Human-Robot Upper Body Gesture Imitation Analysis for Autism Spectrum Disorders

Isura Ranatunga[1], Monica Beltran[1], Nahum A. Torres[1], Nicoleta Bugnariu[2], Rita M. Patterson[2], Carolyn Garver[3], and Dan O. Popa[1]

[1] Dept. of Electrical Engineering, University of Texas at Arlington, 416 Yates Street, Arlington, TX 76011 USA
isura.ranatunga@mavs.uta.edu, popa@uta.edu

[2] University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX 76107, USA
{nicoleta.bugnariu,rita.patterson}@unthsc.edu

[3] Autism Treatment Center, 10503 Metric Drive, Dallas, TX 75243, USA
cgarver@atcoftexas.org

Abstract. In this paper we combine robot control and data analysis techniques into a system aimed at early detection and treatment of autism. A humanoid robot - Zeno is used to perform interactive upper body gestures which the human subject can imitate or initiate. The result of interaction is recorded using a motion capture system, and the similarity of gestures performed by human and robot is measured using the Dynamic Time Warping algorithm. This measurement is proposed as a quantitative similarity measure to objectively analyze the quality of the imitation interaction between the human and the robot. In turn, the clinical hypothesis is that this will serve as a consistent quantitative measurement, and can be used to obtain information about the condition and possible improvement of children with autism spectrum disorders. Experimental results with a small set of child subjects are presented to illustrate our approach.

1 Introduction

Autism Spectrum Disorder (ASD) is a developmental disorder characterized by deficiencies in social interaction, speech, cognition, motor coordination and imitation [4]. According to the Autism and Developmental Disabilities Monitoring (ADDM) Network of the Centers for Disease Control and Prevention (CDC) an average of 1 in 88 children in the US is diagnosed with an ASD [2]. Although the cognitive capacity of individuals in the Autism spectrum vary greatly, most of the individuals have sensorimotor abnormalities. Although Autism was recognized as early as 1943 by Kanner and accepted to have a biomedical origin by the 1980s, there is a lack of quantitative diagnosis tools. Currently ASD diagnosis mainly focuses on qualitative behavior observation which results in imprecise and sometimes arbitrary categorization of individuals in the Autism spectrum [4,14].

G. Herrmann et al. (Eds.): ICSR 2013, LNAI 8239, pp. 218–228, 2013.

Robotic systems have been developed for use in the therapy of individuals in the autism spectrum such as FACE, AuRoRa, Kaspar, Nao, and Keepon [14,15], but many of them do not engage in dynamic gestural interaction in a truly autonomous, interactive manner. Studies show that the appearance of the robot plays an important role in how children relate to and interact with such robots [11, 10, 12] and suggest that imitation and turn-taking are types of interactions useful in motivating and engaging children with ASD. In these types of projects, the interaction capacity of the robot is restricted due to lack of objective criteria to rate imitative gestural Human-Robot Interaction (HRI). A robot called Bandit was used to guide older adults to perform imitative exercises [3]. This project had a robot perform upper body gestures that the subject imitated, performance criteria were related to achievement of target poses. Projects involving robots interacting with individuals, specifically children with ASD tend to be interdisciplinary projects that have a small number of participants, which makes a good clinical conclusion difficult. These projects and others involving the use of robots for assisting humans in a social, collaborative setting can be considered part of a relatively new field called Socially Assistive Robotics (SAR) [14].

Our recent multi-institutional project aims to create the technological and clinical basis for a novel Human-Robot Interaction System - RoDiCA, which will be used in the future as an early diagnostic and treatment tool for children with ASD. The clinical hypothesis is that motor developmental problems including imitation are present and can be used in early identification and diagnosis of ASD. The technological hypothesis is that a humanoid robot can motivate children with ASD to engage in motor activities and these interactions can be analyzed for diagnosis and therapy.

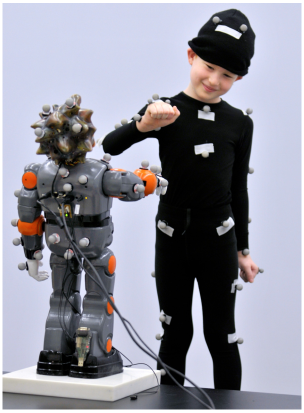

Fig. 1. Zeno and Child during an Experiment - Photo credit: Fort Worth Star-Telegram/Max Faulkner

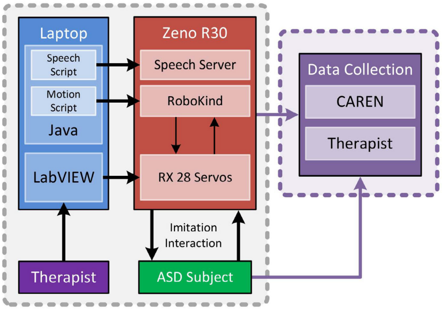

Fig. 2. System diagram

This project builds on our HRI work in [7, 8, 16, 9]. We have integrated the functionalities developed by the NGS Lab at UT Arlington, Hanson RoboKind, and the Osteopathic Heritage Foundation Physical Medicine Core Research Facility (OHFPMCRF) at UNT Health Sciences in a HRI framework that enables us to study the interaction of human subjects with the Zeno robot.

According to Baer et al. [1] any behavior whose structure follows that of a model functionally and is close to the model temporally can be considered imitation. In this respect they are specific about what constitutes imitative behavior; the imitator should be following cues from the model. From this definition it is clear that imitation does not generate the exact behavior that is exhibited by a model. One of the points being that the structure or topography of the action is given more importance than the temporal similarity of the behavior. This means that we can assess the quality of imitation by ignoring some temporal differences in the observed behavior.

The contribution of this paper, is the development of a robotic system which uses Dynamic Time Warping (DTW) as a tool for the diagnosis and treatment of Autism. DTW [5, 13] was used extensively in the speech processing community, and is a dynamic programming algorithm, which gives a distance measure that is locally temporally invariant. It has also been used extensively by the data mining community recently [6]. Here, we use DTW as a similarity measure for comparing arm motions initiated by robot Zeno and imitated by children. Results show that the DTW similarity measure can serve as both a meaningful and objective measure for evaluating the HRI quality. In the future, more trials will test the clinical hypothesis of using DTW as a tool to identify a motor marker for ASD.

This paper is organized as follows: section 2 describes the experimental platform, data collection and provides details about the subjects of the study; section 3 describes the method of data analysis and the development of DTW as a tool to identify a motor marker for ASD; section 4 shows preliminary results from the experiment and the initial reactions from ASD children; section 5 concludes the paper and describes our future work.

2 System Description

Project RoDiCA aims at developing a new motor cortex for the Zeno humanoid robot, as well as data collection and processing components to enable the early detection and treatment of ASD [9]. An overall system diagram of the HRI system including Zeno, the therapist, child and the associated data collection environments is shown in Fig. 2.

2.1 Humanoid Robot

Zeno is a 2 foot tall articulated humanoid robot with an expressive human like face shown in Fig. 1. It has 9 degrees of freedom (DOF) in the upper body and arms, an expressive face with 8 DOF, and a rigid lower body [16, 9]. The robot is capable of moving the upper body using a waist joint, and four joints each on the arms implemented using Dynamixel RX-28 servos. It has a 1.6 GHz Intel Atom Z530 processor onboard and is controlled by an external Dell XPS quad core laptop running LabVIEW [9].

Two different modes of interaction are implemented on Zeno. The first is called 'Dynamic Interaction,' which uses a Kinect sensor and LabVIEW to allow full teleoperation control of the arms and waist DOFs. This allows a therapist to interact with the child through the robot, or allows the child to control the robot directly. The second mode called 'Scripted Interaction,' is based on the Zeno RoboKind software which allows preprogrammed motions and conversations using text to speech software. In this paper we look at the data captured during both scripted and dynamic interaction modes where the child imitates Zeno's motion.

2.2 Data Collection and Testing of Human Subjects

The data collection is performed using a motion capture system from Motion Analysis Corp, Santa Ana, CA, and CAREN, a virtual reality system from Motek Medical BV described in detail in previous work [9]. This system consists of cameras that capture motion at 120 Hz. In this work the human subject and the robot Zeno are both instrumented using 40-50 reflective markers.

For the present pilot study interaction experiments were conducted with Zeno facing the subject as seen in Fig. 1. A scripted routine of gestures was run on Zeno, which the child was prompted to imitate using Zeno's speech functionality. The following 42 second scripted routine from [9] was run on Zeno:

1 Wave hello with right arm [8.3s to 15.8s]
2 "Tummy rub" with right arm [16.8s to 25.3s]
3 Fist bump right with arm [25.7s to 30.3s]
4 Fist bump left with arm [31.0s to 34.5s]
5 Wave goodbye with left arm [34.5s to 42.0s]

For this phase of the analysis, we only used the time series of the joint angles of Zeno's and the subject's motion during a right handed wave. Similarity in motion between child and robot is measured using the joint angles, since Cartesian positions are more difficult to compare due to dimensional and pose differences e.g. robot may be much smaller than child, and rotated in space to face each other.

3 Method

In this section, we describe how the data is processed, describe the DTW algorithm and propose the development of DTW as a tool to identify a motor marker for ASD.

3.1 Data Representation and Inverse Kinematics

In this paper, we record the imitative behavior using the four joint angles of the arm as depicted in Fig. 3. The angles are α and β in the shoulder joint and γ and θ in the elbow joint of Zeno. The data from the motion capture system is in the form of Cartesian joint positions of the shoulder $P_s = (x_s, y_s, z_s)$, elbow $P_e = (x_e, y_e, z_e)$ and the hand $P_h = (x_h, y_h, z_h)$ respectively. The joint angles are then calculated from these positions using the trigonometric equations below.

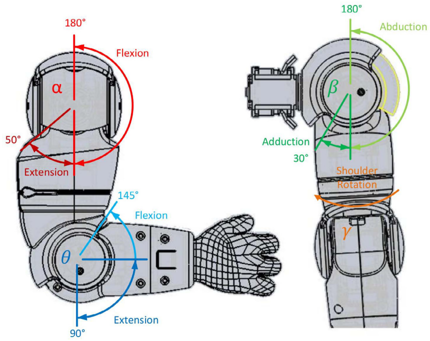

Fig. 3. Zeno's Arm Angles

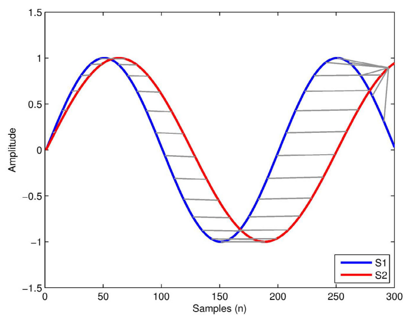

Fig. 4. Example DTW match between signals S_1 and S_2

We define n_c the surface normal to the plane containing P_s, P_e and P_h, $n_c = V_{se} \times V_{eh}$, V_{se} is the vector from P_s to P_e, V_{eh} the vector from P_e to P_h, V_{eh} the vector from P_e to P_h and V_{es} the vector from P_e to P_s.

The following equations are used to obtain the joint angles:

$$\alpha = tan^{-1}\left(\frac{V_{se}(3)}{V_{se}(2)}\right), \tag{1}$$

$$\beta = cos^{-1}\left(\frac{V_{se}(1)}{|V_{se}|}\right), \tag{2}$$

$$\gamma = cos^{-1}\left(\frac{n_i \cdot n_c}{|n_i||n_c|}\right) \mathrm{sgn}(\begin{bmatrix}1 & 0 & 0\end{bmatrix}(n_i \times n_c)), \tag{3}$$

$$\theta = cos^{-1}\left(\frac{V_{eh} \cdot V_{es}}{|V_{eh}||V_{es}|}\right), \tag{4}$$

Where $V_{se}(1)$, $V_{se}(2)$ and $V_{se}(3)$ are the x, y and z components of V_{se} respectively, and n_i is defined as:

$$n_i = R_{\alpha\beta}\begin{bmatrix}0\\0\\1\end{bmatrix} \tag{5}$$

$$R_{\alpha\beta} = \begin{bmatrix} cos(\beta) & -sin(\beta) & 0 \\ cos(\alpha)sin(\beta) & cos(\alpha)cos(\beta) & -sin(\alpha) \\ sin(\alpha)sin(\beta) & sin(\alpha)cos(\beta) & cos(\alpha) \end{bmatrix} \quad (6)$$

Since the imitation between Zeno and the human is mirrored, Zenos right arm is compared to each subjects left arm angle trajectories. This data is then preprocessed by z-normalization, suggested by various researchers as an important step to be performed before running any tests [6]. The z-normalization removes offsets and scaling issues in the data, see following equation:

$$\boldsymbol{s_z} = \frac{\boldsymbol{s} - \mu_s}{\sigma_s} \quad (7)$$

where μ_s and σ_s are the mean and standard deviation of the signal $\boldsymbol{s}$.

3.2 Motion Analysis Using Dynamic Time Warping

There is a need in the Autism research community to obtain quantitative measurements of imitation quality. We propose using the DTW algorithm to obtain a similarity measure between time series joint angle signals. DTW is an established signal processing method that offers a distance measure between signals similar to the Euclidean distance. However, time-warping is applied to signals to align them optimally, prior to taking the difference. Optimal alignment in this context, is the alignment of the signal time samples that makes the total distance between the signals as small as possible. This alignment induces a non-linear mapping between the two signals, e.g. warping of the signals. A good description of the DTW algorithm is given by Keogh et al. [6]. The strength of DTW is in its ability to compare the similarity between signals by ignoring time-delays and uneven time sampling. This situation is very relevant in the context of our problem, since the motion of child and robot experiences both these effects.

Given two signals that are time dependent $\boldsymbol{X} = \{x_1, x_2...x_n\}$ and $\boldsymbol{Y} = \{y_1, y_2...y_n\}$ where $n \in \mathbb{N}$. DTW finds the optimal distance between these two signals using a local distance measure d, we use the euclidean distance as the distance measure $d(\boldsymbol{X}, \boldsymbol{Y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$.

The DTW cost $\mathfrak{D}(n, n)$ can be calculated using the following recursion:

$$\begin{aligned} \mathfrak{D}(i,j) = d(s_i, r_j) + min\{&\mathfrak{D}(i-1, j-1), \\ &\mathfrak{D}(i-1, j), \\ &\mathfrak{D}(i, j-1)\} \end{aligned} \quad (8)$$

DTW is applied to each set of angle trajectories generating a DTW value for each of the four angles. A range of motion is then obtained in order to weigh each angle using equation 9, where $\boldsymbol{X}$ represents a column vector of unnormalized joint angles:

$$W = max(\boldsymbol{X}) - min(\boldsymbol{X}). \quad (9)$$

The combined DTW distance for all four angles is calculated by a weighted average as shown in 10.

$$A_w = \frac{W_\alpha D_\alpha + W_\beta D_\beta + W_\gamma D_\gamma + W_\theta D_\theta}{W_\alpha + W_\beta + W_\gamma + W_\theta}, \tag{10}$$

where,

W_* - weight per joint
D_* - calculated DTW distance for each joint

The pseudo-code used for implementing our version of DTW is shown in Algorithm 1.

Algorithm 1. DTW Cost
H

```
procedure dtwCost(S,R)
    m ←row(S)
    n ←row(R)
    𝔇 ← zeros(n + 1, m + 1)
    ℭ ← 0
    𝔇(1,1) ← 0
    for i ← 2, m do
        𝔇(i,1) ← 𝔇(i-1,1) + d(S(i),R(1))
    end for
    for j ← 2, n do
        𝔇(1,j) ← 𝔇(1,j-1) + d(S(1),R(j))
    end for
    for i ← 2, n do
        for j ← 2, n do
            ℭ ← d(S(i),R(j))
            δ ← min(𝔇(i-1,j),𝔇(i,j-1),𝔇(i-1,j-1))
            𝔇(i,j) ← ℭ + δ
        end for
    end for
    return 𝔇(m,n)
end procedure
```

Fig. 4 shows a typical result when using the DTW algorithm from 1, the gray lines depict the nonlinear map between the signals. It can be seen from the right side of Fig. 4, where many points from S_1 are mapped to a single point of S_2, this is because DTW matches every point of the signals together.

4 Results

A set of clinical experiments with several children were performed to test the robot, capture the motion, and to compare the motion data using DTW. During experiments, children were directed by the robot to follow along performing several hand gestures, such as wave with both hands, tummy rub, fist bump, etc. Each motion was performed three times. The best one was picked for analysis.

Table 1. DTW distance per joint

Subject	**Age**	**Gender**	**Type**	α	β	γ	θ	β θ **Combined**
3	6	Male	Control	0.4421	0.6730	0.2775	0.1145	0.3075
2	6	Male	ASD	1.2128	1.2128	1.1359	0.7232	0.8913
12	9	Male	Control	0.6436	0.0792	0.5253	0.3073	0.2247
4	9	Male	ASD	0.3387	0.6738	0.6899	0.9251	0.7331
13	11	Male	Control	0.6164	0.2267	0.6409	0.2854	0.2666
6	11	Male	ASD	0.4473	0.1881	0.4319	0.3228	0.2757
11	12	Male	Control	1.1141	0.2583	0.5157	0.1159	0.1636
7	12	Male	ASD	0.2350	0.6023	0.4192	0.2091	0.4102

Wave gestures were compared for four different age groups (6, 9, 11, 12) as seen in Table 1, they consist of pairs of control and ASD children recruted from the Dallas Autism Treatment Center. Each age group consists of a control subject and an ASD subject. Representative joint angle trajectories for each DOF are shown in Fig. 5. The angle trajectories of subject 12 gives insight into the significance of each joint in performing each action. Here the elbow angle θ has the best match during the imitation, this is because the elbow performs a periodic motion when executing a wave, so it is easy to script and imitate. β which is the other main angle used in a hand wave is also fairly similar. The difference seen in Fig. 5 for the normalized angles varies more in the human trajectory because it is difficult for humans to keep their shoulder in the exact place over a period of time like the robot. The computation of the DTW for the different subjects performing the wave motion is shown in Table 1 and summarized in Fig. 6. The weights for α and γ in Equation 10 are set to zero to calculate a weighted average using only β and θ, this is because β and θ angles are more representative for the motions performed in this study.

By looking at this data we can see that the combined average for the control subjects are all lower than for the ASD subjects. This shows that DTW in combination with weighting based on range of motion for the most important angles, β and θ, for a hand wave shows a promising method of comparing human to robot imitation. Also, notice by looking at Table 1 the DTW values for β and θ also are consistent for each age group with the exception of β for the group of eleven year olds. This discrepancy is simple due to a better imitation by the ASD subject for β, but as mentioned before β will always have variation due

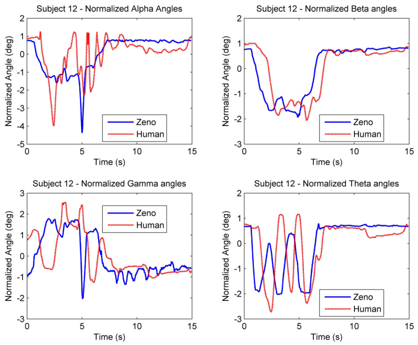

Fig. 5. Normalized Joint Angles

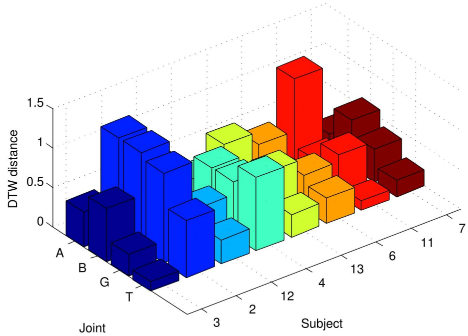

Fig. 6. DTW values

to human nature. However, this inconsistency helps show that the value of the overall performance is not affected by one angle, but instead provides a consistent value based on both angles used. This demonstrates that DTW is a good method to use in motion comparison analysis.

These preliminary results show that the DTW similarity measure can serve as both a meaningful and objective measure for evaluating the quality of imitation behavior. With age all children improve their imitation behavior, but the children with ASD consistently perform worse than their age-match controls. The combined weighted joint average for the wave motion has a higher value in all ASD children, indicating that their imitation of Zenos waving motion is less accurate.

5 Conclusions and Future Work

The contribution of this paper is the development of a robotic system which uses Dynamic Time Warping (DTW) as a tool for the diagnosis and treatment of Autism. This paper describes the development of a robotic system capable of measuring the quality of imitation interaction between a humanoid robot and a human subject. A humanoid robot called Zeno, is used to perform gestures which the human subject imitates. The similarity of the gestures performed is measured using DTW, and used to objectively analyze the quality of the imitation interaction between the human and the robot. The hope is that this type of system will enable consistent objective measurement of the quality of imitation, and can be used to obtain information about the condition and possible improvement of children with ASD.

From the graphs of the wave motion it is clear that the robot motion should be derived from nominal human motion trajectories. The initial data analysis shows that DTW can be a good tool for comparing imitation interaction since it allows the comparison of temporally inexact imitative motion. However, it is not clear what robot motions to select, and what DOFs to compare using DTW in order to best affect ASD treatment.

In the future we will comprehensively evaluate the DTW algorithm in a larger cross-longitudinal study. The motions generated by the Zeno robot will be made more natural and human like. Data collection, natural human motion generation and analysis will be contained in the newly developed motor cortex of the Zeno robot. The imitation script will be modified to increase its ability to engage the children with autism. The analysis of imitation will be extended to facial gestures.

Acknowledgment. This work was supported by the TxMRC consortium grant: Human-Robot Interaction System for Early Diagnosis and Treatment of Childhood Autism Spectrum Disorders (RoDiCA) and by the US National Science Foundation Grant CPS 1035913.

The authors wish to thank David Hanson, Richard Margolin, Matt Stevenson and Joshua Jach of Hanson RoboKind and Intelligent Bots LLC for their help with Zeno R-30. The authors thank Robert Longnecker, Janet Trammell, David Cummings and Carolyn Carr from UNTHSC for their help with the experimental data collection presented in this paper. The authors also thank Abhishek Thakurdesai from the NGS group for help with Zeno programming.

References

1. Baer, D., Peterson, R., Sherman, J.: The development of imitation by reinforcing behavioral similarity to a model. Journal of the Experimental Analysis of Behavior 10(5), 405 (1967)
2. Centers for Disease Control and Prevention: Autism and developmental disabilities monitoring (addm) network (Sep 2012), `http://www.cdc.gov/ncbddd/autism/addm.html`
3. Fasola, J., Mataric, M.: Using socially assistive human-robot interaction to motivate physical exercise for older adults. Proceedings of the IEEE 100(8), 2512–2526 (2012)
4. Geschwind, D.H.: Advances in autism. Annual Review of Medicine 60(1), 367–380 (2009)
5. Itakura, F.: Minimum prediction residual principle applied to speech recognition. IEEE Transactions on Acoustics, Speech and Signal Processing 23(1), 67–72 (1975)
6. Keogh, E., Ratanamahatana, C.A.: Exact indexing of dynamic time warping. Knowledge and Information Systems 7, 358–386 (2005), doi:10.1007/s10115-004-0154-9
7. Rajruangrabin, J., Popa, D.: Robot head motion control with an emphasis on realism of neck-eye coordination during object tracking. Journal of Intelligent & Robotic Systems 63(2), 163–190 (2011), doi:10.1007/s10846-010-9468-x
8. Ranatunga, I., Rajruangrabin, J., Popa, D.O., Makedon, F.: Enhanced therapeutic interactivity using social robot zeno. In: Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments, PETRA 2011, pp. 57:1–57:6. ACM, New York (2011)
9. Ranatunga, I., Torres, N.A., Patterson, R., Bugnariu, N., Stevenson, M., Popa, D.: Rodica: a human-robot interaction system for treatment of childhood autism spectrum disorders. In: Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments, PETRA 2012. ACM, New York (2012)

10. Ricks, D.J., Colton, M.B.: Trends and considerations in robot-assisted autism therapy. In: 2010 IEEE International Conference on Robotics and Automation (ICRA), pp. 4354–4359 (2010); iD: 1
11. Rinehart, N., Bradshaw, J., Brereton, A., Tonge, B.: Movement preparation in high-functioning autism and asperger disorder: A serial choice reaction time task involving motor reprogramming (2001)
12. Robins, B., Dautenhahn, K., Dubowski, J.: Does appearance matter in the interaction of children with autism with a humanoid robot? Interaction Studies 7(3), 509–542 (2006)
13. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on Acoustics, Speech and Signal Processing 26(1), 43–49 (1978)
14. Scassellati, B., Admoni, H., Mataric, M.: Robots for use in autism research. Annual Review of Biomedical Engineering 14(1), 275–294 (2012), doi:10.1146/annurev-bioeng-071811-150036
15. Tapus, A., Peca, A., Aly, A., Pop, C., Jisa, L., Pintea, S., Rusu, A.S., David, D.O.: Children with autism social engagement in interaction with nao, an imitative robot. Interaction Studies 13(3), 315–347 (2012)
16. Torres, N.A., Clark, N., Ranatunga, I., Popa, D.: Implementation of interactive armplayback behaviors of social robot zenofor autism spectrum disorder therapy. In: Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments, PETRA 2012. ACM, New York (2012)