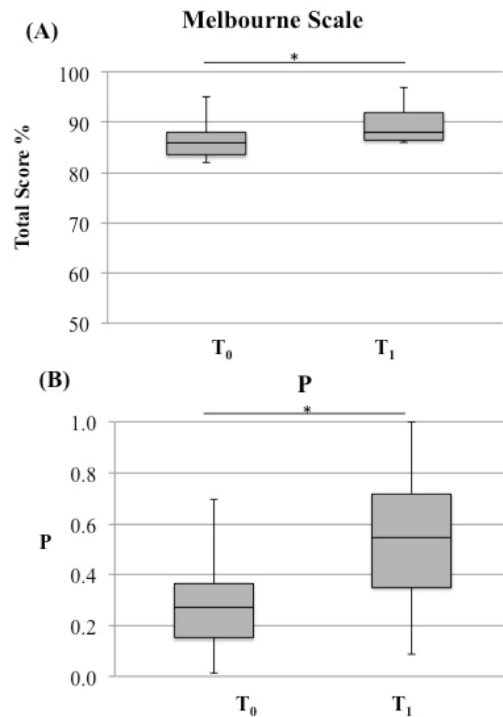
B: results of the group (N= 14) during the first week (T_0), the 12th and 16th days ($T_{1/2}$) and the last week (T_1) of training. Median, I quartile and III quartile are represented as box. The maximum and minimum values are also shown as bars. * represents p<0,05 in the Wilcoxon statistical test.

All the exercises analyzed showed significant variation of P_i over time. P_i allowed to follow the training and to analyze performance variation between the beginning and the end of exercises. Moreover analyzing two subjects with a different disability level, we observed that P_i was able to highlight some differences in the effectiveness of the robotic training. In particular the subject with moderate impairment at baseline improved the performance while mild impaired subject seems not to gain improvement. Other researchers have already assessed that this effect may be due to ceiling for less severely impaired subjects [7].

To conclude, a simple parameter has been proposed in this work. It seems to be a useful tool to follow the training and to give information about the performance on the exercises with minimal additional effort for patients and clinicians. First tests confirm that P variations are coherent with the functional clinical data thus supporting that it could give information on functional recovery. Further investigations are required to confirm these results and to validate the parameter with more clinical scales. Data of an age-matched healthy control group will be acquired in order to have a reference about the maximum performance achievable during the exercises. Moreover a validation with kinematics data would be important in order to compare the parameter with other quantitative and operator-independent data.

Figure 2. Comparison between P related to all the 8 exercises considered and Melbourne scale for a subgroup of 7 patients

A: results of the Melbourne scale for the group (N= 7) before (T_0) and after (T_1) the training. B: P of the group (N= 7) during the first week (T_0) and the last week (T_1) of training.

Median, I quartile and III quartile are represented as box. The maximum and minimum values are also shown as bars. * represents p<0,05 in the Wilcoxon statistical test.

REFERENCES

[1] Lo H S, & Xie S Q 2012. Exoskeleton robots for upper-limb rehabilitation: State of the art and future prospects. *Med Eng Phys*, *34*(3), 261-8.

[2] Fortin C, Ehrmann Feldman D, Cheriet F, & Labelle H. 2011. Clinical methods for quantifying body segment posture: a literature review. *Disabil rehabil*, 33(5), 367-83

[3] Fasoli, S E, Krebs H I, Stein J, Frontera W R, et al 2003. Effects of robotic therapy on motor impairment and recovery in chronic stroke. *Arch Phys Med Rehabil, 84*(4), 477-82.

[4] Brewer B R, McDowell S K, & Worthen-Chaudhari L C 2007. Poststroke upper extremity rehabilitation: a review of robotic systems and clinical results *Top stroke rehabil*, *14*(6), 22-44

[5] Merlo A, Longhi M, Giannotti E, Prati P, et al. 2013. Upper limb evaluation with robotic exoskeleton. Normative values for indices of accuracy, speed and smoothness. *NeuroRehabil*, *33*(4), 523-30.

[6] Randall M, Carlin J B, Chondros P and Reddihough, D 2001. Reliability of the Melbourne assessment of unilateral upper limb function. *Dev Med Child Neurol*, *43*(11), 761-7

[7] Krebs H I, Volpe B T, Ferraro M, Fasoli S, et al. 2002. Robot-aided neurorehabilitation: from evidence-based to science-based rehabilitation. *Top stroke rehabil*, *8*(4), 54-70