Human-Robot Collaboration: Designing a Framework for Intent

Yvonne Coady
University of Victoria
Victoria, BC, Canada
ycoady@uvic.ca

Derek Jacoby
QVirt.com
Victoria, BC, Canada
derekja@qvirt.com

Proceedings of the CHI 2017 Workshop on Amplification and Augmentation of Human Perception, May 07, 2017, Denver, CO, USA.

Abstract
Robotic systems can amplify our ways and means of interacting with our environment. In manufacturing settings, human capabilities of strength and touch in manipulating objects are often increased many orders of magnitude. Seamless collaboration, however, is still cumbersome due to the nature of the feedback loops involved. What if the means of collaboration could shift to intended outcomes rather than detailed processes to achieve them? The proposed framework combines patterns of brain activity, natural language, applied machine learning, and haptic feedback within an augmented reality environment to amplify feedback loops in human-robot communication.

Author Keywords
Development environments and frameworks; human-robot interaction; augmented reality.

Introduction
In robotic systems, *pose adjustment* requires a human operator to refine a robot's position using a *teach pendant* (Figure 1). These fine-grained adjustments are discrete directions, either in joint space or task space, that can be replayed upon command. The use of sensors for mimicking human movements can eliminate some of these painstaking adjustments, these approaches only work if there is a direct 1:1 mapping between human and robotic movements.

Figure 1: A commercially available teach pendant, with a large red stop button on the top right [3].

Cloud Services and Open Source Opportunities

Enhanced Input: open data repositories such as BrainBench [9] for pattern mining, and chatbots using customizable speech services [1].

Enhanced Processing: Open source applied machine learning repositories, such as Tensorflow [6].

Enhanced Feedback: 3D controllers equipped with feedback for touch [5], stimulus [7], and shared, collaborative augmented reality spaces [8].

What if, instead of detailing *how* to manipulate objects in a physical space, we could focus on *what* the intended outcome should be in an augmented reality space? Robots could then "decide" which movements achieve the outcome, keeping the human-in-loop for feedback.

We propose a framework designed to capture *intent* called VARIETI (Virtual and Augmented Reality Integrated Environment for Transmitting Intent)—introducing a layer of virtual or augmented relatiy for humans and robots to express intentions to each other. Specifically, by leveraging cloud services and open repositories, VARIETI supplies pluggins to the Unity Development Environment [4] for enhanced input, processing, and feedback during real-time human-robot collaboration (sidebar). The general architecture for the system is organized as shown in Figure 2.

VARIETI: A Framework for Intent

The community around development for virtual and augmented reality has already created a rich set of plugins for Unity. Our proposed framework adds new dimensions of interaction alternatives necessary to make human-robot feedback loops more seamless. Specifically by, (1) amplifying thoughts and speech into intent, and (2) expressing outcomes in virtual worlds that can be readily transposed into physical settings.

Here we briefly overview how we envision specific representation from commodity devices, cloud services and open source and data repositories for each module: brain activity patterns, speech recognition, applied machine learning, haptic feedback, and shared collaborative augmented reality spaces shown in Figure 3.

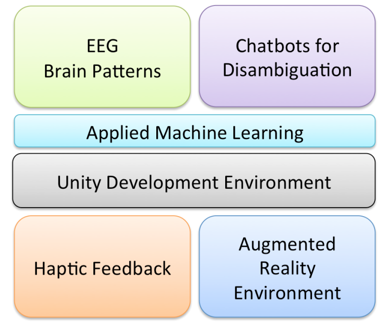

Figure 2: Proposed extensions to the Unity Development Environment include input modules for both brain patterns and speech, processing with a machine learning module, and output through both haptic feedback, and interaction within an augmented reality environment.

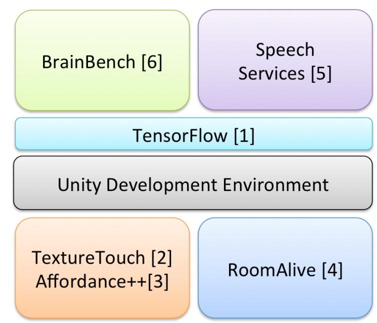

Figure 3: Specific influences for the first generation of VARIETI extensions.

Figure 4: A commercially available device for detecting brain activity used in [9].

Figure 5: A prototype by Microsoft for haptic feedback in a 3D controller [5].

Figure 6: Interaction in RoomAlive with virtual objects for shared augmented reality [8].

Brain Activity Patterns and Customized Speech Recognition
Patterns of brain activity can now be captured in commodity devices (Figure 4) and mined. Web services for collecting and sharing open data sets for pattern mining within natural language processing are becoming more mainstream. Specifically, BrainBench [9] has released several benchmarks as part of web service for a a new system that makes it easy to test semantic models using brain imaging data.

Discoverability challenges with brain pattern and speech input will need to establish what can be used, and when it can be used. We believe these challenges can be met with chatbot disambiguation dialogues, coupled with teach pendant uses. Natural dialogues can add confirmation steps when the system is not certain of the input.

Applied Machine Learning
Though web services contain their own machine learning techniques, open source projects capable of large scale execution across heterogeneous environments are increasingly easy to deploy. TensorFlow [6] uses dataflow graphs to abstract computation to can use cores across machines in a cluster, and within a machine across multiple computational devices, including multicore CPUs, general purpose GPUs.

Lowering the obstacles to adoption of machine learning techniques will allow operators to experiment with new processes that could allow robots to “decide” how to best move from one pose to another. Robots can learn about their environment without requiring programming to do so, and communicate their own intent in augmented reality. New movements can be accomplished by composing different learned movements. Metrics such as energy consumption, future requirements, and even collaboration with other robotic systems can be incorporated into decision making.

Haptic Feedback
Hand-held controllers that render the shape of virtual objects through physical shape displacement have shown promise in VR and AR environments (Figure 5). By moving controllers over virtual surfaces, operators amplify their sense of touch in both virtual and physical environments. These devices have shown increased accuracy in interaction with virtual objects [5].

We envision devices like these enabling operators to feel 3D surfaces the robotic system may be working on. This critical element of feedback is necessary to truly amplify touch in both virtual and physical settings.

Collaborative Augmented Reality Environments
Projection mapping uses video projectors to create a shared space for augmented reality (Figure 6). Users can collaboratively engage with projected content that seamlessly coexists with their shared existing physical environment. Recent work has specifically focused on interaction with virtual objects within human-to-human collaborative tasks [8].

In human-robot scenarios, this enables everything from manipulation of poses, to execution of movements and actions within the augmented portion of the physical space. Intent can be communicated virtually, physically, or some combination of the two.

The Ensemble
Taken together, VARIETI extends the state-of-the-art through plugins for enhanced input, processing, and feedback modules. The framework is designed to easily deploy a shared space for immersive, collaborative, augmented reality based communication through amplified feedback.

Conclusions and Next Steps

In 1955, Morton Heilig aimed to stimulate four of the five senses: sight, hearing, smell and touch with his patented Sensorama Simulator [2] shown in Figure 7. By 1962, without investors, the Sensorama stalled in the prototype stage.

Figure 7: The Sensorama machine with sights, sounds, wind, smells and a rumbling motorcycle seat and vibrating handlebars [2].

The design of the VARIETI framework aims to allow seamless collaboration with robotic systems by amplifying two senses: sight and touch. Our early work will explore a simple feedback loop in VR only, where a model of robotic system is manipulating objects in a virtual space based on expressed intent. The next phase will introduce AR, where physical space is integrated with the virtualized model. In the final phase, intent will be translated to the physical robot in a physical space, driving research questions that overlap the five components of feedback in the system:

- Pattern Mining Brain Activity
- Speech Recognition Services
- Applied Machine Learning
- Haptic Feedback and Stimulus
- Shared Collaborative Augmented Reality

REFERENCES

1. Microsoft Cognitive Services. `www.microsoft.com/cognitive-services/`
2. The Sensorama. `www.engadget.com/2014/02/16/morton-heiligs-sensorama-simulator/`
3. Teach Pendants. `www.pcmag.com/encyclopedia/term/61185/teach-pendant`
4. Unity. `unity3d.com/unity/multiplatform/vr-ar`
5. Hrvoje Benko, Christian Holz, Mike Sinclair, and Eyal Ofek. 2016. NormalTouch and TextureTouch: High-fidelity 3D Haptic Shape Rendering on Handheld Virtual Reality Controllers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 717–728. DOI: http://dx.doi.org/10.1145/2984511.2984526
6. Martín Abadi et al. 2016. TensorFlow: A System for Large-Scale Machine Learning. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*. USENIX Association, GA, 265–283. https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi
7. Pedro Lopes, Patrik Jonell, and Patrick Baudisch. 2015. Affordance++: Allowing Objects to Communicate Dynamic Use. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2515–2524. DOI: http://dx.doi.org/10.1145/2702123.2702128
8. Andrew D. Wilson and Hrvoje Benko. 2016. Projected Augmented Reality with the RoomAlive Toolkit. In *Proceedings of the 2016 ACM on Interactive Surfaces and Spaces (ISS '16)*. ACM, New York, NY, USA, 517–520. DOI: http://dx.doi.org/10.1145/2992154.2996362
9. Haoyan Xu, Brian Murphy, and Alona Fyshe. 2016. BrainBench: A Brain-Image Test Suite for Distributional Semantic Models. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*. 2017–2021. http://aclweb.org/anthology/D/D16/D16-1213.pdf