Supporting Clinicians in Robot-Assisted Therapy for Autism Spectrum Disorder: Creating and Editing Robot Animations with Full-Body Motion Tracking

J. Alan Atherton
Department of Computer Science
Brigham Young University
Provo, UT 84602
Email: j.alan.atherton@byu.edu

Michael A. Goodrich
Department of Computer Science
Brigham Young University
Provo, UT 84602
Email: mike@cs.byu.edu

***Abstract*—Robots show potential to be helpful in therapy for children with autism spectrum disorder. In our experience, clinicians occasionally desire to change the robot behavior to suit the needs of different children. Because clinicians typically lack programming experience, they must currently ask a programmer to program the robots. Robots may be more useful to clinicians if they are able to create and modify actions for the robot. We are designing a user interface based on full-body motion capture to enable clinicians to create and modify animation sequences for robots. Clinicians, computer scientists, and mechanical engineers are collaboratively involved in the design process.**

I. INTRODUCTION

Autism spectrum disorder (ASD) commonly involves communication impairments, limited social interaction, and limited imagination [4]. Researchers are increasingly interested in using robots (like the one pictured in Figure 1) in therapy for children with ASD [3]. Reports indicate that many such children show interest in robots and find them engaging. Most importantly, anecdotal evidence is accumulating that robots can be a pivot to facilitate interaction between the child and clinician in one-on-one interaction [4, 5, 20]. The problem is that robot behavior needs to be changed to accommodate differing needs and as an individual child progresses.

Our goal is to empower *clinicians* to customize robot behavior to suit the needs of each child. The problem is that changing robot behavior by traditional computer programming is a complex process [16]. There are several ways to change robot behavior without the need for traditional computer programming. For example, visual programming and programming by demonstration are methods that require very little training compared to traditional computer programming.

Clinicians in our group who want to create new robot animations at present must call on a programmer. This usually means the suite of robot behaviors is limited and does not change much in between therapy sessions because of the coordination effort required. We hypothesize that it is better to enable clinicians to animate the robot on their own. The trouble with giving clinicians control is that they are not usually experienced in robot programming. Nevertheless, we claim that for robots to achieve their therapeutic potential, clinicians must program them [22].

Fig. 1. **Troy, BYU's humanoid robot**

We divide the programming task into two categories, which we call *animate* and *choreograph*, respectively. Animating a new robot action is similar to character animation, where robot motor commands are specified to create a particular motion. Choreographing means arranging a set of existing robot behaviors in sequences and assigning user inputs to replay the sequences for use in real-time interaction.

We are in the process of collaboratively designing user interfaces to enable clinicians to animate and choreograph robots. Our design group includes professors and clinicians who are involved with therapy for children with autism, mechanical engineering professors and students, and computer science professors and students. We seek to create interfaces that are tuned specifically to the needs of clinicians that we work with. Our goal is to support clinicians so that they can help children with autism.

At present, the interface is designed as two separate components for choreographing behaviors and animating actions, respectively. Previous work has discussed the choreography interface (Giullian et al. [5]). In this paper we discuss ideas for the design of the animation interface. We propose to use a full-body gesture-based method for motion capture and especially for *editing* the captured motion.

II. RELATED LITERATURE

The use of robots in education is an idea that has been studied for a few decades [17]. Children, and especially those with learning disabilities, typically have great interest in technology. In more recent times, researchers have looked at using robots to assist clinicians in teaching children with ASD.

Because this field is so new, the research is in an exploratory stage. Some research looks at driving robot behavior autonomously based on the child's behavior as determined by tracking the child's distance from the robot [2]. Other researchers have looked at loading the robot with some behaviors that are activated when the child or clinician presses a button on the robot itself [4]. Another approach has the robot autonomously play a collaborative game with the child [20].

We want to enable clinicians, who are not typically animation experts, to animate an upper-body humanoid robot. Animating a robot is very much like 3D computer animation with some additional constraints, such as physical dynamics. Lasseter gives an excellent tutorial on how to apply traditional animation principles to computer animation [14]. One key aspect related to our application is timing. Lasseter says, "[timing] is an important principle because it gives meaning to movement…and can even carry emotional meaning." Timing an animation is considered to be a major challenge, especially to non-experts [19]. Terra and Metoyer [19] present a sketch-based method for setting the timing of an animation after key poses have been created.

Systems exist for generating human-like arm and hand gestures without having to animate at a low level (see for example [10, 1]). Such systems are designed for creating virtual conversational agents that give instructions or answer questions. They are good for typical human conversational gestures, including iconic (representing the conversation subject), deictic (pointing), and other more abstract gestures. The limitation of such systems is the library of gestures that are possible. Because our target application needs free-form upper-body animation, such systems are not flexible enough.

Another approach is a sort of high-level scripting language (for example [12]). From the script, a motion planner, often paired with machine learning, creates a motion plan for the character. The advantage of this method is that characters with numerous degrees-of-freedom (for example, full human morphology including fingers and toes) are easier to direct. The disadvantage is that control of the character is nearly entirely algorithmic, and so the user is limited to the capabilities of the algorithm.

Motion capture appears to be a good method to enable non-experts to create animations without having to deal with issues involving timing. Gleicher [6] provides a very good history of motion capture and the challenges it brings. One issue that always arises when using motion capture is how to edit the data. Reasons to edit motion capture data include: reusing an existing animation with modification, creating physically unfeasible motions, correcting imperfections of reality, and adding secondary motion. Because motion capture provides a pose for each frame of the animation, editing the motion proves to be challenging.

Full-body tracking interfaces have existed for some time. The ALIVE system was important as one of the early full-body interactive systems [15]. Although the systems have existed, they are not widespread, and so user interface design for these systems seems to be at an early stage. Many gesture-based interfaces are designed to interpret the human action as a sort of command (for example [9, 13]).

We have not yet encountered a user interface that uses full-body tracking to edit 3D animations. We hypothesize that such an interface, with little training, will enable non-experts to animate robots.

III. CLINICAL REQUIREMENTS

To design our system, we are working with a multi-disciplinary group of researchers and clinicians who work with children with ASD in the BYU Comprehensive Clinic. The researchers are in the Department of Communication Disorders, the Department of Computer Science, and the Department of Mechanical Engineering at BYU. In other papers, we describe how our group designed the robot that we use [18] and clinical requirements for the robot and visual programming interface [5]. Our user interface design was influenced by feedback from the clinicians and other visual programming environments. We are aware of cognitive task analysis methods as in [11], but we chose this focus group approach to design our system at such an early stage. Specifically, we use existing therapies and a vision of potential robot-assisted therapies to construct requirements for a user interface that enables clinicians to program robots [2]. We met as a group several times over the course of a year while developing the robot and user interface.

A. Methods

Clinicians started by explaining what sort of activities they play with the children. Much of what they do in the BYU clinic involves presenting a toy or game to the child, observing how the child plays, and then trying to insert themselves into the play. The toys and games are designed to support specific therapeutic objectives that often involve social interaction. Once the child becomes disengaged, they change activities. The activities are primarily based around imitation and turn-taking [8]. Because of this dynamic environment, the clinician needs to plan multiple contingent strategies ahead of time. The clinician also need to be able to direct robot actions in real-time, as robot perception is still a long way from determining whether children with ASD are engaged in an activity. Another observation from the clinicians is that children with ASD tend to be very interested in technology and gadgets, so whatever the mechanism that controls the robot must be discreet.

In addition to real-time control of the robot, clinicians expressed the need to change the robot's repertoire of actions in order to discourage the child from fixating on the robot's repetitive behavior. As each child progresses in their level of competency, the changeable actions will also allow the clinicians to adapt the therapy to the needs of the child.

B. Concepts

As children develop, clinicians provide varying levels of scaffolding to help them learn [21]. One way clinicians scaffold is to have an assistant help the child hand-over-hand to perform an action [8]. They sometimes invite the child to perform an action without help first, as a way to measure progress.

The clinicians were concerned that extensive interaction with the robot might foster dependence on the robot or bring out undesirable behaviors in the child. To minimize such possible effects, each child interacts with the robot for 10 minutes out of each 50 minute session, as described in [7]. For the remainder of the time, the clinician interacts with the child in the same way they did before introducing the robot into therapy. In addition, clinicians use the robot as a pivot to facilitate interaction between the clinician and child. Because the robot plays a relatively minor role for a short time, the robot does not need complex choreography. Instead, simple plans with a few options are sufficient at this stage in the research.

Clinicians emphasized the need for lively robot animation and vocalization. Such things are needed to maintain engagement, which is important for social development. One particular class of animations clinicians have requested is songs that have hand and arm actions. For these songs, the actions need to be synchronized with the sound. Facial expressions were also considered important, but not to the same degree as energetic action and speech. These features are necessary for some children to attract their attention.

C. Models

As a result of our meetings, we developed a hypothesis for what robot-assisted therapy can provide. Interactions with the robot can engage the child. When these interactions are triadic, the engagement can lead to interaction with other people. Social interaction can lead to joint attention. Join attention fosters social learning, which leads to generalization. While nothing conclusive has yet surfaced, preliminary results are promising [8].

IV. USER INTERFACE IDEAS

The following list is a brief summary of clinical requirements for robot-assisted therapy. Clinicians need to be able to:

- Direct the robot discreetly in real-time,
- Animate new actions and adapt existing actions to each child as their needs evolve, and
- Create short, simple robot choreography that synchronizes actions with vocalizations and facial expressions.

The visual programming interface fills the need for directing the robot, and for part of creating the robot choreography. We must design another interface component to fulfill the need for creating animations with synchronized sound and facial expressions. We believe that a full-body tracking interface for creating and editing animations will be useful to clinicians. At present, we have a few ideas, but not a refined design for how the interface will work.

For robots that have an approximately human morphology (or some subset thereof), we can use motion capture to animate them. The Microsoft Kinect sensor, combined with PrimeSense's NITE full-body tracking, has made marker-less motion capture available at low cost. Integrating this system will save the time that would be required to implement our own full-body motion tracker. A user interface that allows for motion capture and simple editing of the captured motions may be very useful to clinicians.

Our user interface may use a combination of the Kinect sensor and a Nintendo Wii remote for discrete events. The Wii remote simply provides wireless button presses, as the Kinect is not sensitive enough to reliably pick out fine hand gestures while also capturing full body motion. Incorporating the Wii remote is a workaround, as it would be convenient to use hand gestures instead of needing a remote. On the other hand, using a remote for discrete events may be easier to learn than a system with hand-gesture recognition. Other methods that are possible are voice commands, special body gestures, or hand "dwelling" for example. The Wii remote may also prove useful to point at a screen to select menu items if full-body gestures prove to be awkward for such tasks. The basic idea is to use the Kinect sensor to capture motion and poses to use in animation, and the Wii remote for navigating menus.

As an example use case, the user stands in the view of the Kinect sensor, holding the Wii remote, and looking at a computer screen from a distance for visual feedback. Then the user presses a button on the Wii remote to start recording. The user then performs the desired action, and presses a button to stop recording. The user can then press another button to replay the captured and retargeted action on a virtual avatar of the particular robot they are animating. If something is not quite right, the user can edit the motion using a variety of methods.

To navigate through a replay of the captured motion, we propose three possibilities. One is to have the user pose or act out a short sequence that matches some part of the full motion. Another is to "scrub" through the animation by holding a button on the remote, then moving an arm up and down or left and right as if manipulating a slider widget. A third option is to use only the Wii remote buttons, accelerometers, or point tracking.

One method for editing is to correct a problem pose that came from either poor motion tracking or poor acting. The user first navigates to the problem pose using one of the above methods, then presses a button to signal pose correction. Then the user moves to correct this pose and presses a button to confirm the correction. This method for editing is similar to modifying key frames in traditional computer animation. After changing the pose, the surrounding frames in the animation are blended automatically (possibly with some assistance from the user to decide how many surrounding frames to smooth).

A second method for creating and editing motion is to first capture several important poses, then act out the entire motion fluidly. This interaction style is an adaptation of [19] to full-body tracking, instead of mouse or pen-based input. Such interaction allows the user to adjust the poses directly and ensure they look good, and then the complete action demonstrates the timing of those poses. It is not a problem if the complete action is not perfectly performed, as the system will only use the timing from the full performance. The user can demonstrate the action as many times as needed and use the original refined, important poses to get the timing right.

The next step for this project is to continue designing the interface and use existing work to guide and refine the design. Clinician feedback will continue to be a part of the design process. Once we have a reasonable design, we can begin development as we continue to iteratively refine the design.

V. VALIDATION

Once we have designed and developed a user interface, we will need to validate our claims. We can evaluate whether our interface enables clinicians to animate robots by having clinicians animate robots with our interface. One possibility is to describe a sequence of actions on a high level and ask the clinicians to make the robot act accordingly. Each clinician could create several robot animations. This kind of validation is beneficial because the target user pool is the population in the study. We leave validation to future work.

VI. CONCLUSION

Robots can be beneficial to therapy for children with autism spectrum disorder. For robots to reach therapeutic potential, clinicians need to be able to modify robot behavior. We can simplify the task of animating an anthropomorphic robot by giving clinicians supportive tools. Motion capture may enable intuitive generation of complex animations, and algorithmic analysis inside a gesture-based user interface can help to modify the captured animations intuitively and efficiently for non-experts.

ACKNOWLEDGMENTS

We thank Honda Research Institute and the BYU Comprehensive Clinic for their support.

REFERENCES

[1] J. Cassell, H. H. Vilhjálmsson, and T. Bickmore. Beat: the behavior expression animation toolkit. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '01, pages 477–486, New York, NY, USA, 2001. ACM. URL http://doi.acm.org/10.1145/383259.383315.

[2] D. J. Feil-Seifer, M. P. Black, E. Flores, A. B. St. Clair, E. K. Mower, C. Lee, M. J. Matarić, S. Narayanan, C. Lajonchere, P. Mundy, and M. Williams. Development of socially assistive robots for children with autism spectrum disorders. Technical Report CRES-09-001, USC Interaction Lab, Los Angeles, CA, 2009.

[3] E. Ferrari, B. Robins, and K. Dautenhahn. Therapeutic and educational objectives in robot assisted play for children with autism. *Procs 18th International Symposium on Robot and Human Interactive Communication, RO-MAN*, pages 108–114, 2009.

[4] D. Francois, S. Powell, and K. Dautenhahn. A long-term study of children with autism playing with a robotic pet: Taking inspirations from non-directive play therapy to encourage children's proactivity and initiative-taking. *Interaction Studies*, 10(3):324–373, 2009. URL http://dx.doi.org/10.1075/is.10.3.04fra.

[5] N. Giullian, D. Ricks, J. Atherton, M. Colton, M. Goodrich, and B. Brinton. Detailed requirements for robots in autism therapy. In *Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on*, pages 2595 –2602, 2010.

[6] M. Gleicher. Animation from observation: Motion capture and motion editing. *ACM SIGGRAPH Computer Graphics*, 33(4):51–54, 1999.

[7] M. A. Goodrich, M. A. Colton, B. Brinton, and M. Fujiki. A case for low-dose robotics in autism therapy. In *Proceedings of the 6th international conference on Human-robot interaction*, HRI '11, pages 143–144, New York, NY, USA, 2011. ACM. URL http://doi.acm.org/10.1145/1957656.1957702.

[8] M. Hansen. The effect of a treatment program utilizing a humanoid robot on the social engagement of two children with autism spectrum disorder. Master's thesis, Brigham Young University, 2011.

[9] R. Kadobayashi, K. Nishimoto, and K. Mase. Design and evaluation of gesture interface of an immersive walk-through application for exploring cyberspace. In *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pages 534–539, April 1998.

[10] M. Kipp. *Gesture generation by imitation: From human behavior to computer character animation*. Dissertation.com, 2005.

[11] A. J. Ko, M. M. Burnett, T. R. G. Green, K. J. Rothermel, and C. R. Cook. Improving the Design of Visual Programming Language Experiments Using Cognitive Walkthroughs. *Journal of Visual Languages & Computing*, 13(5):517–544, 2002.

[12] S. Kopp and I. Wachsmuth. Model-based animation of co-verbal gesture. In *Computer Animation, 2002. Proceedings of*, pages 252 –257, 2002.

[13] D.Y. Kwon. *A design framework for 3D spatial gesture interfaces*. PhD thesis, ETH Zürich, 2008.

[14] J. Lasseter. Principles of traditional animation applied to 3d computer animation. *SIGGRAPH Comput. Graph.*, 21:35–44, August 1987. URL http://doi.acm.org/10.1145/37402.37407.

[15] P. Maes, T. Darrell, B. Blumberg, and A. Pentland. The alive system: full-body interaction with autonomous agents. In *Computer Animation '95., Proceedings.*, pages 11–18, 209, April 1995.

[16] B.A. Myers. Taxonomies of visual programming and program visualization. *Journal of Visual Languages & Computing*, 1(1):97–123, March 1990.

[17] S. Papert. *Mindstorms: Children, Computers, and Powerful Ideas.* Basic Books, 1980.

[18] D. J. Ricks, M. B. Colton, and M. A. Goodrich. Design and evaluation of a clinical upper-body humanoid robot for autism therapy. In *In Proceedings of the 2010 International Conference on Applied Bionics and Biomechanics*, Venice, Italy, October 14-16 2010.

[19] S. C. L. Terra and R. A. Metoyer. A performance-based technique for timing keyframe animations. *Graph. Models*, 69:89–105, March 2007. URL http://portal.acm.org/citation.cfm?id=1224804.1224945.

[20] J. Wainer, K. Dautenhahn, B. Robins, and F. Amirabdollahian. Collaborating with Kaspar: Using an autonomous humanoid robot to foster cooperative dyadic play among children with autism. In *Proc. 2010 IEEE-RAS International Conference on Humanoid Robots*, Sheraton Nashville Downtown, Nashville, TN, USA, December 2010.

[21] D. J. Wood, J.S. Bruner, and G. Ross. The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2):89–100, 1976.

[22] M. Yim. Astronauts must program robots. In *AAAI 2006 Spring Symposium: To Boldly Go Where No Human-Robot Team Has Gone Before. Menlo Park, California, USA: AAAI*, 2006.