The 23rd IEEE International Symposium on Robot and
Human Interactive Communication
August 25-29, 2014. Edinburgh, Scotland, UK,

Auto-Evaluation of Motion Imitation in a Child-Robot Imitation Game for Upper Arm Rehabilitation

Arzu Guneysu, Recep Doga Siyli and Albert Ali Salah [1]

Abstract— **The purpose of this study is fusing play-like child robot interaction with physiotherapy in order to achieve upper arm rehabilitation by motivating the child. The proposed system is not intended to substitute for the physiotherapist, but to assist them in their therapeutic tasks by encouraging the child's participation in the activity. Recognizing the imitation performance of the child and supporting him/her with feedback for drawing the child's attention and motivating the child to imitate the robot is crucial. This study concentrates on automatically evaluating the upper body actions of the child during an imitation based physical therapy. For quantifying the performance of the child, two measures were considered: Range of Motion (RoM) and Dynamic Time Warping (DTW) distance. In our initial experiments, eight healthy children were asked to stand in front of a Kinect sensor and to mimic the actions of the humanoid robot Nao, which consist of shoulder abduction, shoulder vertical flexion&extension and elbow flexion. The proposed evaluation measure is verified as a reliable measurement according to Intraclass Correlation Coefficient (ICC) through comparison with evaluations of five physiotherapists as ground truth. The degree of consistency among our ratings and the physiotherapist ratings is between %76 and %96 for different motions.**

I. INTRODUCTION

In Socially Assistive Robotics (SAR), the robots provide assistance to the user through an effective interaction where the purpose is to achieve measurable progress in rehabilitation, learning, etc. [1]. One of the promising areas where SAR can provide solutions is physical rehabilitation, where the therapist coaches a patient to repeatedly use the affected limb(s) without any physical contact. Recent studies showed that utilizing a humanoid robot in physical rehabilitation is feasible [2] [3].

Motivating children during the rehabilitation is crucial, since physical therapy sessions include repetitive tasks [2]. Monotonous actions cause decrease in performance of the behavior of the child due to the reduced attention. Furthermore, repetitions decrease continuous attention for the current activity and the motivation for the repetitive exercise. Since playing a game is a concrete way for children to communicate and express themselves more naturally [4], children's motivation to participate in the activity can be improved by game-like child-robot interaction. A toy-like humanoid robot can be a playmate that fuses play and rehabilitation techniques in a physiotherapy session in an effective and entertaining way [2]. A study by Belpaeme et al. indicated that physical robots can effectively engage the attention of young users and that adaptation to user characteristics can be a useful tool in supporting sustained interaction [5].

In order to achieve an interactive engagement that catches the child's attention and to motivate the child to imitate the robot in a rehabilitation session, the robot should recognize the actions of the child and serve feedback accordingly. Calculating similarity between the motions of the child and robot is crucial in order to provide a proper feedback in any physical therapy.

In this study we propose two measures to evaluate the upper arm gestures of a child during a robotic physiotherapy coaching session. Compared to the other studies in literature, instead of recognizing the type of action performed by the child, our proposed system is attempting to automize the evaluation process by mimicking the decision of physiotherapists.

II. RELATED WORK

Humanoid robots have been used by various researchers for rehabilitation purposes. Choe et al. showed the feasibility of utilizing a humanoid robot in stroke rehabilitation while investigating a multidisciplinary therapy approach. In robot mediated physical therapy the aim is to enhance ranges of motion, to improve dexterity and use of impaired arm in daily activity [6]. Fridin et al. proposed a Robotics Agent Coacher for Cerebral Palsy motor function, which encourages participation of the child to the therapy activity in a game-like human-robot interaction [3]. Their proposed architecture includes a Nao [7], a computer and a Kinect sensor. Wainer et al. used an autonomous humanoid robot, KASPAR, in a collaborative game with autistic children, where the children seemed to see their robotic partner as being more interesting and more entertaining than their human partner [8].

Brooks et al. proposed to use computer vision techniques such as Motion History Imaging, edge detection and Random Sample Consensus (RANSAC) to quantify upper arm rehabilitation metrics (the range of motion and the peak angular velocity) for children during child robot play integrated therapy [2]. In the initial experiments, they evaluated their methodology with abduction and adduction movements by directly asking healthy subjects to perform the actions and recorded these movements via a web-cam [2]. Strohrmann et al. performed a longitudinal study with children that need physiotherapy [9]. The children performed 10 predefined motor tasks such as turning around cards, picking up small objects and climbing stairs. The movements were monitored

[1] Arzu Guneysu, Recep Doga Siyli and Albert Ali Salah are with the Department of Computer Engineering, Bogazici University, Istanbul, Turkey `arzu.guneysu@boun.edu.tr`, `dogasiyli@gmailcom`, `salah@boun.edu.tr`

978-1-4799-6765-0/14/$31.00 ©2014 IEEE

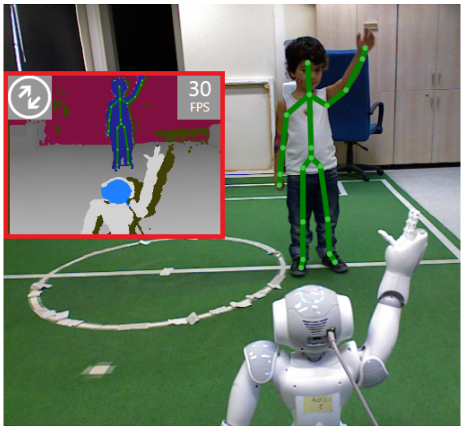

Fig. 1. An example of Kinect sensor view

using wearable sensors and the video recordings were collected for evaluation of therapists. Zannatha et al. presented a stroke rehabilitation system for upper limbs using Kinect, an interactive virtual environment, a humanoid robot and devices producing ergonometric signals [10]. The system architecture they created allows to integrate monitoring programs to evaluate the progress of a patient by a human.

Upper body gesture imitation analysis was also used for early detection and treatment of autism. Ranatunga et al. used a humanoid robot to perform interactive upper body gestures which the child can imitate and recording is done via a motion capture system [11]. Similarity of performances of the child and robot was measured using Dynamic Time Warping.

Our proposed system is based on a physiotherapy session conducted with child under the guidance of a robot. The robot performs the actions, and the child imitates the robot. The success of imitation is measured by processing the Kinect Skeleton data of the child and grading the similarity. Initial experiments have been carried out with healthy children in an upper body physiotherapy imitation scenario. To evaluate the data, five physiotherapists gave ground truth annotations and, our discrete evaluation for each motion is compared with their ratings by using Intraclass Correlation Coefficient (ICC). Details of ICC are given in V-A.

III. EXPERIMENTAL DESIGN

Our initial experiments are conducted with eight healthy children, aged between 3 to 11 and the subjects imitated three different motions performed by the Nao involving only upper-body limbs. Each of the three actions were performed by Nao three times in a circular manner. As the imitation scenario was being carried out, the motions were recorded by a Kinect sensor which was right above and 30 cm. behind the robot, looking at the child and saving depth and RGB data. An example of Kinect sensor view can be seen in Fig.1.

A. Ranges and Types of Motions

The performed motions are shoulder abduction, shoulder vertical flexion&extension and elbow flexion. Human movements are described in three dimensions, based on three planes of motion that pass through the human body, namely sagital, frontal and transverse planes, respectively [12]. The motions are explained briefly below:

1) Shoulder Abduction: Bringing the arms up sideways on transverse plane (See Fig. 2).

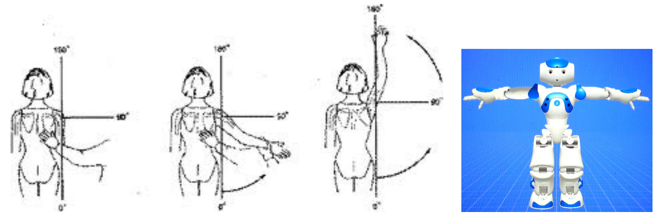

Fig. 2. Shoulder Abduction

2) Shoulder Vertical Flexion&Extension: Raising the one arm up straight forward and then lowering it down straight backward on the sagital plane (See Fig. 3).

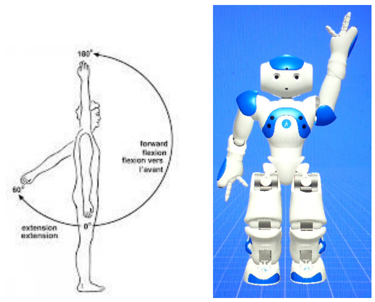

Fig. 3. Shoulder Vertical Flexion&Extension

3) Elbow Flexion: Bringing the lower arms to biceps on both sagital and transverse planes (See Fig. 4).

B. Ground Truth Evaluations

The recorded RGB videos of the children were watched and evaluated by three physiotherapists and two intern physiotherapists who work in a rehabilitation facility. Each of the three motions were performed by an adult and saved as "The Baseline Motion" to be compared with the children's performances. Each action of the children was rated on a five-level Likert scale, representing a very weak and a very strong imitation, respectively. Angular data and the plane of the motion are the main criteria for the evaluation process.

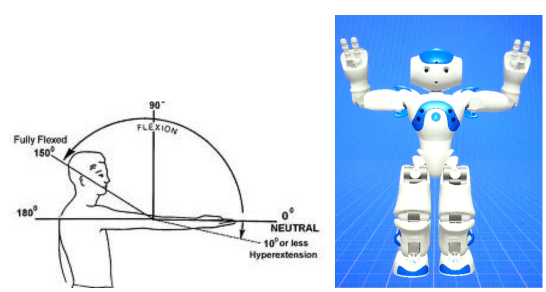

Fig. 4. Elbow Flexion

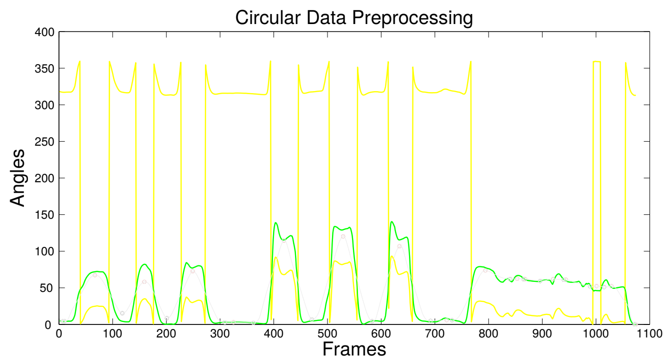

Fig. 5. Circular Data Preprocessing

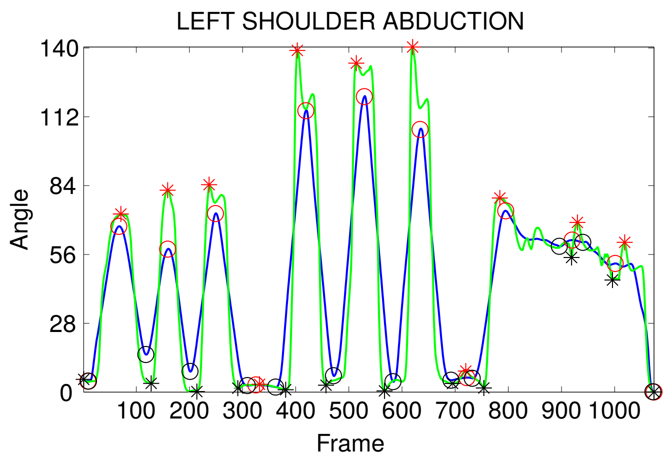

Fig. 6. Peak Detection and Motion Segmentation

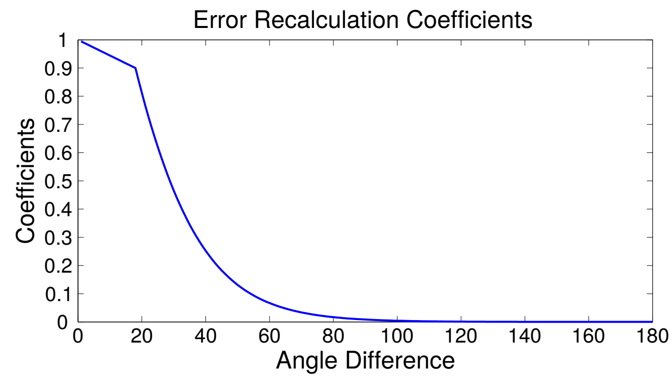

Fig. 7. Similarity Error Recalculation Coefficients

IV. DATA PROCESSING

A. *Circular Data Preprocessing*

Angular data are between 0 and 360 degrees. In order to process it as a signal we proposed a way to shift it such that it becomes a continuous wave. In this process the important part is to calculate which angle to select to apply shifting. This angle is selected by coding the motion as clock-wise and counter-clockwise per frame by looking at the sine, cosine and derivatives of these angular values. An example input and output of this process can be seen in Fig. 5 where the circular data are shown by a yellow line and preprocessed circular data are shown by green line.

In order to get rid of the unnecessary small angle fluctuations due to local minimas and local maximas, we smoothed the data using a moving average filter with a window size of 50 frames as a preprocessing step.

B. *Peak Detection and Alignment*

The local peak values can be detected easily from the smoothed data by using motion directions, which are either clockwise or counterclockwise. The peak values are then aligned to the unsmoothed data to find the real peak points.

C. *Automatic Motion Segmentation*

Our data are composed of motions such that the initial pose of each motion is the same as the last pose. Hence, we are expecting a motion to be completed between three peaks that creates a curve, where either a local minimum will be followed by a local maximum and ending with a local minimum or a local maximum followed by a minimum and ending with a local maximum. The whole motion data of a child segmented into sub-motions using this automatic procedure, an example of which can be seen in Fig 6, where preprocessed data and smoothed data are drawn by green line and blue line, respectively. Red/black circles on smoothed data are maximum/minimum peak points, while red/black stars on preprocessed data are the aligned versions of the peaks.

D. *Dynamic Time Warping*

In order to compare each motion of the child with a baseline motion Dynamic Time Warping was implemented using a window size of three frames. Detailed information about DTW can be found in [13]. Compared motions are curve shaped as described in Section IV-C. The sequences of two motions are warped nonlinearly in the time dimension to determine a measure of their similarity. The change between warped and initial sequences are then analyzed for evaluation of motion imitation.

E. *Evaluation of Motion Imitation*

Similarity of a base motion sequence and a given segmented motion is calculated using two different measures: a similarity error calculated via DTW and a recall measure calculated from the range of angles that are passed by the baseline motion and the child's motion. The similarity error is expected to be low, whereas the recall measure is expected to be high for high similarity.

1) Similarity Error: The similarity error is initially calculated with frame by frame similarity values of each aligned motion via DTW algorithm. Then it is recalculated by penalizing larger amount of angle differences more, while tolerating differences in ten percent of the area range in the corresponding motion type. The penalty function is displayed in Fig. 7.

2) Recall Measure: The angular range that is covered by the patient during a gesture is an important measure in physiotherapy for functional disabilities. Taking this factor into consideration, the angle range covered separately by the baseline motion and child's motion are extracted. The possible range of angles is determined for relevant joints as we explained in Section III-A. We classified each angle as "in range" or "not in range" by assigning 1 and 0 respectively. Accuracy, precision, recall and fall-out measures are calculated. For this study, we only used the recall measure in our evaluation process. The recall measure represents how much of angular area covered by the baseline motion, is also covered by the child's motion in the selected frame window. It is calculated by the following equation:

$$Recall = \frac{TP}{TP + FN} \tag{1}$$

where

- True Positive(TP): Range of angles passed both by the baseline motion and the child's selected area of motion which can be represented as 2.

$$|Rangeof\,Angle_{baseline} \cap Rangeof\,Angle_{child}| \tag{2}$$

- False Negative(FN): Range of angles that are not passed by the child's motion but passed by the baseline motion which can be represented as 3.

$$|Rangeof\,Angle_{baseline} \setminus Rangeof\,Angle_{child}| \tag{3}$$

A grade from 0 to 1 is automatically given by the system and then converted to 5 point scale in order to be comparable to the physiotherapists' evaluations. In the grading phase, only the main joint that is related to the motion is considered and similarity in right and left joint angles are averaged. Since the baseline data include three trials for each action, each trial of the child is compared with these three motions.

The motion of the child is graded by choosing his/her best performance since functionality is the main objective in upper-arm rehabilitation. In Fig. 8 the first row shows the similarity of each automatic segmented motion of the child to the three baseline motions. Base motion-1, motion-2 and motion-3 represent the three trial of the baseline for the shoulder abduction. Similarities are represented via gray scale where white and black color means maximum and minimum similarity, respectively. Second row shows the mean similarity over three comparisons for each segmented motion of the child. It is the comparison of the abduction motion with the all motion trials of the child and the first three motions are the shoulder abduction trials of the child so the three motions in the beginning of the second row in the graph are whiter. The darker areas represent where the abduction motion (motion 1) is not detected. There is another whiter area at the beginning of third motion trials since the elbow flexion starts with shoulder abduction. While starting to an elbow flexion a person should bring the arms up sideways then start to move the lower arms to the biceps. The last two rows show processed right and left joint data of the child correspondingly. In the third and forth rows first three curves correspond to the first motion trials, it is followed by a waiting time and then three trials of second motion. Passing from the second motion to the third motion is fast so the second motion trials are closely followed by the third motion trials.

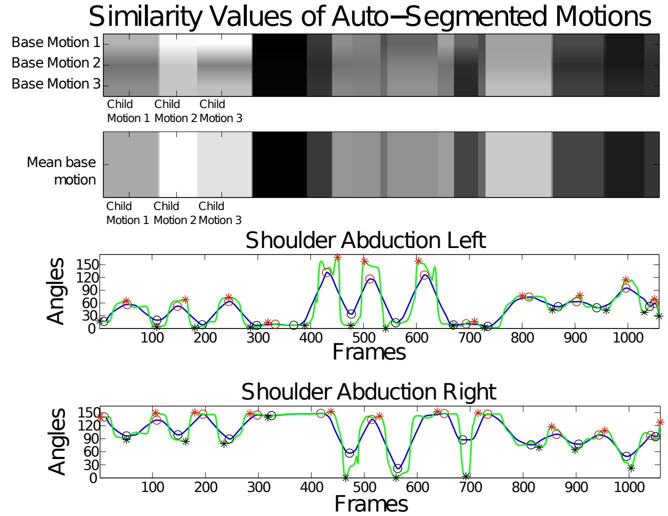

Fig. 8. Similarity Values of Auto-Segmented Motions of the 4th Subject with three Shoulder Abduction Motions of Baseline Data

V. EVALUATION & RESULTS

A. Intraclass Correlation Coefficient

The proposed evaluation is compared with other physiotherapists' evaluations by calculating average agreement values via the Intraclass Correlation Coefficient (ICC), which is a measure of the reliability of ratings. The theoretical formula for the ICC is as in equation 4 where $s^2(w)$ is the pooled variance within subjects, and $s^2(b)$ is the variance of the trait between subjects [14]. However, since the true values are not known, formula is modified with estimates of the two parameters from sample data. For N Raters who evaluate k subjects, $s^2(b)$ is estimated from the subject's mean rating across the raters who rate the subject. Each mean rating is subject to sampling variation deviation from the subject's true trait level. Since the actual mean ratings are often based on a few ratings, these deviations are acceptable. Correction for this error variation should also be estimated. When all subjects have k ratings, $s^2(w)$ equals the average variance of the N ratings of each subject.

The same set of N raters rating each subject case corresponds to a fully-crossed $Rater$ x $Subject$, 2-way ANOVA design in which both Subject and Rater are separate effects. Rater is considered as a random effect; this means the N raters in the study are considered a random sample from a population of potential raters. ICC estimates the reliability of the larger population of raters [14].

In order to compare our evaluation results with physiotherapist evaluations we calculated the agreement value between all raters. In each run we leave one evaluator out and get the average agreement value. The ICC algorithm is also run between all raters (including our proposed system), which is the "All included" case in Table I. It can be concluded from Table I that,our proposed system of evaluation is taken out, agreement level slightly increases for the second and

TABLE I
AGREEMENT VALUES OF DIFFERENT COMBINATIONS OF EVALUATORS FOR RATING OF EACH MOTION

Raters	Motion1	Motion2	Motion3
All included	0.7935	0.8961	0.952
Intern1 out	0.8193	0.8669	0.9344
Intern2 out	0.7148	0.8853	0.9357
Physiotherapist1 out	0.7334	0.8578	0.9384
Physiotherapist2 out	0.7725	0.8675	0.9417
Physiotherapist3 out	0.7666	0.8853	0.9356
System out	0.759	0.9003	0.9665

the third motions, but not for the first motion. We observe that the difference is within the limits of change observed by taking a physiotherapist out from the group. The reason of the high agreement for the first motion evaluation is that the main and only criteria of the evaluation of the first motion by the physiotherapists is one joint angle that we also considered in our proposed system. On the other hand the second and third motion evaluations of physiotherapists include more joint values than we considered . Since we did not integrate all joint values while calculating the similarity, the agreement between other evaluators and our system decreased slightly. Another reason of this is that the second and third motions are harder and complicated so that they were misperceived by children and during the evaluations, the therapists compensated that misunderstanding by giving higher grades to the children while the proposed system considered the angles only.

$$ICC = \frac{s^2(b)}{s^2(b) + s^2(w)} \tag{4}$$

.

VI. CONCLUSION

We proposed an automatic off-line procedure to evaluate the motions of a child who imitates actions of a robot by using a discrete measure for similarity with an expected motion. For preprocessing, a method is proposed for circular-data representation in order to process it as a signal composed of curves. In data analysis, after automatically segmenting motions of the child, Dynamic Time Warping is applied to each motion sequence in order to measure similarity with the expected motion. Evaluation is done by combining this similarity and recall measure. According to the therapists' feedbacks, we did not take the latency of the motion into account while calculating the grades, since it may be an indicator of different things like attention deficit, boredom, or corresponds to lack of perception. Velocity and symmetry of the motion are also not considered since the main goal on these therapies is to increase the functionality, which completing the motion properly in its full range. Evaluations made by our proposed methods were compared with gradings of three physiotherapists and two intern physiotherapists via Intraclass Correlation Coefficient by calculating average agreement values. Results show that evaluation of the first motion is similar to the evaluation of physiotherapist where it slightly differs in other motions. The main reason is that the second and third motion performances of the children are graded generously by physiotherapists since they are complex and maybe misperceived by the children.

VII. FUTURE WORK

As a future work, joint similarities that are not taken into account will be considered for complex motions. In order to eliminate mis-perception of complex motions, they will be combined with functional activities in the future experiments, like reaching to a pen behind the neck in order to accomplish elbow flexion motion. In order to give proper feedback to the child while performing the action, the procedure will be done online.

ACKNOWLEDGMENT

The authors would like to thank Prof. Dr. H. Levent Akın for allowing us to use the Nao robot and the Artificial Intelligence Laboratory facilities in the Department of Computer Engineering at Boğaziçi University during the experiments. We would also like to thank Assist. Prof. Dr. Devrim Tarakcı, and Physiotherapist Emine Dilek Kurbaloğlu for allowing us to make observation in their physiotherapy lessons and all physiotherapists for taking part in our study.

REFERENCES

[1] D. Feil-Seifer and M. J. Mataric, "Defining socially assistive robotics," in *Rehabilitation Robotics, 2005. ICORR 2005. 9th International Conference on*. IEEE, 2005, pp. 465–468.

[2] D. A. Brooks and A. M. Howard, "Quantifying upper-arm rehabilitation metrics for children through interaction with a humanoid robot," *Applied Bionics and Biomechanics*, vol. 9, no. 2, pp. 157–172, 2012.

[3] M. Fridin, S. Bar-Haim, and M. Belokopytov, "Robotics agent coacher for CP motor function (rac cp fun)," in *workshop on robotics for neurology and rehabilitation, San Francisco, CA*, 2011.

[4] G. L. Landreth, *Play therapy: The art of the relationship*. CRC Press, 2012.

[5] T. Belpaeme, P. E. Baxter, R. Read, R. Wood, H. Cuayáhuitl, B. Kiefer, S. Racioppa, I. Kruijff-Korbayová, G. Athanasopoulos, V. Enescu *et al.*, "Multimodal child-robot interaction: Building social bonds," *Journal of Human-Robot Interaction*, vol. 1, no. 2, pp. 33–53, 2012.

[6] Y.-k. Choe, H.-T. Jung, J. Baird, and R. A. Grupen, "Multidisciplinary stroke rehabilitation delivered by a humanoid robot: Interaction between speech and physical therapies," *Aphasiology*, vol. 27, no. 3, pp. 252–270, 2013.

[7] *Aldebaran Robotics*. [Online]. Available: http://www.aldebaran-robotics.com/en/

[8] J. Wainer, K. Dautenhahn, B. Robins, and F. Amirabdollahian, "A pilot study with a novel setup for collaborative play of the humanoid robot kaspar with children with autism," *International Journal of Social Robotics*, vol. 6, no. 1, pp. 45–65, 2014.

[9] C. Strohrmann, R. Labruyère, C. N. Gerber, H. J. van Hedel, B. Arnrich, and G. Tröster, "Monitoring motor capacity changes of children during rehabilitation using body-worn sensors," *Journal of neuroengineering and rehabilitation*, vol. 10, no. 1, p. 83, 2013.

[10] J. M. Ibarra Zannatha, A. J. M. Tamayo, A. D. G. Sánchez, J. E. L. Delgado, L. E. R. Cheu, and W. A. S. Arévalo, "Development of a system based on 3d vision, interactive virtual environments, ergonometric signals and a humanoid for stroke rehabilitation," *Comput. Methods Prog. Biomed.*, vol. 112, no. 2, pp. 239–249, Nov. 2013. [Online]. Available: http://dx.doi.org/10.1016/j.cmpb.2013.04.021

[11] I. Ranatunga, M. Beltran, N. A. Torres, N. Bugnariu, R. M. Patterson, C. Garver, and D. O. Popa, "Human-robot upper body gesture imitation analysis for autism spectrum disorders," in *Social Robotics*. Springer, 2013, pp. 218–228.

[12] C. C. Norkin and D. J. White, *Measurement of joint motion: a guide to goniometry*. FA Davis, 2009.
[13] M. Müller, "Dynamic time warping," in *Information Retrieval for Music and Motion*, editor, Ed. Springer Berlin Heidelberg, pp. 69–84.
[14] *Intraclass Correlation and Related Methods*. [Online]. Available: http://www.john-uebersax.com/stat/icc.htm