Natural Interaction Design of a Humanoid Robot

François Ferland,[†] Dominic Létourneau,[†] Arnaud Aumont,[†] Julien Frémy,[†] Marc-Antoine Legault,[†] Michel Lauria,[*] François Michaud[†]

[†]Université de Sherbrooke
[*]University of Applied Sciences Western Switzerland (HES-SO)

Designing robots that interact naturally with people requires the integration of technologies and algorithms for communication modalities such as gestures, movement, facial expressions and user interfaces. To understand interdependence among these modalities, evaluating the integrated design in feasibility studies provides insights about key considerations regarding the robot and potential interaction scenarios, allowing the design to be iteratively refined before larger-scale experiments are planned and conducted. This paper presents three feasibility studies with IRL-1, a new humanoid robot integrating compliant actuators for motion and manipulation along with artificial audition, vision, and facial expressions. These studies explore distinctive capabilities of IRL-1, including the ability to be physically guided by perceiving forces through elastic actuators used for active steering of the omnidirectional platform; the integration of vision, motion and audition for an augmented telepresence interface; and the influence of delays in responding to sounds. In addition to demonstrating how these capabilities can be exploited in human-robot interaction, this paper illustrates intrinsic interrelations between design and evaluation of IRL-1, such as the influence of the contact point in physically guiding the platform, the synchronization between sensory and robot representations in the graphical display, and facial gestures for responsiveness when computationally expensive processes are used. It also outlines ideas regarding more advanced experiments that could be conducted with the platform.

Keywords: Human-robot interaction, robot design, physical interaction, telepresence, computational resource management.

1. Introduction

Valli (2008) defines natural interaction "in terms of experience: people naturally communicate through gestures, expressions, movements, and discover the world by looking around and manipulating physical stuff; the key assumption here is that they should be allowed to interact with technology as they are used to interact with the real world in everyday life, as evolution and education taught them to do." *Natural* is thus synonymous with *normal interactions in everyday life*.

For the last 15 years, our research laboratory has been interested in the design of natural interactive mobile robots. These platforms evolved from small robots, such as an autonomous rolling robot used in child development studies (Michaud, Laplante, et al., 2005; Michaud, Salter, Duquette, & Laplante, 2007; Salter, Michaud, & Larouche, 2010) and a teleoperated interactive platform used as an imitation agent with children with low-functioning autism (Duquette, Michaud, & Mercier, 2008), to a telepresence robot for home care assistance (Michaud, Boissy, et al., 2010) and a robot that can interact autonomously with people in open settings (Michaud, Côté, et al., 2007). In these studies, the experimental settings were conducive to natural interaction: Toddlers discovered how to interact with a rolling robot; rehabilitation professionals evaluated whether they could conduct clinical evaluations through a telepresence robot; an interactive autonomous robot got to face the challenges of going to a conference as a regular attendee. These studies allowed us to develop and evaluate the robots holistically; that is, by considering the robot at a human-robot-context systems level (Lee et al., 2009; Nelson & Stolterman, 2003) in order to identify technological specifications and requirements that can affect human-robot interaction (HRI). This led to the development of innovative technologies and algorithms for natural interaction, such as the following:

Differential Elastic Actuators (DEAs) for safe and compliant actuation (Lauria, Legault, Lavoie, & Michaud, 2008). A DEA acts as an active elastic element that can inherently absorb shocks, perceive the forces from the environment on the robot, and control the forces applied in response. This capability, also known as interaction control (Buerger & Hogan, 2005), is essential for a robot that interacts naturally with people in open settings, because physical contact with the robot must be sensed and processed appropriately. DEAs are conceptually similar to Series Elastic Actuators (SEAs) (Robinson, 2000; Williamson, 1995), which are used on the Meka B1, Cody (Jain & Kemp, 2009), and Nexi-MDS platforms, but use a differential coupling (harmonic drive) instead of a serial coupling of a high impedance mechanical speed source (an electrical DC brushless motor) and a low impedance mechanical spring (a passive torsion spring). A non-turning sensor connected in series with the spring measures the torque output of the actuator. This results in a more compact and simpler solution for controlling mechanical elasticity and viscosity in accordance with an admittance control scheme (Buerger & Hogan, 2005; Hogan, 1985), and the performance is similar to that of SEAs.

ManyEars, a sound source localization, tracking, and separation system (Valin, Michaud, & Rouat, 2007; Valin, Yamamoto, et al., 2007) that uses an array of eight microphones. People interacting with a robot with voice generation capabilities naturally talk back to the robot, as observed, for instance, in Michaud, Laplante, et al. (2005). ManyEars provides an enhanced speaker signal for improved speech and sound recognition in real-world settings. It can simultaneously localize and track up to four sound sources and separate three, all in reverberant and noisy conditions. Only a very limited set of HRI platforms, such as the Nexi-MDS, which has a four-microphone array and a wearable microphone, and the Snackbot (Lee et al., 2009), which uses the Acoustic Magic microphone array, are equipped - with limited capabilities - with artificial audition systems for interacting with people. ManyEars is also the framework on which the HARK open source software (Nakadai, Okuno, Nakajima, Hasegawa, & Tsujino, 2008) is built.

As observed during the holistic design of Snackbot (Lee et al., 2009), the design of an interactive robot is extremely difficult, even when conventional technologies are used. The main challenge is that there is as much to learn from the interaction between the integrated technologies as

from the embodied human-robot interaction, in addition to interdependent effects. In the field of human-computer interaction, "interaction design" is the practice of designing digital products, environments, systems, and services for human-computer interaction (Cooper, Reimann, & Cronin, 2007). Because additional factors must be taken into consideration in embodied interaction, the question then becomes how "interaction design" should be done in the context of natural HRI.

This paper presents one avenue that we are exploring with the design of IRL-1 (Michaud, Ferland, Létourneau, Legault, & Lauria, 2010), a humanoid platform with an expressive face, an orientable head, an eight-microphone array, two arms with 4 degrees of freedom (DOF) each, and grippers, which are all mounted on an omnidirectional, non-holonomic mobile platform. DEAs are used for interactive control of the 4-DOF arms and active steering of the mobile platform, and ManyEars is used to provide artificial audition capabilities. The technologies designed to enable natural interaction were developed independently, and thus may not function seamlessly when integrated onto the same platform. By conducting feasibility studies (also referred to as case studies in Tsui and Yanco (2009)) with the robot in different interaction scenarios, we studied the capabilities of the platform to facilitate this integration following an incremental and iterative design process, along with evaluating ideas of interaction scenarios before conducting in-depth HRI studies. These feasibility studies allowed us to identify, investigate, and validate the influences of IRL-1's specific capabilities (such as the ones provided by DEAs and ManyEars) on the dynamics of interacting with the platform. After presenting IRL-1 and its natural interaction capabilities, this paper describes feasibility studies examining the ability to physically guide the platform, the integration of directed vision and audition for telepresence, and the influence of integrating software processes under the constraint of limited computing resources on the robot's responsiveness in HRI. In addition to demonstrating the capabilities of the platform, observations and insights about design and evaluation of their current and future integration work are provided.

2. IRL-1

Figure 1 shows IRL-1, an omnidirectional platform with two compliant arms and a humanoid torso. Each of these elements provides important capabilities in terms of natural HRI.
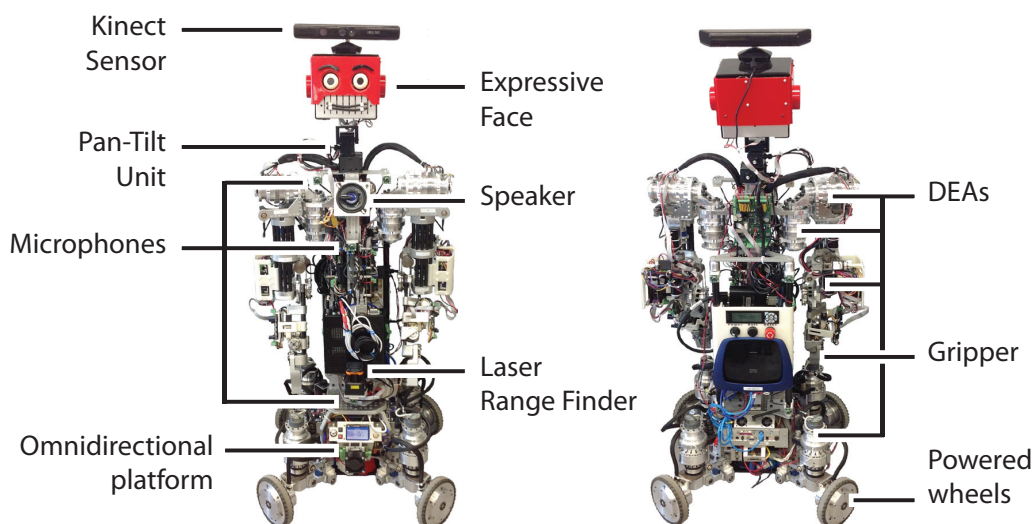

Figure 1. Photographs of the front and back of IRL-1. The robot has 30 DOF and 55 sensors.

Omnidirectionality is the ability to move in all directions without changing orientation, which is useful when moving in tight spaces or crowded areas. Instead of using Mecanum wheels as does the Cody robot (Jain & Kemp, 2009), or powered caster wheels as do PR2 (Wyrobek, Berger, Loos, & Salisbury, 2008) and Rollin' Justin (Borst et al., 2009), omnidirectionality on IRL-1 is provided by DEA-steered ($\pm 90°$), powered wheels with a lateral offset from their attachment points. They are lighter, more precise, and mechanically simpler than Mecanum wheels, and they provide a built-in horizontal suspension system that is absent when powered caster wheels are used. This also makes physical interaction possible from almost any point on the platform, as long as the force applied is not parallel to all of the wheels' propulsion axes. By having off-centered steerable wheels, a force applied at a wheel's contact point can be measured as torque on its steering axis. By combining these torque measurements into a global model of forces applied to the chassis, the platform can react to an external force by moving in its direction (Frémy, Michaud, & Lauria, 2010). The use of off-centered wheels also lowers the height of the chassis, and a passive vertical suspension made of four Rosta springs is used to connect the steerable wheels to the platform's chassis, thus helping to keep the wheels in contact with the ground on uneven surfaces. The platform has a 34 kg payload and can reach a maximum velocity of 1.47 m/s.

DEAs are also used to provide compliance to IRL-1's arms. Each arm is attached to the torso and has 4 DOF (three in the shoulder and one in the elbow). Impedance control of each joint enables an infinite combination of arm behaviors, from zero impedance for free movement with gravity compensation to high stiffness constraining the arms to precise positions or ranges of movement. The arms can sustain impacts with humans or objects, and they can be controlled from low to high admittance with gravity compensation (Legault et al., 2008). A gripping tool serves as a hand at the end of each arm. IRL-1's arms provide sufficient capabilities for HRI studies, even though they have a smaller number of DOF than other compliant mobile manipulation platforms, such as Meka B1, Cody (Jain & Kemp, 2009), Nexi-MDS, Rollin' Justin (Hirzinger, Sporer, Schedl, Butterfass, & Grebenstein, 2004; Ott et al., 2006), and PR2 (Wyrobek et al., 2008).

IRL-1's torso is equipped with one loudspeaker, a Hokuyo UTM-30LX laser range finder (30 m range, $270°$ angle and 25 ms/scan, 0.4 m from the ground), one Kinect motion sensor (providing a 3D point cloud with color for a 640×480 image with an angular field of view of $57°$ horizontally and $43°$ vertically, up to 5 m and at 30 frames per second) and a robot head installed on a pan-tilt unit (FLIR PTU-D46-17, actuated using step motors, pan $\pm 128°$, tilt $[-17, +25]°$). Figure 2 illustrates facial expressions that the robot can make using nine servo-motors (four for the mouth, three for the eyes, and two for the eyebrows). These are not as advanced as the facial expression capabilities of platforms like Meka B1 and Nexi-MDS, but they are sufficient to convey the intended emotion. IRL-1's robot head has been used in other HRI studies (Carter, Scheutz, & Schermerhorn, 2009; Rich, Ponsleur, Holroyd, & Sidner, 2010; Shayganfar, Rich, & Sidner, 2012).

IRL-1 integrates a broad set of features that could be beneficial for HRI studies, with 30 DOF (8 DOF on the mobile base, 10 DOF for the arms and grippers, 11 DOF for the head and the neck, and the loudspeaker) and 55 sensors (position encoders for all 29 actuated DOF plus 16 torque sensors for the ones actuated using DEA, the laser range finder, the Kinect sensor and the eight microphone array). The robot is equipped with two Mini-ITX computers, both running Linux. The first one, a 2.0 GHz Core 2 duo processor with 2 GB of RAM is located in the mobile base and hosts processes related to the robot's base, such as motion control, simultaneous localization and mapping (SLAM), and path planning. The second one, a 2.67 GHz Core i7 quad core processor with 4GB of RAM, is located in front of the torso. It hosts the arm controllers, head expression and orientation, sensor systems such as the base laser range finder, Kinect camera and microphone array, and high-level cognitive processes. The remaining hardware architecture includes 20 distributed controller modules for sensing, power, and low-level control, communicating with each other through a 1 Mbps CAN
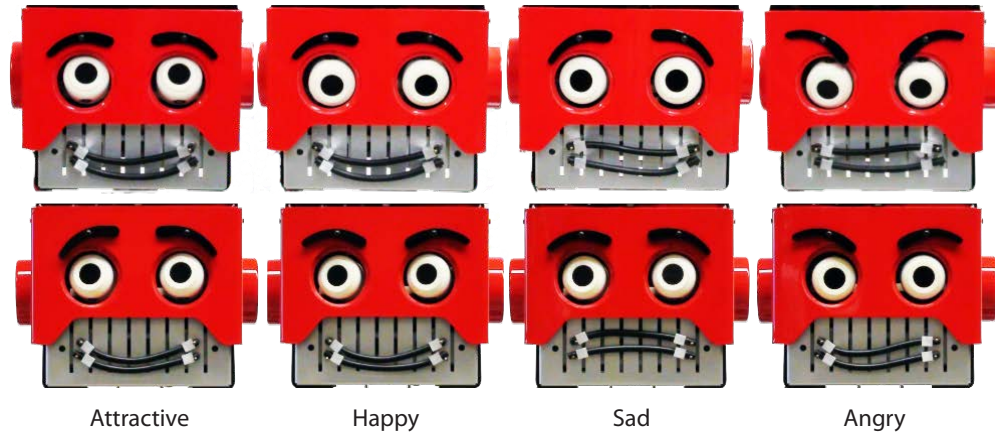
| Attractive | Happy | Sad | Angry |

Figure 2. Photographs of IRL-1's facial expressions: attractive, happy, sad, and angry. The robot's facial expressions are programmed to switch between variations of a single emotion every five seconds.

bus (Michaud, Létourneau, et al., 2005). Nickel-metal hybrid batteries provide power to the mobile base for around half an hour of autonomy at maximum speed or for about 2 hours when the robot is immobile. The torso is powered by a separate pair of 10 Ah LiFePO$_4$ battery packs, providing up to 10 hours of autonomy when the robot's arms are stationary.

3. Physical Interaction Feasibility Study

Physical interaction with a small robot can be relatively simple, intuitive, and natural, as has been observed with children and toddlers (Michaud, Laplante, et al., 2005; Michaud, Salter, et al., 2007). But for large and heavy robots such as IRL-1 (1.4 m high and around 70 kg), an active method is required for the robot to react to human pushing or pulling. Remote controls are commonly used to operate such robots. However, leading the robot directly by the hand or arm has been shown to be more desirable in many ways (Chen & Kemp, 2010, 2011). In that regard, the contact points for direct physical interaction (DPI) should not be limited to specific parts of the robot. In previous work, we demonstrated that this capability can be achieved with our omnidirectional base (Frémy et al., 2010), and here we investigate whether it is possible with IRL-1's upper torso installed on the base.

Figure 3 illustrates the experimental setup. The objective is to make the robot rotate 180° by going through a three-point turn trajectory under two conditions:

- Physical interaction: Participants can push or pull the robot directly, and the applied forces sensed by DEAs are used for active steering of the omnidirectional base. IRL-1's arms are slightly extended to keep its grippers at 0.45 m from the shoulders. Torque measurements from DEAs on the arms are not exploited to detect physical interaction, but pushing or pulling the arms or the grippers can generate enough force on a wheel's contact point to be detected by the DEA-steered powered wheels.
- Remote control: The left analog control stick commands both linear velocity in X and angular velocity around Z. Participants were free to move around to get a better view of the robot's movement.
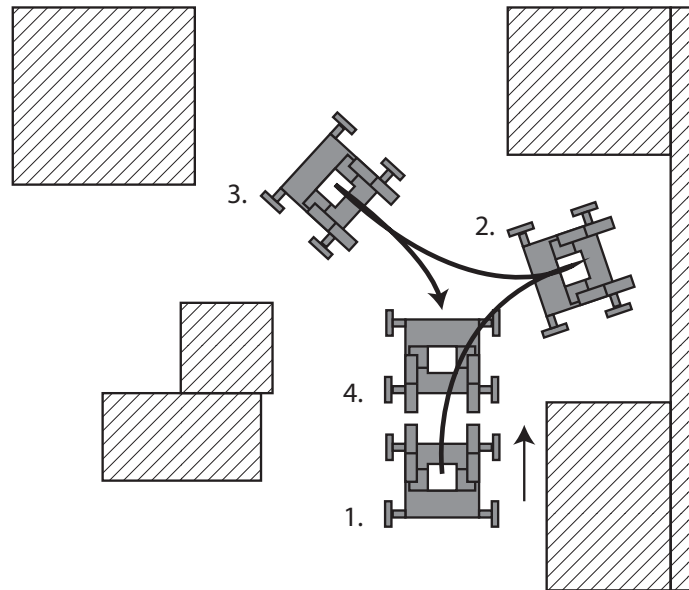
Figure 3. Physical interaction experiment involving a three-point turn trajectory.

The experiments took place in a cluttered office space of 10 m². Instructions for the robot and the remote control were given to the participants, and they had time to practice with the control conditions before performing the experiment. A convenience sample of six male engineers (aged from 22 to 26 years) were permitted five attempts with each control method to perform the maneuver while the task completion time and the number of collisions were measured.

Participants performed faster with the remote control ($\bar{x} = 23.93$ s, $\sigma = 10.98$ s) compared to physical interaction ($\bar{x} = 34.67$ s, $\sigma = 14.99$ s). This can be explained by the fact that the maximum velocity can easily be reached using the remote control. Also, participants were more careful not to damage the robot when interacting physically with it, and more collisions occurred when participants used the remote control (three versus none). This last observation concurs with Chen and Kemp (2010, 2011); that is, collisions occurred significantly less often when participants performed their tasks with a DPI instead of a gamepad remote control.

Comments from the participants revealed that while it was easy to move the robot on a straight path, it was difficult to send the mobile base on a curved path or to switch from a left-handed curve to a right-handed one. Torque measurements from the steering axes' actuators can be noisy and often contain a constant bias difficult to predict and mostly caused by the weight of the robot's torso itself. While we could dynamically estimate the center of gravity of the robot based on all of its joint positions, unmonitored suspension travel would render this estimate almost useless for predicting the actual effect of the weight of the robot's upper body on the force-guiding algorithm. A minimal torque threshold of 0.800 Nm was set to compensate for this bias: While this effectively masks noise and bias from the torque measurements, it explains why more force must be applied to move the platform. This is especially apparent when rotating the platform by physically interacting with its arms: The longest moment arm to apply torque around the robot's Z axis passes through the shoulders, which even when actuated with high impedance absorb part of the torque. Suspension travel and force absorption by the elasticity of the arms can be avoided by applying external forces

directly to the mobile platform's steering actuators, but these contact points are lower and not as easily accessible. Therefore, for a platform where elasticity and force-controlled actuation are present both at the locomotion and arm levels, the contact point for physical interaction will influence the performance. To improve sensitivity to physical interaction with the arms, we can use position and torque measurements directly from the arms' DEAs to examine how these can be used together with the torque measurements from the base to allow smooth, precise, and responsive physical interaction with the robot. Since there is an interaction between the two compliant systems (the omnidirectional base and the arms), it will also be important to measure the contact point with people who are naturally interacting with the platform. Upgrading the suspension of the omnidirectional base and using compliant actuators that do not exploit springs as the elastic elements (Fauteux, Lauria, Heintz, & Michaud, 2010) would also help. In spite of these limitations, two participants mentioned that they found pushing the robot more intuitive than using the remote control, and no comments were made to the contrary. One of these participants noted that the remote control scheme was very similar to common car-racing videogames (which made it easy for him to understand), but the learning curve still appeared to be easier with physical interaction.

4.   Augmented Teleoperation Feasibility Study

Natural interaction in robot telepresence can be addressed from two perspectives: from the perspective of the person interacting with the robot at the remote location or from the perspective of the distant operator. Using humanoid robots such as IRL-1 for telepresence applications provides capabilities such as motion, gesture, force sensing, and sound source localization that go beyond what is available on more conventional telepresence platforms, such as Rovio, Vgo, TiLR, Giraffe, RO-7i, Sparky Jr., Texai and QB (Guizzo, 2010), or our telepresence robot for home care assistance (Michaud, Boissy, et al., 2010). For the distant operator, the main challenge of telepresence is understanding an environment in which one is not physically present (Nielsen, Goodrich, & Ricks, 2007). Situation awareness (Endsley, 1988; Labonté, Boissy, & Michaud, 2010; J. C. Scholtz, Antonishek, & Young, 2005; J. Scholtz, Young, Yanco, & Drury, 2004) and cognitive load (Labonté et al., 2010; Nielsen et al., 2007) are primary concerns, especially with a humanoid robot with many sensors and DOF. Assuming that the remote operation station is a computer, the challenge then becomes how these capabilities can be appropriately represented so that the operator can exploit the information and abilities provided by the platform. Evaluating natural interaction from the perspective of the person interacting with the robot can be addressed afterwards.

Using an egocentric viewpoint (seeing the world from the robot's perspective) is usually better for navigating and avoiding obstacles, while an exocentric viewpoint (observing the world from an external perspective) gives the operator a better understanding of the environment's structure (Ricks, Nielsen, & Goodrich, 2004). In previous work, we developed and evaluated interfaces with different viewpoints (Ferland, Pomerleau, Le Dinh, & Michaud, 2009; Michaud, Boissy, et al., 2010). We conducted a comparative study with 37 novice operators between 1) a video-centric display, 2) an augmented reality display (superimposing the video stream on a 3D virtual model of the environment), and 3) a mixed-perspective display providing an exocentric viewpoint (the 3D model) in the center of the display and an egocentric view (the video feed), with a reference between both perspectives (the robot position) (Labonté et al., 2010). The second modality was revealed to be the most usable and most effective interface in terms of completion time and number of commands, especially for women, people older than 30 years old, or people working on computers less than 22 hr per week executing moving tasks. The third modality turned out to be preferable in situations requiring precise navigation but is sensitive to the robot's localization accuracy, which influences the alignment of the 3D model with object features in the video stream. In a follow-up pilot study, we experimented with an adjustable ego/exocentric 3D display, allowing operators to change viewpoints

as desired (Ferland et al., 2009). From a test population of 13 participants, this functionality was found to be very useful, and experienced videogamers performed better, making the robot move at maximum speed. The growing popularity of videogames with the general population is certainly a factor to take into consideration in the design of teleoperation interfaces. In 1998, 3% of the Canadian population over 15 years of age spent an average of 1 hr 48 min per week playing video games. In 2010, these figures grew to 6% and 2 hr 20 min (Béchard, 2011).

Therefore, we began our exploration of IRL-1 as a telepresence platform by extending our work on the adjustable ego/exocentric 3D display (Ferland et al., 2009) and investigating the added benefits and challenges of omnidirectionality and directed perception. This interface combines a 3D reconstruction of the environment, using either laser range finder readings, with a video projection method or a point cloud (produced by a stereoscopic camera or a Kinect sensor), and it allows the operator to switch seamlessly between egocentric and exocentric viewpoints. Exocentric viewpoints are necessarily limited by the camera's field of view (FOV). With the Kinect sensor mounted on IRL-1's orientable head, 3D video projection no longer had to stay aligned with the front of the robot: It could change according to the orientation (pan and tilt) of the head. Figure 4 illustrates the concept: (a) is the case when the virtual camera is aligned with the robot's orientation and head, located slightly behind the robot to visualize a representation of robot as in Nielsen et at. (2007) and Ferland et al. (2009); (b) represents the case of an exocentric viewpoint that is independent of the robot's head orientation, and (c) is when the virtual camera is aligned with the robot's head but not necessarily with the robot's orientation. This capability keeps the virtual camera aligned with the Kinect FOV, thus simplifying the virtual camera's pan and tilt control. The virtual target is always at a 5 m distance from the robot, but the distance between the virtual camera's position and the robot can be adjusted, effectively acting as zoom control.
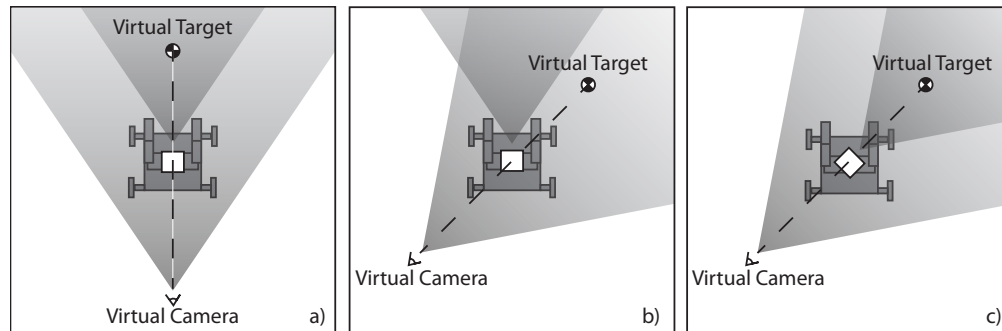

Figure 4. Top view of possible exocentric viewpoints: a) viewpoint aligned with the robot's orientation and head position; b) viewpoint independent of the robot's orientation and head position; and c) viewpoint aligned with the robot's head position but not with the robot's orientation.

To rapidly prototype and evaluate this concept, we implemented a simple experimental setup using rviz, which is the visualization tool provided with Willow Garage's Robot Operating System (ROS) (Quigley et al., 2009), as shown in Figure 5. An articulated, semi-transparent model of IRL-1 is rendered at the bottom. Laser range readings are displayed as points with colors, from red to blue, based on the distance of the perceived objects along the robot's X axis. The point cloud generated by and visible from the Kinect camera is shown in the middle of the screen. Sounds located by ManyEars are displayed as red arrows pointing in the direction of the sound sources. The feasibility study involved teleoperating the robot in the cluttered environment, given no specific objectives, using one of two modes:
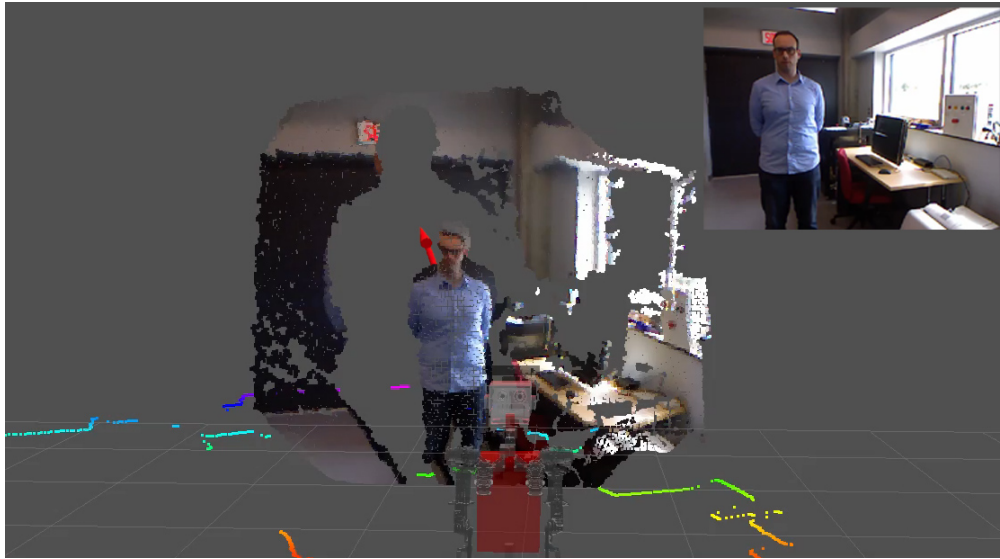
Figure 5. Telepresence prototype interface within rviz. The image captured by the Kinect camera is shown in the top right corner. The red arrow, partially masked by the Kinect point cloud, shows the direction of a detected sound source.

- **Mode 1**: IRL-1's head is always oriented forward, with no pan or tilt controls, but the virtual camera's position and target can be rotated around the robot's base, as in Figure 4(b). This mode corresponds to our previous work on adjustable ego/exocentric 3D displays (Ferland et al., 2009).
- **Mode 2**: The virtual camera's position and target still rotate around the robot's base, and its head orientation follows, within its limits, the virtual camera's orientation, as in Figure 4(c). If the virtual target is beyond a soft angle limit in any direction ($\pm30°$), the robot's base reorients itself toward the virtual target. This new display modality exploits the added capabilities of IRL-1 by associating the user's virtual camera with the robot's motor control.

A convenience sample of seven male engineers, aged 22 to 35 years, teleoperated IRL-1 using each mode for up to 10 minutes.

From these experiments, we observed that Mode 2 has considerable potential. Orienting the Kinect while the base is moving helped locate nearby obstacles, especially those that could not be detected by the laser range finder because of its height. In Mode 1, objects on the ground at close range were not part of the Kinect's FOV, thus increasing the importance of the laser range readings for navigation. On the other hand, latency in the Kinect data feed became apparent when the pan-tilt unit state was updated on the display slightly earlier than the Kinect's point cloud. The Kinect data stream used considerable bandwidth and was subject to large variations in network latency for each point cloud. The result was that the Kinect point cloud was displayed temporarily out of alignment with the robot's head. In our implementation, no specific precautions were taken to minimize the effect of the network latency's variation when executing user commands and displaying perceived information. As mentioned in Kelly et al. (2011), teleoperation becomes even more difficult in environments surrounded by dense obstacles, with network latency and high speed requirements.

While IRL-1's maximum velocity is not significantly higher than in our work on the adjustable ego/exocentric 3D display (0.5 m/s vs. 0.3 m/s), IRL-1's head motion produces rapid changes to the content of the video feed and the virtual camera's position that need to be compensated for. Motion prediction with a velocity-driven dynamic model (Kelly et al., 2011) and wave variables (Munir & Book, 2002) is currently being considered to compensate for latencies that can occur between graphical representation of the robot's state, and perception.

We also learned that customization options should be available to the user. For instance, the threshold angle that triggers the robot's base movement when adjusting the pan angle of the virtual camera could be provided as a movement sensitivity setting. A participant complained that his commands, which were frequently very close to this movement threshold, resulted in jerkiness in both the robot's movement and the camera display. Allowing users to set this threshold would be preferable. Finally, an interesting idea for a test case would be to have participants engage in a teleoperated pursuit task and to measure the number of collisions, the distance between the robot and the target, and the time that the target remains visible on the display.

5. Computing Resource Management Feasibility Study

The design of a highly sophisticated mobile robot requires the integration of capabilities such as navigation, localization and mapping, vision and audio processing, tracking and recognition, graphical or natural interfaces, and planning and reasoning using different abstraction levels (Michaud, Côté, et al., 2007). Integrating many capabilities onto the same platform increases the level of complexity for controlling the robot and, incidentally, the load on the robot's computing resources. While technological progress in computer design expands the limits every year, we can also safely assume that software requirements will continue to increase in complexity for improved perception and reasoning capabilities. The use of offboard computing resources, such as web-enabled robots (Tenorth, Klank, Pangercic, & Beetz, 2011) or cloud computing robotics, is an option. However, this approach is subject to latency and reliability issues, and its use is limited to locations with Internet connectivity.

Humans solve the problem of limited computing resources by selective attention. In psychology, selective attention has been studied according to two schools of thought (Bundesen & Habekost, 2004):

- Early selection theory (Broadbent, 1958) proposes that human perceptual capacities are limited, and unimportant stimuli need to be filtered before they overload our analyzing capabilities.
- Late selection theory (Deutsch & Deutsch, 1963) proposes that our perceptual analysis capabilities are actually unlimited and that selection occurs later based on the perceived importance of the stimuli.

More recent work by Lavie (1996) attempts to unify both theories by showing that while our perceptual resources are limited, we try to exploit them to their maximum. Early selection would then only be necessary under a significant perceptual load.

We are currently investigating the implementation of selection attention mechanisms in our Hybrid Behavior-Based Architecture (HBBA) (Michaud, Ferland, et al., 2010), which is illustrated in Figure 6. Driven by data processed by Perception Modules, Behavior-Producing Modules (BPMs) are distributed processes used to generate commands sent to the robot's actuators using an action selection strategy (e.g., priority-based) (Pirjanian, 2000). BPMs are activated and configured according to the Intentions of the system and evaluated in the Intention Workspace from concurrent Desires generated by the Motivation modules. Similar to BPMs, Motivations are distributed processes manifesting Desires for the accomplishment or the termination of Intentions, which are associated with
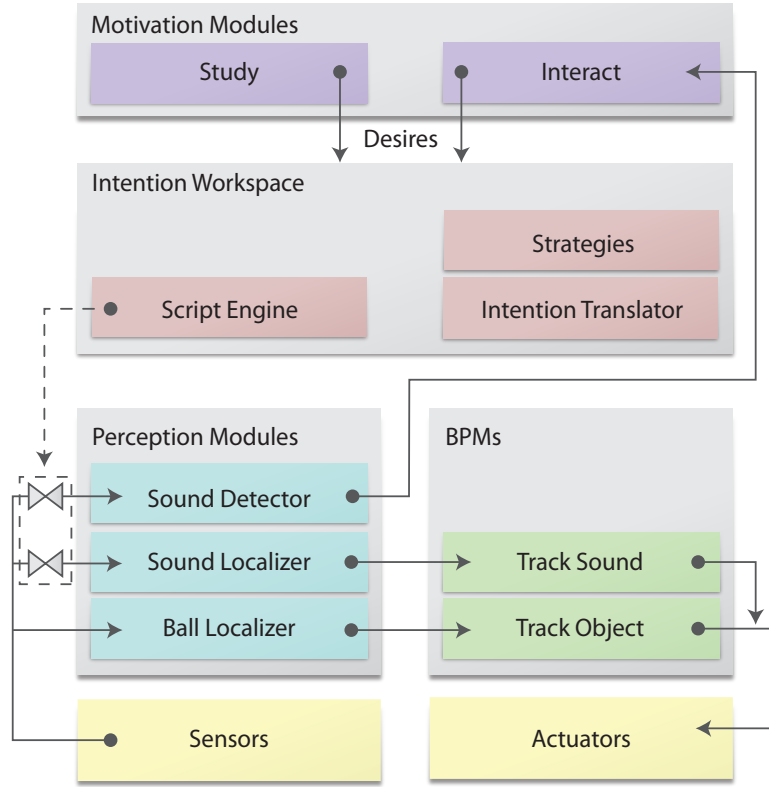
Figure 6. HBBA conceptual framework with experiment-specific modules.

the activation and configuration of BPMs and with the modulation of Perception Modules. Intentions are therefore used for early selective attention by activating or deactivating filters (represented by the ⋈ symbol at the input of two Perception Modules) that modulate the flow of information from sensors to perceptual modules, or Late Selection by activating and configuring BPMs.

In a previous experimental setup to validate this framework, IRL-1 was programmed to track the 3D position of a blue ball with a Kinect camera and a tracking algorithm implemented with the Point Cloud Library (PCL) (Rusu & Cousins, 2011). Meanwhile, a person was trying to distract IRL-1 by hand clapping or by speaking loudly. In this case, Study is the Motivation Module making the robot track the blue ball (using the Ball Localizer Perception Module and the Track Object BPM), while Interact makes the robot responsive to sounds by orienting its head toward the sound sources (using ManyEars for Sound Localizer and the Track Sound BPM). To test our early selective attention mechanism, we constrained the available CPU resources so that these BPMs could not be executed concurrently in real time: Either sound sources were detected after a delay of multiple seconds, or the ball was tracked at rates of less than 1 Hz. A less computer-intensive Perception Module called Sound Detector was therefore added to monitor loud noises using a single microphone located on the front of the torso. Unlike Sound Localizer, Sound Detector can stay activated when no loud noise is perceived without consuming significant resources, allowing IRL-1 to be more responsive while tracking the blue ball. When a loud noise was detected, the average delay between the beginning of the hand clap sound and the movement of the head pan-tilt unit was 1015 ms ($\sigma = 294$ ms),
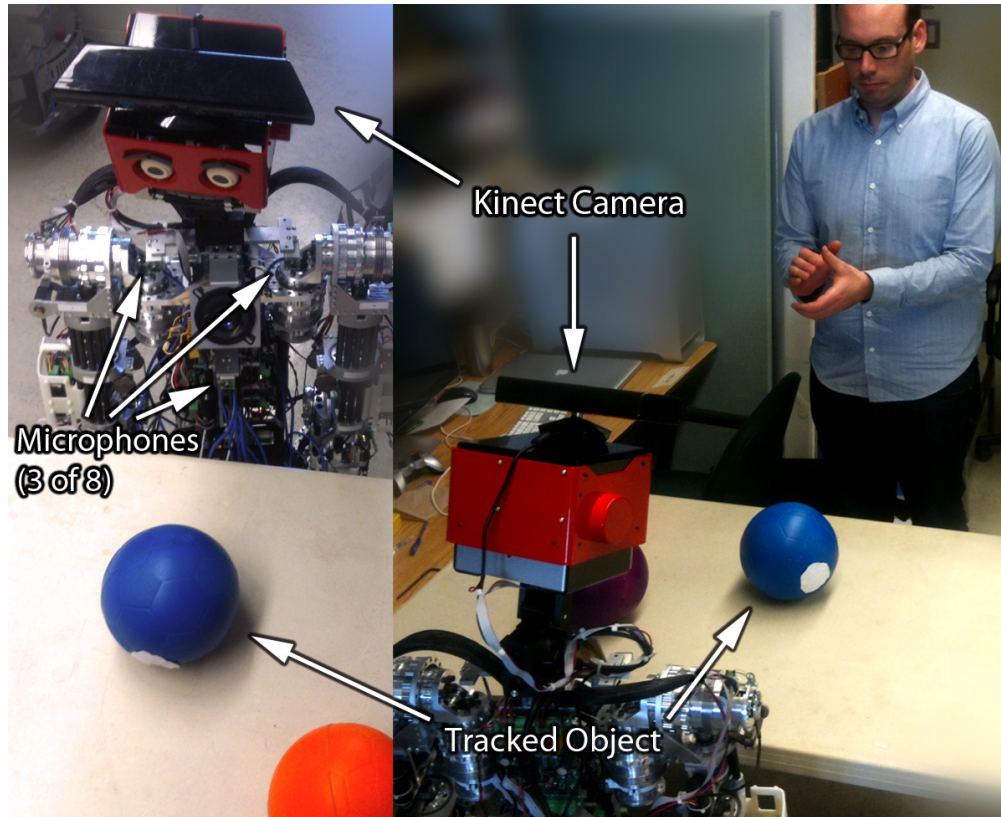
Figure 7. Experimental setup for the feasibility study on resource management.

compared to 692 ms ($\sigma = 122$ ms) when Sound Localizer was the only running process. This additional 323 ms delay was mostly caused by the time required by ManyEars to process the audio stream buffer of the last recorded second to locate the sound source.

To determine if people interacting directly with the robot would consider the additional delay acceptable, we asked a convenience sample of eight male engineers (aged 22 to 35 years) to distract the robot and to indicate when a) a delay is noticeable, b) the delay is considered uncomfortable, and c) the delay is considered unacceptable. Artificial delays from 0 to 3000 ms with 100 ms increments were generated and added to the base response time. Participants were asked to base their evaluation on their own expectations of the robot and of the task, and not to compare it to a human. Before conducting the experiments, participants were allowed to interact with IRL-1 to find the required sound level for reliably distracting its attention from the blue ball.

Figure 8 illustrates the results. Noticeable additional delays are mainly identified between 400 and 600 ms (for 7 out of 8 participants), while for delays considered uncomfortable and unacceptable the ranges are much larger. This suggests that the overhead delay of 323 ms is not a variable that should influence natural HRI experiments.

Comments from participants provided interesting insights for the preparation of future experiments. For instance, when hand clapping was not loud enough to be detected, a participant took this as a sign that the robot was busy and that the ignored interruption was not simply caused by
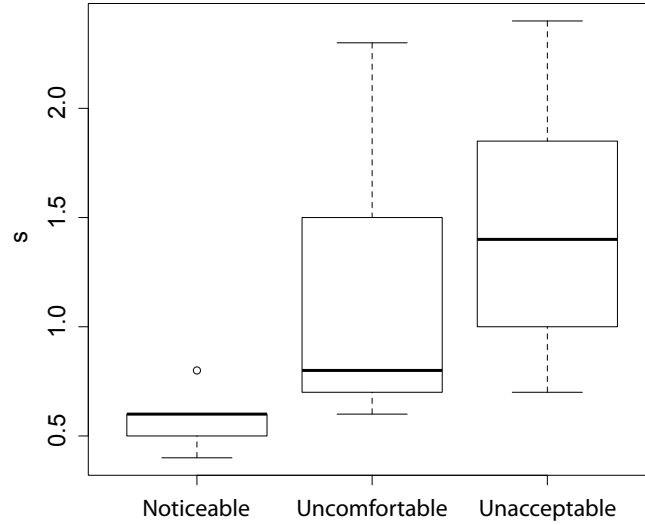
Figure 8. Noticeable, uncomfortable and unacceptable delays for IRL-1's responsiveness to sound distraction.

a malfunction. In that regard, IRL-1's facial expressions revealed an interesting side effect. Facial expressions on IRL-1 were programmed to switch between variations of a single emotion every five seconds. This results in IRL-1's eyes moving slightly at regular intervals, which caused some participants to ask if this should be interpreted as a sign of IRL-1 being successfully distracted by the sound. They were told that it was not the case because IRL-1's head had not moved, but some participants indicated that it made the robot appear more responsive. This may be a simple and natural gesture for making the robot respond rapidly to the detection of a sound before ManyEars can provide a location for the detected sound. Such an addition would be in line with the observation that external observers feel that an *attentive* robot, which moves its head toward the interlocutor in a two-person, turn-taking conversation, is considered to be more natural than a *distracted* robot (moving its head with a 500 ms delay) (Trafton, Bugajska, Fransen, & Ratwani, 2008). Finally, a participant commented that he would have said a much lower delay was unacceptable if he had been trying to disturb a person instead of a robot; he explained that he did not really expect contemporary robots to react to his interruptions. It may therefore be interesting to investigate people's expectation for interaction responsiveness in relation to a robot's appearance, capabilities, and role.

6.   Discussion and Future Work

The HRI perspective addressed in this paper relates to the concept of "natural interaction design", which we could define as the practice of designing technologies and capabilities for human-robot interaction. Pursuing the long-term objective of developing robots, algorithms, and technologies for natural HRI, experiments conducted over the years in different HRI contexts (e.g., children, home care assistance, interactions in public spaces) have led us to design new technologies, such as DEAs and ManyEars, that we have integrated into IRL-1, an omnidirectional, force-sensing,

compliant humanoid robot. The feasibility studies presented in this paper are part of an iterative and incremental process aimed to improve understanding of how the integration of the advanced motion, perception, and processing capabilities influence each other in natural interaction. Experimenting with new technologies brings new challenges, and we believe that the feasibility studies conducted with IRL-1 are an important stepping stone to evaluating design and evaluation considerations for safe and efficient progress toward more advanced integration and HRI studies.

The feasibility studies explored distinctive capabilities of IRL-1: the ability to be physically guided by perceiving forces through DEAs used for active steering of the omnidirectional platform; the integration of vision, motion and audition for an augmented telepresence interface; and the influence of delays as computationally expensive processes (such as ManyEars) are integrated in responding to user requests. In addition to demonstrating how these capabilities can be used in HRI, this paper illustrates how their design and evaluation are intrinsically interrelated. For instance, the contact point at which IRL-1 is physically guided influences interaction, and this should be taken into consideration both for the design (by also using the arms' DEAs to measure applied forces) and the evaluation (by observing where users guide the platform). The orientation of IRL-1's head (and therefore the Kinect), independent of the platform motion, can make network delays visible in the graphical display, making this an important consideration in the design and evaluation of telepresence interfaces for humanoid robots. And as reasonable delays can be reached using a selective attention mechanism as computationally expensive processes are integrated on IRL-1, simple but responsive facial gestures may be beneficial in natural interaction, and should be investigated further.

As the design of IRL-1 progresses, many additional experiments can be imagined, such as the use of facial expressions to communicate information about motion constraints in physically guiding the platform; the study of user preferences in controlling mobility and viewpoints and of representing forces and sound classification on the graphical display; and vocal interaction with IRL-1 as it navigates in a cluttered environment (which requires the addition of SLAM capabilities, which are computationally expensive), while responding in a timely fashion to environmental conditions and user requests. In developing new technologies and conducting trials in such a wide set of conditions, we believe that advancing the state of the art in natural HRI is an interesting avenue to address the fundamental integration challenge in the field of artificial intelligence (Brachman, 2006), making robots more versatile, meaningful, and rational machines.

Acknowledgments

References

Béchard, M. (2011). *General social survey - 2010 overview of the time use of Canadians* (Tech. Rep. No. 89-647-XWE). Ottawa, ON: Statistics Canada.

Borst, C., Wimböck, T., Schmidt, F., Fuchs, M., Brunner, B., Zacharias, F., et al. (2009). Rollin' Justin – Mobile platform with variable base. In *Proceedings of the IEEE International Conference on Robotics and Automation* (pp. 1597–1598). Anchorage, AK: IEEE, http://dx.doi.org/10.1109/ROBOT.2009.5152586.

Brachman, R. (2006). (AA)AI more than the sum of its parts. *AI Magazine*, *27*(4), 19–34.

Broadbent, D. E. (1958). *Perception and communication*. London, UK: Pergamon Press.

Buerger, S. P., & Hogan, N. (2005). Impedance and interaction control. In T. R. Kurfess (Ed.), *Robotics and Automation Handbook* (pp. 19.1–19.24). Boca Raton, FL: CRC Press.

Bundesen, C., & Habekost, T. (2004). Attention. In K. Lamberts, R. L. Goldstone, & R. Goldstone (Eds.), *Handbook of cognition* (pp. 105–129). London, UK: SAGE.

Carter, K., Scheutz, M., & Schermerhorn, P. (2009). A humanoid-robotic replica in USARSim for HRI experiments. In *IROS Workshop on Robots, Games, and Research.* St. Louis, MO: Retrieved from http://hrilab.tufts.edu/publications/carterscheutzschermerhorn09rgr.pdf.

Chen, T. L., & Kemp, C. C. (2010). Lead me by the hand: Evaluation of a direct physical interface for nursing assistant robots. In *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 367–374). Osaka, JP: ACM, http://dx.doi.org/10.1109/HRI.2010.5453162.

Chen, T. L., & Kemp, C. C. (2011). A direct physical interface for navigation and positioning of a robotic nursing assistant. *Advanced Robotics*, *25*(5), 605–627, http://dx.doi.org/10.1163/016918611X558243.

Cooper, A., Reimann, R., & Cronin, D. (2007). *About Face 3: The essentials of interaction design*. Indianapolis, IN: Wiley.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*(1), 80–90.

Duquette, A., Michaud, F., & Mercier, H. (2008). Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism. *Autonomous Robots*, *24*(2), 147–157, http://dx.doi.org/10.1007/s10514-007-9056-5.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society 32nd Annual Meeting* (pp. 97–101). Anaheim, CA: Human Factors & Ergonomics Society.

Fauteux, P., Lauria, M., Heintz, B., & Michaud, F. (2010). Dual differential rheological actuator for high performance physical robotic interaction. *IEEE Transactions on Robotics*, *26*, 607–618, http://dx.doi.org/10.1109/TRO.2010.2052880.

Ferland, F., Pomerleau, F., Le Dinh, C., & Michaud, F. (2009). Egocentric and exocentric teleoperation interface using real-time, 3D video projection. In *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 37–44). San Diego, CA: ACM, http://dx.doi.org/10.1145/1514095.1514105.

Frémy, J., Michaud, F., & Lauria, M. (2010). Pushing a robot along – A natural interface for human-robot interaction. In *Proceedings of the IEEE International Conference on Robotics and Automation* (pp. 3440–3445). Taipei, TW: IEEE, http://dx.doi.org/10.1109/ROBOT.2010.5509584.

Guizzo, E. (2010). When my avatar went to work. *IEEE Spectrum*, *47*(9), 26–32.

Hirzinger, G., Sporer, N., Schedl, M., Butterfass, J., & Grebenstein, M. (2004). Torque-controlled lightweight arms and articulated hands: Do we reach technological limits now? *International Journal of Robotics Research*, *23*(4-5), 331–340, http://dx.doi.org/10.1177/0278364904042201.

Hogan, N. (1985). Impedance control: An approach to manipulation: Part II - Dynamics systems measurement control. *Journal of Dynamic Systems, Measurement, and Control*, *107*(1), 8–16, http://dx.doi.org/10.1115/1.3140713.

Jain, A., & Kemp, C. C. (2009). Pulling open novel doors and drawers with equilibrium point control. In *Proceedings of the IEEE-RAS International Conference on Humanoid Robotics* (pp. 498–505). Paris, FR: IEEE, http://dx.doi.org/10.1109/ICHR.2009.5379532.

Kelly, A., Chan, N., Herman, H., Huber, D., Meyers, R., Rander, P., et al. (2011). Real-time photorealistic virtualized reality interface for remote mobile robot control. *International Journal of Robotics Research*, *30*(3), 384–404, http://dx.doi.org/10.1177/0278364910383724.

Labonté, D., Boissy, P., & Michaud, F. (2010). Comparative analysis of 3D robot teleoperation interfaces with novice users. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, *40*, 1331–1342, http://dx.doi.org/10.1109/TSMCB.2009.2038357.

Lauria, M., Legault, M.-A., Lavoie, M.-A., & Michaud, F. (2008). Differential elastic actuator for robotic interaction tasks. In *Proceedings of the IEEE International Conference on Robotics and Automation* (pp. 3606–3611). Pasadena, CA: IEEE, http://dx.doi.org/10.1109/ROBOT.2008.4543763.

Lavie, N. (1996). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology*, *21*(3), 451–468, http://dx.doi.org/10.1037/0096-1523.21.3.451.

Lee, M. K., Forlizzi, J., Rybski, P. E., Crabbe, F., Chung, W., Finkle, J., et al. (2009). The Snackbot:

Documenting the design of a robot for long-term human-robot interaction. In *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 7–14). San Diego, CA: ACM, http://dx.doi.org/10.1145/1514095.1514100.

Legault, M.-A., Lavoie, M.-A., Cabana, F., Jacob-Goudreau, P., Letourneau, D., Michaud, F., et al. (2008). Admittance control of a human centered 3 dof robotic arm using differential elastic actuators. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 4143–4144). Nice, FR: IEEE, http://dx.doi.org/10.1109/IROS.2008.4651039.

Michaud, F., Boissy, P., Labonté, D., Brière, S., Perreault, K., Corriveau, H., et al. (2010). Exploratory design and evaluation of a homecare teleassistive mobile robotic system. *Mechatronics*, *20*(7), 751–766, http://dx.doi.org/10.1016/j.mechatronics.2010.01.010.

Michaud, F., Côté, C., Létourneau, D., Brosseau, Y., Valin, J.-M., Beaudry, É., et al. (2007). Spartacus attending the 2005 AAAI conference. *Autonomous Robots*, *22*(4), 369–384, http://dx.doi.org/10.1007/s10514-006-9014-7.

Michaud, F., Ferland, F., Létourneau, D., Legault, M., & Lauria, M. (2010). Toward autonomous, compliant, omnidirectional humanoid robots for natural interaction in real-life settings. *Paladyn Behavioral Robotic Journal*, *1*(1), 57–65, http://dx.doi.org/10.2478/s13230-010-0003-3.

Michaud, F., Laplante, J., Larouche, H., Duquette, A., Caron, S., Letourneau, D., et al. (2005). Autonomous spherical mobile robotic to study child development. *IEEE Transactions on Systems, Man, and Cybernetics*, *35*(4), 471–480, http://dx.doi.org/10.1109/TSMCA.2005.850596.

Michaud, F., Létourneau, D., Arsenault, M., Bergeron, Y., Cadrin, R., Gagnon, F., et al. (2005). Multi-modal locomotion robotic platform using leg-track-wheel articulations. *Autonomous Robots*, *18*(2), 137–156, http://dx.doi.org/10.1007/s10514-005-0722-1.

Michaud, F., Salter, T., Duquette, A., & Laplante, J.-F. (2007). Perspectives on mobile robots used as tools for pediatric rehabilitation. *Assistive Technologies*, *19*(1), 21–36, http://dx.doi.org/10.1080/10400435.2007.10131863.

Munir, S., & Book, W. (2002). Internet-based teleoperation using wave variables with prediction. *IEEE/ASME Transactions on Mechatronics*, *7*(2), 124–133, http://dx.doi.org/10.1109/TMECH.2002.1011249.

Nakadai, K., Okuno, H. G., Nakajima, H., Hasegawa, Y., & Tsujino, H. (2008). An open source software system for robot audition HARK and its evaluation. In *Proceedings of the IEEE-RAS International Conference on Humanoid Robots* (pp. 561–566). Daejeon, KR: IEEE, http://dx.doi.org/10.1109/ICHR.2008.4756031.

Nelson, H. G., & Stolterman, E. (2003). *The design way*. Englewood Cliffs, NJ: Educational Technology Publication.

Nielsen, C., Goodrich, M., & Ricks, R. (2007). Ecological interfaces for improving mobile robot teleoperation. *IEEE Transactions on Robotics*, *23*(5), 927–941, http://dx.doi.org/10.1109/TRO.2007.907479.

Ott, C., Eiberger, O., Friedl, W., Bäuml, B., Hillenbr, U., Borst, C., et al. (2006). A humanoid two-arm system for dexterous manipulation. In *Proceedings of the IEEE International Conference on Humanoid Robots* (pp. 276–283). Genoa, IT: IEEE, http://dx.doi.org/10.1109/ICHR.2006.321397.

Pirjanian, P. (2000). Multiple objective behavior-based control. *Robotics and Autonomous Systems*, *31*(1-2), 55–60, http://dx.doi.org/10.1016/S0921-8890(99)00081-0.

Quigley, M., Gerkey, B., Conley, K., Faust, J., Foote, T., Leibs, J., et al. (2009). *ROS: An open-source robot operating system.* Kobe, JP: Retrieved from http://robotics.stanford.edu/ ang/papers/icraoss09-ROS.pdf.

Rich, C., Ponsleur, B., Holroyd, A., & Sidner, C. L. (2010). Recognizing engagement in human-robot interaction. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction* (pp. 375–382). Osaka, JP: ACM, http://dx.doi.org/10.1109/HRI.2010.5453163.

Ricks, B., Nielsen, C., & Goodrich, M. (2004). Ecological displays for robot interaction: a new perspective. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 2855–2860). Sendai, JP: IEEE, http://dx.doi.org/10.1109/IROS.2004.1389842.

Robinson, D. (2000). *Design and Analysis of Series Elasticity in Closed Loop Actuator Force Control*. (Doctoral thesis, Massachussetts Institute of Technology, Cambridge, Boston. Retrieved from http://dspace.mit.edu/handle/1721.1/54838).

Rusu, R., & Cousins, S. (2011). 3D is here: Point Cloud Library (PCL). In *Proceedings of*

the IEEE International Conference on Robotics and Automation (pp. 1–4). Shangai, CN: IEEE, http://dx.doi.org/10.1109/ICRA.2011.5980567.

Salter, T., Michaud, F., & Larouche, H. (2010). How wild is wild? A taxonomy to characterize the 'wildness' of child-robot interaction. *International Journal of Social Robotics*, *2*(4), 405–415, http://dx.doi.org/10.1007/s12369-010-0069-4.

Scholtz, J., Young, J., Yanco, H., & Drury, J. (2004). Where am I? Acquiring situation awareness using a remote robot platform. In *Proceedings of the IEEE Conference on Systems, Man and Cybernetics* (pp. 2835–2840). The Hague, NL: IEEE, http://dx.doi.org/10.1109/ICSMC.2004.1400762.

Scholtz, J. C., Antonishek, B., & Young, J. D. (2005). Implementation of a situation awareness assessment tool for evaluation of human-robot interfaces. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, *35*(4), 450–459, http://dx.doi.org/10.1109/TSMCA.2005.850589.

Shayganfar, M., Rich, C., & Sidner, C. L. (2012). A design methodology for expressing emotion on robot faces. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 4577–4583). Vilamoura, PT, http://dx.doi.org/10.1109/IROS.2012.6385901: IEEE.

Tenorth, M., Klank, U., Pangercic, D., & Beetz, M. (2011). Web-enabled robots – Robots that use the web as an information resource. *Robotics & Automation Magazine*, *18*(2), 58–68, http://dx.doi.org/10.1109/MRA.2011.940993.

Trafton, J. G., Bugajska, M. D., Fransen, B. R., & Ratwani, R. M. (2008). Integrating vision and audition within a cognitive architecture to track conversations. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction* (pp. 201–208). Amsterdam, NL: ACM, http://dx.doi.org/10.1145/1349822.1349849.

Tsui, K. T., & Yanco, H. A. (2009). Towards establishing clinical credibility for rehabilitation and assistive robots through experimental design. In *Proceedings of the Robotics: Science and Systems Workshop on Good Experimental Methodology in Robotics*. Seattle, WA: Retrieved from http://robotics.cs.uml.edu/fileadmin/content/publications/2009/tsui-yanco-camera-ready-clinical_validity.pdf.

Valin, J.-M., Michaud, F., & Rouat, J. (2007). Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering. *Robotics and Autonomous Systems Journal*, *55*(3), 216–228, http://dx.doi.org/10.1016/j.robot.2006.08.004.

Valin, J.-M., Yamamoto, S., Rouat, J., Michaud, F., Nakadai, K., & Okuno, H. (2007). Robust recognition of simultaneous speech by a mobile robot. *IEEE Transactions on Robotics*, *23*(4), 742–752, http://dx.doi.org/10.1109/TRO.2007.900612.

Valli, A. (2008). The design of natural interaction. *Multimedia Tools and Applications*, *38*, 295–305, http://dx.doi.org/10.1007/s11042-007-0190-z.

Williamson, M. (1995). *Series elastic actuators*. (Master's thesis, Massachussets Institute of Technology, Cambridge, Boston. Retrieved from http://dspace.mit.edu/handle/1721.1/36966).

Wyrobek, K., Berger, E., Loos, H. Van der, & Salisbury, K. (2008). Towards a personal robotics development platform: Rationale and design of an intrinsically safe personal robot. In *Proceedings of the IEEE International Conference on Robotics and Automation* (pp. 2165–2170). Pasadena, CA: IEEE, http://dx.doi.org/10.1109/ROBOT.2008.4543527.

Authors' name and contact information: François Ferland (francois.ferland@usherbrooke.ca), Dominic Létourneau (dominic.letourneau@usherbrooke.ca), Arnaud Aumont (arnaud.aumont@usherbrooke.ca), Julien Frémy (julien.fremy@usherbrooke.ca), Marc-Antoine Legault (marc-antoine.legault@usherbrooke.ca), François Michaud (francois.michaud@usherbrooke.ca), Department of Electrical Engineering and Computer Engineering, Université de Sherbrooke, 2500, boul. de l'Université, Sherbrooke (Qc) Canada, J1K 2R1. Michel Lauria (michel.lauria@hesge.ch), University of Applied Sciences Western Switzerland (HES-SO), Geneva, Switzerland.