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

Graded Cueing Feedback in Robot-Mediated Imitation Practice for Children with Autism Spectrum Disorders

Jillian Greczek, Edward Kaszubski, Amin Atrash, and Maja Matarić
University of Southern California
3710 McClintock Ave, Room 423
Los Angeles, CA 90089-0781
1-213-740-6245
greczek@usc.edu, ekaszubski@gmail.com, atrash@usc.edu, mataric@usc.edu

ABSTRACT
We performed a study that examined the effects of a humanoid robot giving the minimum required feedback – graded cueing – during a one-on-one imitation game played children with autism spectrum disorders (ASD). 12 high-functioning participants with ASD, ages 7 to 10, each played "Copy-Cat" with a Nao robot 5 times over the span of 2.5 weeks. While the graded cueing model was not exercised in its fullest, using graded cueing-style feedback resulted in a *nondecreasing trend in imitative accuracy* when compared to a non-adaptive condition, where participants always received the same, most descriptive feedback whenever they made a mistake. These trends show promise for future work with robots encouraging autonomy in special needs populations.

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

General Terms
Experimentation, Human Factors.

Keywords
Socially assistive robotics, graded cueing, autism spectrum disorders (ASD), Nao robot, patient autonomy

1. INTRODUCTION
Socially assistive robots have the potential to augment therapy and rehabilitation by providing personalized care at any time and for as long as is needed. Studies are beginning to show how robots can invoke behavior change in humans over long-term interactions [11]. In this work, a model of the occupational therapy technique of *graded cueing* is presented as a general framework for long-term health behavior coaching. Although the model was not exercised in its fullest in this study, promising results came from this varied feedback, and interesting observations were obtained about child-robot interaction.

This work was sponsored by NSF IIS-0803565, NSF IIS-1139148, and CNS-0709296.
RO-MAN'14, August 25–29, 2014, Edinburgh, Scotland, UK.

Graded cueing is a process of behavior shaping that uses increasingly specific cues, or prompts, to help improve people's skills at everyday tasks during recuperative therapy [2]. It is typically used in treatments for individuals who have lost skills, such as through a brain injury or stroke, or need to learn new skills, such as social skills of individuals with autism spectrum disorders (ASD). In graded cueing, a therapist asks a patient to perform a task, then prompts the patient with increasing specificity based on how much the patient struggles with the task. The goal is to increase both patient task performance and autonomy in performing the task, through minimal therapist intervention.

We applied the graded cueing framework in a "Copy-Cat" imitation game played between a Nao robot and a child with ASD. This work is motivated by the evidence that children with ASD are often behind in their development of imitative behavior [12], and that practicing through repeated interactions with a therapist can improve imitation abilities [14]. The model approximates those effect with a socially assistive robot in order to use technology to broaden access to ASD therapy.

2. RELATED WORK
Humanoid robots are good candidates for ASD therapy because of their simplified social signals and ability to give predictable, concrete feedback. Such feedback has been found to be effective in teaching social behavior to children with ASD [13]. There is a growing practice of robots being used in ASD interventions. Robins et al. found that four children with ASD imitated a doll-like robotic toy, often without any initial prompting; they attributed this to the robot's "simpler" physical appearance [15]. Duquette et al. found that a humanoid robot elicited more shared attention between two children with ASD than a human mediator [4]. Ferrari et al. presented the IROMEC robot as a social mediator in interactive play between a child with ASD and a parent, teacher, or therapist [7].

Larson found that children with ASD prefer concrete feedback, such as lights, colors, and sounds, which can be measured and quantified [13]. Ingersoll reported that multimodal feedback is more effective than any single feedback modality alone [10]. Because of this, the Nao robot in the presented study uses lights, colors, and sounds as feedback modalities to indicate to the participants whether they have imitated correctly. The Nao robot in particular has previously been used in a several studies with children with ASD, such as the ARIA system [6]. Fujimoto et al.

conducted a study where a child with ASD imitates a robot, but the focus was on modeling human motion rather than affecting child imitation behavior [8].

This work extends previous work by Feil-Seifer and Matarić, which introduced the use of graded cueing applied to socially assistive robots interacting with children with ASD [5]. In that study, graded cueing was implemented as a finite state machine in the context of a Simon Says imitation game of arm postures. Other work has focused on evaluation techniques for child imitation during robot-mediated ASD interventions. Hwang et al. submitted an interaction scenario for robot-mediated imitation training for children with ASD and discussed how to evaluate the success of an intervention that used a humanoid robot [9]. Quantification of child imitation of robot movement is also described in a study by Bugnariu et al. using the ZENO Robokind robot [3]. That study showed a rigorous comparison of dynamic imitation using a highly instrumented environment. The requirement for participants to wear sensors is always a concern with human subjects studies, particularly with sensitive populations such as children with ASD. Our study was conducted with static poses, but in a real classroom environment, using a Microsoft Kinect and no requirement for wearable sensors.

3. METHODS

Tasks that can be taught and reinforced by graded cueing must be able to be broken down into discrete steps. In this study, the task consisted of a single step: simple imitation. The user's imitation of the robot's arm pose, consisting of two arms, is considered one step, a binary decision of whether the user has copied correctly or not. A more complex task could consist of multiple steps or be broken down into a series of subtasks (e.g., one per arm).

Each discrete task step is accompanied by a series of increasingly specific user prompts. These prompts are ranked by specificity, and each subsequent prompt provides more guidance to the user on how to complete the task. In therapeutic applications, a therapist using graded cueing would adjust the level of his or her prompting based on each patient's ability to complete the task. Thus, the prompting is neither frustrating nor patronizing to the patient, and facilitates the increase in the patient's autonomy at completing the task increases over time, ideally resulting in no more need for prompting at all. To reflect this strategy, our graded cueing computational model seeks to minimize the number and specificity of required user prompts, while maximizing user success at completing the task.

3.1 Model

This work presents a general framework for graded cueing, based on a probabilistic model of first prompt choice. The model consists of N states, one for each level of the user's task ability. For example, specificity level 4 is the best action for user ability level 4. Importantly, prompt specificity is inversely proportional to user ability – users with lower ability require more specific prompting, while users with higher ability require less specific prompting. An N+1 state (P0) represents the success task state.

During execution, the robot uses the child's responses to maintain a distribution over perceived ability level for the child. The model selects appropriate actions based on a policy calculated over this distribution. The goal of the model is to select the *first* prompt to give the user depending on the user's ability level. From that initial prompt, any subsequent prompt required is the next most specific prompt. Prompt specificity and ranking is task- and domain-specific. An example from our study is shown in Section 3.2.

To compensate for the few available data points often seen in human-robot interaction contexts, particularly with regards to children with ASD, a probabilistic model was developed to adapt quickly to the changing user state. Probabilistic models have been shown to be effective tools for representation in assistive technologies [1].

To achieve rapid adaptation, a Bayesian approach was used, wherein the model maintains a distribution over the states, is updated based on user responses, and selects actions by sampling from the distribution. However, there were not enough data points in our study, and the participants were too proficient at the task to fully exercise the model. Consequently, we focus on the implementation of the model and compare the results of graded cueing to constant feedback.

3.2 Implementation

The probabilistic model of first prompt choice was implemented as part of a "Copy-Cat" imitation game played between a Nao robot and a child with an ASD. In the game, the robot poses standing up and using both of its arms, and asks the child to mirror its pose, using phrases such as "Can you copy me?" The robot's poses are randomly chosen from 5 different arm positions, one position per arm, resulting in 25 total possible poses (Figure 1). Because the difficulty of individual poses in relation to other poses was unknown, in this study, each pose was considered equally difficult to imitate. The Nao robot's text-to-speech default voice was used for all robot utterances.

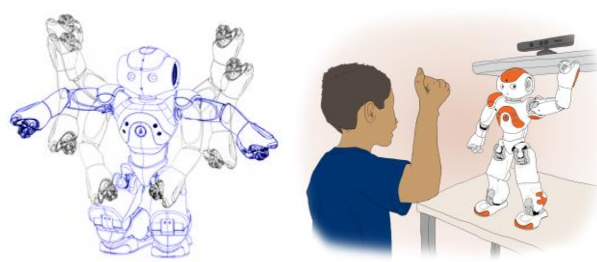

Figure 1: Robot arm poses.

If the child successfully copies the robot's pose within a tolerance level, the robot gives positive verbal feedback (e.g., "Good job," "That's right"), nods, and flashes its eyes green. If the child does not successfully copy the robot's pose, the robot begins prompting, using the starting prompt given by the model for the particular user. During the very first interaction, the starting prompt specificity is P1, the least specific level of prompting, but for each subsequent session, the starting prompt is determined by the probabilistic model of graded cueing. In this way, each child's model of first prompt choice becomes slightly different depending on their performance at the imitation game. After the first prompt, if the child requires further prompting, the robot responds with the next prompt among four levels of prompt specificity used for this simple imitation task:

P0. No prompts given (Success)

P1. Words (e.g. "Are you sure that's right?")

P2. Words + Gesture ("Look again at your left arm" + wiggle left arm)

P3. Specific Words + Gesture (“Bend your right arm more” + wiggle right arm)

P4. Specific Words + Specific Gesture (“Your arms look like this… Move them like this.” + imitation of child’s pose)

If the robot reaches the final prompt and the child still does not imitate successfully, the robot responds with “Let’s try another pose,” and moves on to the next arm pose.

Because children with ASD can have great variation in imitation ability, care was taken to include another condition for those already beyond the level of the “Copy-Cat” game. If the child correctly imitated 3 robot poses in a row, the robot entered “fever” mode by saying “You’re good at this!” and changing its eye color. For as long as the child continued correctly imitating the robot’s poses, the robot’s eyes changed to a different color of the rainbow (orange, yellow, cyan, blue, magenta) with each round. This condition was added to keep children already good at the game from losing interest and potentially causing a false decline in performance.

3.3 Measuring Imitation

The real-time comparison of human and robot poses (joint positions) is a nontrivial computational problem. In this study, existing ROS package arm_pose_recognizer, part of the USC Interaction Lab ROS repository, was used to compare the distances between human and robot joints to measure pose accuracy using the Kinect sensor [16]. The Kinect was chosen because it is inexpensive and portable. Although the Nao robot is humanoid, its proportions are not the same as human arms, and arm_pose_recognizer takes this into account, normalizing limb length so that the two poses can be accurately compared. Given a list of robot and human arm joints to be compared, a "unit humanoid" is constructed for each humanoid using the actual 3-DOF angles between links, but using link lengths of 1. The two unit humanoids are then aligned at a root frame (in this case the neck), and the distance between corresponding joints is calculated. Additionally, given a three-dimensional per-joint variance,, the algorithm in this package is able to return a list of per-joint error vectors in R^3 encoded as the number of standard deviations from the target joint position along each spatial dimension. The measurement of accuracy used by the package is in standard deviations in the X, Y, and Z directions determined by a data collection with adults that was used in its creation. This is a measure of distance, although the notation used uses standard deviations. After in-lab piloting with adults, a cutoff point for “correct” imitation was determined empirically to be 6 standard deviations for hands and 7 for elbows, which had a higher overall variance during testing as well as in the study data. The increased variance in elbow position versus hand position could be due to the nature of human movement or the nature of the arm poses, many of which are distinguished from each other by arm bend.

3.4 Validation Study

We conducted a study to validate the graded cueing model with a group of elementary-aged children with ASD. 12 participants were recruited from an ASD-only class of students between the ages of 7 and 10. The study was conducted over 2.5 weeks (18 days) with each participant receiving two 15-minute sessions per week, resulting in 5 sessions of interactive play per participant. Each session consisted of 10 rounds (10 robot poses, one pose per round), that the child was asked to imitate.

The participants were divided into two groups: 1) adaptive, receiving graded cueing feedback (3 male, 3 female); and 2) non-adaptive control group, receiving constant feedback (5 male, 1 female). Participants in the non-adaptive condition received P4 feedback from the robot up to 4 times in a row for each pose they attempted to imitate. The non-adaptive condition was chosen as a comparison because it represents giving more information, but not encouraging user autonomy in completing the task, compared to the graded cueing condition.

The experimental setup consisted of the Nao robot placed on a table in front of a (seated or standing) child so that it was approximately at the child’s eye-level. Because the poses were considering arms only, the children had the option to sit or stand in front of the robot in each session. A Kinect was placed on a tripod behind the robot and was used for sensing the child’s pose. Due to the restrictions of the classroom space, the experimenters were on the side of the robot, but still in view of the child. Experimenters explained to the participants to stand or sit in front of the robot and interact with it. After this, experimenters sat down and looked away from participants at all times and did not answer any participant questions. Participants were brought to the classroom by an aide, who remained in the back of the room observing for the child’s safety and comfort. The aide also answered any questions the child had that were not about the interaction. For all additional questions, the participants were left to figure out what the robot meant by themselves.

Figure 2: Experimental setup.

Although the sample size is too small to produce statistical significance, the results of the study were expected to show (1) an increase in the average correctness of the child’s pose, measured as a decrease of the child’s deviation from the robot’s pose; (2) a decrease in the level of prompt specificity; and (3) a decrease in the number of prompts the child required for successful imitation of the robot.

4. RESULTS

While the underlying structure of the graded cueing feedback was designed to be adaptive, the adaptation was not exercised in this study. Overall, the participants had little difficulty with the imitation task. Because participants "got it in one" for the majority of poses, the probabilistic model of prompt choice did not get enough data to adjust any participant's starting prompt beyond P1. However, rounds that did require prompting saw different behavior on average from the graded cueing group than the control group. This suggests that *varied* feedback was qualitatively preferred by the participants in this study. Graded cueing feedback, if it does encourage autonomy, did not result in any worse performance than the most specific feedback.

4.1 Prompt Specificity and Performance

The group that received graded cueing feedback required 80 total prompts (counting multiple prompts in the same round) during the entire study, while the non-adaptive group required 111 total prompts. The total number of prompts required by each participant over all interaction sessions is shown in Figure 3.

The average number of prompts for the non-adaptive group was higher than that of the graded cueing group, but, as shown, that average is greatly affected by the large number of prompts required by participant D. Every time prompting was required, with the exception of 2 rounds from participant D, it was required at all four levels of prompting before moving to the next pose. In these two instances, only 3 prompts were required, in this case P4 repeated 3 times, before successful imitation.

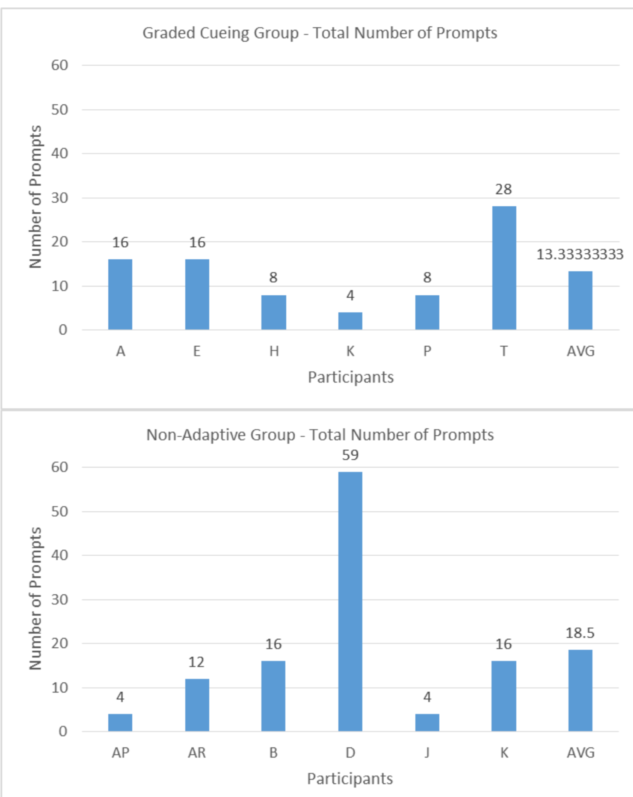

Figure 3: Total number of prompts required per participant, as well as the average (AVG) for each group.

Figure 4 shows the distribution of prompting over each session. If the participant did not require prompting during a session, meaning they got every pose correct on the first try, then no bar appears on the graph. Ideally, the number of prompts required should decrease over time for each participant, but that was not the case for all participants, likely due to the short duration of the study. The same is true for the variance in imitation error over all sessions – variance decreases over time for some participants, but not for others, resulting in no discernable trend. Additional statistical analyses were performed, but no additional statistical significance or other meaningful results were found.

The inherent limitations of the Kinect sensor to noise and lighting conditions must be considered as having influenced these results. The fourth session, for example, was a particularly sunny day, which resulted in reduced sensor accuracy.

Because of the small size of the study, statistical significance was not expected. Instead, a detailed review of two participants that make up the majority of the prompting data is provided in the following section.

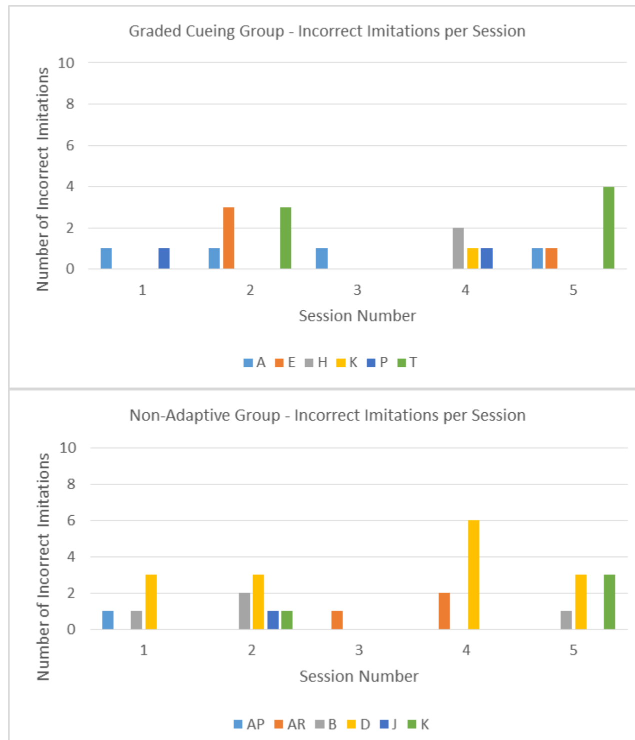

Figure 4: The number of rounds requiring any level of prompting by the robot over each session.

4.2 Case Studies: Participants T and D

As shown in Figures 3 and 4, the majority of rounds that required prompting took place during interaction with two participants, T in the graded cueing group and D in the non-adaptive group. Because most of the prompting data are from these two participants, they are now examined in more detail.

4.2.1 Participant T in the Graded Cueing Group

In the graded cueing group, participant T required the most prompts, each of which resulted in unsuccessful imitation after all four levels of prompt specificity. During each unsuccessful imitation, however, participant T's performance *did not get worse* with each subsequent prompt, as shown via examples in Figure 5. Although the cutoff point was for any one value in the X, Y, or Z direction, the average error for all three directions for each participant's left and right hands and elbows is shown in the following figures for ease of viewing.

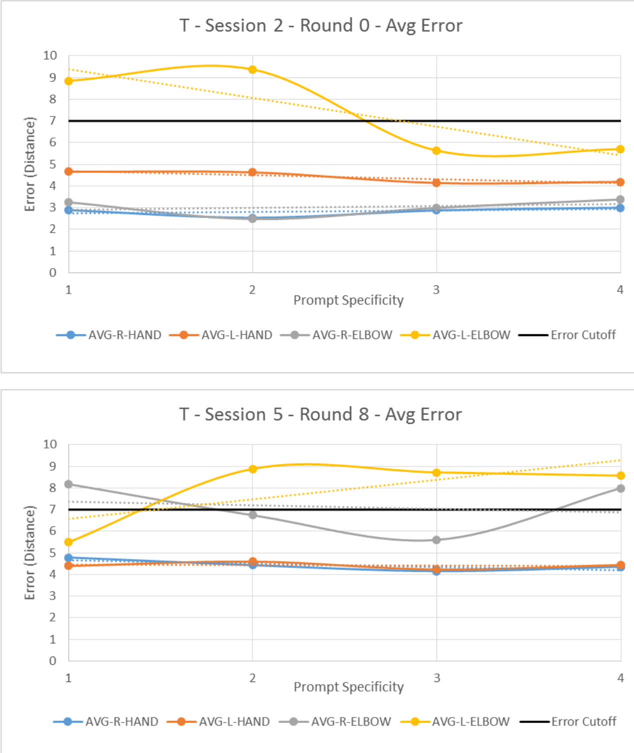

Figure 5: Representative example rounds from participant T. The black line represents the cutoff for successful imitation.

Although a larger and more detailed study is merited, these trends in the graded cueing group are promising for future work. As a whole, participants who received graded cueing did not show a decline in imitation accuracy within each round they received prompting, sometimes even increasing in accuracy.

4.2.2 *Participant D in the Non-Adaptive Group*

In the non-adaptive group, participant D required 59 prompts over the entire study, making up a large portion of the prompting data. This participant clearly had the most trouble with imitation, and received the most specific feedback for each mistake. In spite of receiving the most information, participant D did not increase in imitation ability within rounds. However, participant D did not always get worse, as shown in Figure 6, and also had the only two rounds where anyone required fewer than four prompts, but those successes were likely an accidental result of exploring the position space, as they were both at the end of the fourth session and both showed a steady increase in average error with each prompt.

4.3 Eye Gaze, Vocalization and Frustration

The data were annotated post-study for incidences of participants' eye gaze toward the robot, vocalizing, and indicating their frustration (either vocally or by making a frustrated gesture). All participants consistently focused their gaze on the robot for the duration of the interaction.

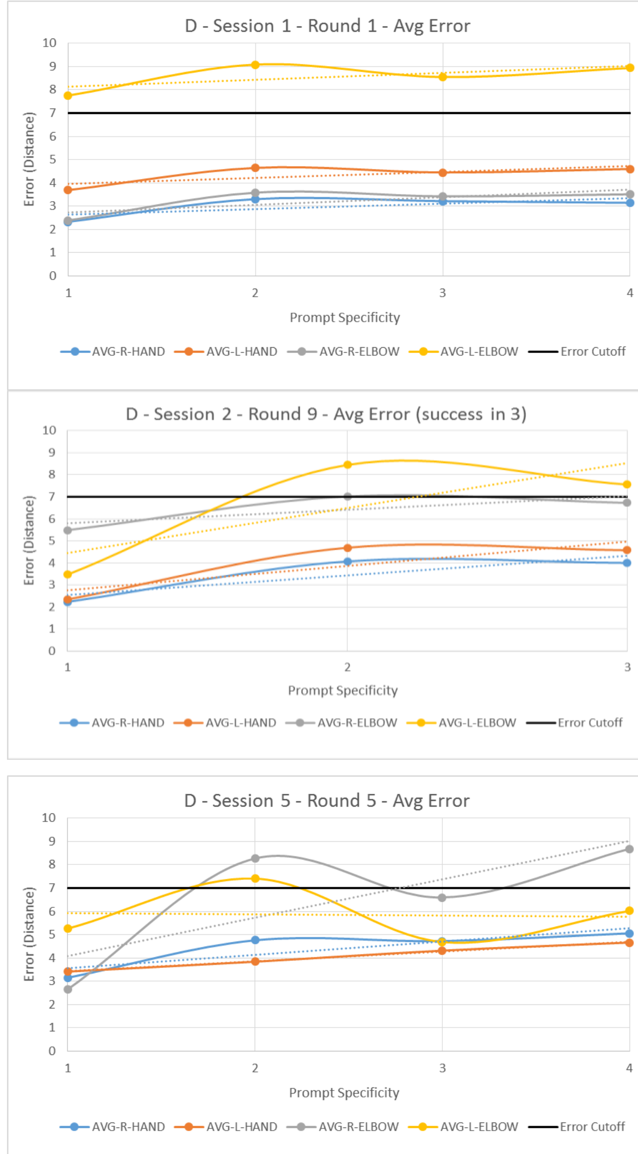

Figure 6: Representative example rounds from participant D. Session 2, Round 9 shows an example of an accidental success, as the error rate continues to increase with each prompt.

There was little vocalization from participants, either toward the robot or toward the teacher's aide or experimenters, with one exception. In the non-adaptive group, participant J repeated everything that the robot said in each session. Although aware that copying the robot's speech was not required, participant J did so anyway "because it was fun". This was a useful validation that the synthesized speech used by the robot was being properly understood by the participants. Participant J made few errors in repeating the robot's speech, and no errors specifically when repeating prompts said by the robot.

Participants in the non-adaptive group showed 4 instances of frustration with the robot over all sessions, while the graded cueing group showed only 2 instances. However, frustration was not necessarily correlated with poor performance at the task, suggesting that another benefit of graded cueing was in the qualitative experience for the participants.

5. DISCUSSION

The results of this study suggest that varied feedback may be more effective, and less frustrating, than fully descriptive feedback in a child-robot imitation game for children with ASD. Because graded cueing feedback did not result in a decline in imitation accuracy, using feedback that encourages autonomy in performing a task will be considered in future work with children with ASD, as well as in other domains.

The most variance in pose accuracy came from elbow position rather than hand position, which implies that the threshold for correctness might need to be raised, and the robot's feedback should focus on elbow position more specifically. While position information worked well for hands, measuring elbow angle may be a better metric in future work due to the high variance of position. Conversely, there were also instances where the cutoff point might have been too lenient, where imitation was counted as correct while participants' arms were in the correct positions, but too close to their bodies. These instances were not anticipated in the feedback model, and will need to be added in future work, possibly as a comparison of joint angles.

6. CONCLUSION

This work presented a study of graded cueing feedback for an imitation game played between a Nao robot and a child with ASD. This study is part of ongoing work in socially assistive robotics to create adaptive, personalized systems where humanoid robots can help people with special needs to incorporate new health and therapeutic behaviors into their everyday lives. While the graded cueing model was not exercised in its fullest, these preliminary results show a trend of nondecreasing imitation accuracy, dmonstrating promise for varied and potentially minimal feedback in increasing user autonomy through robot-mediated intervention.

7. ACKNOWLEDGMENTS

We thanks Dr. Sharon Cermak for lending her expertise in the implementation of the graded cueing feedback and model. This work is supported by NSF IIS-0803565, NSF IIS-1139148, and CNS-0709296.

8. REFERENCES

[1] Atrash, A., Kaplow, R., Villemure, J., West, R., Yamani, H., Pineau, J.: Development and Validation of a Robust Speech Interface for Improved Human-Robot Interaction. *International Journal of Social Robotics (IJSR)* 1(4), 345–356 (2009).

[2] Bottari, C.L., Dassa, C., Rainville, C.M., Dutil, É.: The IADL Profile: Development, content validity, intra- and interrater agreement. *Canadian Journal of Occupational Therapy* 77(2), 90–100 (2010).

[3] Bugnariu, N., Young, C., Rockenbach, K., Patterson, R. M., Garver, C., Ranatunga, I., … Popa, D. (2013). Human-robot interaction as a tool to evaluate and quantify motor imitation behavior in children with Autism Spectrum Disorders. In *2013 International Conference on Virtual Rehabilitation (ICVR)* (pp. 57–62). IEEE.

[4] Duquette, A., Michaud, F., Mercier, H.: Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism. *Autonomous Robots* 24(2), 147–157 (2007).

[5] Feil-Seifer, D.J., Mataric, M.J.: A Simon-Says Robot Providing Autonomous Imitation Feedback Using Graded Cueing. In*: Poster paper in International Meeting for Autism Research (IMFAR),* Toronto, On, Canada (May 2012).

[6] Esubalew, T., Member, S., Lahiri, U., Swanson, A. R., Crittendon, J. A., Warren, Z. E., … Member, S. (2013). Robot-Mediated Intervention Architecture (ARIA) for Children With Autism. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, *21*(2), 289–299.

[7] Ferrari, E., Robins, B., Dautenhahn, K.: Therapeutic and educational objectives in robot assisted play for children with autism. In*: IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2009),* pp. 108–114. Toyoyama International Conference Center, Japan (September-October 2009).

[8] Fujimoto, I., Matsumoto, T., Silva, P.R.S., Kobayashi, M., Higashi, M.: Mimicking and Evaluating Human Motion to Improve the Imitation Skill of Children with Autism Through a Robot. *International Journal of Social Robotic (IJSR)* 3(4), 349–357 (2011).

[9] Hwang, J., Kim, M.-G., & Cho, K. (2013). Developing interaction scenarios of robot-mediated imitation training for children with Autism Spectrum Disorders. In *2013 IEEE Ro-Man* (pp. 616–620). IEEE.

[10] Ingersoll, B., Schreibman, L., Tran, Q.H.: Effect of sensory feedback on immediate object imitation in children with autism. *Journal of Autism and Developmental Disorders* 33(6), 673–683 (2003).

[11] Kidd, C.D., Breazeal, C.: Robots at home: Understanding long-term human-robot interaction. In: *Int. Conf. on Int. Rob. Sys. (IROS),* Nice, France (2008).

[12] Larson, J.C.G., Mostofsky, S.H.: Motor Deficits in Autism. *Autism: A Neurological Disorder*, 231–247 (2006).

[13] Larson, M.J., South, M., Krauskopf, E., Clawson, A., Crowley, M.J.: Feedback and reward processing in high-functioning autism. *Psychiatry Research* 187(1-2), 198–203 (2011).

[14] Nadel, J.: Imitation and imitation recognition: Functional use in preverbal infants and nonverbal children with autism. In: Meltzoff&, A.N., Prinz, W. (eds.) *The Imitative Mind: Development Evolution, and Brain Bases*, pp. 42–62. Cambridge University Press (2002).

[15] Robins, B., Dautenhahn, K., Boekhorst, R.T., Billard, A.: Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills? *Universal Access in the Information Society* 4(2), 105–120 (2005).

[16] https://code.google.com/p/usc-interaction-software/.