




The design of Lil'Flo, a socially assistive robot for upper extremity motor assessment and rehabilitation in the community via telepresence

Journal of Rehabilitation and Assistive Technologies Engineering
Volume 8: 1–26
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20556683211001805
journals.sagepub.com/home/jrt



Michael J Sobrepera^{1,2} , Vera G Lee³  and Michelle J Johnson^{2,3,4} 

Abstract

Introduction: We present Lil'Flo, a socially assistive robotic telerehabilitation system for deployment in the community. As shortages in rehabilitation professionals increase, especially in rural areas, there is a growing need to deliver care in the communities where patients live, work, learn, and play. Traditional telepresence, while useful, fails to deliver the rich interactions and data needed for motor rehabilitation and assessment.

Methods: We designed Lil'Flo, targeted towards pediatric patients with cerebral palsy and brachial plexus injuries using results from prior usability studies. The system combines traditional telepresence and computer vision with a humanoid, who can play games with patients and guide them in a present and engaging way under the supervision of a remote clinician. We surveyed 13 rehabilitation clinicians in a virtual usability test to evaluate the system.

Results: The system is more portable, extensible, and cheaper than our prior iteration, with an expressive humanoid. The virtual usability testing shows that clinicians believe Lil'Flo could be deployed in rural and elder care facilities and is more capable of remote stretching, strength building, and motor assessments than traditional video only telepresence.

Conclusions: Lil'Flo represents a novel approach to delivering rehabilitation care in the community while maintaining the clinician-patient connection.

Keywords

Rehabilitation robotics, telemedicine, community-based rehabilitation, socially assistive robotics, human robot interaction, telerehabilitation, social robotics, robot design, upper extremity rehabilitation, medical robotics

Date received: 31 October 2020; accepted: 19 February 2021

Introduction

There is a growing need to provide motor rehabilitation in rural communities and other resource-scarce areas. Many patients in rural areas are currently underserved by physical and occupational therapists.^{1,2} Due to a lack of local therapists, these patients must travel long distances to reach a center of excellence to receive care. For patients, this requires a caretaker, often a family member, to take time off work or take time off school. There is a need to improve care for both pediatric and adult patients who reside in rural areas; we focus on the pediatric patients in this work. Two large pediatric patient populations that are in need of life long rehabilitation are cerebral palsy (CP) and

¹Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania, Philadelphia, PA

²General Robotics, Automation, Sensing & Perception Laboratory, Department of bioengineering, Rehabilitation Robotics Laboratory, University of Pennsylvania, Philadelphia, PA

³Department of Bioengineering, Rehabilitation Robotics Laboratory, University of Pennsylvania, Philadelphia, PA

⁴Department of Physical Medicine and Rehabilitation, University of Pennsylvania, Philadelphia, PA

Corresponding author:

Michael J Sobrepera, Rehabilitation Robotics Laboratory, University of Pennsylvania, 1800 Lombard St, Office 19, Philadelphia, PA 19146, USA.
Email: mjsobrep@seas.upenn.edu



brachial plexus birth palsy (BP). CP is the most common motor disorder in young children, occurring in 2–3 per 1000 live births.³ It results in motor disorders of the upper and lower limbs that are often accompanied by other impairments in sensation, perception, cognition, communication, and behavior.^{4,5}

Brachial plexus birth palsy, also known as perinatal brachial plexus injury, obstetric brachial plexus injury, and simply brachial plexus (BP) has a different mechanism and pathology than CP, but similarly affects upper extremity function. BP occurs as a result of damage to the brachial plexus nerves during delivery, and occurs in about 1 per 1000 births.⁶ Outcomes from brachial plexus injuries can vary from self-resolving to long-term disability, requiring rehabilitation, potential surgery, and long-term management.⁷

CP and BP patients need rehabilitation care and medical services beyond the hospital, in community-based settings. In addition to local clinics, pediatric patients may have some form of care in their school or a local primary care clinic. In a specialized school, they may have access to an in-house therapist; elsewhere, they may have access to a traveling therapist who visits them for a short period each week. The remainder of their rehabilitation is often done at home with the aid of family members.

Rehabilitation is often a long-term need for CP and BP patients, and as a result compliance and motivation is often an issue. A lack of compliance may lead to decreased functional outcomes. Although compliance can be monitored when the child is with a therapist, it is more challenging at home. Routine diagnostics and assessments are critical to ensure that patients are receiving the best treatment for their condition and current state.⁸ This is especially important for rehabilitation, where patients with the same pathology can have divergent manifestations and challenges in their daily lives.⁹ Frequent assessment of function can give the care team a good indication of a patient's progress and could provide a method to motivate patients, and is therefore an important aspect of ongoing rehabilitation.

One way to address clinician shortages in medically underserved areas is through telemedicine: the remote application of medicine using telecommunications, which generally include two-way video and audio.¹⁰ For example, evidence has shown that range of motion in adolescents can be measured effectively via telemedicine, and subjects may prefer telemedicine visits for certain tasks.¹¹ For clarity, we define telemedicine as specifically using remote audio and video (i.e. telepresence) for medicine and allow electronic health care (e-health) to define the larger scope of any medicine practiced using remote electronic communications. Telemedicine enables patients to receive care in the

community, their local clinics, schools, homes, etc. This can help overcome barriers to care due to travel and scheduling, potentially increasing the number of interactions.

This method has its limitations. Interactions over telepresence are not as rich as in-person therapy. Limits from the field of view, resolution, projection of three dimensions onto two, and network latency can decrease the sense of presence of the remote person and hamper their spatial reasoning.^{12,13} The lack of a present person could lead to a reduction in patient motivation. When combined with less clear instructions through telepresence, where seeing what a clinician wants a patient to do can be more challenging than in person, patient compliance might be reduced. If patients are unmotivated and non-compliant, the clinician may not see the motions needed to accurately assess the patient and the patient may not get the personal benefit from the interaction.

One solution to augment the richness of remote interactions is to use Socially Assistive Robots (SAR).¹⁴ SARs combine the space of assistive robots, which give aid or support to human users (e.g. traditional rehabilitation robots, wheelchair robots, mobility aides, etc.) and socially interactive robots, whose main task is to interact with people socially. The SAR is therefore designed to create close and effective interaction with human users in order to give assistance, leading to measurable progress in convalescence, rehabilitation, learning, etc. SARs could be coupled with a telepresence system and a computer vision system to improve remote health care delivery by providing a physically present social entity for patients to interact with. The SAR could act as a mediator in the interaction, thus providing a richer experience. In addition to providing presence, by operating locally to the patient, a number of other challenges with video based telepresence could be addressed. The entire robot would be visible to the patient, solving field of view challenges. And the robot would be able to act smoothly even in poor network communications, when the video with the clinician would be degraded.

To explore the idea of using a SAR with telepresence, we have designed two generations of systems. Based on our initial prototype ("Flo", seen in Figure 1),^{15,16} discussions with clinicians, and the literature, we developed a series of design requirements for our second iteration, "Lil'Flo". The goal for the Lil'Flo system is to help understand how pediatric patients with motor impairments requiring telerehabilitation, the application of telemedicine for rehabilitation, can be assessed efficiently, cheaply, and accurately. The system enables the examination of two critical components: the telepresence system for communication with



Figure 1. The first version of our social robot on a telepresence system, Flo, composed of a Nao robot attached to a VGo telepresence robot.

the patient and the perception system for understanding the patient's motor function.

Our long-term study goal is to be the first to rigorously compare classical telepresence-based assessment and the use of a social robot as a mediator in telepresence. The ideas of social robots and telepresence are not novel, nor is the study of remote versus physically present agents. The innovation of this project will be to use a social robot as a third agent, physically present with the patient, to mediate telepresence. Many other systems attempt to temporarily replace the therapist with the social robot. Our goal is to extend the reach of the clinician into the community and improve their ability to remotely interact with their patients and accurately track their rehabilitation progress. The social robot would therefore act as a bridge for pathways of communication lost over telepresence (Figure 2). We hope patients will be more motivated and compliant in performing their activities at home and will be able to

receive instruction on improving their rehabilitation. This model for interactions has driven the design decisions for Lil'Flo.

This paper reviews the related literature, the design requirements for the Lil'Flo robotic system, the design developed, and future directions.

Related literature

Rehabilitation robots that are able to gesture, communicate, motivate, comfort, and teach exercises have been tested in pediatric and geriatric interactions in and out of hospital environments.^{17–23} Experiences with these robots have been positive, and the robots have been shown to promote engagement and adherence to prescribed exercise. However, to our knowledge, they have not yet been tested as a tool to augment telemedicine as a third agent to interact with the patient and therapist.

SARs

A number of social robots have been developed for upper extremity rehabilitation, from which we take inspiration. The Nao-Therapist project initially developed a custom robotic bear named Ursus^{24–26} and has now moved to a Nao robot,²⁷ which is used for upper extremity rehabilitation for pediatric patients. The system uses a Microsoft Kinect sensor to track patients, allowing the robot to autonomously play games with them. It can both demonstrate and correct poses in a pose mirroring game and in a pose sequence recall game. In a longitudinal study of the system, with 13 subjects participating in on average 11.6 sessions of approximately 24 min each,²⁸ all stakeholders, clinicians, parents, and children, found the system useful and wanted to continue to use it.

RAC CP Fun is another Nao-based robotic platform designed to engage with preschool students who have CP.¹⁸ The robot can play various games and motivate physical activity. The interactions with the system were designed to build off the motor learning literature, with an emphasis on giving feedback to the patient. The robot interacts by singing songs, changing its position relative to subjects, and providing feedback. Fridin et al. compared outcomes of using the robot between typical children and CP children, finding that the CP group exhibited a higher level of interactions as measured by the child-robot interaction measurement index which relies on eye contact as well as various facial, body, and vocal expressions of emotion.^{18,29}

Another Nao-based system is Zora, which is commercially available. It has been tested on a cohort of children with disabilities and has been shown to improve the quality of care.³⁰ However, it was reported

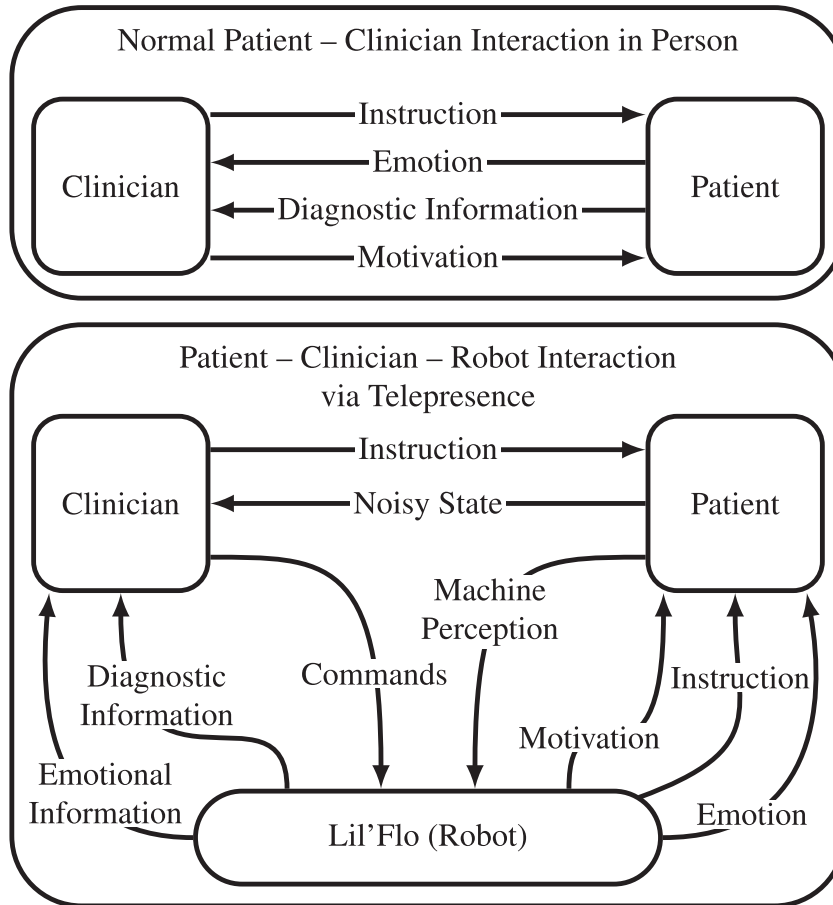


Figure 2. A comparison of our theory of interactions between patients and clinicians in-person and patients, clinicians, and the robot in telepresence interactions. Without the robot in the telepresence interaction, there is a loss of communication channels compared to the in-person interaction.

that the software required to operate the system was labor intensive for clinicians.

The idea of socially assistive robots as mediators for interactions has been explored for in-person therapy of children with behavioral disorders such as autism spectrum disorder, where direct human-human interaction can be challenging. The Milo and Kaspar robots are two examples of this with sizable deployments in the clinic.^{31–33} Neither are designed for neuromotor rehabilitation.

SARs can easily become complicated systems, it is important to consider how to make systems approachable by being simple and affordable. The CosmoBot system is a good example of how simplifying problems can lead to effective systems. CosmoBot is a small toy-like space robot, integrated into “Cosmo’s Learning Systems”. It has arms with a single shoulder degree of freedom, an actuated mouth, an actuated head, and the ability to drive around. It interacts with patients through a button board, accelerometers placed on the patient, and 3rd party interface devices

(e.g. joysticks, buttons). During a 16-week longitudinal study with interactions once a week with four subjects aged 4–10 with CP, it was shown that the system itself was robust and easy to use.³⁴ Patients were engaged and excited to play with the robot throughout the length of the study. The system was marketed for a few years by AT KidSystems. Even with its limited number of degrees of freedom in its arms and torso, it was still able to motivate patients to work on their rehabilitation goals. By using the trackers on the patients’ bodies, the robot was able to both interact and collect objective data throughout the study.

Tega is a small smartphone-powered robot designed primarily for education, helping students to develop language skills through interactive storytelling.¹⁹ It has a design that is supposed to be cute and approachable with five degrees of freedom, allowing it to bob up and down, twist, lean, and look up and down. Tega is inappropriate for most physical and occupational rehabilitation techniques, as it has no limbs. However, it is worth appreciating for its emotional expressiveness,

based on principles from animation, and its relatively low cost, using a cellphone for both its face and computational power.

Recent studies have proposed the use of virtual agents with digital embodiment in immersive environments for assessment and rehabilitation. Augmented and virtual reality (AR and VR) are methods used to create these immersive environments. VR has previously been explored for utility in motor rehabilitation for stroke patients suffering from upper extremity impairments. In a meta-analysis of 12 studies on the effect of using VR in rehabilitation of stroke patients, it was found that there was a 15% improvement in arm motor impairments and 20% improvement in arm motor function when using VR rehabilitation compared to traditional rehabilitation from pooled observational studies.³⁵ Evidence also suggest that VR can lead to high levels of engagement and motivation. The technology is relatively affordable and accessible.

Triandafilou and colleagues developed a 3D, networked multi-user VR environment for home therapy after stroke where physically remote users are able to interact via avatars whose movements are controlled by the kinematics of the user, retrieved from Kinect cameras located in the home.^{36,37} Their recent clinical trial indicated that stroke survivors were able to adhere to the rehabilitation treatment and that arm displacement during a session was a reasonable metric to track progress over the game. Challenges remain in the complexity of interacting within such systems. Further, it is not clear how the effect of having an entire interaction in a 3D world projected through a 2D computer monitor compares to other alternatives where 3D embodiment exists.

Elor et al. used a VR system to gamify constraint induced therapy for patients with hemiparesis in Project Star Catcher, and found that it was motivating and a cost-effective form of therapy.³⁸ Project Star Catcher focuses on single users, and does not attempt to embody the user or any avatars, instead immersing the player, using a head mounted VR display, in a new reality, catching stars. Their work demonstrates that head mounted displays and off the shelf controllers can be used to provide immersive environments to people with motor disability while measuring kinematics for progress tracking.

Liu et al. leverage embodiment in a virtual environment to help with rehabilitation from gait asymmetry. Rather than projecting an avatar of another entity, they project an avatar of the patient walking,³⁹ allowing the patient to see their own gait. They showed that this resulted in better gait speed and more optimal gait lengths.

These applications suggest that a digital embodiment as well as an immersive world both on a screen

and via a head mounted display can be successful for both rehabilitation and assessment of patients remotely.

But how do virtual systems compare to physically embodied ones? The social robot community has studied questions of robot presence. Fridin and Belokopytov demonstrated that when comparing an in-person robot and robot projected onto a screen, pediatric subjects interacted significantly more with the in-person robot.⁴⁰ Bainbridge et al. showed that physical presence is important for trust and motivation, especially for uncomfortable tasks,⁴¹ and Kiesler et al. showed that subjects co-located with a physical robot were more engaged with it and acted more ideally around it (as measured by following diet advice) when compared to a virtual agent.⁴² A recent study similarly found that stroke patients prefer using SARs for rehabilitation therapy over virtual agents shown on a screen, reporting higher engagement levels and exercise performance with the physical robot.⁴³ Mann et al. demonstrated that subjects were more likely to trust, be engaged with, and follow instructions from a robot giving instructions and asking questions, compared to the same interface on a tablet.⁴⁴ In fact, a study has shown that simply adding a SAR onto a pre-existing neurorehabilitation device as a third agent for motivation and engagement purposes has the potential to increase performance results.⁴⁵

Both virtual reality systems and SAR-augmented telepresence systems such as Lil'Flo are worthy of investigation for motor rehabilitation uses. There is not yet clarity in the literature on user preferences and outcomes with the two modalities, especially as they both continue to advance.

Telepresence systems

There have been many telepresence systems developed. The simplest telepresence can be achieved using cellphones, tablets, or computers with a screen, camera, and Internet connection. More advanced systems use a screen on a stick morphology with a screen and camera mounted on a mobile base, allowing the remote operator to drive the telepresence interface. Still more advanced are systems that utilize actuation of the screen based on where the remote operator is looking.^{13,46} Some systems also have appendages which can be moved to show operator intent⁴⁶ or to achieve a goal.

Dodakian et al. presented a system for stroke patients, using a custom tabletop game system attached to a computer for both individual activities and activities monitored by telepresence.⁴⁷ The system did not use a robot, but by making the games physical, compliance was increased. Using the games with

intermediate telepresence provided a way to maintain motivation and provide ongoing assessment. What we propose is different from these: the telepresence operator is presented in a traditional screen-on-a-stick fashion, while the robot is an independent social entity which can play with subjects and demonstrate actions of interest.

It may be of interest to extend features of Lil'Flo to VR by bringing remote clinicians into patients' virtual environments on an in world screen while the patient interacts with an autonomous or remote controlled avatar.

Objective assessment

A focus on accessible metrics has been missing from the socially assistive rehabilitation robotics and telerehabilitation literature. In order to make telemedicine feasible from the clinician's perspective, tools are needed to assess the patient's function quickly and easily. As shown in the Shriners Hospital Upper Extremity Evaluation (SHUEE)⁴⁸ and Assisting Hand Assessment (AHA),⁴⁹ assessment can be done by graders watching recorded video but requires expert graders and is time-consuming, limiting the availability of the test and increasing its cost.

There is a large literature of metrics for upper extremity function that can be evaluated using robots like the MIT-Manus.⁵⁰ These generally rely on the kinematics of subjects, but do not provide the outputs that clinicians expect to work with and require expensive fixed hardware to be used. By extending from this work to use low-cost mobile hardware with computer vision that leverages modern machine learning techniques and working to generate metrics which are intuitive to clinicians, the ability to assess patients could be improved.

Design requirements for Lil'Flo, a socially assistive robot for upper extremity motor assessment and rehabilitation via telepresence

The requirements for the Lil'Flo project are presented below. We discuss our initial prototype and the evolution of the design requirements that lead to those we used in Lil'Flo robotic system.

Our original prototype

The first version of our social robot system was comprised of two individual pieces of hardware: the VGo base and the NAO humanoid (Figure 1). The systems did not interface with each other and were controlled through separate software, which made setup of the

robot and system operation difficult, even for the researchers familiar with it. For further experimentation with design, we wanted a more modular system than what the NAO and VGo offered. We found that although the NAO is highly programmable and easy to use, it is hard to modify, and when it breaks, hard to maintain. A computer and cameras are permanently built into the NAO and are unable to be upgraded. The VGo is user-friendly in its default configuration but offers no programmatic interface to extend its capabilities and adding additional hardware is non-trivial.⁵¹ We found from testing that the NAO platform lacked sufficient sensors to perceive patient interaction and again its cameras are not upgradeable. The NAO has a static face with eyes that can change color; it has been claimed that this, with the pan tilt of the head, is sufficient to convey emotion. But we felt that further facial expressiveness was needed.

The sizing of the system also proved to be difficult, as it is too tall for most pediatric encounters but has a small humanoid for adult encounters. Because of how the structure of the two robots fit together, the NAO's center of gravity sat in front of the VGo's center. To counteract this imbalance, an extra weight was added on the back of the VGo, which significantly increased the mass of the system and raised its center of mass, making the system less stable. We used the torso-only version of the NAO, which is no longer sold, and faced challenges with the robot overheating during use. Although newer models have improved on the overheating problem, multiple fellow researchers and collaborators have reported that it can still be a challenge. The NAO/VGo combination was also costly: The current full body NAO is cheaper than when the original prototype was built, but still costs \$9000 USD pre-tax. The VGo robot costs \$4000 USD or \$6000 USD with Verizon 4G LTE with a dock. There is an optional service contract which is around \$400 USD depending on the term of the contract. Additional modifications would then be needed to add sufficient cameras and compute power, adding further cost. This type of integration requires significant engineering effort and requires compromises⁵¹ which we did not believe were logical to undertake. The cost of components is not necessarily important for research, but by proving that low cost components can be effectively used, an argument can be made that a larger impact can be achieved post technology translation.

The primary takeaways from our initial demonstrations and surveys with the first version of our social robot were that clinicians viewed the robot as a social entity, although they did not find it to be as helpful as we had hoped given its lack of modularity, difficulty in setup, and high cost.¹⁵

Design requirements for Lil'Flo

In developing our new prototype, we decided to rethink some of our prior design assumptions leveraging our new data. Our primary design requirements became that the robot be low-cost, flexible and adaptable to testing, expressive, programmable, and sized for pediatric as well as adult patients. We decided to split our concerns into two sister platforms: one for adult populations (Big'Flo) and one for pediatric populations (Lil'Flo). We are focusing first on the pediatric system, with plans to port the design learnings to a larger adult version with similar capabilities, but a larger size and more adult-oriented aesthetic.

Interaction to support motor function. During motor rehabilitation interactions, there are three primary categories of motion that we are seeking to drive: motions that exhibit the range of motion of the patient, motions that allow observation of the kinematics of the subjects' movement (what does the position vs. time profile of motions look like), and bi-manual motions. It is important that the patient understand what they are being asked to do so that they can attempt the activities. Because different patients will be at different levels of function, the pacing of activities has to be adjustable to the patient on the fly. Ideally, the difficulty of activities should also be adjustable.

Because we are interested in upper extremity rehabilitation and demonstrating movements with a humanoid robot, it is important that the humanoid has arms. At a minimum, the arms need to be able to perform shoulder flexion, shoulder extension, shoulder adduction/abduction, shoulder internal/external rotation, and elbow flexion/extension, these are the primary motion planes affected by brachial plexus injuries. For cerebral palsy, wrist supination and even hand manipulation are relevant as well, but our clinical partners have told us that useful testing can be done without them.

With this platform, we plan for all interactions to be seated, which means the humanoid should appear to be seated as well. Although there are some exercises that can be done with the legs while seated, the focus is upper extremity work, and so legs are not critical.

Interaction to increase engagement. The literature supports the idea that social robots improve engagement and elicit social interactions that keep both pediatric and geriatric patients engaged.^{21,28,52-54} In rehabilitation, motivation is linked to motor function improvement, thus we believe that the use of robots with the ability to interact will be beneficial for patient outcomes. In order to be motivational and build meaningful interactions with subjects, the robot should appear friendly.

The humanoid portion of our system is designed to interact with patients socially, and an important component of social communication is facial expression. The design of a face can affect the perceived trustworthiness, likability, and friendliness of a robot.⁵⁵ Speech is also a necessary component for social robot communication.

Robot as a peer. The goal for the robot is for the humanoid to interact with the patient as a peer. The robot should not act as a doctor, therapist, or nurse, but as a friend who is playing games and doing activities. To have mass appeal, the robot should be somewhat ambiguous in its identity: for the system to be approachable to a wide variety of patients, it should not be strongly male or female nor be strongly of one race nor creed. In order to avoid the uncanny valley, the robot should not be overly human.

Ensuring that the robot is interesting, intelligent, sociable, able to communicate, and helpful are important. Engineers therefore must consider what parameters are needed to ensure these features are realized. It is important to highlight that none of this needs to be authentic, the interactions can be a show, as long as they are convincing.

Portability. In our new system, we aimed to reduce the weight of the system and lower the center of mass. To allow the system to be tested and used in different locations, it needs to be portable without too much difficulty. It should therefore be liftable by a single person. Within the space which it is deployed, the system should be mobile under its own power. This allows the remote operator to move the robot without assistance, which is important to the feasibility of telepresence to expand access to care. Mobility also improves the sense of presence for the remote operator.⁵⁶

Easy setup. We received feedback from clinicians that such a robotic system needs to be easy to setup. Although this prototype is not envisioned for unsupervised clinical deployment, ease of setup helps prevent errors by reducing the cognitive load on the operators. Additionally, because the robot is designed to work in medical care situations, it must be easy to clean.

Modularity. It is important that the mechanical design is modular at multiple scales. At the system level, the humanoid needs to have a head, face, and arms that can be modified and replaced. For the benefit of research, this modularity allows the system to be used as a platform for developing and testing a variety of designs. For example, being able to add extra degrees of freedom to arms, add lighting to the chest, or change

the style of face is useful. Modularity is also useful for repairability and upgradeability. At the platform level, to test the efficacy of different systems, namely the humanoid, it is important that they can be removed.

Remote operation. Given the goals of the system, enabling remote operations and telepresence are clearly necessary. It is important to demonstrate that the telepresence-related technologies work well and reliably, but the interface presented to the remote operator does not need to be fully refined at the research stage of development, so long as the lack of refinement does not impair function.

For Lil'Flo's intended uses, one hour of runtime is required at a minimum, to allow the system to run for the length of a study. For true clinical deployment, a much longer run time would be desired. It is not necessary for the robot motors to be particularly powerful, as the system does not walk or lift anything other than its own arms, relying instead on social interactions to drive rehabilitation goals. It is important that they be sized large enough that they do not overheat when moving the arms.

Programmable by clinicians. Most interactions with this prototype will be programmed and managed by researchers, but both for their benefit and as a proof of concept for clinician-controlled operations, it is important that both programming and operation be as simple and user-friendly as possible. From this, direct constraints were developed for both programming and operation: all operations should be done through a single interface, and interactions should be able to be developed through a graphical interface. Similarly, it is important that the underlying code is easy to understand and leverages existing frameworks to enable future development. The end user should not need to be aware of, competent with, or think about the underlying code during operation.

It is important that users, whether clinicians or researchers, do not need to install any non-standard software. Many robotics systems use Linux-based software. However, Linux is not well-received in healthcare facilities. For general adoption of the system, the technology must be as easy as possible for clinicians to learn to use and require minimal IT overhead.

Provide assessment information. An important component of ongoing rehabilitation is continuous assessment. Assessment allows clinicians to update the rehabilitation regimen, help give ongoing instruction, and motivate patients. We believe that one of the key use cases for telepresence systems with social robots will be intermittent assessments, while patients complete most of their rehabilitation at home. Doing assessments

remotely is possible. The Shriners Hospital Upper Extremity Evaluation (SHUEE)⁴⁸ and Assisting Hands Assessment (AHA)⁴⁹ are two examples of assessments that are done using recorded video. Although recorded video can be used to perform assessments, they are time-consuming and must be graded by multiple clinicians to make assessments objective. Clinicians whom we have talked to describe having to watch videos multiple times to accurately assess the patient. At the same time, many assessments fail to be objective or gain a sense of objectivity by significantly reducing the dimensionality of the test. An overview of clinically relevant assessments of upper extremity function are provided by Gilmore et al.⁵⁷ and Wagner et al.⁵⁸ It is important that the system we are building provides the infrastructure for exploring remote objective assessment and that the interactions which we design elicit actions that are valuable for assessment.

Improve quality of interactions for patients. It is important that all of the design is done with the goal of improving the quality of interactions for patients. Improving outcomes is important, but so is focusing on how the patient feels during interactions. This is implicit in the prior design requirements. Within the goals of increasing motivation and engagement with SARs is a natural goal that patients will enjoy the interactions more. The added goal of Lil'Flo, above and beyond prior SARs for upper extremity motor rehab, is to give patients access to both the robot and clinician without needing to travel as far as they might otherwise. The trifecta of goals: better outcomes, enjoyable interactions, easier access will enable a system which is patient focused.

The prototype of Lil'Flo

Figure 3 shows the current generation of the robot with and without the humanoid. Since Lil'Flo is designed to be relevant for upper extremity rehabilitation, it is designed with an anthropomorphic form to allow the system to demonstrate human motion naturally. Because we are targeting pediatric patients, it has proportions similar to those of a child through its arms and torso, with a simple head, reminiscent of a toy, giving emphasis to its face. The robot supports shoulder flexion/extension, shoulder adduction/abduction, internal/external rotation, and elbow flexion/extension. It is built using motors from an XYZ Robotics Bolide robot. The shell around the motors is designed to make the robot more aesthetically pleasing and structurally sound. The robot is designed to be mounted onto a mobile base on which it will appear seated.

The first step in designing the robot was to develop sketches of what the system might look like. Sketches

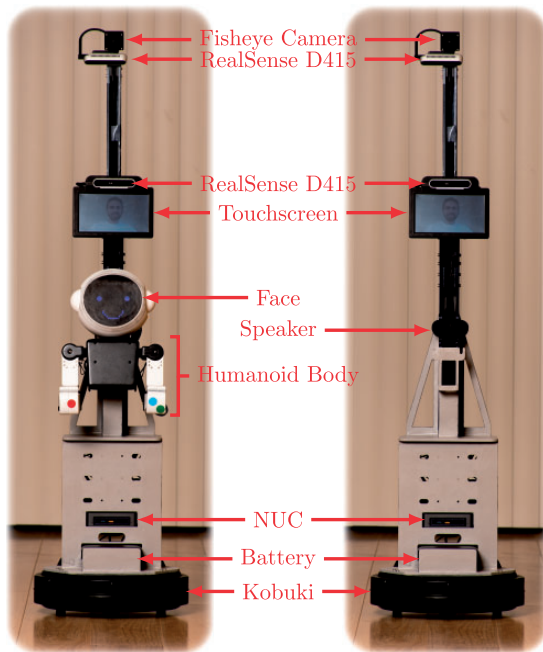


Figure 3. Our second generation socially assistive robot with telepresence, Lil’Flo, constructed using smart servos, a custom frame, and a custom exoskeleton. On the right, the same system, without the humanoid.

generally fell into the categories of spaceman, toy/doll, and anime (Figure 4). As we continued to refine our ideas, the three major design themes evolved into a spaceman, animal, and child theme (Figure 5). We worked to simplify these ideas, developing primarily with the spaceman theme.

We felt that making a generic concept which still had a geometrically interesting form would appeal to the greatest number of people. The color scheme is primarily white and black (Figure 6), providing good contrast along the arms to easily see the components and differentiate body parts. The goal of the neutral color scheme is to give the system a neutral sentiment and allow the face and motions to drive the emotional state, as well as to prevent the robot from seeming to be from any human group.

Body and arms

Lil’Flo uses motors, a control board, and the chest of the XYZ Robotics Bolide robot, a commercially available, affordable, edutainment robotics platform (Figure 6). The Bolide’s motors are serially controlled and can provide digital feedback. However, the Bolide is not appropriate for upper extremity rehab in its native form, lacking appropriate placement of degrees of freedom and having an exposed skeleton. To rectify

this, we have developed a custom exoskeleton for Lil’Flo, which has a visually pleasing exterior, and is designed to minimize weight and assembly steps, allowing easy maintenance and experimentation on the robot’s form (Figure 7). The shells have internal structures which give them rigidity while remaining lightweight. The design leverages the rigidity of the motor casings where possible. The motors are placed as high up the kinematic chain as possible, one in the chest, two in the upper arm, and one in the forearm, to improve performance. The motors are fully encased and pinch points are minimized. The hard exterior makes it possible to wipe down the robot, but it is not sealed to fluids or dust.

To produce a test ready system, the parts were 3D printed in an ABS like material using fused deposition modeling. The parts were sanded, filled with high build primer, and painted with multiple coats of semi-gloss paint. We felt it important that the parts look production grade to prevent low quality fabrication from being a confounding variable in testing.

The system of motors is controlled by custom software, exposing it to the robot operating system (ROS). The interface works over a serial connection to a microcontroller which performs the real time operations to interpolate the motors over motions. The microcontroller communicates with the motors via a second serial connection. This allows the arm’s movement to be captured, visualized, and controlled in ROS.

The Bolide system provides better than hobby grade servos which are digitally controlled with integrated low-level controllers and can provide feedback to the higher level system. These motors were selected because they are cheap, serially controllable servos that can generate enough torque for our system (stall torque of 25 kg-cm), sufficient to wave the arms around.

Head and face

Our second physical prototype for the head can be seen in Figure 8. The prototype highlighted the importance of having a screen on the face which shows only the eyes and mouth while hiding the internal mechanics and gave direction to the correct proportions for the head. During prototyping, an informal straw poll between a head with and without ears showed that the ears were preferred, as they gave a sense that the robot could hear.

The current head is broken into three major components, a front shell with a translucent urethane panel, the back, and an inner skeleton which holds the LED matrices (Figure 7). The front and back sections connect with a seam that follows along the back of the ears and above the top of the head, making a clean line

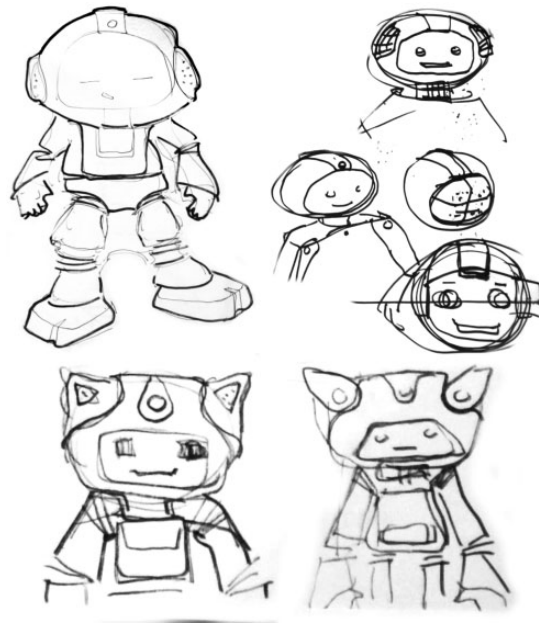


Figure 4. Early concept sketches for the robot showing a spaceman like concept, a toy/doll concept, and an anime like concept.

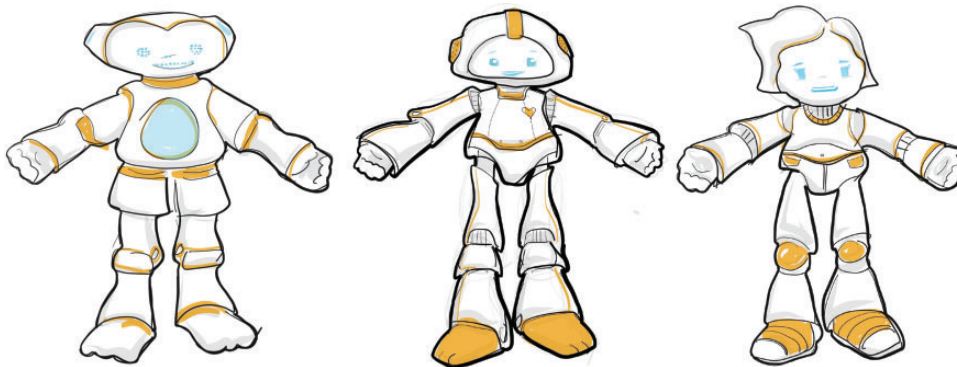


Figure 5. More refined sketches showing an animal, spaceman, and child theme.

which could be interpreted as a hairline. The head is 3D printed out of white ABS like material and sanded.

Because the face is often the center of attention for human–human interactions, we want the flexibility to alter the facial expression on the fly. Initially, LCD screens were explored to create the face, but it was challenging to get the geometry of the head and face to work with a single large screen while being affordable and having a bright screen. LED matrices solved these challenges. The primary compromises with using LED matrices are that they are single color and have low resolution. The other option for making a digital face is to use a projection based system, as Quori uses.⁵⁹ The projection system can handle complex geometries and provide good brightness. However,

fitting a projector and reflector in a small platform is challenging and costly.

We ultimately designed a head with a dark translucent face and variable brightness LED dot matrices behind the surface (Figure 6). Three LED dot matrices, one for each eye and one for the mouth, are controlled from a Teensy microcontroller which provides power to the matrices and communicates with them over serial. The microcontroller then communicates with the remainder of the system and receives power via a USB connection. The LED matrices allow us to alter the expressions of the robot while keeping the face simple from an aesthetic and maintenance perspective.

To achieve the dark face, a translucent black urethane screen is molded into a thermoplastic 3D printed

shell of the head. Molding is used rather than other techniques to allow the screen to have a geometrically interesting shape, with a curved forehead, and to allow the piece to seamlessly exist within the rest of the head. This is done by making a positive of the face and screen together and pouring a mold to match it (Figure 9). This yields a mold which fits the front of the face and defines the shape of the screen (Figure 9). The 3D printed face is then placed into the mold and a clear urethane which has been mixed with a black colorant is added (Figure 9).

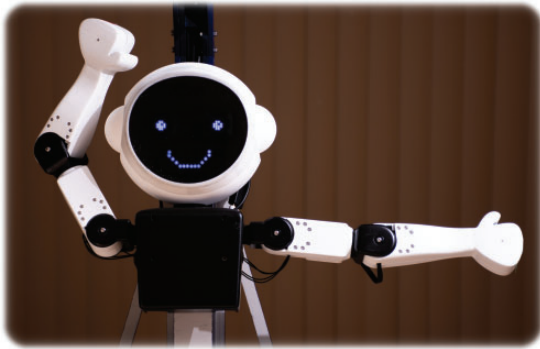


Figure 6. A close-up shot of the Lil'Flo humanoid robot. The color scheme uses white arms with black joints and a black chest. The head is white, with a dark screen backed by blue LED lights in a matrix.

By using a dark face, we provide good contrast to the head with clean lines, highlighting the face as a point of attention. The dark face also allows us to hide the internal structure of the head while allowing good transmission of light from the internal LEDs, leading to good facial feature visibility.

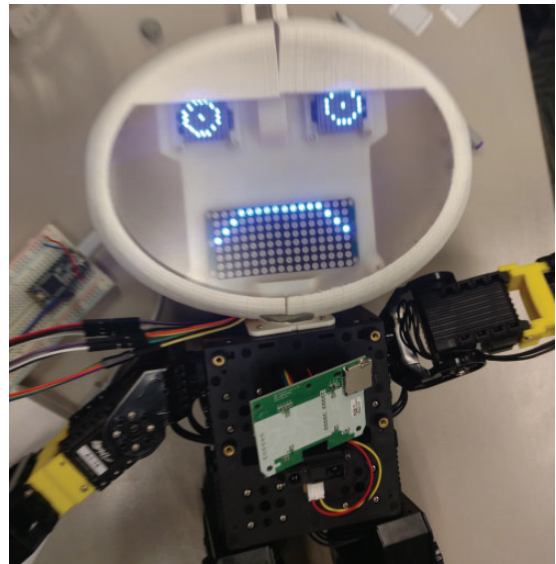


Figure 8. The second physical prototype of the head for the robot. The head is too wide for the body and the clear screen exposes all the internal components.

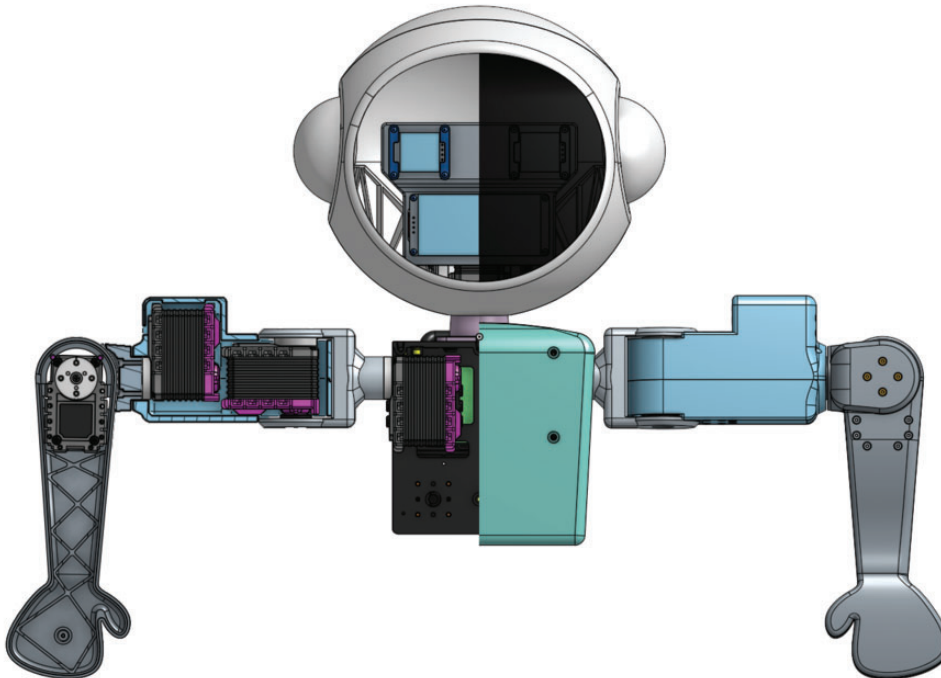


Figure 7. The computer-aided design of the humanoid with a cutaway to show the internals of the arms and torso, allowing the orientation of the motors to be seen, and head, allowing the internals of the head, including the display screens to be seen. The same motors are used throughout the design. Wires are not shown. Fasteners are hidden in the cut out region.

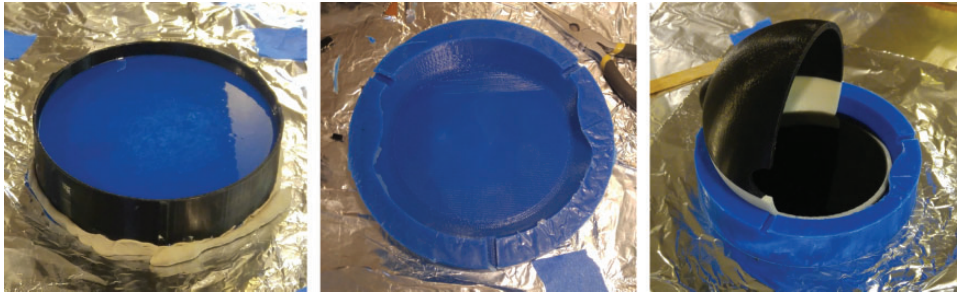


Figure 9. On the left, making a mold to make the face screen. In the middle, the completed mold for the face screen, which is used to mold the translucent front screen. On the right, the translucent front screen being molded into a prototype 3D printed model of the head. The material being used is a clear urethane with black colorant added.

Finally, a series of patterns for the LED matrices had to be designed. Eye and mouth proposals were developed, taking inspiration from emoticons and general facial expressions. These were distilled down to mouth/eye combinations which we felt would be most relevant to our use cases. Some of the eye sets are directional (i.e. able to look left, right, up left, down, center, etc.), others only have a single direction.

Base

The base of the robot is built with an off-the-shelf Kobuki mobile robot platform with a custom set of risers to hold the computer, humanoid, cameras, and screen. The Kobuki was selected because it is affordable and integrates well with the Robot Operating System (ROS). It also provides mounting points to build custom systems off. The structure of the remainder of the frame is constructed of MDF, which was laser cut, glued, filled, smoothed, and painted. The paint uses a primarily gray color in satin with blue accents. There are three distinct areas of the base: The first section is the bottom, which attaches to the Kobuki, houses the computer, a USB switch, a battery, and excess wiring. The bottom area also has mounting holes that were used to explore camera placement. The middle area mounts permanently to the base and holds the humanoid. The humanoid hangs from this section, held by gravity and friction, allowing it to be easily removed. The middle area also houses the screen with one of the cameras. The screen is held within a custom 3D printed casing that also provides a mount for a camera. The top section is mounted to the middle section using screws. This was added to the design based on feedback from clinical partners to allow a better field of view for cameras by getting them higher up. The top section holds two cameras.

The base height was selected after some trial and error to make a stable system. The middle of the face of the robot is around 72 cm off the ground, the middle

of the screen is about 90 cm off the ground, and the top of the robot is about 132 cm off the ground. For interacting with people who are sitting, this produces a system which is low to the ground, but higher than a Nao robot on the ground which has been used in multiple other SAR focused rehabilitation studies. When sitting, we have found these heights to be on the short side of comfortable for an adult. We did not want the humanoid to tower over a subject and because the system is designed for children, being on the shorter side may actually be an advantage.

System architecture

The entire system runs on an Intel NUC with a solid state drive and 16 gigabytes of RAM. Connected to the NUC are a pair of Intel RealSense D415 cameras and a USB 3.0 powered hub (Figure 10). The hub is connected to a 180 degree fisheye camera by ELP, a small speaker which uses USB for both power and input, a touchscreen panel, the robot's face, the robot's body, and the Kobuki base. The NUC is also connected by HDMI to the touchscreen panel. The screen is a 7 inch 800 × 480 pixel TFT screen with a resistive touch overlay. The screen is driven by a driver chip, which provides the HDMI and USB interfaces. The NUC has a dual microphone array built into its front panel, which is exposed through the robot base.

Audio is captured by the microphone array built into the NUC. Because the NUC is placed about half-way up the base, this provides reasonably good sound quality. Audio for Lil'Flo's voice and the remote operator are provided through a speaker behind Lil'Flo's head. The remote operator is shown on the TFT display.

Cameras. Video is captured by two Intel RealSense cameras and a fish-eyed camera. The Intel RealSense D415 cameras are used because they give reasonably

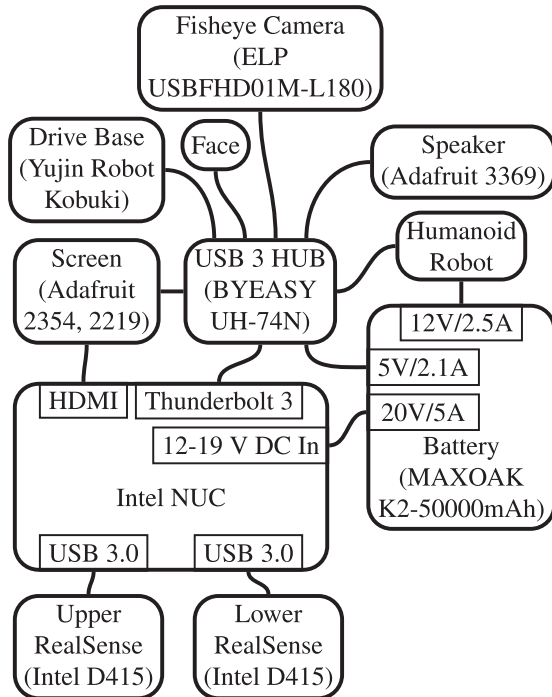


Figure 10. The system's primary components. The NUC has two USB 3 ports on its back panel as well as a Thunderbolt/USB 3 type-c connector, a HDMI port, and a power port. The RealSense cameras are each connected to an independent USB port to guarantee that they have sufficient available bandwidth. A powered USB hub is connected to the type-c port. The USB hub, NUC, and motors within the humanoid are all powered by a MAXOAK 185 Wh/50,000 mAh battery. The screen uses a USB connection for both power and to provide touch input to the computer. The Kobuki base is only connected via USB. It maintains an independent power system. The microphone is built into the NUC and so not shown. The speaker, face, and fisheye camera all use USB for both power and data.

good depth results, when compared to similar cameras, but do most of the processing on board. They are low-cost, low-power consumption, light-weight, compact, and have a minimum z-range less than 0.3 m. Because the RealSense cameras use true stereo, they can tolerate uses in environments with multiple other cameras and with strong infrared light, such as near windows. The D415 places all of its imagers on a single rigid plane, allowing them to stay calibrated to each other. The sensors use a rolling shutter, but for low speed human tracking, that is sufficient. Because of the relatively narrow field of view for each sensor, they have a small angle between each pixel, giving fine resolution, which is useful for pose tracking. They also use RGB+IR sensors for their depth stereo pair, allowing them to more easily find features in color rich environments.

A challenge of building the system was getting cameras placed in such a way that they could see the subject

at a variety of distances. In order to tolerate both full range of motion activities and activities touching the robot, across a range of ages, the field of view needed to be broader than what one camera could provide. We initially placed one camera above the screen and one lower on the base to serve the two different fields of view. However, looking at peoples hands from below was awkward for analysis and even more awkward for people interacting with the robot (Figure 11, first image). We explored placing two cameras above the screen, but that did not give good coverage of the space where people would be touching the robot, both near their hands and head (Figure 11, second image). During a demonstration, our clinical partners suggested that we needed a larger field of view than we might expect to accommodate the mobility of children, even when seated. They recommended we move at least one camera higher up or behind the robot. To accommodate this, we designed an additional mast for the base which screws on. The mast adds 29 cm of camera height from the top of the screen. At the same time, testing with the RealSense cameras had shown them to be good for interactions with a single person, but did not provide sufficient situational awareness for driving or when additional people to the side of the robot tried to interact. So we added a fisheye webcam quality camera from ELP. We examined placing all cameras at the top of the mast, which allows for easy image stitching, but does not provide an expanding field of view at distance, nor does a good job of covering the head area for tall subjects close to the robot (Figure 11, third image). We settled on placing the fisheye camera at the top of the mast with one RealSense camera and the other RealSense above the screen (Figure 11, fourth image). This configuration gives good coverage of the arms and hands for interactions close to the robot from the RealSense on the mast, good coverage of subjects at distance for full arm range of motion activities, with the ability to back up to scale the field of view to the subject, and good general visibility through the fisheye camera. The downside of this configuration is that there is a blind spot between interactions which are near and far away and the images would be challenging to stitch together, so must either be treated separately or brought together as point clouds. Because of this, we chose not to synchronize the cameras, which is a feature available in the D415. Synchronizing the cameras can place more load on the processor and USB bus, but is desirable if images are being stitched together.

Software

Figure 12 describes the software stack. The entire software stack is built to interface with the Robot

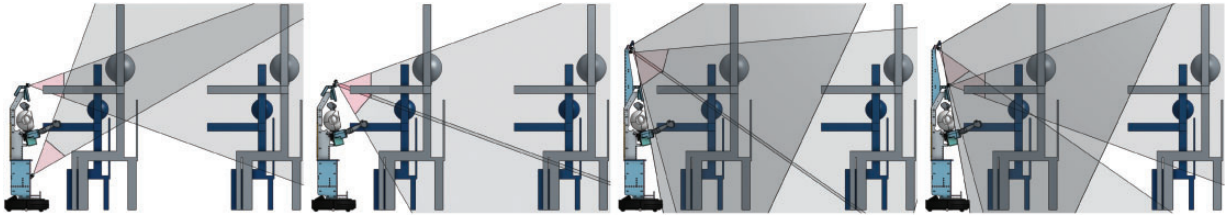


Figure 11. Four different configurations for the cameras. The gray areas show the field of view. In all configurations, two Intel RealSense D415 RGB+D cameras are included. In the last two configurations, an ELP 180 degree fisheye camera is also shown. The depth cameras have a red area shown, which shows an area where depth cannot be recovered. There are two sizes of example subjects shown, an approximate small six year old and an approximate large eleven year old.⁶⁰

Operating System (ROS).⁶¹ ROS was selected because it has become the de facto standard for robot integration. By using ROS, the development of the system can take advantage of a significant amount of work which has been done by others and help contribute components to the community. ROS also provides an opinionated and easy to use system for integrating complex systems. The software stack which we have developed is separated into a series of different components. This infrastructure also makes it relatively easy to operate in a simulation mode for development and testing. The robot currently runs on Ubuntu 16 with ROS Kinetic and the development machines on Ubuntu 18 with ROS Melodic.

Our system primarily uses a plays and scripts style of operating under traded control,⁶² allowing the operator to easily manage the complexity of the system. This comes with somewhat limited control from the operator, which is mitigated through options to control the system more directly as explained below. Although we have seen a lot of progress towards it, autonomy remains out of reach for reliable human-robot interaction, so using human-in-the-loop systems can allow complex applications like ours to be feasible.⁶² The clinician can provide perception and reasoning that a robot cannot yet provide.

The face is controlled by a pair of ROS nodes, one for handling the serial interface with the Teensy microcontroller mounted in the head and one for handling the face state and commands. The control node exposes a series of services which allow the face, brightness, and eye direction to be set, as well the available faces to be requested. When there is a change made to the active face, the control node publishes the new state. The available faces and eyes are stored in a JSON file. When the system is in a simulation mode, the control node runs alone.

There is a single node to interface with the humanoid. It communicates over serial, with the microcontroller located within the humanoid, based on the XYZ robotics software. Individual poses are loaded from the computer onto the microcontroller to build

a dictionary of available poses, and then sequences are loaded as matched pairs of pose IDs and times. The ROS node exposes an action server which receives a series of joint targets which define either a single pose to move to or a series of poses to move through. Sequences of poses are interpreted to create linear motion in joint space between successive targets on each joint. So if the left arm is commanded to a position at one second and five seconds and the right arm is commanded to a position at two, three, and five seconds, the left arm will interpolate from its position at start to the first position at one second and from there to the second position at five seconds, ignoring the joints on the right arm, which will be interpolating independently. By using repeated targets, this gives complete flexibility of motion. For example, if the left arm and right arm should move together to one second, the right arm should move to another pose by three seconds while the left arm stays still, and then both arms should move together to poses at five seconds, then right arm would be given targets for its first pose at time one second, second pose at three seconds, and third pose at five seconds while the left arm would be told to go to its first pose at one second, its first pose again at three seconds, and its second pose at five seconds.

There are a series of software checks which ensure that the commands are delivered successfully to the microcontroller. Once the sequence is successfully sent, the action server provides feedback to the calling system on motion progress. In addition, the robot provides frequent status updates to the controller on the current position of the arms, which the control node publishes. These messages are used to maintain the state of the robot and to allow the poses to be saved.

There is a core package which handles launching the entire system, maintaining a database for the robot, and running games on the robot. The database stores poses, sequences of poses, utterances (things the robot can say), and game buckets in a SQLite architecture. JSON is used to expand the standard SQLite data types to handle arrays of various types. The poses are stored

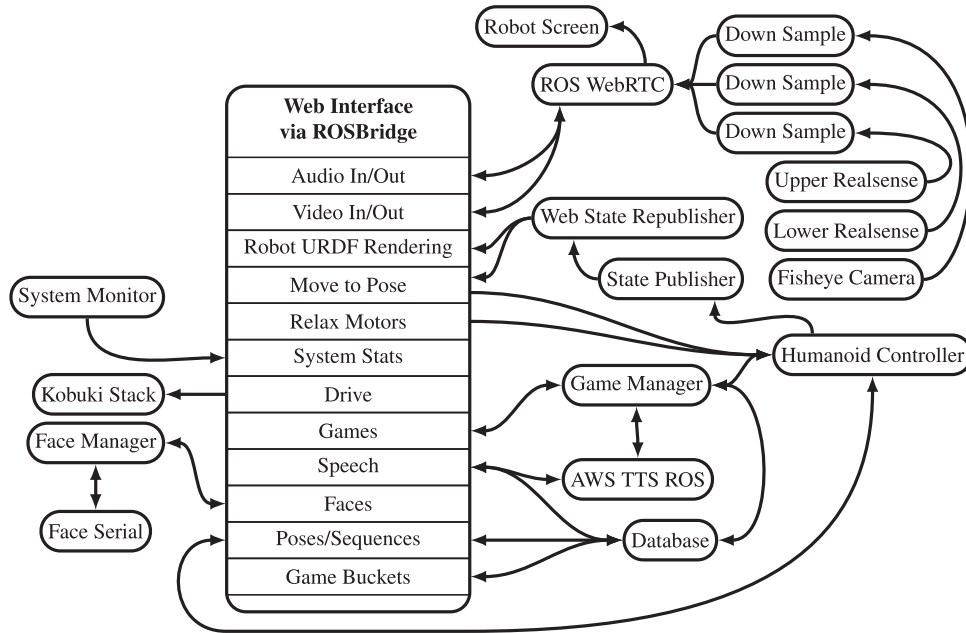


Figure 12. The nodes and their connectivity running on the robot. The entire stack is extremely complex, with many messages passing between nodes, so here we are showing a simplified “logical connection map”, omitting the connections to the recording system, which records from nearly every node, and hiding the intermediate servers along connections between the robot and operator’s interface.

with an id, description, and joint name/angle pairs. The poses are side agnostic and can be selected for either side by consuming applications. The pose sequences are stored with an id, description, sequence of times, targeted arms, and pose ids (references to the poses table). Access to the database is exposed through a series of services which perform error checking on row creations and modifications. The database can be searched or directly indexed and modifications can be made through direct index.

Image and audio capture. Data is read from the RealSense cameras using the ROS RealSense package, which generates output seen in Figure 13. We use the D415’s onboard vision processor to generate depth data and only publish the depth and color feeds. This saves bandwidth and processing power. The depth feeds run at 1280×720 pixels, which is the value recommended for pose extraction by Intel. The color feeds also run at 1280×720 pixels, which is less than their maximum 1920×1080 , a compromise to save processor overhead and storage space. All feeds run at 30 frames per second. We do not align the depth to the color or generate point clouds at the time of capture, to minimize processing. These steps are completed as part of a post processing pipeline.

The fisheye camera video is captured using the `videostreamcv` package, which was the only available package for webcam/usbcam capture that did not

require significant modification to other software running on the computer and which ran at the full 30 frames per second. The fisheye camera runs at 640×480 pixels. It is capable of running at up to 1280×1024 , but the quality of the imaging optics are not sufficient to justify the compute and storage costs of running at that higher image size.

There is an image processing node for each color video feed which downsamples and republishes the feed at a lower resolution, 300×168 pixels for the D415s and 320×240 pixels for the fisheye camera. All cameras are also throttled to 15fps in the down sampling pipeline. This allows the feeds to be sent to the web interface at a lower resolution, saving encoding processing and bandwidth. The full resolution and frame rate feeds are saved in bag files for future processing.

There is a node that runs to capture audio. This node simply captures audio from the microphones and publishes it within the system.

Speech synthesis. There is also a speech synthesizing node. The robot’s voice is synthesized using Amazon Polly through the `AmazonWebService(AWS) RoboticsTTS-ROSPackage`. Other methods of synthesizing voice were considered as well, including solutions from Microsoft Azure, Google Cloud Platform, Nuance, Voicery, and Acapela. We chose the AWS solution because it is low cost, with zero startup

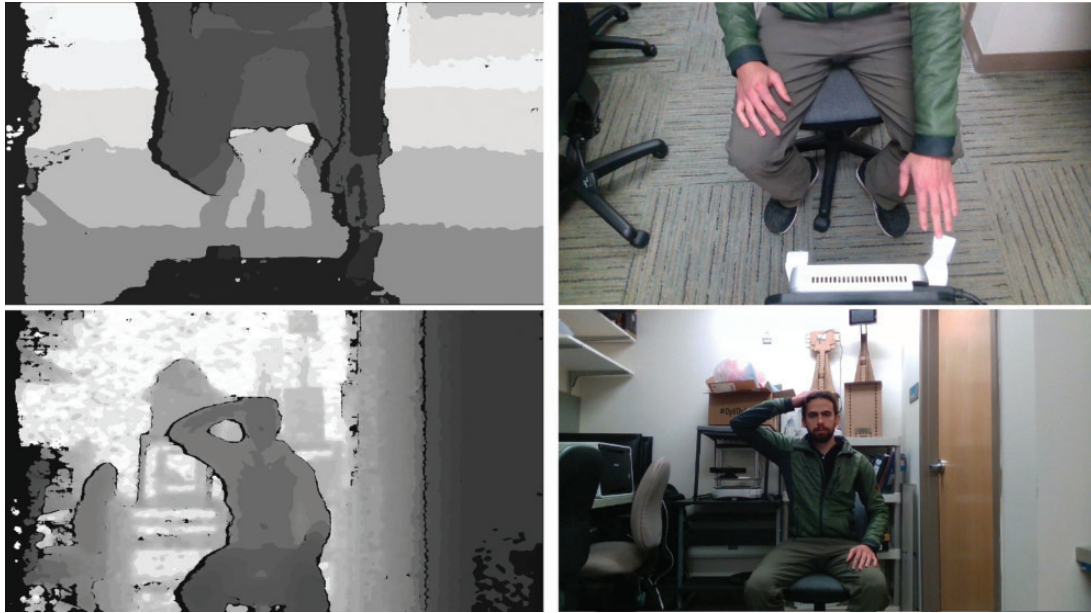


Figure 13. Example output from the RealSense D415. On the first row, the upper camera looking down on a hand touching the robots hand. On the bottom a person touching their head, captured by the lower RealSense. On the left are visualizations of the depth data and on the right the color images. The demonstration subject is a member of the research team.

costs, flexible, produces high quality utterances, has a rich API, and had the extra ease of a preexisting ROS integration.

Settling on using the AWS Polly system, we chose the Salli voice. Initially we used the Ivy voice, but with time, found it to be overly childlike, somewhat annoying, and difficult to understand for some team members. The Salli voice is warm and soothing, easy to understand, and weakly aged. In order to allow the system to perform in challenging network situations, we modified the AWS ROS TTS node to incorporate caching, caching audio files to the disk with a separate SQLite database to maintain references.

Robot exercise games. There are currently two exercise game types implemented, a Simon Says game type and a target touch game. In Simon Says, the robot first reads out a set of instructions, which inform subjects that they should do what the robot says and does if it says Simon Says. It then demonstrates steps for the subject to perform, giving them instructions on what to do. For example, “Simon Says touch the top of your head with your right hand”. Thirty percent of the actions are repeated without a “Simon Says” command. The order of the steps is randomized. In the target touch game, the robot moves its hands to different poses asking subjects to touch the colored dot on the robot’s hand with a specific subject hand ten times.

The custom game runner manages the operation of games. Games exist in two parts: a game type and a

game bucket of steps. Game steps can be either a pose to move the left arm to, a pose to move the right arm to, a pose to move both arms to, or a sequence of poses. In all cases, the target is a reference to a database object, either a pose or a sequence. Each step also has text associated with the step and an optional time parameter which can be used to set the length of time the robot should use to reach a pose or modify all the times in a sequence to complete the sequence in the specified time. The games are abstractions in code which are given a bucket of step definitions and return a sequence of actions with the fully expanded pose or sequence as a series of joint targets with times and a speech component. This allows additional games to be added as simple middleware with a clear API and no need to think of the underlying robot mechanics.

To begin a game, any other node can send the game runner a message to load a game of either Simon Says or target touch, with a bucket of steps. The game runner processes the game and waits for a command to begin. The game runner publishes two topics, one for feedback and one with command options and listens for commands. The commands can include: start to begin the game, next to go to the next step in the game, repeat to repeat the prior step, congratulate which will make the robot say a congratulatory sentence randomly selected from a bucket of options, try again which will make the robot say an utterance from a bucket of options that communicate that the subject should try again and then run the prior step, quit game

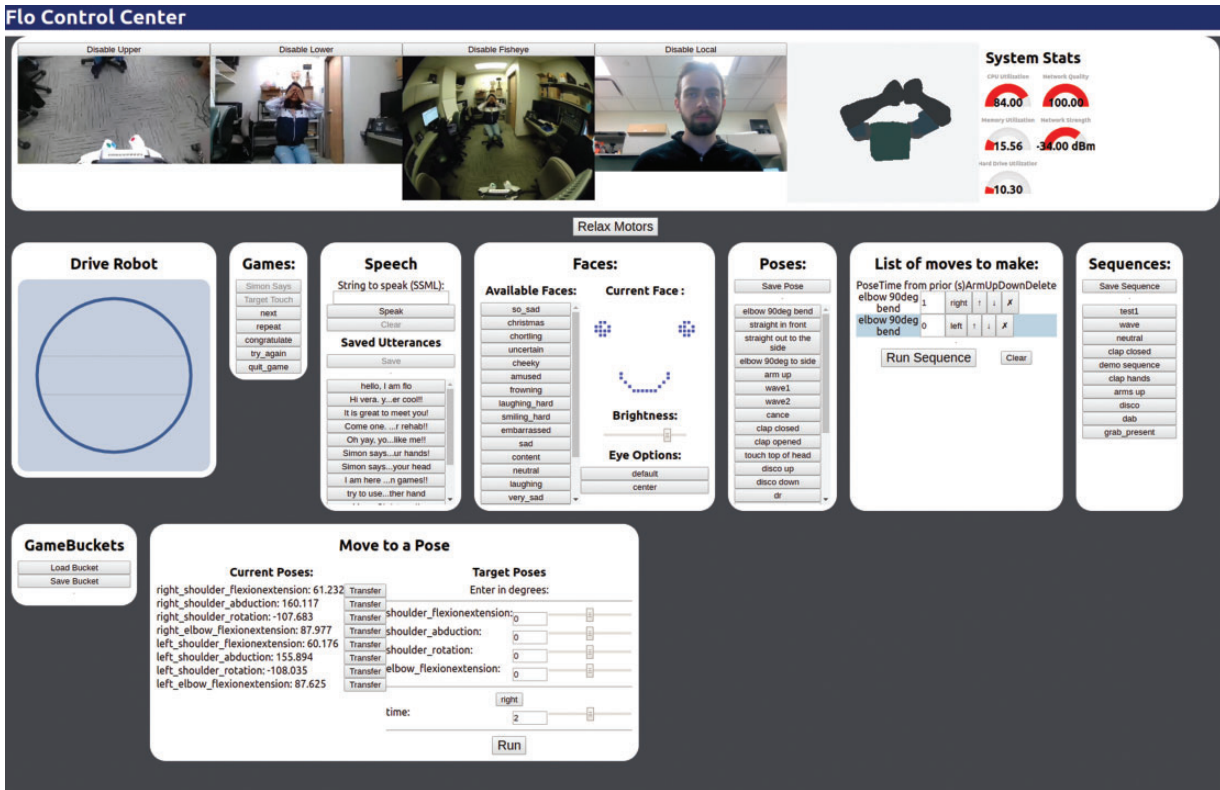


Figure 14. The web interface for controlling the robot. On top are the two video feeds from the RealSense cameras, the feed from the fisheye camera, and the feed from the local camera. There is a rendering of the robot in its current state and system stats for the robot computer showing the CPU utilization, memory utilization, hard drive utilization, network quality, and network signal strength. Below are modules for driving the robot, running games, making the robot speak, controlling the robot's face, saving and running poses and sequences, creating game buckets, and manually moving the robot. Not shown: a number of the buttons for saving and loading provide pop-ups with further options to achieve functionality. The operator and demonstration subject are both members of the research team.

which will exit the currently running game, and finish game which exits the game upon completion.

Remote operation. The entire system is designed to be controlled remotely. This is done through a custom web interface (Figure 14) written in typescript, using React. The web interface uses the rosbridge suite on the robot to gain access to actions, services, and topics provided by the rest of the stack, the tf2 web republisher to get access to the transformations which define the robot, ros3djs to show a rendering of the robot, and WebRTC ROS to send video from the robot to the web and back, all from the Robot Web Tools project.⁶³ Sitting between the robot and the front end is a custom server stack running on NodeJS with PostgreSQL, Redis, and NGINX, all within Docker containers. The server manages logins and access to the robot and passes data between the web interface and robot.

At the top of the interface is an area which shows outputs to the operator: the three video feeds from the

robot, the video feed of the operator which is being sent to the robot, the rendering of the robot's pose, and system stats for the robot's computer. This allows the operator to see where the robot is. The downward facing camera provides a good view of ground obstacles and the fisheye camera allows a wide field for situational awareness. The rendering of the robot state is done from sensors in the motors, not from the planned state, so it provides quality feedback to the operator. The system stats are important, especially the network stats, for working over telepresence, where driving out of network range can leave the robot stranded. All the video is transferred using WebRTC and the video of the operator is captured using the browser based getUserMedia stack.

There is then an area of controls for the robot. There is a button to relax the motors, which is useful for programming poses by moving the robot in person. The drive console presents a circle for the different directions the robot can drive, the user clicks with their mouse to drive the robot. The games module

connects to the game manager to select games and run them by dynamically displaying the available game commands and sending them back to the server. A speech module allows the user to arbitrarily type speech for the robot to say, save speech which the robot has said, and select prewritten text. The faces module allows selection of the stored faces, shows the currently displayed face, allows the brightness of the LEDs to be set, and allows the direction of the eyes to be set. The poses, list of moves to make, and sequences modules all work together to define new poses based on the current robot's pose and define, edit, and run sequences as a list of poses with targeted completion times and arms. The game buckets module allows the user to load game buckets for editing or create new game buckets. Finally, the move to a pose module allows the remote user to manually move the robot through software, this is useful for training new poses in simulation.

To display the video from the operator on the robot's screen, a small custom program was built which displays the remote video topic using OpenCV. The display runs full screen and on top of all other windows, ensuring that the experience for the patient is not disturbed by pop ups or notifications.

ROS is used to launch the system through a cascade of launch files. There are two entry points to the launch chain, one to run in simulation and one to run on the robot. The simulator launch file simply launches the standard stack with some added parameters. In practice, the system is started by running a script on the robot (which runs on system start) that starts tmux, starts a roscore, launches the system, starts the connections to the web server, and starts convenience panes to allow setting the robot's volume, monitoring detailed performance, and setting the default audio device. At that point, a user need only login to the website hosted by the web server and select the robot to control (from the ones which they have been granted access to), and they are ready to operate the robot. For lower level administrative access, system managers with appropriate access can ssh into the robot and view the status of the internal system directly.

Communication security. Although this is still a research platform, it is important to ensure that data which is being transmitted is protected, both for the protection of study participants and to demonstrate feasibility. The system uses industry standard security practices to ensure the protection of remote connections. The entire remote interface is served over an HTTPS connection. The servers hosting the system are protected with passwords and firewalls. Users login using passwords which are stored in a database as salted hashes using bcrypt. All video communication is done using

WebRTC which is encrypted end to end and cannot be read, even if intercepted by a third party. The messaging needed to initiate the WebRTC connections is done using encrypted web sockets. These layers of security ensure that video and audio which are transmitted between the remote operator and the robot are visible only at the robot and by the operator.

Data collection

All relevant data is recorded in rosbags, this includes all the video feeds and their respective information, all the commands to run the robot which operate via ROS topics, the mobile base information, the robot pose information, and logging output. The rosbags are split every one minute to keep them small and prevent loss of data from corruption at system startup and shutdown. Recording data remains a challenge. Recording the image streams in their raw format leads to large bag files, approximately 18 GB per minute. If the image streams are recorded in their compressed format, the bag files are around 3.1 GB per minute. If further lz4 compression is used during recording, the bag file size is about 2 GB, however compressing the depth images costs significant processing power, leading to dropped frames. If the video files are recorded compressed, but the depth is left uncompressed, the size is about 4.3 GB, even with lz4 compression. Our current system has approximately 2 TB that can be devoted to bag storage, meaning that we can store approximately 110min of data using raw video streams and raw depth streams. For research purposes, this is sufficient. Although using compression allows storing more data, it has proven to lead to unreliable recording.

Interaction

Using all the tools outlined above, we can construct how an envisioned interaction proceeds. In order to use the robot, the robot must be turned on, which launches the software stack. A remote operator can then login to the website, select the robot, and connect through their web browser. The operator can then navigate the robot to the patient and introduce themselves and Lil'Flo. The operator can ask the patient if they want to play a game with Lil'Flo, select a game type, and select a game bucket relevant to the patient. As the patient plays the game with Lil'Flo, the operator has direct control of the progress of the game, it is a single button click to go to the next step, repeat the prior step, and/or give feedback. The operator retains the flexibility to type in custom text for the robot to speak and, at any time, send the robot to a different pose to probe the patient further. As the patient struggles or does

well, the operator can change the face on the robot to express empathy for their struggle and joy for their success.

As interactions progress over time, the operator may choose to modify the game buckets available for different patients. All the while the system will be collecting and logging data on the interactions and how the patients are moving.

Design evaluation

As we have progressed through the design process, we have evaluated the success of our designs against our design requirements.

Upper arm workspace evaluation

To understand the viability of the new arm design, the range of motion of the robot's joints were measured using the encoder built into the motors. The arm motions were compared to standard human joint range of motions for the shoulder and elbow. Measurements from the constructed robot of range of motion show shoulder flexion: 3.6 radians, shoulder extension: 3.9 radians, shoulder abduction: 3.0 radians, shoulder adduction: 0.1 radians, shoulder internal rotation: 2.4 radians, shoulder external rotation: 3.2 radians, elbow flexion: 1.5 radians, elbow extension: 0.68 radians. The limits are imposed by both physical contact between components and length limits on the wires connecting the motors. The range of motion of the arms provide coverage of human range of motion except for shoulder adduction and elbow flexion. Elbow supination and pronation are not present in the design, nor are any wrist motions. The shoulder internal/external rotation is done near the elbow joint, instead of near the shoulder. We believe that this is sufficient to meet the mechanical requirements for creating interactions to support motor function.

Face expression evaluation

To test the head alone, we isolated it from the body and presented it to 10 healthy subjects. The results are reported fully in our prior publication.⁶⁴ The robot acted in a static mode, dynamic mode, and an iterating mode. Subjects first experienced the head/face in both the static and dynamic modes, while being asked questions about it, both open-ended and on a Likert scale. The subjects were then shown all the faces on the robot and gave the first thoughts that they had. The questions comparing the various face designs and some of the open-ended statements were particularly interesting and helped guide the final design of the robot.

Based on the results we culled the set of faces which we had developed down to a final set, shown in

Figure 15 which clearly showed an emotion. The data suggested that a limited face, like the ones shown, can convey gross emotions clearly.

From the open-ended questions during the interview, a number of important points were made. The electronics on our system suffered from interference, which caused some flickering in the eyes and face. The subtle flickering was enough that a number of subjects mentioned it as being detrimental to their interaction with the robot, leading to a redesign of the power system to use higher quality wires in twisted data and power pairs with quality connectors going to the microcontroller and matrices.

Demonstrations of robot. We have used the robot for demonstrations to multiple hundreds of children, typically functioning, as part of outreach efforts, and they often respond to the face. When the face changes from being happy to sad, there is often an audible "awww" from the audience, indicating that there is some empathetic connection with the robot or at the least an understanding of the emotion the robot is conveying.

Overall hardware and software

We achieved a low cost system. The total cost of the system is much lower than our previous iteration. The motors cost about \$400 USD, the computer cost a total of \$600 USD, the battery cost \$130 USD, the Kobuki base cost \$500 USD, the pair of RealSense cameras cost \$300 USD, the fisheye camera cost \$45 USD, and the screen and controller cost \$78 USD. There is also a cost for the various cables, primer, raw materials, and paint used. Many of the parts were either laser cut or 3D printed, which is not cheap. If these designs were to be fully translated, these components would be cheaper and higher volume production techniques would be used.

The current hardware design and software integration effort is a significant improvement over our prior design in terms of robustness and modularity. The humanoid, which is the central component of the system is much more robust and more modular than our prior system and easier to maintain. In over a hundred hours of testing, its motors have never failed from over heating. However, the motors used suffer from communication challenges and in a future iteration should be upgraded. Future directions would improve the motors and allow for more robust, faster, and multitasking communication over the USB. Another achievement is a long run time for experiments. The system easily runs for one and a half hours when all systems in the stack are running and an operator is connected. Internal storage limits how long the robot can capture data.

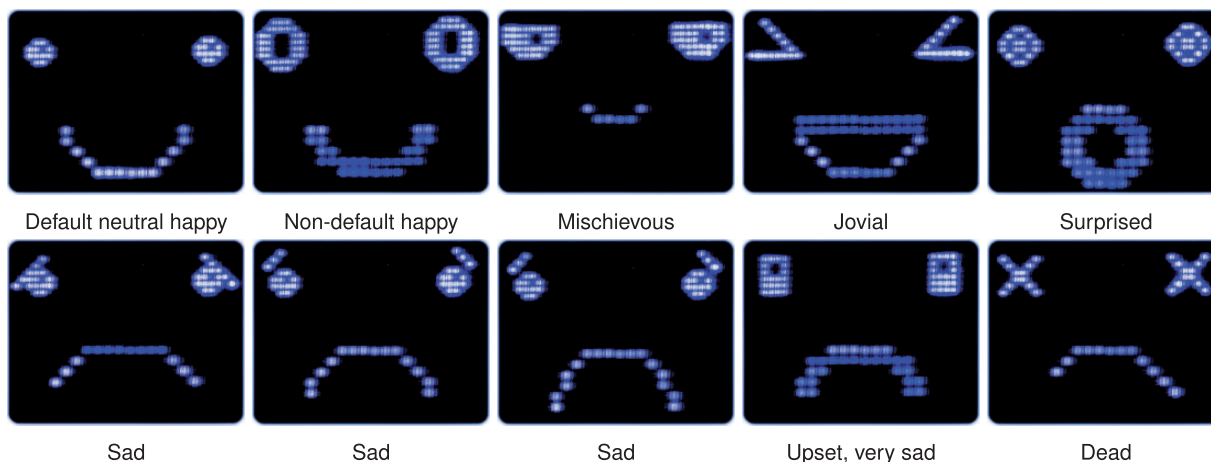


Figure 15. Faces which we found had clear expressions of emotion to subjects in our pilot study. (a) Default neutral happy, (b) non-default happy, (c) mischievous, (d) jovial, (e) surprised, (f) sad, (g) sad, (h) sad, (i) upset, very sad and (j) dead. Source: reproduced with permission from Sobrepera et al, 2019.⁶⁴

We have achieved a lightweight and portable system. Given the building materials, the system is light weight at only 9.8 kg and portable. It can be easily carried by one person, although its height can make it awkward. One future desire is to investigate the impact of the height on patient's trust: would a shorter system may make the patient feel more confident or would being taller may make the patient more likely to listen to instructions.⁶⁵

The software stack we have developed enables telepresence and remote communication. The web interface does a good job of controlling the robot and demonstrating how easily such a system can be controlled. Although useful for rapid development, using ROSBridge to communicate over a network outside the robot is non-ideal, preventing scaling and redundancy. Although the use of ROS presents some challenges for security and efficiency, it allows the system to be built upon and iterated. By using an on robot firewall and secure server for communications, security challenges have been addressed.

Driving can be challenging for a first time user, but the final set of camera positions does not make it overly onerous. For long term deployment, autonomous navigation would be preferred, the combination of ROS and the RGB+D cameras should make the addition of that functionality possible.

Virtual usability testing

To begin to understand how therapists might receive Lil'Flo and how they would expect it to impact the care of their patients we surveyed 13 clinicians who work in rehabilitation and have seen Lil'Flo in person in an urban setting in Pennsylvania. The survey was

conducted as part of a study approved by the University of Pennsylvania Institutional Review Board. The survey asked a series of demographic questions, questions on COVID-19 and rehab, and experiences with telehealth. The survey then presented a video overview of Lil'Flo (<https://youtu.be/OHybatsjzog>) to remind respondents of the system's features. Respondents then answered a series of questions on the utility of Lil'Flo.

We received responses from eight physicians in physical medicine and rehabilitation, one physician who is a consult from medicine, one physical therapist, two occupational therapists, and one clinical neuropsychologist. Respondents have a minimum age of 26 years and maximum of 68 years with average 35.8 years. They range from 1 to 17 years of experience, averaging 5.9 years of experience. Eight respondents are female and 6 male; 12 have worked in a rehab center, 4 in an outpatient facility, 6 in an inpatient facility, 3 in a hospital for children, and 5 in a general hospital.

The clinicians were asked to respond to three questions about the robot: (1) "How interested would you be in using the Lil'Flo system?" on a scale from "Not at all interested" (0) to "Very Interested" (100); (2) "What locations do you think Lil'Flo could be deployed in?" as a multiple choice question with the 10 possibilities: rural outpatient clinics, rural inpatient clinics, elder care facilities, schools, patient homes, community centers, urban inpatient clinics, urban outpatient clinics, none, or other; and (3) "Which types of activities do you believe you could do with each type of tool?" with three tools: non-video remote interaction, video-call remote interaction, and Lil'Flo remote interaction, across 11 activity types: motor assessments, stretching,

strength building, ADL practice, cognitive assessments, cognitive exercises, environmental adaptation, orthotics assessment/prescription, discussions about surgery, discussions about radiology results, and medical prescriptions. For question 3, clinicians could elect to say that the activity was not possible remotely.

When asked “How interested would you be in using the Lil’Flo system?” on a scale from “Not at all interested” (0) to “Very Interested” (100), using a continuous slider, six responded with a value greater than 50 (Figure 16). For a novel technology, we find these results encouraging.

When asked “What locations do you think Lil’Flo could be deployed in?” as a multiple choice question with the 10 possibilities shown in Table 1, the clinicians selected rural inpatient clinics at the highest frequency with 10 responses followed by elder care facilities with 9 responses. It appears the clinicians see the utility of Lil’Flo for serving rural communities. Interestingly, although Lil’Flo was designed for pediatric patients, 69% of the clinicians viewed Elder care facilities as a location where Lil’Flo could be deployed. When Lil’Flo was initially conceived, inpatient urban clinics, where 62% of the clinicians believed Lil’Flo could be deployed, were not considered a target. However, as COVID-19 has changed the nature of medicine, there may now be a stronger argument for deployment in areas where travel is not a concern.

Finally, when asked “Which types of activities do you believe you could do with each type of tool?”, Lil’Flo was selected more often than traditional video based telepresence for motor assessments, stretching, and strength building (Figure 17). These are the three activities presented which most closely align with the rehab goals Lil’Flo is designed for. For each of these three types of activities 3–4 of the clinicians indicated that they could not be done remotely. This indicates that these tasks are challenging remotely, and perhaps serves as validation towards working on these issues. Orthotics Assessments/Prescription was rated by seven of the respondents as not being possible remotely, which makes sense given the hands on nature of that task set. For prescribing medicines, discussing surgery, and discussing radiology results, one more clinician felt that traditional video would be better than Lil’Flo. It isn’t clear for these cases how the Lil’Flo system would be different than traditional video, although the humanoid is not designed to add value to these types of interactions. Traditional video was also rated as being usable for environmental adaptation by two more clinicians than the Lil’Flo system. This makes sense, given that carrying around a video interface for environmental adaptation is likely easier than a driving robot. For ADL practice an equal number of clinicians felt that Lil’Flo and traditional telepresence would be

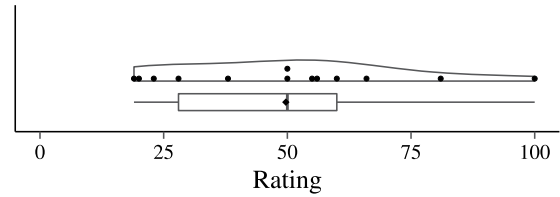


Figure 16. Ratings by clinicians on how interested they would be in using Lil’Flo, from not at all (0) to very (100). Shown as a violin plot on top with actual responses as dots and a boxplot on the bottom with mean shown as a diamond.

Table 1. Responses from 13 rehab clinicians to the question “What locations do you think Lil’Flo could be deployed in?” when presented with the shown 10 options.

Clinical locations	Count of responses
Rural inpatient clinics	10
Elder care facilities	9
Rural outpatient clinics	8
Urban inpatient clinics	8
Community centers	6
Urban outpatient clinics	5
Schools	4
Patient homes	3
None	0
Other	0

usable. We had expected that Lil’Flo would receive more positive responses than traditional telepresence for this category. And finally, for cognitive exercises, one more clinician felt that Lil’Flo would be usable compared to traditional telepresence, perhaps suggesting a future direction of investigation.

Comparison to existing systems

Compared to existing systems, Lil’Flo offers a number of unique attractive characteristics. Compared to the Nao based systems, the humanoid component of Lil’Flo has an expressive face, it is light weight, easy to maintain and modify with few parts outside of the motors, and easy to integrate into a larger ROS based system. More generally, the entire platform presents a novel concept, using social robots to augment telehealth. This stands in contrast to existing systems which are meant to either interact with the patient and therapist together or interact with the patient alone. The platform is also differentiated in that it carries its cameras on board. Many other systems require an external camera.

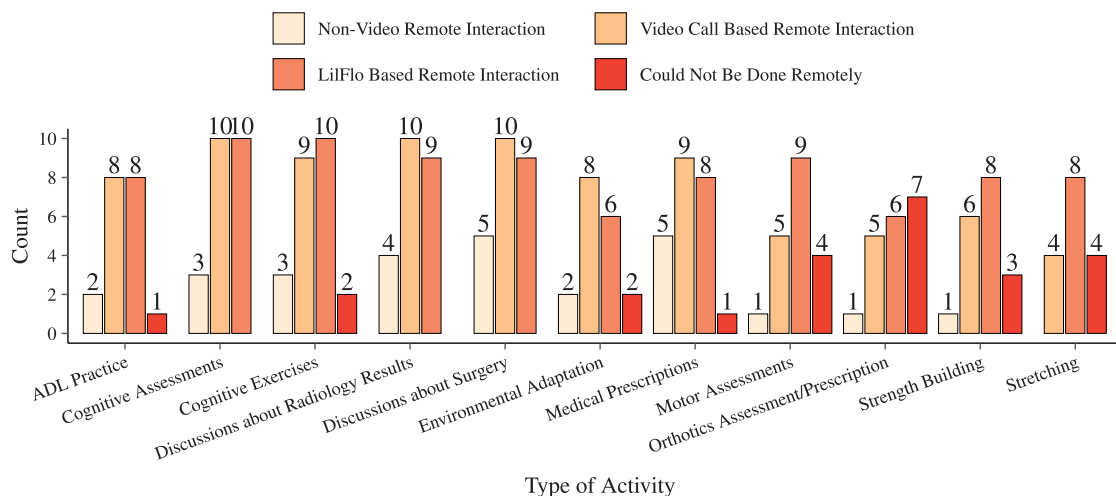


Figure 17. Clinical activities which 13 clinicians selected when prompted with: “Which types of activities do you believe you could do with each type of tool? Non-video remote communication includes: phone calls, text messages, email, instant messages, and other types of communication which allow you to interact with patients from afar, without using video”.

Future work

We are beginning to test the system with CP, BP, and typical subjects. Subjects play a Simon Says and target touch game in three conditions: the social robot with telepresence, the telepresence system without the social robot, and in person face to face with the operator. Subjects are assigned an order of conditions in a block random fashion. In all conditions the same games and script are used. Subjects are measured using a box and block test, grip strength test, and color trails test. Surveys and video grading, both automatic and manual, are being used to measure subject engagement, enjoyment, understanding, and compliance. The data which is collected will be used to help develop automated assessment tools.

In addition we are working on a pilot deployment study with the system where clinicians will operate the robot to interact with their patients. They will use the robot to play one of the existing games and then do activities which they would do during a normal interaction, using Lil’Flo. Both the clinicians and the patients will be surveyed to understand the quality of interactions from both sides.

There are a number of changes which we have planned for the system. We want to improve the mechanical design inside the head of the robot to make it more stable, improve the tint of the screen, and improve the finish of the head’s printed parts. We would like to eventually switch to using dynamixel motors within the arms to improve their controllability. Doing so would also make it easier to add lighting in the body and arms, which could run off the dynamixel bus. We would like to use a larger

screen, which would also require a more capable base, and would likely come with an increase in system size and cost.

As the system continues to mature, we expect to automate more of its functionality. The plays and scripts model used currently is sufficient for research. However, we expect to find clinicians wanting the system to handle more of the control load. The pilot deployment study we are working on will provide clarity on this. In addition, further automation will allow the system to be more robust to network disturbances by allowing interactions to continue more smoothly during communication dropouts. Automation could potentially further improve access by allowing one clinician to simultaneously treat multiple patients with the robot running autonomously while the clinician moves from patient to patient virtually.

Ethical considerations

Increasing autonomy of robots in healthcare settings often raises ethical issues. In general there are questions about the biases implicit in data based algorithms that will eventually drive most robots.⁶⁶ And some are concerned that robots will displace healthcare workers or simply be a distraction to care.⁶⁷ However, failing to use robotic technologies to improve patient care could be considered irresponsible and in itself presents ethical concerns. The Lil’Flo robot described here melds telepresence, robotics, and computer vision to allow more patients to access healthcare while promoting patient–clinician interactions. In doing so, we hope to address some of the shortage of clinicians and caretakers in the

rehab and care spaces. In view of the growing shortage of rehabilitation healthcare workers, we believe these technologies offer a solution to clinicians to help clinicians avoid compromising care as populations in need of care grow.

Beyond the ethics of care, in the current climate of growing concern over personal security and privacy there is concern that collecting, aggregating, and learning from large amounts of subject data could make some subjects uncomfortable and their data vulnerable to hackers. It is therefore imperative to both communicate clearly to subjects/patients what is being collected and how it is being used as well as taking every precaution to safeguard any data which we have. It is not however clear how to safely share data within the research community to accelerate development. In general, Human Subject Ethic Committees do not allow the publication of identifying data. To address this challenge, we ask all of our subjects for media and information releases above and beyond what is included in the standard consent. Subjects are welcome to participate in the study without agreeing to such releases. If they choose to opt-in, then their data will be compiled into a data set which can be released to other researchers to help accelerate development of novel algorithms. Balancing the competing needs of privacy and compliance against research progress can be difficult, but is important.

Discussions around the ethics in these spaces are growing as media coverage grows. It is important that roboticists and clinicians think about the ethical issues implied in the technologies which they are developing and using, try to understand the feelings of their subjects and patients, and be present in the public discussions on these issues.

Conclusion

We have created a telepresence robot which can interact with patients in the communities where they live, learn, work, and play. To help bridge the gap between telepresence interactions and in person ones, we have attached a humanoid robot to our system. The humanoid can play games with patients and demonstrate movements to them, as a peer. This is designed to engage subjects and help them better understand the activities which they need to do, while keeping the clinician closely involved.

Compared to other systems which are available and described in the literature, we believe that our system is unique. It addresses the challenges which we set out to tackle and will enable us to examine how telepresence and social robotics can be used together for telerehabilitation. In a small survey of a cohort of clinicians who have seen Lil'Flo in person, clinicians expressed their

belief that Lil'Flo could be deployed in rural inpatient clinics, elder care facilities, rural outpatient clinics, and urban inpatient clinics. The clinicians were on whole non-negative in their interest in using Lil'Flo. And finally, the clinicians believed more frequently that Lil'Flo could be used for motor assessments, strength building, and stretching than traditional video based telepresence. Testing for more thorough validation of the system is underway.

Interested researchers can view the code to run the entire system at <https://www.med.upenn.edu/rehabilitation-robotics-lab/lilflo-code.html> and the mechanical design at <https://www.med.upenn.edu/rehabilitation-robotics-lab/lilflo-cad.html>.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Funding for this project has been provided by the Department of Physical Medicine and Rehabilitation in the Perelman School of Medicine at the University of Pennsylvania, by the School of Engineering and Applied Science at the University of Pennsylvania and by the Eunice Kennedy Shriver National Institute Of Child Health & Human Development of the National Institutes of Health under Award Number F31HD102165. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Guarantor

MJJ

Contributorship

MJJ conceived the original idea for a telepresence-based social robot. MJS led the design, fabrication, and software integration of the system. VGL aided with the design and fabrication of the base. MJS led the composition of this manuscript with support from VGL and MJJ.


Acknowledgements

Many people have contributed to this project. We want to especially thank Enri Kina and Carla Diana for their work on the design of the robot. We also want to thank the many clinicians who have contributed their opinions to the system, especially Dr. Laura Prosser and Dr. John Chuo.

ORCID iDs

Michael J Sobrepera  <https://orcid.org/0000-0003-2609-0430>

Vera G Lee  <https://orcid.org/0000-0002-4665-6270>

Michelle J Johnson  <https://orcid.org/0000-0002-7964-0304>

References

1. Lin V, Zhang X and Dixon P. Occupational therapy workforce in the United States: forecasting nationwide shortages. *Pm R* 2015; 7: 946–954.
2. Zimbelman JL, Juraschek SP, Zhang X, et al. Physical therapy workforce in the United States: forecasting nationwide shortages. *Pm R* 2010; 2: 1021–1029.
3. Surveillance of Cerebral Palsy in Europe. Surveillance of cerebral palsy in Europe: a collaboration of cerebral palsy surveys and registers. *Dev Med Child Neurol* 2000; 42: 816–824.
4. Rosenbaum P, Paneth N, Leviton A, et al. A report: the definition and classification of cerebral palsy April 2006. *Dev Med Child Neurol* 2007; 49: 8–14.
5. Graham HK, Rosenbaum P, Paneth N, et al. Cerebral palsy. *Nat Rev Dis Primers* 2016; 2: 15082.
6. Abzug JM, Mehlman CT and Ying J. Assessment of current epidemiology and risk factors surrounding brachial plexus birth palsy. *J Hand Surg Am* 2019; 44: 515.e1–515.e10.
7. Abzug JM and Kozin SH. Evaluation and management of brachial plexus birth palsy. *Orthop Clin North Am* 2014; 45: 225–232.
8. Steiner W, Ryser L, Huber E, et al. Use of the ICF model as a clinical problem-solving tool in physical therapy and rehabilitation medicine. *Phys Ther* 2002; 82: 1098–1107.
9. Jiandani MP and Mhatre BS. Physical therapy diagnosis: how is it different? *J Postgrad Med* 2018; 64: 69–72.
10. Howard IM and Kaufman MS. Telehealth applications for outpatients with neuromuscular or musculoskeletal disorders: telehealth applications for outpatients. *Muscle Nerve* 2018; 58: 475–485.
11. Abel KC, Baldwin K, Chuo J, et al. Can telemedicine be used for adolescent postoperative knee arthroscopy follow-up? *JOPA JBJS* 2017; 5: e26–e24.
12. Johnson S, Rae I, Mutlu B, et al. Can you see me now? How field of view affects collaboration in robotic telepresence. In: *Proceedings of the 33rd annual ACM conference on human factors in computing systems. CHI'15*. Seoul, Republic of Korea: ACM, pp.2397–2406.
13. Sirkin D, Venolia G, Tang J, et al. Motion and attention in a kinetic videoconferencing proxy. In: Hutchison D, Kanade T, Kittler J, et al. (eds) *Human–computer interaction – INTERACT 2011*. Berlin, Germany: Springer Berlin Heidelberg, pp.162–180.
14. Feil-Seifer D and Matarić MJ. Defining socially assistive robotics. In: *9th international conference on rehabilitation robotics, 2005. ICORR 2005*. Chicago, IL: IEEE, pp.465–468.
15. Johnson MJ, Sobrepera MJ, Kina E, et al. Design of an affordable socially assistive robot for remote health and function monitoring and prognostication. *Int J Progn Health Manag*. Special issue PHM for human health and performance, www.phmsociety.org/node/2548 (2019, accessed 3 March 2021).
16. Wilk R and Johnson MJ. Usability feedback of patients and therapists on a conceptual mobile service robot for inpatient and home-based stroke rehabilitation. In: *5th IEEE RAS/EMBS international conference on biomedical robotics and biomechanics*. New York, NY: IEEE, pp.438–443.
17. Alemi M, Meghdari A, Ghanbarzadeh A, et al. Effect of utilizing a humanoid robot as a therapy-assistant in reducing anger, anxiety, and depression. In: *2014 second RSI/ISM international conference on robotics and mechatronics (ICRoM)*. Tehran, Iran: IEEE, pp.748–753.
18. Fridin M and Belokopytov M. Robotics agent coacher for CP motor function (RAC CP fun). *Robotica* 2014; 32: 1265–1279.
19. Westlund JK, Lee JJ, Plummer L, et al. Tega: a social robot. In: *2016 11th ACM/IEEE international conference on human–robot interaction (HRI)*. Christchurch, New Zealand: IEEE, pp.561.
20. Pulido JC, González JC, Suárez-Mejías C, et al. Evaluating the child–robot interaction of the NAOTherapist platform in pediatric rehabilitation. *Int J Soc Robotics* 2017; 9: 343–358.
21. Fasola J and Matarić MJ. A socially assistive robot exercise coach for the elderly. *J Hum Robot Interact* 2013; 2: 3–32.
22. Robinson H, MacDonald B and Broadbent E. The role of healthcare robots for older people at home: a review. *Int J Soc Robotics* 2014; 6: 575–591.
23. Torta E, Werner F, Johnson DO, et al. Evaluation of a small socially-assistive humanoid robot in intelligent homes for the care of the elderly. *J Intell Robot Syst* 2014; 76: 57–71.
24. Calderita LV, Manso LJ, Bustos P, et al. THERAPIST: towards an autonomous socially interactive robot for motor and neurorehabilitation therapies for children. *JMIR Rehabil Assist Technol* 2014; 1: e1.
25. Suárez Mejías C, Echevarría C, Nuñez P, et al. Ursus: a robotic assistant for training of children with motor impairments. In: Pons JL, Torricelli D and Pajaro M (eds) *Converging clinical and engineering research on neurorehabilitation*. volume 1. Heidelberg, Germany: Springer Berlin Heidelberg, pp.249–253.
26. Calderita LV, Bustos P, Suárez-Mejías C, et al. THERAPIST: towards an autonomous socially interactive robot for motor and neurorehabilitation therapies for children. In: *2013 7th international conference on pervasive computing technologies for healthcare and workshops*. Venice, Italy: IEEE, pp.374–377.
27. González JC, Pulido JC, Fernández F, et al. Planning. Execution and monitoring of physical rehabilitation therapies with a robotic architecture. In: Cornet R, Stoicu-Tivadar L, Hörbst A, et al. (eds) *Digital healthcare empowering Europeans. Proceedings of MIE2015, 26th medical informatics in Europe conference, MIE 2015*. volume 210. Madrid, Spain: IOS Press, pp.339–343.
28. Pulido JC, Suarez-Mejias C, Gonzalez JC, et al. A socially assistive robotic platform for upper-limb rehabilitation: a longitudinal study with pediatric patients. *IEEE Robot Automat Mag* 2019; 26: 24–39.
29. Fridin M. Kindergarten social assistive robot: first meeting and ethical issues. *Comput Hum Behav* 2014; 30: 262–272.

30. van den Heuvel RJF, Lexis MAS and de Witte LP. ZORA robot based interventions to achieve therapeutic and educational goals in children with severe physical disabilities. *Int J Soc Robotics* 2020; 12: 493–504.
31. RoboKind LLC. Meet Milo!, www.robokind.com/robot-s4autism/meet-milo (accessed 3 March 2021).
32. Huijnen CAGJ, Lexis MAS and de Witte LP. Matching robot KASPAR to autism spectrum disorder (ASD) therapy and educational goals. *Int J of Soc Robotics* 2016; 8: 445–455.
33. Wood LJ, Robins B, Lakatos G, et al. Piloting scenarios for children with autism to learn about visual perspective taking. In: Giuliani M, Assaf T and Giannaccini ME (eds) *Towards autonomous robotic systems*. Cham, Switzerland: Springer International Publishing, pp.260–270.
34. Brisben AJ, Safos CS, Lockerd AD, et al. The CosmoBot™ system: evaluating its usability in therapy sessions with children diagnosed with cerebral palsy, 2005; 12: 14–16. Available at: <https://doi-org.proxy.library.upenn.edu/10.1145/1052438.1052453>.
35. Saposnik G, Outcome Research Canada (SORCan) Working Group and Levin M. Virtual reality in stroke rehabilitation: a meta-analysis and implications for clinicians. *Stroke* 2011; 42: 1380–1386.
36. Triandafilou KM, Tsoupikova D, Barry AJ, et al. Development of a 3D, networked multi-user virtual reality environment for home therapy after stroke. *J Neuroengineering Rehabil* 2018; 15: 1–13.
37. Thielbar KO, Triandafilou KM, Barry AJ, et al. Home-based upper extremity stroke therapy using a multiuser virtual reality environment: a randomized trial. *Arch Phys Med Rehabil* 2020; 101: 196–203.
38. Elor A, Teodorescu M and Kurniawan S. Project star catcher: a novel immersive virtual reality experience for upper limb rehabilitation. *ACM Trans Access Comput* 2018; 11: 1–25.
39. Liu LY, Sangani S, Patterson KK, et al. Real-time avatar-based feedback to enhance the symmetry of spatiotemporal parameters after stroke: instantaneous effects of different avatar views. *IEEE Trans Neural Syst Rehabil Eng* 2020; 28: 878–887.
40. Fridin M and Belokopytov M. Embodied robot versus virtual agent: involvement of preschool children in motor task performance. *Int J Hum Comput Interact* 2014; 30: 459–469.
41. Bainbridge WA, Hart JW, Kim ES, et al. The benefits of interactions with physically present robots over video-displayed agents. *Int J Soc Robotics* 2011; 3: 41–52.
42. Kiesler S, Powers A, Fussell S, et al. Anthropomorphic interactions with a robot and robot-like agent. *Social Cognition* 2008; 26: 169–181.
43. Vasco V, Willemsse C, Chevalier P, et al. Train with me: a study comparing a socially assistive robot and a virtual agent for a rehabilitation task. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 2019; 11876: 453–463.
44. Mann JA, Macdonald BA, Kuo IH, et al. People respond better to robots than computer tablets delivering health-care instructions. *Comput Human Behav* 2015; 43: 112–117.
45. Céspedes N, Múnera M, Gómez C, et al. Social human-robot interaction for gait rehabilitation. *IEEE Trans Neural Syst Rehabil Eng* 2020; 28: 1299–1307.
46. Adalgeirsson SO and Breazeal C. MeBot. A robotic platform for socially embodied presence. In: *2010. 5th ACM/IEEE international conference on human-robot interaction (HRI)*, HRI'10. Osaka, Japan: IEEE, pp.15–22.
47. Dodakian L, McKenzie AL, Le V, et al. A home-based telerehabilitation program for patients with stroke. *Neurorehabil Neural Repair* 2017; 31: 923–933.
48. Davids J, Peace L, Wagner L, et al. Validation of the shriners hospital for children upper extremity evaluation (SHUEE) for children with hemiplegic cerebral palsy. *J Bone Joint Surg* 2006; 88: 326–333.
49. Krumlinde-Sundholm L and Eliasson AC. Development of the assisting hand assessment: a rasch-built measure intended for children with unilateral upper limb impairments. *Scand J Occup Ther* 2003; 10: 16–26.
50. Tran V, Dario P and Mazzoleni S. Kinematic measures for upper limb robot-assisted therapy following stroke and correlations with clinical outcome measures: a review. *Med Eng Phys* 2018; 53: 13–31.
51. Tsui KM, Norton A, Brooks DJ, et al. Iterative design of a semi-autonomous social telepresence robot research platform: a chronology. *Intel Serv Robotics* 2014; 7: 103–119.
52. Scassellati B, Admoni H and Matarić M. Robots for use in autism research. *Annu Rev Biomed Eng* 2012; 14: 275–294.
53. Malik NA, Hanapiah FA, Rahman RAA, et al. Emergence of socially assistive robotics in rehabilitation for children with cerebral palsy: a review. *Int J Adv Robot Syst* 2016; 13: 135.
54. Westlund JMK, Park HW, Williams R, et al. Measuring children's long-term relationships with social robots. In: *Proceedings of the 17th ACM conference on interaction design and children. IDC'18*. Trondheim, Norway: ACM, pp.207–218.
55. Kalegina A, Schroeder G, Allchin A, et al. Characterizing the design space of rendered robot faces. In: *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction – HRI'18*. Chicago, IL: ACM, pp.96–104.
56. Rae I, Mutlu B and Takayama L. Bodies in motion: mobility, presence, and task awareness in telepresence. In: *Proceedings of the SIGCHI conference on human factors in computing systems, CHI'14*. Ontario, Canada: ACM, pp.2153–2162.
57. Gilmore R, Sakzewski L and Boyd R. Upper limb activity measures for 5- to 16-year-old children with congenital hemiplegia: a systematic review. *Dev Med Child Neurol* 2010; 52: 14–21.
58. Wagner LV and Davids JR. Assessment tools and classification systems used for the upper extremity in children

- with cerebral palsy. *Clin Orthop Relat Res* 2012; 470: 1257–1271.
59. Weng Y, Specian A and Yim M. Low cost optical mechanical system for human robot interaction. In: *Proceedings of the ASME 2018. International mechanical engineering congress and exposition*. Volume 4A: dynamics, vibration, and control. Pittsburgh, PA: ASME, pp.V04AT06A029–1–V04AT06A029–9.
 60. Fredriks AM, van Buuren S, van Heel WJM, et al. Nationwide age references for sitting height, leg length, and sitting height/height ratio, and their diagnostic value for disproportionate growth disorders. *Arch Dis Child* 2005; 90: 807–812.
 61. Quigley M, Gerkey B, Conley K, Faust J, Foote T, Leibs J, Berger E, Wheeler R and Ng AY. “ROS: an open-source Robot Operating System,” in Proc. Open-Source Software workshop of the International Conference on Robotics and Automation (ICRA), 2009.
 62. Goodrich MA, Crandall JW and Barakova E. Teleoperation and beyond for assistive humanoid robots. *Rev Hum Factors Ergon* 2013; 9: 175–226.
 63. Toris R, Kammerl J, Lu DV, et al. Robot web tools: Efficient messaging for cloud robotics. In: *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. Hamburg, Germany: IEEE, pp.4530–4537.
 64. Sobrepera MJ, Kina E and Johnson MJ. Designing and evaluating the face of lil’Flo: an affordable social rehabilitation robot. In: *2019 IEEE 16th international conference on rehabilitation robotics (ICORR)*, 2019. Toronto, Canada: IEEE, pp.748–753.
 65. Rae I, Takayama L and Mutlu B. The influence of height in robot-mediated communication. In: *2013 8th ACM/IEEE international conference on human–robot interaction (HRI)*. Tokyo, Japan: IEEE, pp.1–8.
 66. Howard A and Borenstein J. The ugly truth about ourselves and our robot creations: the problem of bias and social inequity. *Sci Eng Ethics* 2018; 24: 1521–1536.
 67. Tuisku O, Pekkarinen S, Hennala L, et al. “Robots do not replace a nurse with a beating heart”: the publicity around a robotic innovation in elderly care. *ITP* 2019; 32: 47–67.