The 2009 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 11-15, 2009 St. Louis, USA

Therapeutic-Assisted Robot for Children with Autism

P. Ravindra S. De Silva, Katsunori Tadano, Azusa Saito, Stephen G. Lambacher, and Mastake Higashi

Abstract— **In this paper, we propose a therapeutic-assisted robot for children with autism to ameliorate their skill of joint attention. The robot conducts a goal-directed based interaction to establish engagement between the child and robot in order to establish a beneficial learning environment for autistic children. An unsupervised Mixture Gaussian-based cluster method is proposed to detect the child's intention in real time to process the goal-directed task smoothly. The novelty of this approach is that does not require the use of any training data or a trained model to detect the child's intention. Our autonomous robotic system is tested with several autistic children at a School for the Disabled in Nagoya, Japan. The results of the initial interaction showed that the children enjoyed interaction with and feedback from the robot, which confirmed that the robot can be used as mediator or an object of joint attention. The unsupervised approach was able to detect the children's intention at every time segment to process the goal-directed task with a higher accuracy rate. The results of the goal-directed task showed that the proposed interaction was highly effective in enhancing their joint attention. Since most of the children attempted to imitate the robot's gestural behaviors and used a variety of learning patterns to attend to the robot's fingered object in the environment to obtain joint attention with robot.**

I. INTRODUCTION

Both computer-based and robotic systems have been used to show that it is possible to create a learning environment to improve the skills of autistic children in the social interaction and understanding. Children diagnosed with autism suffer from Autistic Spectrum Disorder (ASD) to varying degrees. ASD typically impairs a child's non-verbal and verbal communication skills, making it difficult for them to maintain eye contact, to understand other people's emotions, intentions, and behaviors. New research has been carried out in the use of interactive technologies [1][2] for the purpose of creating a beneficial learning environment as a therapeutic device to enhance autistic children's impaired skills [3][4]. These systems have focused on developing an interactive scenario to enable autistic children to engage in and enjoy technologies to improve their natural social interaction [5][6].

Michud [7] developed a variety of mobile robots for children with autism to encourage them to establish social interaction with a robot. The children acquired a simplified, safe, and predictable environment to enjoy a robot's feedback and interactions. Nadel [8] developed a robotic system which is capable of having a simple imitation interaction using a turn-taking game to encourage social interaction. Most existing robotic systems are used for establishing child-robot interaction for attaining social interaction with a robot [9][7]. Recently, autistic researchers have investigated the factors leading to impaired skills in non-verbal communication. Siller & Sigman [10] have discovered that a lack of skill in joint attention (JA) has a strong effect on the impaired non-verbal skills of autistic children. An incapability of following a partner's gaze, head turning, and eye contact have been disclosed as factors leading to a lack of joint attention skills, since the improvement of a child's joint attention has become fundamental point to improve their non-verbal communication skills, which are essential for social learning.

The design of desirable interactive scenarios (contexts) is a big challenge. In particular, it is difficult to infer and generalize scenarios to increase the impaired skills of autistic children [2]. Indeed, each child has different impaired skills and their own interests, which are difficult to understand unless their behaviors are observed long-term. Robins [11] conducted various interactive scenarios for autistic children to examine their patterns of engagement, interest, and joint attention. This study attempted to build interaction between a child, caregiver, and robot for the purpose of increasing the child's social skills. The robot expressed both a variety of physical and auditory motions to obtain the child's attention, and a caregiver pointed out the robot's performance to achieve the child's joint attention. Here, the robot was used as the object of joint attention. Each of the experiments was recorded as video, with the behaviors of the child corresponding to the scenarios later being analyzed. But a study by [12] suggested that it is important to consecutively increase the complexity of toys, e.g., the interaction. Robins [11] also points out that it is important to adapt or provide feedback from the robot by considering the current behaviors of the child. In this study, the robot lacked the ability to trace the child's eye gaze data or to recognize their behavior in real time, which caused the robot to gradually change the complexity of the interaction by considering the child's performances and behaviors.

Most existing robotic systems have explored the use of a toy-based robot as a social partner for children with autism. In recent decades, there has been a need to develop a robotic system capable of having complex interaction patterns and feedback for processing therapeutic intervention. An important feature in the therapeutic process is for the robot to autonomously change the interaction and provide feedback according to the behaviors of the child [12]. In this sense, the robot has to recognize the current attention of the child according to their eye gaze information.

P. Ravindra S. De Silva, Katsunori Tadano, Mastake Higashi are with Computer Aided Design Lab, Toyota Technological Institute, Japan (ravindra, sd07411, higahsi)@toyota-ti.ac.jp

Azusa Saito is with the Department of Software Engineering,University of Aizu, Japan d8082201@u-aizu.ac.jp

Stephen G. Lambacher is with the Aoyama Gakuin University, Japan slambacher@si.aoyama.ac.jp

Our approach is toward developing a child joint attention through a robot-assisted therapeutic process based on interactions between the child, the teacher, and the robot. Considering the aforementioned challenges, we propose an interactive scenario with an unsupervised clustering method to detect a child's intentions in real time for the purpose of processing the interactions without the use of training data. The proposed robotic system is helpful for use with any autistic children experiencing varying levels of impaired joint attention. The proposed unsupervised approach is capable of detecting a child's intention when the child has a complex eye gaze pattern without the use of a trained model. The robot adjusts to various interactive styles, provides feedback, and gradually changes its interaction pattern based on the performance of the child's joint attention.

II. Scenario for Robot-Assisted Interaction

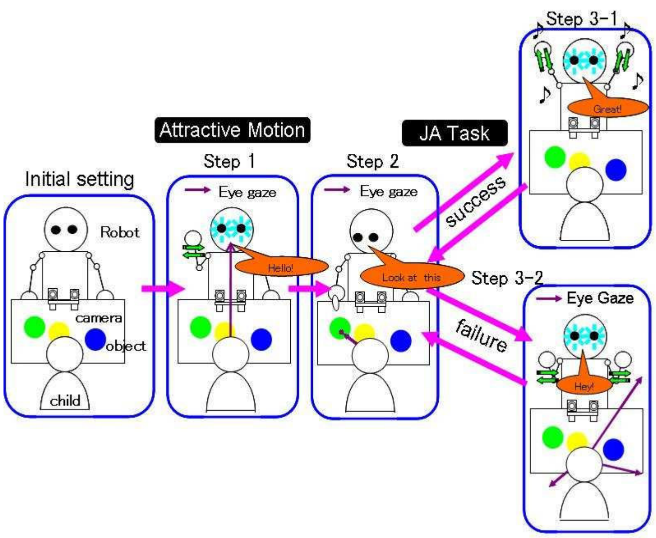


Fig. 1. The steps of the proposed interactive scenario for improving the JA skills of autistic children.

Tove [13] found that children feel awkward in maintaining eye contact in human-child interactions. However, while they attended to interactions with the robot, they did not seem to have any difficulties looking into artificial eyes (i.e., the robot's eyes) - the reason being that humans used several communication channels (gesture, voice, eye, and facial) when they communicated, while the robot had fewer communication channels with which to easily understand for autistic children. Thee factors we can used to create a better learning environment through robotic systems. Indeed, we can create an interaction-based robotic system as a therapeutic device for children with autism.

Our learning process is based on interactions between the teachers, child, and robot (Figure 1). In the initial stage, however, the robot performs a dance motion, music, and vocal interaction to arouse the child's interest. During this stage, the teacher also interacts with the child to escalate the child's interest. The robot interaction is divided into two phases: an attraction phase and a joint attention (JA) task phase. During the first phase, the robot gives a formal welcome by calling out each child's name, introducing itself, dancing to music, and administering a type of auditory interaction to attract the children's attention. During the process, the robot records the child's eye gaze data and uses an unsupervised approach to detect the child's intention during every time segment. After the robot attracts a child's interest, it gradually increases the complexity of the interaction by introducing the JA task to the child.

The JA task is based on following process: first, the robot fingers an object on a table and the child must attend to a similar object that has been pointed out by the robot. In the initial stage, a teacher also participates in the interaction. If the child does not understand the process, the teacher provides support to help the child attend to the object. If the child achieves JA with the robot, the robot expresses a joyful motion and asks the child to perform additional tasks. If the child is unable to achieve JA with the robot, the robot repeats a similar interaction. During each stage, the child's eye gaze data is recorded along with the JA performance, including the time it took the child to look at an object and the time during which the child was attracted to the robot. According to the child's JA performance, the robot attempts a different kind of trial by changing the direction of the object's position.

III. Tracking a Child's Attention from Eye-gaze Vector

Our approach traces eye-gaze information from a camera located on the table (fixed camera), which is more useful to trace their eye gaze information continuously instead of using the robot's eye camera. The reason for this is that during the interaction the robot must move its head, which makes it impossible to trace child's eye gaze behavior continuously. This type of strategy must always be used when dealing with disabled children. The robot reacts as it obtains eye gaze information from the robot's eye camera, rotating its head and gestures toward to the child's eye gaze direction. To estimate the eye gaze information, images from two cameras were analyzed in real-time, and reference features were set up on the face. Estimated at a rate of 60 Hz, these features are used for tracking the corners of the eyes and mouth. The FaceLab system uses two cameras to help estimate the 3D coordinates of each feature. Verification of the gaze direction is based on both head pose and position of the irises of the eyes. By calculating the position of the iris, the horizontal and vertical rotations can be estimated, i.e., the gaze direction. Instead of considering both the left and right eye data separately, we were interested in estimating just a single eye gaze vector to represent both the left and right eye data. With the position of right eye iris as $(P_{RE} = (x_{RE}, y_{RE}, z_{RE}))$ and the position of the left eye iris as $(P_{LE} = (x_{LE}, y_{LE}, z_{LE}))$, the average position of irises is estimated as $P_E = (P_{RE} + P_{LE})/2$. We used P_E as an origin of the gaze vector. The angles of the irises were estimated through the average pitch and yaw of both the left and right irises' angles: $[\theta_P = (\theta_{RP} + \theta_{LP})/2]$,

$[\theta_Y = (\theta_{RY} + \theta_{LY})/2]$, where the right iris pitch θ_{RP}, left iris pitch θ_{LP}, right iris yaw θ_{RY}, and left iris yaw θ_{Ly}. Then, the vector gaze is estimated as

$$V = (-\sin\theta_Y \cos\theta_P, \sin\theta_P, -\cos\theta_Y \cos\theta_P)^T \quad (1)$$

In the current scenario, we were interested in estimating the child's attention on a shared visual space (table environment). The crossing point of the eye gaze vector and table space were considered as the current attention of the child, which is estimated through

$$P_{cross} = P_E + V. \quad (2)$$

The eye gaze was used to acquire a sufficient variation in eye gaze data $(P_{cross(x)}, P_{cross(z)})$, which helped to precisely detect the current intention of the child.

A. An Unsupervised Approach to Detect a Child's JA to the Object

Existing approaches for obtaining a robot's joint attention utilize a considerable amount of a caregiver's data in order to construct a learning algorithm to detect a caregiver's intention [14][15][16]. Moreover, these approaches are only capable of detecting a caregiver's intention using simulated head pose patterns. Nevertheless, the training and creating of an accurate model for a complex eye gaze pattern is still difficult, because a complex eye gaze pattern may contain a combination of several eye gaze patterns that attend to several objects. Consequently, it is still difficult to construct boundaries for model parameters in order to classify a child's object of interest according to eye gaze data. A difficult task in developing a robotic system for children with autism is the need to collect eye gaze data to train a recognition model of the child's intention. The reason is that it is difficult to provide instructions and for the child to remain at the same place long enough to obtain training data from them. Also, due to the unpredictability of a child's eye gaze patterns, it is difficult to train a machine-learning algorithm to detect their intentions. In light of these realities, our approach segments the eye gaze data for the purpose of predicting the current state of attention based on a Mixture Gaussian-based unsupervised clustering. The novel part of this research is the development of an unsupervised approach to detect children's intentions with the performance of their joint attention without using any training data from the autistic children.

Our proposed approach is described in Figure 2. At the present time, our approach segments the eye gaze data at every 30 frames. Each of the segments is considered as separate to apply the unsupervised approach to detect a child's intention. We used cross points $(P_{cross(x)}, P_{cross(z)})$ of the child's eye gaze vector in the table space. The cross point data includes an initial number of clusters and are used as input data in the agglomerative clustering algorithm. The estimated clusters are considered as a component of the Mixture Gaussian, and the parameters (mean and variance)

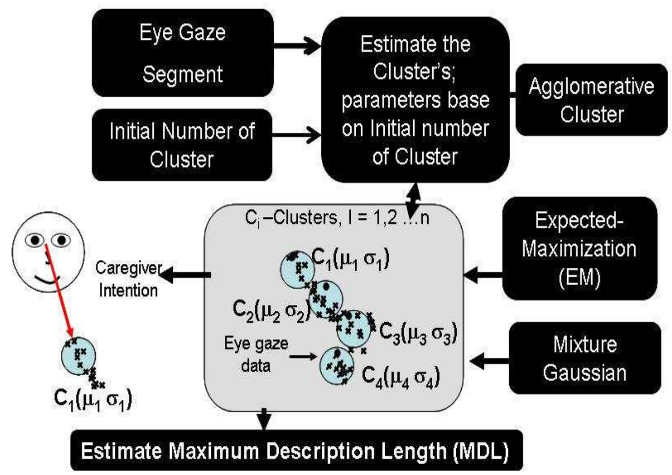


Fig. 2. The proposed unsupervised Mixture Gaussian Cluster for detecting a child's attention from the eye gaze segment, $C_1(\mu_1, \sigma_1)$, represent the mean (μ_1) value and standard deviation (σ_1) for cluster 1.

are estimated using the Expectation-Maximization (EM). Consequently, we repeat the same procedure, merging the close clusters to find the optimum number of clusters for a given eye gaze segment. To abridge the number of clusters, all pairs of clusters (the mean value of the cluster) are compared in order to find the closest clusters. If the distance between the mean values of the clusters (ℓ_{ij}) is less than the threshold distance ($\ell_{ij} < \delta$), these groups become linked and are merged into a single group. The distance between the mean values are computed through the following equation, where μ_i and μ_j are the mean values of cluster C_i and C_j. Each of the steps is used to estimate the Minimum Description Length (MDL), and the minimum of the MDL is used to decide the number of clusters that are optimum to the child's eye gaze.

$$\ell_{ij} = \{d(\mu_i, \mu_j) : \mu_i \epsilon C_i, \mu_j \epsilon C_j\} \quad (3)$$

The MDL can be estimated by the following equation, where $P_{y_n|x_n}(y_n|k, \lambda)$ is the probability density function for the eye gaze data, L is the number of continuously valued real numbers required to specify the parameter λ, M is the number of the dimension in an eye gaze vector, and N is the number of eye gaze frames in the segment.

$$MDL(K, \lambda) = -log(P_{y_n|x_n}(y_n|k, \lambda)) + \frac{1}{2} L log(NM) \quad (4)$$

The probability density function $P_{y_n|x_n}(y_n|k, \lambda)$ is estimated using the following steps. Suppose Y is the $2-$*dimensional* random vector (cross point in the table, $(P_{cross(x)}, P_{cross(z)})$) that can be the model using a GM distribution. Let us assume that the current model contains K clusters. Then, the following parameters are necessary for estimating the parameter of the k^{th} cluster: π_k is the probability that a cross point data (eye gaze vector) has cluster k, μ_k is the dimensional mean vector for cluster k, and R_k is the 2×2 covariance matrix for cluster k. In addition, when K denotes the number of the cluster, we use the notation of π, μ, and R to denote the parameter sets $(\pi_k)_{k=1}^K$, $(\mu_k)_{k=1}^K$, and $(R_k)_{k=1}^K$. K and $\lambda = (\pi, \mu, R)$ denote the set of all parameters for defining the MG distribution. Here, each of the clusters is annotated to predict the child attention according to the eye gaze data, which are crossed with the table space. Suppose $Y_1, Y_2 ... Y_N$ (where $Y_i = (P_{cross(x_i)}, P_{cross(z_i)})$) is N sample size of the cross point from the information class (selected time interval data) of interest. Also, assume that for each cross point Y_i the cluster of that cross point is given by the random variable X_n, which is usually unknown but useful for analyzing the problem. Here, we will assume that each cluster has a multivariate Gaussian distribution, which is the probability density function for the cross point Y_n, assuming that $X_n = k$ is given by

$$P_{y_n|x_n}(y_n|k,\lambda) = \frac{1}{(2\pi)^{M/2}}|R_k|^{-1/2} \\ exp(-\frac{1}{2}(y_n - \mu_k)^t R_k^{-1}(y_n - \mu_k)) \tag{5}$$

We do not know the cluster X_n of each sample. Therefore, to compute the density function of Y_n given the parameters λ, we must apply the definition of the conditional probability and sum over k:

$$P_{y_n}(y_n|\lambda) = \Sigma_{k=1}^K P_{y_n|x_n}(y_n|k,\lambda)\pi_k \tag{6}$$

Then, the log likelihood value can be estimated considering the entire sequence by

$$\log P_y(y|K,\lambda) = \Sigma_{n=1}^N \log \Sigma_{k=1}^K P_{y_n|x_n}(y_n|k,\lambda)\pi_k, \tag{7}$$

where $Y = (Y_n)_{n=1}^N$.

Each of the clusters is characterized by the mean and variance values. The robot predicts the current state of child attention by referring to the mean value of the cluster at each time segment without any training data.

IV. Recognition of a Child's Joint Attention

After recognizing the child's attention in the environment, the robot needs to combine that information with the robot's fingered object to check the performance of the child's joint attention. To accomplish the above task, our approach creates a static geometrical model for each of the objects (Figure 3). When the child reaches for an object in the environment, eye gaze data is obtained around the area of the interested object with a d radius circle. Outwardly, each of the models has a static geometrical model that is established with the d (currently we used d value as 10cm, which is based on the preliminary experimental results) radius circle. The recent intention of the child (mean value of the cluster) is compared with the objective models at each time segment. If the current mean value is inside the object's model, then we can predict that the child has obtained joint attention with the robot. However, the robot uses the following procedure to check whether the mean value of the cluster is inside of the

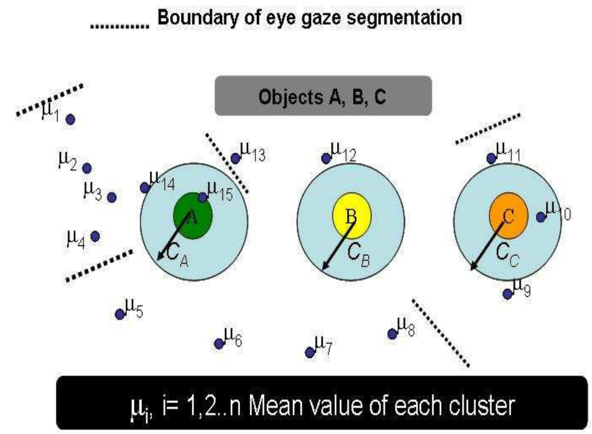


Fig. 3. The figure depicts the Mixture Gaussian-based unsupervised cluster for detecting the child's intention at each time segmentation. $\mu_i, i = 1, 2, ..n$ represents the mean values of the cluster, which are considered as the child's intention. Finally, the child's intention is compared with the object model to recognize the child's joint attention, gray hatching representing the geometrical boundary for objects.

object model: suppose the point of the mean value (child's intention) is represented as $p = (P_{cross(\mu_x)}, P_{cross(\mu_y)})$, then the center of the circle (defined geometrical model for object) is represented as $C = (c_x, c_y)$. Later, the distance (d_{pc}) of a given mean point and center of the circle are computed. The robot uses the following rule to decide the child's attention to the robot's fingered object: if $d_{pc} <= r$, $p \epsilon C_p$ and C_p, the cluster contains more than 30 frames.

V. Experiment Setup

The main focus of this research was to develop a variety of modalities for an autonomous robot to interaction with autistic children in real time. In particular, we propose an unsupervised approach to detect a child's intention with their joint attention (performances) without the use of any training data or trained model. The objective of the following experiment was to assure the effectiveness of proposed interactive scenario and performance of the proposed unsupervised modalities to improve their joint attention when interacting with robot. The interactive scenario attempts obtain a child's JA with a robot, providing leverage in improving their JA with a human, as well as help them to improve their social communication skills.

In order to carry out the experiment to examine the effectiveness of the assistive robot as a therapeutic device for autistic children, collaboration was arranged with the Tempaku School for the Disabled. The experiment included five students, two teachers, two operators, and a HOAP-3 robot. Each of the children had Autistic Spectrum Disorder (ASD), and ranged in age from ten to eleven years old. All participants had been receiving intensive behavioral intervention from the school using specially designed educational techniques. These techniques were tailored to the child's needs and level of behavioral disabilities. All the children had problems with language, possessed a limited use of non-verbal communication, and avoided social interaction and communication with other humans. In addition, all the children lacked well-developed joint attention skills and were observed to inconsistently respond to a teacher's instruction for joint attention. Each of the children participated in three interaction trials with the robot. The robot interaction was divided into two phases: an attraction phase and a JA task phase.

A. First Phase of Robot's Interaction with the Child

In the initial phase of the interaction, the robot expresses an interactive dance with vocals to attract the child's attention. The robot's attractions are very important to motivate the children to engage in interaction. The camera system enabled 60 frames per second during the robot's attraction, with a total of approximately 900 frames of child eye gaze data being recorded. A typical problem with autistic children is their inability to have eye contact and interaction with others. During the attraction, the robot records the child's eye gaze data by keeping tract of how long the frames are of the child looking at the robot, the number of frames the child's eye gaze data is not recorded, and when the children looks away from the robot. Figure 4 shows the above information of every student who participated. According to the overall results, the students spent from between 72.7%- 88.7% of the time being attracted by the robot. During the interaction, a student was sometimes too far from the camera's range so that the time system was unable to record their eye gaze data. However, the total percentage of such cases was still quite low (3.6%-15.6%). The highest number of errors (15.6%) occurred due to the child being away from the camera range. These statistics are quite remarkable, as they show that our combined design of motion, music, and vocal interactions of the robot was capable of successfully attracting the students' attention. The children had eye contact with the robot and the robot was effectively utilized as a mediator or object of JA. According to their behavioral history at the disabled school, when teachers performed assessment-based teaching, they failed to receive the children's attention. A robot's attractive points (vocal, music, and motion) help to attract children's attention, and continued engagement helps to create a better learning environment for autistic children.

B. Robot Engaged in a Joint Attention Task with a Child

After attracting the child, the robot tries to motivate the child and gradually increases the complexity of interaction as it introduces the joint attention task to the child. The robot creates a simple task for the child to achieve JA: the robot first points to an object in the environment to motivate the child to achieve joint attention to that object. According to the JA performance of the child, the robot changes the position of the object to achieve JA from different directions in the environment. In each of the trials, the robot provides attractive feedback (motion and vocal) to the child to continually engage them in interaction. To examine the performance of the proposed unsupervised algorithm, we counted how many times the robot could detect the object a child was looking at. Sixteen interactive trials were conducted. Each student typically performed three (3 * 4 students = 12 trials) trials, and there were four times when the children did not look at the object. From the 16 trials, the unsupervised approach predicted 12 trials with a 75% accuracy rate, which was quite remarkable when dealing with the autistic children in real time.

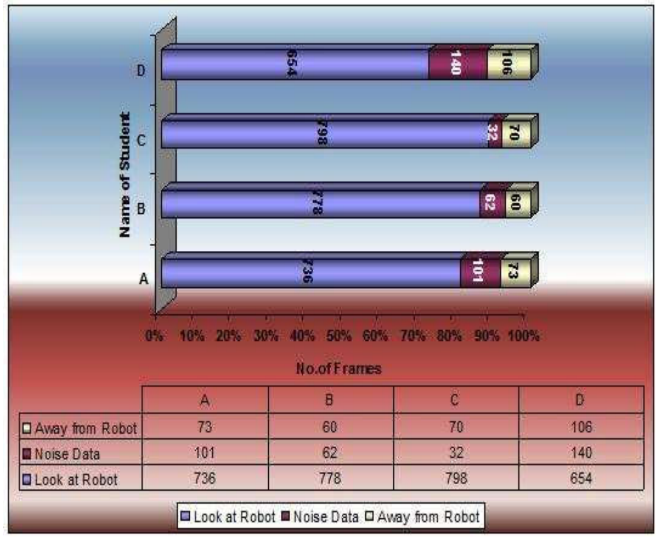


Fig. 4. Results showing the number of frames the children either looked at or away from the robot, including noisy data during the robot's attraction (during the first phase of the interaction).

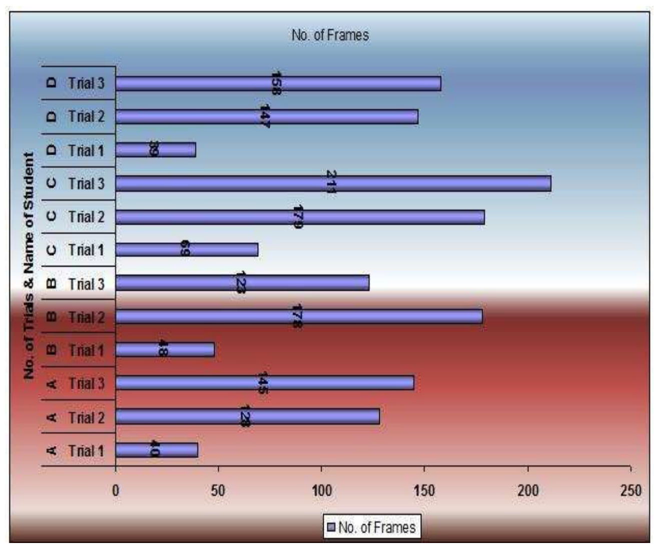


Fig. 5. A depiction of the number of times the children had joint attention with robot during each trial.

The robot required 11 seconds to process the JA task. Figure 5 shows how long (frames) the child looked at the fingered objects of the robot. The results show that all the students spent less time (number of frames) in the initial trial. However, in the next two trials their times gradually increased (number of frames) except for student B during $trial2$. The key point to consider here is that when the children have adapted to the robot, they more comfortably exposed themselves and enhanced their JA skills while being engaged with robot.

As shown in Figure 6, during the initial stage, the teacher guided the students in obtaining joint attention with the robot. However, in the final trial, the child was able to imitate the robot's finger, which would have verified understanding of the robot's hand direction in obtaining joint attention with the robot. The results indicated that the robot's fingering gesture was dependable as a tool to trigger JA.

Figure 7 shows the eye gaze pattern as the child D attends to the robot's preferred object to achieve JA, with these patterns numbered in seven steps. In the first step, the child looks at the robot and attempts to understand the robot's request (regions 1 and 2). Next, the child follows the direction of the robot's hand toward an object while maintaining this gaze for some time (regions 3 and 4). We again return to the robot's space in order to reconfirm the robot's request, but we cannot show the data here as they belong to a different coordinate space (i.e., the robot's space). The child once again follows the robot's hand (shown as numbers 5 and 6) to reach the robot's preferred object in number 7. This was an interesting result in the experiment. Indeed, the child was able to learn how to obtain joint attention with the robot while being engaged with it.

VI. Conclusion

We developed a variety of modalities with interactive scenarios to realize the effectiveness of our proposed assistive robot. The main purpose of the study was to recognize a child's intention with performance of JA attention through an unsupervised approach. Our proposed unsupervised approach was useful in detecting the above information without the use of any training data and trained model. It can be a great challenge to obtain training data from disabled children. But our proposed robotic system was capable of interacting with the children without training the system. A secondary purpose of our research was to understand what kind of interactive scenario is most useful to improve their enjoyment with JA when they interact with robot. This kind of experience should be helpful for children to improve their JA with humans as they learn how to cope with society. The step-by-step and long-term interactions are essential for them to obtain these skills. The results of our experiment, as is discussed below, revealed that it was possible to use a therapeutic system as robot for children with autism.

The results showed that our proposed approach was capable of detecting the children's intentions with a high accuracy rate. The first phase results of the interaction showed that for a majority of the time (72.7%- 88.7%) the children paid attention to the robot. Even the eye gaze data patterns revealed that the children spent most of the time looking at the robot's facial area, and when the robot displayed a motion, the child paid attention to the robot's movements. These results indicate that the interactions with the robot were capable of motivating the children to increase their eye contact and engage the robot more robustly. The data shown in Figure 5 reveal the children spent a short amount of time looking at the object indicated by the robot during the initial trial of the JA task. However, once adjusting to the robot, the children spent a considerable amount of time looking at the objects indicated by the robot, resulting in JA with the robot. An interesting eye gaze pattern was discovered with child D in Figure 7. First, the child attended to the robot's request by looking at the robot's fingered object on the table, since the child's eye gaze indicated an initial glance at the robot's face and hand, with attention then given to the direction of the robot's arm. The child looked sequentially at the object, made repeated attempts to clarify the robot's arm direction, and finally attended to the object indicated by the robot. The eye gaze pattern data revealed that the children learned how to adjust their eye gazes in order to have joint attention with the robot.

VII. Acknowledgments

We would like to extend our sincere appreciation to the Tempaku School for the Disabled for their partnership in developing our assistive robot. This research has been supported by both the Grant-in-Aid for Young Scientists (B)(19700477) from the Japan Society for the Promotion of science (JSPS) and the Grant-in-Aid for Sustainable Research Center of the Ministry of Education, Science, Sports and Culture of Japan.

References

[1] O. Grynszpan, J. C. Martin, and N. Oudin, "On the annotation of gestures in multimodal autistic behaviour," in *The 5th International Workshop on Gesture and Sign Language based Human-Computer Interaction*, 2003, pp. 25–33.

[2] M. Davis, K. Dautenhahn, C. L. Nehaniv, and S. D. Powell, "Touchstory: Towards an interactive learning environment for helping children with autism to understand narrative," in *ICCHP*, 2006, pp. 785–792.

[3] B. Robins, K. Dautenhahn, R. Boekhorst, and A. Billard, "Robotic assistants in therapy and education of children with autism: Can a small humanoid robot help encourage social interaction skills," *Special issue of Universal Access in the Information Society*, vol. 4, no. 2, pp. 105–120, 2005.

[4] D. Strickland, L. Marcus, K. Hogan, G. Mesibov, and D. McAllister, "Using virtual reality as a learning aid for autistic children," in *Proceedings of the Autism France Third International Conference on Computers and Autism*, 1995, pp. 119–132.

[5] D. Strickland, "A virtual reality application with autistic children," *PRESENCE: Teleoperators and Virtual Environment*, vol. 5, no. 3, pp. 319–329, 1996.

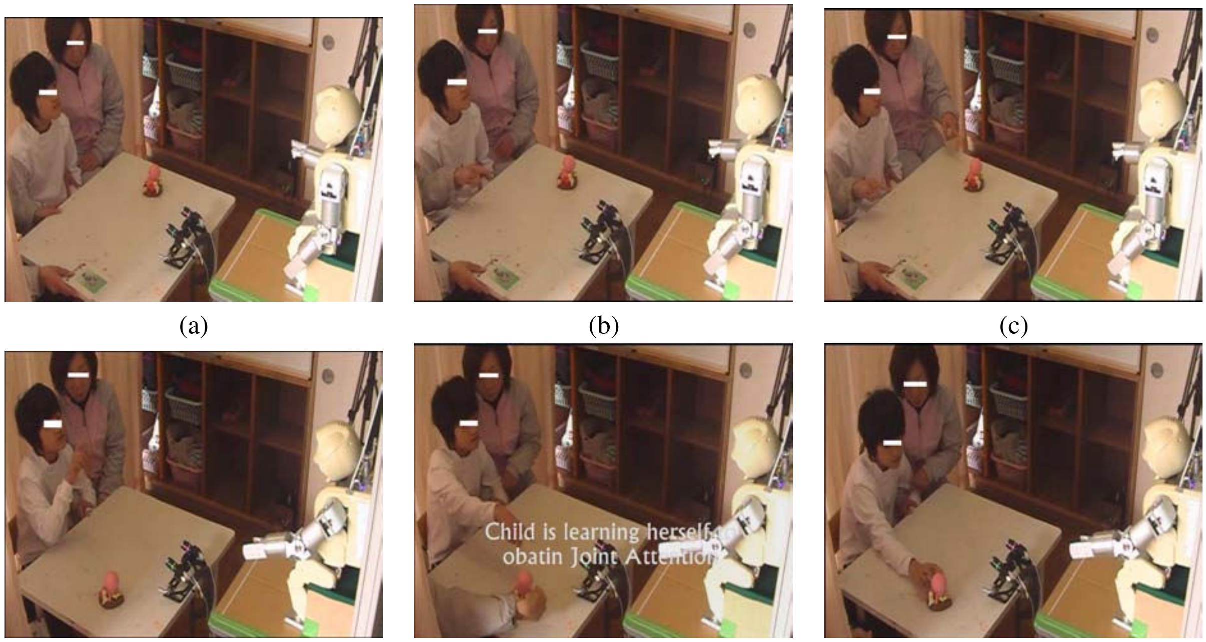
[6] D. Murray, *Autism and information technology: therapy with computers*. London: Autism and learning: a guide to good practice, David Fulton Publishers, 1997.

[7] F. Michaud and C. Thberge-Turmel, "Mobile robotic toys and autism socially intelligent agents - creating relationships with computers and robots," *Kluwer Academic Publishers*, pp. 125–132, 2002.

[8] J. Nadel, C. Guerini, A. Peze, and C. Rivet, *The evolving nature of imitation as a format of communication*. London: Imitation in Infancy, Cambridge University Press, 1999.

[9] I. Werry, K. Dautenhahn, B. Ogden, and W. Harwin, "Can social interaction skills be taught by a social agent? the role of a robotic mediator in autism therapy," in *CT '01: Proceedings of the 4th International Conference on Cognitive Technology*. London, UK: Springer-Verlag, 2001, pp. 57–74.

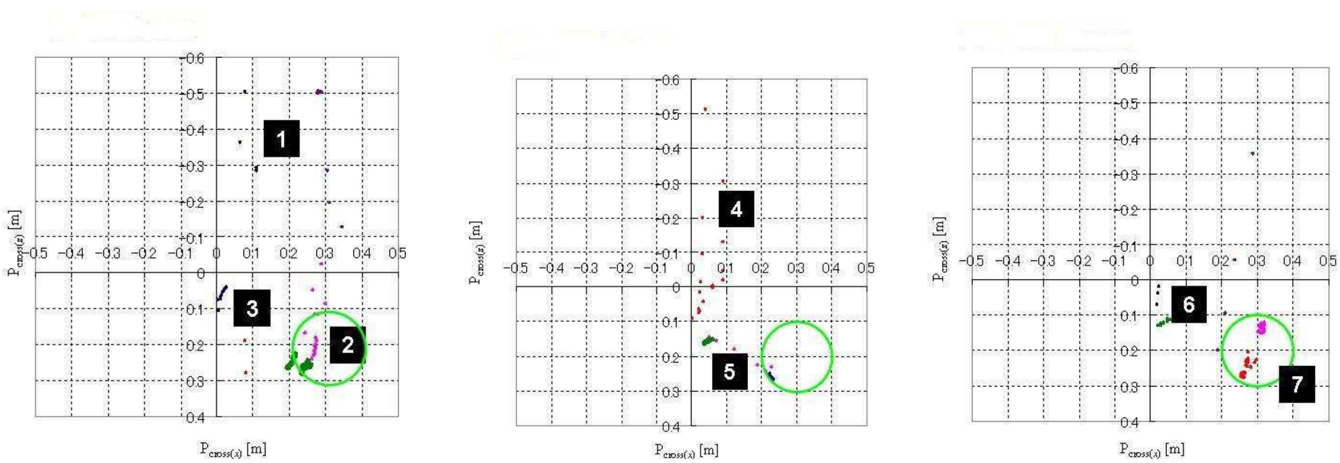
[10] S. M. and S. M., "The behaviors of parents of children with autism predict the subsequent development of their children's communication," *Journal of Autism and Developmental Disorders*, vol. 32, no. 2, pp. 77–89, Apr 2002.



(a) (b) (c)

(d) (e) (f)

Fig. 6. A teacher guiding a child to obtain joint attention with the robot in trial 1 (images of (a), (b), and (c)). In the final trial, child was able to learn how to obtain joint attention with the robot while being engaged with it (images (d), (e), and (f)).



(a) (b) (c)

Fig. 7. The child is learning how to attend to an object to achieve JA with the robot. The eye gaze path is shown with the child paying attention to the robot's request (see region 1). Region 2 shows the child approaching the object requested by the robot. The child looks at the robot a second time to verify the robot's fingered object (regions 3 and 4), and it finally obtains JA with the robot (regions 5, 6, and 7). The figure shows only the eye gaze behavior on the table. However, to obtain the remainder of eye gaze path, the eye gaze information was combined on the robot space (we defined a plane parallel to the robot and estimated the intersection position on that plane and the child's eye gaze vector). To predict and define the above regions, we combined the eye gaze information on the robot space with the table space. (a) Frames 1-100, (b) frames 101-200, and (c) frames 201-300.

[11] B. Robins1, P. Dickerson, P. Stribling, and K. Dautenhahn, "Robot-mediated joint attention in children with autism: A case study in robot-human interaction," *Interaction Studies*, vol. 5, no. 2, pp. 161–198, Nov 2004.

[12] C. Ferrara and S. D. Hill, "The responsiveness of autistic children to the predictability of social and nonsocial toys," *Journal of Autism and Developmental Disorders*, vol. 10, no. 1, pp. 51–57, 1980.

[13] T. Blomgren and A. Tenggren, "Robots as an instrument for (re)habilitation of autistic children," in *Technical Report*. Malmo University, 2002.

[14] Y. Nagai, M. Asada, and K. Hosoda, "Developmental learning model for joint attention," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2002, pp. 932–937.

[15] H. Kozima and H. Yano, "A robot that learns to communicate with human caregivers," 2001. [Online]. Available: citeseer.ist.psu.edu/kozima01robot.html

[16] M. Doniec, G. Sun, and B. Scassellati, "Active learning of joint attention," in *IEEE/RSJ International Conference on Humanoid Robotics (Humanoids 2006)*, 2006.