

The Design of Lil’Flo, an Affordable Socially Assistive Robot for Telepresence Rehabilitation

Michael J. Sobrepera^{1,2}, Michelle J. Johnson^{1,3}

¹University of Pennsylvania, ²Department of Mechanical Engineering, ³Department of Rehabilitation Medicine

INTRODUCTION

Cerebral Palsy (CP) occurs in 2 to 3 per every 1000 live births, making it the most common motor disorder in young children [1]. There is no cure for CP. However, disciplined therapy can improve outcomes. There is currently a shortage of therapists available to treat CP, a lack of objective measures for the upper extremities (UEx) to evaluate progress of patients, a mismatch of location between therapists and their patients, and high costs associated with therapy. In many ways CP is a good representation of many classes of disorders which impair the UEx. To begin to address the need for an affordable quantitative diagnostic tool in pediatric rehabilitation which can be used both in-person and remotely, the UPenn Rehab Robotics lab is developing a low cost socially assistive robot (SAR) [2] (Lil’Flo) to aid in remote, semi-autonomous, and fully autonomous assessment and treatment of upper extremity impaired pediatric patients. The robot will include a humanoid form seated on a mobile base (fig. 1) which can engage with children in low force contact, acting as a companion to the patient. It will be able to interact with patients and demonstrate activities to facilitate assessment via telepresence based remote control. For example, patients could be asked to play a game of “Simon Says” with the robot or touch various targets on the robot. The system will analyze patient behavior to determine the level of function of the patient and appropriate treatment paths forward. As a final gold standard, we envision a robot which could naturally “play” with patients as an avenue for diagnostics and therapy. There has been previous work in this space, most notably by the NAOtherapist (formerly Therapist, Ursus) project [3-5] and the RAC CP Fun project [6-7]. We have also explored the use of SARs in the elder care space using our Flo robot (fig. 1) [8].

A critical component of this project will be a perception system which can evaluate patient UEx function. The perception system will deliver information to the clinician which they cannot receive via their senses due to the barrier of telepresence. For example the range of motion of the patients joints, the maximum velocities of the joints, the emotional state of the patient, and a set of overarching diagnostic scores. It will also be useful in both non-remote and remote interactions, to deliver objective measures which can be difficult to obtain via traditional means, making it useful beyond strictly robotic applications, as a purely observational system.

The robot and perception system together will hopefully help to improve information transfer in telepresence interactions, which can break down due to the narrow amount of communication which is possible via video and audio alone. This idea is summarized in figure 2.

HARDWARE DESIGN

Although robotic systems exist which could be used to achieve some of our goals, they fail to meet all our design requirements: 1) low cost for maximum impact 2) expressive face for social connectivity 3) easily modifiable hardware 4) mobile for remote deployments. So, we are designing our own hardware to meet those requirements. We are also interested in showing, especially for telepresence encounters, the benefits of using a humanoid robot form over using a robot with only a screen. This will justify translational work in the telepresence space allowing the project to have the broadest impact in the shortest possible time. It will be easier to accomplish this with a custom system.

We currently have a mobile robot from VGo with a removable Aldebaran Nao Robot torso mounted on the robot, which has been used for initial validation work [8]. This platform has been useful for the speed with which we can test with it. However, it fails to demonstrate the true potential of an SAR in the space, due to its high cost. When



Figure 1: On the left, the system which we developed to explore the use of socially assistive robots in elder care facilities, it is a Nao robot on a VGo telepresence base. On the right, our work in progress for the lower cost Lil’Flo system.

our low cost robotic platform is complete, it will replace the Nao/VGo system at an expected cost ratio of 1:5. Both systems can be seen in figure 1.

To help accelerate work on the design of our system we began with off the shelf robotic components. An XYZ Bolide robot as the humanoid and an iRobot Create as the base. The Bolide was chosen for its independently controllable motors with digital feedback. We have modified it to have proportions of a young child, extending its limbs significantly. We have also designed and are iterating on a custom head which can display emotions via LED screens. Work is still underway to develop a skin to cover the robot's internals.

We decided to make Lil'Flo approximately half of the height of the original Flo robot. This allowed us to use a low cost mobile robot without having to worry about the system tipping over. This also makes sense for the target population of children, who will always be seated during robot interactions for safety.

Although the iRobot Create was a useful first mobile base, we are currently transition to an iCleo Kobuki, which is a more robust more easily controlled differential drive mobile base. In the long term, we are interested in exploring a holonomic base and whether the additional costs inherent in one are justified by improved patient interactions.

PERCEPTION SYSTEM DESIGN

Current techniques for evaluating UEx function are inadequate. Measures which are judged by clinicians are subjective, leading to inter and intra clinician variability. Further, the cost and availability of high skilled clinicians prevents all patients from being evaluated sufficiently frequently. Measures based on motion capture (MoCap) systems are high cost and markers on the skin can lead to non-representative behavior by patients. To enable remote telepresence and autonomous robot interactions with patients, as shown in figure 2, and to improve the objectivity of the diagnostic space in general, it is necessary to develop computational techniques for analyzing patient function with high repeatability on low cost hardware. It is known that neurological damage alters UEx motion, making it less smooth, with lower maximum velocities, smaller ranges of motion, etc [9-10]. We believe that these changes in motion could be detected by a computer vision based pipeline.

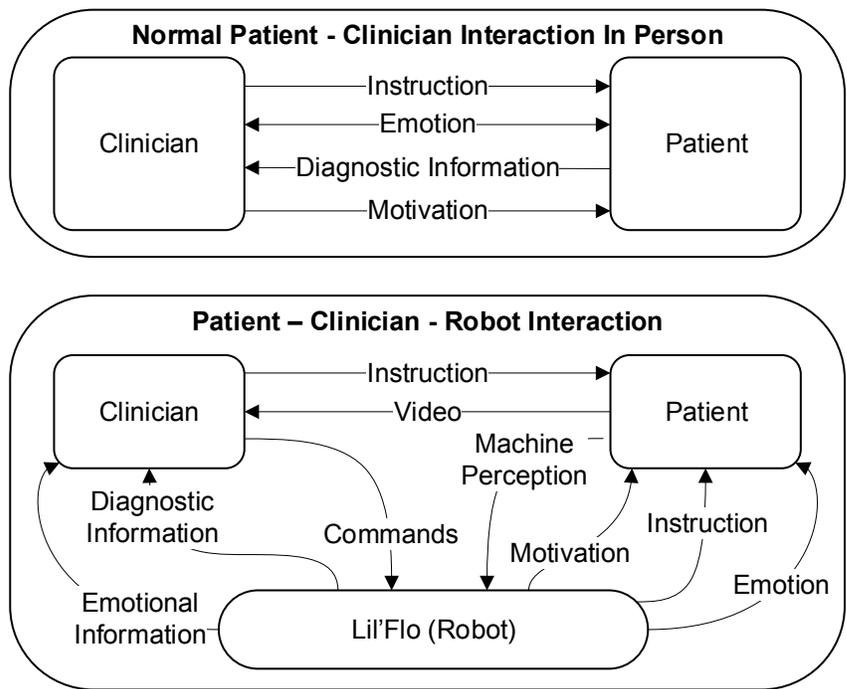


Figure 2: A comparison of interactions between patients and clinicians in person and patients, clinicians, and the robot in telepresence interactions.

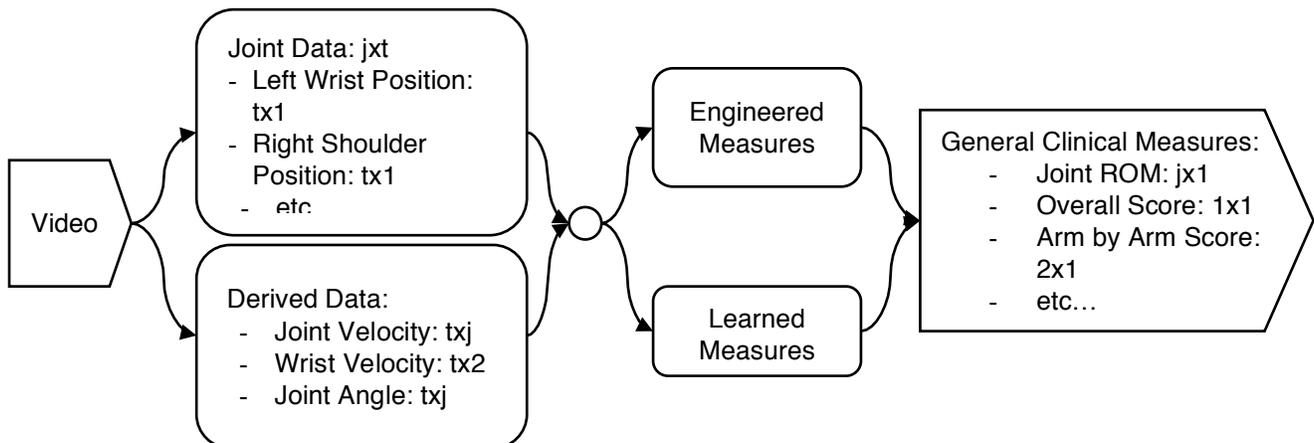


Figure 3: Block diagram of the perception system showing the progression from video input to general clinical measures being output to the clinician.

The advent of the Microsoft Kinect and related low cost RGBD sensors led to technology demonstrations capturing subject motion to determine kinematic measures such as velocity and range of motion [11] and even as grading tools for existing tests such as the SHUEE [12]. Our lab has performed work looking at precision contact measurement of wrist kinematics for diagnostics in adult stroke and CP patients [13].

At the same time, the computer vision community has progressed greatly in recent years. Advents in machine learning, especially deep convolutional neural networks, have enabled recognition of objects to become a low speed solved problem. More recent novel network architectures such as stacked hourglass networks [14], Faster R-CNN [15], and part affinity fields [16] have made pose estimation from video realizable. Building on this work, progress has been made towards reconstruction 3D joint positions from 2D joint positions [17-18]. These methods rely on models of the human form which are learned to represent the general shape of a person. Widespread adoption of UAVs has also led to further developments in the RGB+D sensor space with much of the low-cost side of the market moving to stereo based designs.

We are currently working to develop the pipeline for performing objective evaluations of upper extremity impaired pediatric patients as overviewed in figure 4. The first step of the pipeline will be to extract 3D time series joint positions (TSJP) from RGB video. In initial testing, we have used part affinity fields [16] and stacked hourglass networks [14] to extract the 2D joint locations of subjects doing a series of range of motion activities. We have then leveraged [17] to estimate the full 3D pose of the subject. The results can be seen in Figure 4. Given that our target population’s physiology is atypical, we will have to adjust these methods to develop on the fly bespoke models of each patient, as opposed to relying on the statistical models of an average person which are currently used. As the project progresses, we will also explore opportunities and challenges for augmenting the RGB pipeline with low cost depth data, for example from the Intel RealSense line of cameras. In addition to using known kinematic measures of function which can be extracted from TSJP, as we gather more data, we will develop a learned model using machine learning techniques to attempt to classify patients, with the option to use known medical information and the instructed patient activity/“optimal” motion as priors.

To perform the necessary computations for the perception system a small computer (Intel NUC 7 i5 BNK) has been placed on the robot. It will do the low cost initial computations to decrease the video size to be processed. For example, segmenting out the subject and performing facial recognition. The selected and compressed video will then be streamed to a server for complete processing in a high-powered compute environment with distributed GPUs. The onboard computer will also be used for eventual navigation tasks as well as voice synthesis and robot control.

CONCLUSION

Our hope in working on this project is to address gaps in care by extending the geographic reach of clinicians by using telepresence for rehab. By using a humanoid robot as a social agent for communication to and motivation of patients we expect to show that care is improved when compared to telepresence on a screen alone. To further augment telepresence interactions, we are developing a perception system which can deliver back to the clinician metrics which they can likely not garner via a video feed alone. We expect that this system will have implications beyond telepresence interactions by allowing for low cost objective testing of upper extremity function in pediatric patients. In the long term, we hope that these two technologies can be the basis for the development of more autonomous systems which will allow clinicians to focus on the most demanding components of therapy, leaving the more monotonous tasks to robots, enabling lower costs and increased patient interaction time leading to overall improved outcomes.

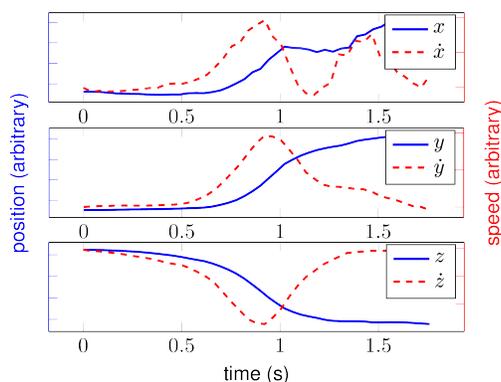


Figure 4: Data collected during a pilot test with a healthy subject, with the subject touching the back of their head. Data was collected using a standard uncalibrated cell phone camera and processed using a stacked hourglass network [14] followed by Zhou’s method for 3D reconstruction [17] and cleaned with dropout rejection. On top, is the x-direction, in the middle, is the y-direction, and on the bottom, is the z-direction. Solid blue lines are positions in arbitrary coordinates, dashed red lines are speeds in arbitrary coordinates. As can be seen, the primary axes of motion, y (1.5 unit change) and z (.6 unit change), show smooth movements with velocities approximating the bell curve predicted by Hogan [9]. The x-axis which underwent much

REFERENCES

- [1] Cans, C. (2000). Surveillance of cerebral palsy in Europe: a collaboration of cerebral palsy surveys and registers. *Developmental Medicine & Child Neurology*, 42(12), 816-824.
- [2] Feil-Seifer, D., & Mataric, M. J. (2005, June). Defining socially assistive robotics. In *Rehabilitation Robotics, 2005. ICORR 2005. 9th International Conference on* (pp. 465-468). IEEE.
- [3] Calderita, L. V., Manso, L. J., Bustos, P., Suárez-Mejías, C., Fernández, F., & Bandera, A. (2014). THERAPIST: Towards an autonomous socially interactive robot for motor and neurorehabilitation therapies for children. *Journal of Medical Internet Research*, 16(10), e1. <https://doi.org/10.2196/rehab.3151>
- [4] Calderita, L. V., Bustos, P., Suárez-Mejías, C., Fernández, F., & Bandera, A. (2013). THERAPIST: Towards an autonomous socially interactive robot for motor and neurorehabilitation therapies for children. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th International Conference on* (pp. 374-377). IEEE.
- [5] Mejías, C. S., Echevarría, C., Nuñez, P., Manso, L., Bustos, P., Leal, S., & Parra, C. (2013). Ursus: A robotic assistant for training of children with motor impairments. *Converging Clinical and Engineering Research on Neurorehabilitation*, 1, 249-253.
- [6] Fridin, M., & Belokopytov, M. (2014). Robotics agent coacher for CP motor function (RAC CP Fun). *Robotica*, 32(8), 1265-1279.
- [7] M. Fridin, S. Bar-Haim, M. B. (2011). Robotics Agent Coacher for CP motor Function (RAC CP Fun). *Workshop on Robotics for Neurology and Rehabilitation*.
- [8] Wilk, R., & Johnson, M. J. (2014, August). Usability feedback of patients and therapists on a conceptual mobile service robot for inpatient and home-based stroke rehabilitation. In *Biomedical Robotics and Biomechanics (2014 5th IEEE RAS & EMBS International Conference on* (pp. 438-443). IEEE.
- [9] Hogan, N. (1984). An organizing principle for a class of voluntary movements. *Journal of Neuroscience*, 4(11), 2745-2754.
- [10] Nelson, W. L. (1983). Physical principles for economies of skilled movements. *Biological cybernetics*, 46(2), 135-147.
- [11] Matthew, R. P., Kurillo, G., Han, J. J., & Bajcsy, R. (2014, September). Calculating Reachable Workspace Volume for Use in Quantitative Medicine. In *ECCV Workshops (3)* (pp. 570-583).
- [12] Rammer, J. R., Krzak, J. J., Riedel, S. A., & Harris, G. F. (2014, August). Evaluation of upper extremity movement characteristics during standardized pediatric functional assessment with a Kinect®-based markerless motion analysis system. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE* (pp. 2525-2528). IEEE.
- [13] Lott, C., & Johnson, M. J. (2016, August). Upper limb kinematics of adults with cerebral palsy on bilateral functional tasks. In *2016 IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)* (pp. 5676-5679). IEEE.
- [14] Newell, A., Yang, K., & Deng, J. (2016, October). Stacked hourglass networks for human pose estimation. In *European Conference on Computer Vision* (pp. 483-499). Springer International Publishing.
- [15] Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(6), 1137-1149.
- [16] Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2016). Realtime multi-person 2d pose estimation using part affinity fields. *arXiv preprint arXiv:1611.08050*.
- [17] Zhou, X., Zhu, M., Leonardos, S., Derpanis, K. G., & Daniilidis, K. (2016). Sparseness meets deepness: 3D human pose estimation from monocular video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4966-4975).

- [18] Bogo, F., Kanazawa, A., Lassner, C., Gehler, P., Romero, J., & Black, M. J. (2016, October). Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image. In *European Conference on Computer Vision* (pp. 561-578). Springer International Publishing.