Classifying a Person's Degree of Accessibility from Natural Body Language During Social Human–Robot Interactions

Derek McColl, Member, IEEE, Chuan Jiang, and Goldie Nejat, Member, IEEE

Abstract—For social robots to be successfully integrated and accepted within society, they need to be able to interpret human social cues that are displayed through natural modes of communication. In particular, a key challenge in the design of social robots is developing the robot's ability to recognize a person's affective states (emotions, moods, and attitudes) in order to respond appropriately during social human-robot interactions (HRIs). In this paper, we present and discuss social HRI experiments we have conducted to investigate the development of an accessibilityaware social robot able to autonomously determine a person's degree of accessibility (rapport, openness) toward the robot based on the person's natural static body language. In particular, we present two one-on-one HRI experiments to: 1) determine the performance of our automated system in being able to recognize and classify a person's accessibility levels and 2) investigate how people interact with an accessibility-aware robot which determines its own behaviors based on a person's speech and accessibility levels.

Index Terms—Accessibility-aware behaviors, affect classification, automated body pose recognition, human-robot interaction (HRI), social robots.

I. INTRODUCTION

H UMAN–ROBOT interaction (HRI) involves investigating the design and performance of robots which are used by or work alongside humans [1]. These robots interact through various forms of communication in different real-world environments. Namely, HRI encompasses both physical and social interactions with a robot in a broad range of applications, including cognitive rehabilitation [2],

Manuscript received August 15, 2014; revised July 8, 2015 and December 17, 2015; accepted January 14, 2016. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada, in part by the Canada Research Chairs Program, and in part by the Ontario Graduate Scholarship for Science and Technology.

The authors are with the Autonomous Systems and Biomechatronics Laboratory, Department of Mechanical and Industrial Engineering, University of Toronto, Toronto M5S3G8, ON, Canada (e-mail: derek.mccoll@mail.utoronto.ca; nejat@mie.utoronto.ca). teleoperation of uninhabited air vehicles [3], search and rescue [4], prosthetics [5], and collaborative manipulation tasks [6].

Our own research in this field is centered on the development of human-like social robots with the social functionalities and behavioral norms required to engage humans in natural assistive interactions such as providing: 1) reminders; 2) health monitoring; and 3) cognitive training and social interventions [7]–[10]. In order for these robots to successfully partake in social HRI, they need to be able to recognize human social cues. This can be achieved by perceiving and interpreting the natural communication modes of a human, such as body language, paralanguage (intonation, pitch, and volume of voice), speech and facial expressions.

It has been shown that changes in a person's affect are communicated more effectively with nonverbal behavior than verbal utterances [11]. A significant amount of research has focused on creating automatic systems utilized for identifying affect through paralanguage (see [12]) and facial expressions (see [13]). Our work focuses on recognizing a person's affect through body language. Body language displays are very important for communicating human emotional states [11]. For example, Walters and Walk [14] found that emotion recognition from images of posed static postures with the face and hand expressions obscured is as accurate as emotion recognition of facial expressions. Schouwstra and Hoogstraten [15] conducted a study with stick figures with varying head and spinal positions in which they asked college students to infer emotional states from the positions. Their findings indicate a significant relationship between emotion, and head and spinal positions.

The majority of automated systems that have been developed have primarily focused on classifying a person's affective state from dynamic body gestures, i.e., [16]–[20]. Only a few automatic body language-based affect recognition techniques consider static body poses and postures, i.e., [21]–[24]. For example, in [21], a database of manually segmented joint rotations of individuals playing sports-themed video games was created with a motion capture system. The joint data corresponded to postures representing affect after winning or losing scenarios. A multilayer perceptron was used to recognize four affective states: 1) triumphant; 2) defeated; 3) concentrating; and 4) frustrated. In [22], a recognition system based on facial features obtained from a camera, posture information from a pressure sensing chair, pressure information from a pressure sensitive mouse, and skin conductance from a wireless sensor was able to predict if a child would become frustrated during a problem solving activity on a computer with a 79% recognition rate. In [24], adaptive resonance theory neural networks were used for affective pose recognition via five specific Kinect SDK static skeleton poses for the affective states of frustration, disagreement, confusion, anger, and shyness. The recognized affective states were compared with the affective states determined from verbal information to identify an actor's overall affective state during a drama improvisation scenario. This information was used to determine the behavior of a virtual agent interacting with actors.

Our own research in this area focuses on the development of an automated affect from static body language classification system to be used during social HRI [25], [26]. Bull [27], Coulson [28], Mehrabian [29], and Davis and Hadiks [30], [31] have all determined that static body language is an important source of information for affect. and directly contributes to the understanding of how affect is expressed through body language. Furthermore, changes in static body language can induce changes in affective states [32]. One main advantage to using static body language is that a person usually displays these unconsciously and unintentionally, and therefore, they are natural and not forced. Ekman and Friesen's work [33] found that in communicative situations, body language can be a dominant source of information regarding gross affect states between two interactants. In addition, work by Mehrabian [29] has shown a relationship between the body positioning of a communicator and his/her attitude toward an addressee. Thus, it is important that during social HRI, a robot has the ability to identify and categorize human displays of static body language with the aim of improving engagement during such interaction through its own appropriate display of behaviors.

In this paper, we investigate the integration of our automated affect from body language recognition system for social robotic applications capable of interpreting, classifying, and responding to natural body language during HRI. Our proposed system is capable of 3-D human body language identification and categorization by utilizing the RGB and 3-D data of the Kinect sensory system for segmentation of upper body parts and 3-D pose estimation using a reverse tree structure body model. Once a 3-D body pose has been identified, it is used to classify the person's affect. Herein, affect is determined by an individual's degree of accessibility toward a social robot. A person's degree of accessibility refers to his/her psychological state which includes affect and cognitive states. Namely, accessibility refers to an individual's level of openness and rapport toward another during dyadic social interactions [34]. Previous research has found a significant relationship between an individual's accessibility and his/her body pose [30]. We have developed an automated accessibility recognition system that utilizes and adapts the position accessibility scale of the nonverbal interaction and states analysis (NISA) [30], [34] to identify an individual's degree of accessibility utilizing his/her trunk and arm orientations toward a robot. NISA was originally designed and verified as a manually coded scale to determine a person's degree of accessibility with respect to another person during conversations, interviews, and therapy sessions [30], [31].

Two unique HRI experiments are presented in this paper. The first investigates the performance of our robot integrated automated system in being able to recognize and classify a person's accessibility levels during HRI. Namely, we compare the performance of our system with respect to an existing commercially available body tracking software. The second experiment uniquely investigates how people actually interact with a robot which explicitly uses the identified accessibility levels throughout the social interaction to determine its own behaviors. To do this, we compare our proposed accessibility-aware robot with a nonaccessibility-aware robot, i.e., a robot that does not respond to a user's affective body language.

II. HUMAN BODY LANGUAGE RECOGNITION DURING HRI

In general, body language has been categorized into four distinct classes [35].

- 1) Emblems: Gestures that have a direct verbal translation.
- 2) Illustrators: Movements that are directly tied to speech.
- 3) *Regulators:* Gestures that maintain and regulate a conversation, such as to tell a person to hurry up, repeat, continue, etc.
- 4) *Adaptors:* Body language that conveys emotions or performs bodily actions.

To date, several robots have been developed to understand human emblematic gestures as input commands, i.e., [36]-[38]. For example, in [36], the Jido robot utilized stereo cameras and a multiobject tracking particle filter for tracking a user's head and two-hand pointing gestures, which are used to indicate an object location to the robot. The robot would then pick up an object, place an object in a location or travel to a location. In [37], a 2-D camera, skin-color region extractor, and a hidden Markov model (HMM) were used to recognize 13 arm gestures, including arms up, out, or down, as input commands for a small robot. The robot then mimicked the arm poses by moving its arms in the same manner as those displayed by the person. In [38], a time-of-flight (TOF) camera was used with an HMM to recognize emblematic arm gestures (i.e., one arm up or two arms out) to control the navigation of an iRobot PackBot in order to have the robot follow behind a person or explore its surroundings for a door frame.

A handful of systems have also been developed to identify a person's affective state from body language during HRI [39], [40]. In [39], manually labeled videos of children playing chess with an iCat robot were taken from a 2-D camera in the environment and analyzed after the HRI sessions to determine engagement in the activity. Machine learning techniques trained on geometric features of the torso, e.g., lean angle, slouch factor, were used to identify engagement. In [40], a 2-D color camera, oriented to capture a front view of a person's upper body, was used to determine human affective (happy, angry, sad, or polite) hand movements. Skin color segmentation and geometric region marking were used with motion tracking to determine the Laban movement features of weight, space, flow, and time. The method is proposed for HRI applications. Preliminary experiments, without a robot, showed that sad, happy, and angry hand movements were identified from strumming a guitar.

In general, robots have yet to be developed that directly interpret adaptor style body language for identifying a person's affect during social interactions in order to determine their own expressive assistive behaviors. Since this type of body language is considered key in revealing a person's emotions or attitudes, it would be beneficial for a robot to perceive, interpret, and respond to adaptors while interacting in social HRI to create more engaging interactions.

We aim to develop and integrate a sensory system that allows a social robot to effectively recognize a person's affective nonverbal behaviors during real-time social HRI by autonomously identifying and categorizing a person's adaptor style body language. Through the use of this sensory system a robot will be able to provide task assistance using its own appropriate expressive behaviors in response to a user's affective body language. Our goal is to implement a noncontact body language identification and categorization system capable of determining affect based on a person's upper body language. Body language is defined, in this paper, as static body poses exhibited by an individual during HRI. NISA [34] is utilized to identify an individual's level of accessibility toward a robot based on his/her body language.

This research builds on our previous work [25], [26] which, similar to the aforementioned literature review, has focused on the development of a sensory system for post-analysis of human affect from adaptor style body language. Furthermore, it does not consider the use of human affect to determine a social robot's behaviors during HRI. In [25], a thermal camera and a 3-D TOF sensor were utilized to determine the accessibility levels of a person interacting with a teleoperated social robot. Manually segmented sensor data was used to identify body poses via an ellipsoid model and heuristic rules. Accessibility levels were categorized from these static poses using NISA. The system obtained an accessibility level recognition rate of 78%. In [26], the sensory system in [25] was replaced with the Kinect sensor which provided 2-D and depth images of a person, and sensor data segmentation was automated. Utilizing the new system resulted in an increased accessibility level recognition rate of 86%. However, both sensory systems still required an environment that only consisted of the person interacting with the robot, which is not realistic for many real-world interaction scenarios.

In this paper, we incorporate a robust automated recognition and classification system, using the Kinect sensor, for determining the accessibility levels of a person during oneon-one social HRI. The system can identify the interactant from cluttered realistic environments using a statistical model, and geometric and depth features. Static body poses are then accurately obtained using a learning method. This system is integrated into our socially assistive robot Brian 2.1 (Fig. 1) to allow the human-like robot to uniquely determine its own accessibility-aware behaviors during noncontact one-on-one social interactions in order to provide task assistance to users.



Fig. 1. Socially assistive robot Brian 2.1 and its Kinect sensor.

III. AUTOMATED ACCESSIBILITY FROM BODY LANGUAGE CLASSIFICATION TECHNIQUE

The recognition of body language is challenging as there exists many configurations in a high dimensionality search space. This task is made more difficult when it is intended for a robot that engages in real-time social HRI using only onboard sensors. Herein, we describe our automated accessibility recognition and classification system that identifies a person's static body poses utilizing sensory information from the Kinect sensor. The proposed approach utilizes both a Kinect 2-D color image, to identify exposed skin regions, and Kinect depth data to generate a 3-D ellipsoid model of a person's static pose.

A. Kinect Sensor

Our research presents the first application of the Kinect sensor for human accessibility recognition and categorization during social HRI. The affordable Kinect sensor consists of a 2-D CMOS color camera and a depth imager, both with resolutions of 640×480 pixels. To obtain depth information, a pattern of spots is projected onto a scene using an IR light source and captured with a CMOS IR detector. The depth of a point in the scene is calculated by measuring the horizontal displacement of a spot in the projected pattern [41]. The operating range of the depth sensor is approximately 0.5-4 m. The Kinect sensor was calibrated for this paper utilizing a 3-D checkerboard pattern consisting of the light squares raised with respect to the dark squares. The sensor is incorporated onto the upper torso of Brian 2.1's platform to provide sensory information for identifying a person's static body pose in a noncontact manner (Fig. 1).

B. Human Static Body Poses

A person can display a diverse range of static body poses during interaction. These static poses contain information regarding variations in the person's stress, rapport, involvement, and affective quality and intensity [30]. The poses that are identified in this paper are adapted from the position accessibility scale of the NISA [30]. In order for NISA to consider a pose to be static, it must be held for at least 4 s [34]. Static body positions are an arrangement of trunk orientations and leans, and arm positions which we utilize to identify a person's accessibility level toward a robot (Table I).

TABLE I STATIC BODY POSE ORIENTATIONS AND LEANS

Trunk Orientation	Trunk Angle with Respect to Robot (absolute angle)	
Away (A)	> 15°	
Neutral (N)	$3^{\circ} - 15^{\circ}$	
Toward (T)	$0^{\circ} - 3^{\circ}$	
Trunk Lean	Trunk Position with Respect to Robot	
Upright	Both shoulders are aligned with the hips	
Forward	Both shoulders are shifted in front of the hips	
Right	Right shoulder is tilted to the side past the right hip	
Left	Left shoulder is tilted to the side past the left hip	
Back	Both shoulders are shifted back from the hips	
Arm Orientation	Arm Arrangements with Respect to Robot	
Away (A)	Arms are further away than the upper trunk	
Neutral (N)	Arms are beside the upper trunk	
Toward (T)	Arms are nearer than the upper trunk	

C. Multimodal Static Body Pose Estimation Approach

The architecture of our multimodal static body pose estimation approach is shown in Fig. 2. 2-D images and depth data acquired by the Kinect sensor are used by the human body extraction and body part segmentation modules to first extract a person from the background and then to identify each specific body part. The body parts are used to identify static body poses via the static pose identification module. The reverse tree structure ellipsoid model module then determines the 3-D poses of each of the body parts in these static poses. Lastly, the body pose estimation module determines the orientations and leans of each static body pose.

1) Human Body Extraction and Pose Initialization: We aim to utilize our social robot in a large variety of indoor locations, including large public/semi-public areas, such as retirement homes, office buildings, museums, and shopping malls, which may consist of cluttered interaction environments as well as the potential of having other people located around the interaction scene. In order to extract the Kinect 3-D data of the person interacting with the robot from the scene, we have developed a technique that utilizes a combination of mixture of Gaussians (MOGs) [42], connected component analysis [43], and head and shoulders contours [44]. A statistical model of the environment is generated by creating an MOG for each pixel of a Kinect depth image utilizing multiple training depth images of the scene (without people), prior to the interaction scenario between the person and Brian 2.1. During the oneon-one interactions, pixel values that have a probability of less than 0.1% of belonging to the statistical model of the scene are investigated further with connected component analysis (as they can potentially represent persons in the scene). Groups of pixels, i.e., connected components, that share edges or corners with each other while having similar depth values are identified. A connected component that is able to be fit with a head and shoulders contour is classified as a person. Finally, the person who is closest to the robot during interaction is identified as the current user. This technique of extracting a person from the depth data of the scene is robust to moving objects and people in the background of the scene. Additionally, utilizing the aforementioned calibration technique, the correspondence



Fig. 2. Multimodal static body pose estimation system architecture.

between the depth imager and the 2-D camera is known and hence, background noise can also be removed from the 2-D images, isolating only the user in the 2-D images. Pose initialization is performed utilizing anthropometric information to estimate waist and hip heights and locations utilizing the same technique presented in [25].

2) Body Part Segmentation: For each extracted human body, the head and lower arms are segmented first, followed by the lower and upper trunks and finally, the upper arms. Skin color information from the Kinect 2-D images is utilized to detect the head and lower arms. We choose to use skin color to segment these body parts as they are easily exposed. The lower arms are readily exposed if a person is wearing a short sleeve shirt or can be by rolling up long sleeves to approximately the elbows. This requirement is consistent with other skin tracking systems for robotic applications that also have clothing requirements [36], [37], [45]. An YCbCr pixel-by-pixel color space range technique [46] is utilized to identify skin regions. This technique is robust to varying illuminations and has also been shown to work for a large range of skin colors [46].

Based on the skin color identification results, a binary image is generated to isolate skin regions, i.e., Fig. 3. In general, NISA body poses displayed by a person generate between 1 and 3 skin regions in each binary image.



Fig. 3. Example poses for the five skin region configurations. (a) Head and lower arms. (b) Head and crossed arms. (c) Head and two arms touching. (d) Lower arm and arm touching head. (e) Both lower arms and head touching.

Each skin region can be identified as one of five different lower arm and/or head configurations: 1) head; 2) lower arm; 3) crossed arms; 4) one arm touching the head or two arms touching; and 5) both lower arms touching the head. Five normalized geometric features are identified for each skin region in order to autonomously classify the region into the aforementioned configurations via a learning technique. These features include: 1) the number of pixels within the region; 2) the location of the centroid of the region; 3) the number of pixels along the perimeter of the region; 4) the eccentricity of the skin region; and 5) the expansiveness of the region. Descriptions and formulations for these features are shown in Table II. Regions with less than N_n pixels are considered to be noise and are removed from the binary image.

The five features are then utilized to classify each skin region. The WEKA data mining software [47] was utilized to determine the most appropriate machine learning technique to utilize for classifying head and/or lower arm configurations. A tenfold cross-validation was performed utilizing learning techniques from each of the following classes: 1) probabilistic (e.g., Naïve Bayes); 2) linear (e.g., logistic regression); 3) decision trees (e.g., random forest); 4) lazy learning (e.g., k-nearest neighbor); 5) meta-classifiers (e.g., AdaBoost with base classifiers such as Naïve Bayes and decision stump); 6) neural networks (e.g., multilayer perceptron); and 7) nonlinear models (e.g., support vector machines). The optimal parameters for each learning technique were found utilizing a grid search strategy. The feature vectors used for comparing the techniques were obtained from the skin regions of 300 static poses displayed by 11 different individuals during social HRI experiments. The AdaBoost technique with a Naïve Bayes base classifier [48] had the highest recognition rate of 99.3% and has been implemented in our architecture.

Once all the skin regions have been classified, the regions containing multiple body parts are further segmented to identify individual parts. Namely, crossed arms are segmented along the major axis of an ellipse fit to the skin region, while configurations 4 and 5 are separated into arm and head regions utilizing a Delaunay triangulation technique [25]. Segmentation examples are presented in Fig. 3. Once the lower

TABLE II Normalized Skin Region Features

Feature	Description	Formulation	
Area	Normalized number of pixels contained within a skin region.	$NA_{i} = \frac{N_{i}}{\sum_{i=1}^{R} N_{i}},$ where <i>R</i> is the total number of skin regions, and <i>N_i</i> is the number of pixels in the <i>i</i> th skin region.	(1)
Centroid Location	Normalized (x,y) pixel coordinates of the geometric center of a skin region.	$CL_{i} = \left(\frac{\sum_{j=1}^{N_{i}} x_{i,j}}{T_{x}N_{i}}, \frac{\sum_{j=1}^{N_{i}} y_{i,j}}{T_{y}N_{i}} \right),$ where $x_{i,j}$ and $y_{i,j}$ are the x and y coordinates of the j^{th} pixel of the i^{th} skin region, and T_{x} and T_{y} are the total number of pixels along the width and height of the image respectively.	(2)
Perimeter	Normalized number of pixels along the border of a skin region.	$NP_{i} = \frac{P_{i}}{\sum_{i=1}^{R} P_{i}},$ where P_{i} is the number of pixels along the perimeter of the i^{th} skin region	(3)
Eccentricity	The ellipse-like shape of a skin region.	$E_i = \sqrt{1 - \frac{b^2}{a^2}},$ where <i>a</i> and <i>b</i> are the lengths of the major and minor axes of an ellipse fitted to the <i>i</i> th skin region.	(4)
Expansivenes	Normalized (x,y) pixel coordinates of two opposite vertices s $V_{i,l}$ and $V_{i,2}$, of the skin region's axis- aligned bounding box.	$V_{i,1} = \left(\frac{\min_{j}(x_{i,j})}{T_{x}}, \frac{\min_{j}(y_{i,j})}{T_{y}}\right),$ $V_{i,2} = \left(\frac{\max_{j}(x_{i,j})}{T_{x}}, \frac{\max_{j}(y_{i,j})}{T_{y}}\right)$	(5) (6)

arms and head are identified, the upper arms and lower and upper trunks are identified utilizing the corresponding 3-D and 2-D Kinect data [25].

3) Static Pose Identification: The aforementioned segmentation technique is applied to every tenth frame captured by the 60 Hz Kinect sensor. Bounding boxes are identified around each of the seven identified body parts, and their size and centroids are tracked to determine a static pose (a pose held for at least 4 s). Once a static pose has been recognized, ellipsoids are fit to the segmented 3-D data.

4) Reverse Tree Structure Ellipsoid Model: An iterative moment analysis procedure is utilized to fit ellipsoids to the 3-D data of the segmented body parts [26]. A full 3-D upper body ellipsoid model is created by connecting the seven ellipsoids at specific joints utilizing a reverse tree structure [26]. Once the overall ellipsoid model is generated, the ellipsoid parameters are then used by the body pose estimation module to determine the static body pose orientations and leans.

D. NISA Static Body Pose Classification

The identified static body poses provide input regarding a person's interest in the interaction with a robot as well as his/her openness and involvement. We utilize the position

TABLE III Accessibility Levels

Trunk Orientation	Accessibility Level	Arm Orientation	Finer- Scaling
Lower/Upper trunk: N/T or		Т	12
T/N combined with forward or upright leans T/T with all	IV	Ν	11
possible leans		А	10
Lower/Upper trunk: N/T or		Т	9
T/N except positions with	III	Ν	8
forward or upright leans		А	7
Lower/Upper trunk: N/N,		Т	6
N/A, A/N , A/T , T/A , with all	II	Ν	5
possible leans		А	4
L		Т	3
with all possible loops	Ι	Ν	2
with all possible leans		А	1

accessibility scale to determine the static body poses of a person, as defined by trunk and arm configurations, in relation to the robot's position. This information is then used with the position accessibility scale of NISA to identify a person's accessibility level toward the robot. NISA states that the larger and more central a change in pose is, the more pivotal it is in representing a change in the interaction [34]. Since positions are important markers of naturalistic behaviors, the orientation of a static body pose of one person relative to another person is linked to his/her degree of psychological openness, rapport, and emotional involvement. Table III presents the static body pose accessibility classification as a function of the trunk and arm patterns.

The position accessibility scale is comprised of four distinct levels, ranging from level I (least accessible) to level IV (most accessible). Each level is characterized by the orientation patterns [away (A), neutral (N), or toward (T)] of the lower and upper trunks and the trunk lean direction (i.e., forward, upright, left, right, and back) with respect to the robot. Each level is divided into three sublevels utilizing the A, N, or T arm orientations as defined in Table I. The finer position scaling for the arm orientations is coded on a 12-point scale with respect to the trunk orientations, where 1 represents least accessible and 12 is most frontally oriented and toward.

IV. PERFORMANCE COMPARISON STUDY

One-on-one HRI experiments with Brian 2.1 were performed in an office setting to determine the performance of our multimodal static body pose estimation approach in identifying a person's accessibility levels toward the robot.

A. Design

Eighteen participants, aged 19 to 35 ($\mu = 24$, $\sigma = 5.33$) participated in the study. During the interactions, a human operator used the Wizard of Oz technique to teleoperate Brian 2.1 from a remote location away from the interaction scene. The operator controlled both the verbal and nonverbal (gestures and facial expressions) interaction capabilities of the robot in real-time. Each participant interacted with the robot in four different interaction stages: 1) introduction stage, where the robot would introduce itself to the participant;



Fig. 4. Example trunk orientations/leans and arm orientations using joint locations provided by the Kinect body pose estimation technique. (a) Trunks: A, arms: N. (b) Trunks: T with a sideways lean, arms: N. (c) Trunks: T, arms: T.

2) instruction stage, where the robot provided the instructions to assemble a picnic table; 3) memory stage, where the robot engaged the participant in a memory game activity; and 4) repetitive stage, where the robot repeated the same behavior for 5 min. Participants were not directed to display any particular body poses while interacting with Brian 2.1. Each participant naturally implemented various static body poses for recognition and classification into accessibility levels toward Brian 2.1. Sensory information from the Kinect sensor was analyzed using the proposed automated accessibility classification system.

The overall performance of the proposed system is determined by comparing the identified static body poses to the poses identified from the Kinect SDK [49]. Furthermore, the accessibility levels obtained from our proposed approach are also compared to the accessibility levels of the poses obtained using the Kinect SDK. The baseline for the aforementioned comparison was obtained from assessments by an expert coder trained in NISA.

B. Kinect SDK Body Pose Estimation Approach

The Kinect SDK utilizes a random decision forest and local mode finding to generate joint locations of up to two people from depth images [50]. The person closest to the robot is identified as the user. We have developed a technique to identify the static body pose orientations and leans in order to determine accessibility levels from the Kinect SDK joint locations during social HRI. To do this, the upper trunk is defined as the plane formed by connecting the points corresponding to the joints of the right and left shoulders and the spine (middle of the trunk along the back). The lower trunk is identified as the plane formed by connecting the points of the left, right, and hip center joints. The lower and upper trunks are shown in Fig. 4. The relative angle between the normal of each plane and the Kinect camera axis is then used to determine an individual's lower and upper trunk orientations with respect to the robot. The position of the left and right shoulders relative to the left and right hips and the angle between the normals of the planes are used to determine the lean of the trunk.

The arm orientations are determined by the relative distances between the lower arms and the upper trunk. Namely, when the average distance of the lower arms is closer to/further from the robot than the upper trunk, the arms are classified



Fig. 5. Example static poses. (a) Accessibility level IV pose. (b) Accessibility level III pose. (c) Accessibility level II pose. (d) Accessibility level I pose.

as T/A. When the lower arms have the same average distance from the robot compared to the upper trunk, the arms are classified as N. Once the trunk and arm configurations are determined, NISA is utilized to identify a person's degree of accessibility with respect to the robot. Examples of trunk/arm orientations and trunk leans are shown in Fig. 4.

C. Accessibility Baseline Coding

To investigate the reliability of both body pose estimation techniques in an HRI setting, an expert trained in NISA coded the accessibility levels of the identified static body poses. The coder was provided with a 2-D image from the Kinect 2-D color camera of each static pose during the HRI experiments. The expert coder then identified both the accessibility level and finer-scaling level for each static pose.

D. Results and Discussions

Overall, the participants displayed 223 different static poses during the experiments. Fig. 5 shows four example poses displayed during the aforementioned interaction experiments with Brian 2.1. Columns (i) and (ii) present the 2-D color and 3-D data of the segmented static poses. The body part segmentation results and the corresponding ellipsoid models obtained from the multimodal pose estimation are shown in columns (iii) and (iv). Last, the multi-joint models obtained from the Kinect SDK body pose estimation approach are presented in column (v).

The poses in Fig. 5 consist of the following: 1) hands touching in front of the trunk with the upper trunk in a toward position and the lower trunk in a neutral position; 2) one arm touching the other arm which is touching the head while leaning to the side with the upper trunk in a neutral position and the lower trunk in a toward position; 3) arms crossed in front

TABLE IV ACCESSIBILITY LEVEL RESULTS

	Multimodal Static Body Pose Estimation Approach		Kinect Body Pe Estimation Tec	ose hnique
Body	Accessibility	Finer-	Accessibility	Finer-
Pose Fig.	Level	Scaling	Level	Scaling
5(a)	IV	12	IV	11
5(b)	III	9	II	6
5(c)	II	6	II	6
5(d)	Ι	2	Ι	2

of the trunk with both trunks in neutral positions; and 4) arms at the sides with both trunks in away positions. The accessibility levels of these poses based on the estimation approaches are presented in Table IV.

Overall, for the multimodal static body pose estimation approach, the ellipsoid body models created using the 3-D and 2-D sensory data from the Kinect sensor very closely matched the participants' static body poses exhibited during the interactions. This is observed by comparing the ellipsoid models to the 2-D images and 3-D depth information in Fig. 5. The average processing time for recognition and classification of the body poses was 89 ms: 53 ms for body part extraction and segmentation, 29 ms for ellipsoid model fitting, and 7 ms for accessibility level classification. It should be noted that variations in the skin colors of the different participants did not influence body part segmentation or ellipsoid model generation. For the multimodal pose estimation approach, occluded body parts were estimated by utilizing the ellipsoid parameters of an occluded part from previous frames as well as the current ellipsoid locations and parameters of adjoining body parts. Blue ellipsoids indicate occluded body parts in Fig. 5. In Fig. 5(d), the blue ellipsoids represent the upper and lower right arms. Parameters of ellipsoids representing the same body parts can change between static poses due to the indirect ellipsoid model approach, namely, the parameters for each ellipsoid are reformulated for every new pose. During the one-on-one HRI experiments, a participant's sleeves would occasionally slide up and down his/her arms, resulting in the multimodal technique segmenting shorter or longer arm ellipsoids. However, this change in sleeve length did not influence any finer scaling accessibility classification results. A small change in sleeve length will also result in a small change in the size of the resulting lower arm ellipsoid and an even smaller change to its centroid position used to identify arm orientations. The latter change is in the same order of magnitude as the depth resolution of the Kinect sensor [51], which has an average resolution of 0.7 cm over the interaction distances of 1.2 to 1.8 m of the participants.

It can be seen from Fig. 5 that with these body poses, the Kinect SDK body pose estimation does not accurately identify the correct poses of the arms. For example, in the Kinect 3-D multi-joint body model of Fig. 5(a) the hands are not clasped in the multi-joint model. In Fig. 5(b), the right arm is not touching the left arm and the left arm is not touching the head and in Fig. 5(c) the participant's arms are not crossed. The random

TABLE V Performance Comparison Statistics

	Coder		
Technique	Accessibility Level Matches	Finer-Scaling Matches	
Multimodal static body pose estimation technique	88%	86%	
Kinect body pose estimation technique	63%	57%	

decision forest used by the Kinect body pose estimation algorithm was trained on over one million sample images; hence, it is dependent on a finite number of training images [50]. It is not possible for a finite training set to include all possible poses and body shapes of all individuals. Additionally, it has not been designed specifically for body language recognition, but rather entertainment scenarios [50]. Hence, we postulate that the pose errors identified above were due to these factors. Although the multimodal static body pose estimation approach requires an initialization pose, the Kinect body pose estimation technique currently needs both the head and shoulders to be visible with the elbows at a lower height than the shoulders during initialization in order to create the necessary body contour to isolate a participant from the background 3-D data, allowing for multiple initial poses.

1) Classification Comparison: The expert coder's ratings of accessibility levels were then compared to the results obtained from the ellipsoid model of the multimodal technique and those obtained from the Kinect 3-D multi-joint body model (Table V). Our own multimodal pose estimation approach had classification rates of 88% for the overall accessibility levels and 86% for the finer-scaling coding with respect to the coder, while the Kinect body pose estimation technique had only 63% and 57%, respectively. The main reason the Kinect body pose estimation approach had lower classification rates was that it could not easily distinguish between body parts in the depth data when the arms were in contact with other body parts.

The strength of agreement between the accessibility levels obtained by the expert coder and the two pose estimation techniques was measured by applying Cohen's kappa to all 223 poses. Cohen's kappa was determined to be 0.78 for the multimodal approach, which characterizes the strength of agreement to be substantial, and kappa was 0.31 for the Kinect body pose estimation approach which has a fair strength of agreement [52].

V. ACCESSIBILITY-AWARE INTERACTION STUDY

The objective of the second set of social HRI experiments was to investigate users' accessibility levels as related to Brian 2.1's behaviors during assistive scenarios between a person and the robot. We compare two robot behavior types to determine if an accessibility-aware emotionally responding robot influences the overall interactions with individuals: 1) the robot determines its assistive behaviors based on the state of the activity, herein defined as the nonaccessibilityaware behavior type and 2) the robot determines its assistive behaviors based on the accessibility level of the person as well as the state of the activity, herein defined as the accessibilityaware behavior type. We used the multimodal pose estimation technique to identify participant accessibility levels during HRI due to its aforementioned higher performance results for the proposed application. These experiments were conducted using two assistive scenarios: 1) robot tutor (RT) and 2) robot restaurant finder (RRF). The RT scenario consisted of Brian 2.1 engaging a participant in memory and logic games. In the RRF scenario, the robot assisted a person in choosing a restaurant to go to for dinner.

A. Participants

Twenty-four participants, ranging in age from 20–35 ($\mu = 24.7$, $\sigma = 4.4$), participated in the study. Each participant interacted with the robot twice, once with each behavior type with one week between interactions. During each interaction Brian 2.1 would perform both assistive scenarios. Participants were not informed that the robot would have different capabilities during each interaction. A counterbalanced design was used where half the participants interacted with the accessibility-aware robot first, while the others interacted with the nonaccessibility-aware robot first. The order of assistive scenarios (RT or RRF) was also counterbalanced.

B. Interaction Scenarios

1) Robot Tutor Interaction: The RT interaction was designed as a cognitively stimulating activity to encourage logical thinking and to practice language and math skills. The interaction consisted of four main stages: 1) greeting/introduction; 2) a double letter word game; 3) logic questions; and 4) a word linking game. During the greeting/introduction stage, the robot introduced itself, the purpose of the interaction and its intended functionality as a social motivator for the interaction. During the double letter word game, the robot asked the participants to come up with two related words, one of which needs to have two consecutive identical letters. The participant and robot took turns playing this game by finding appropriate pairs of words. The robot would start the game by explaining how to play and also provide an example, i.e., apples and oranges.

The logic questions were designed to test a participant's ability to extract meaning from complex information. The robot asked three logic questions. An example logic question was "What is the number of degrees between the hands of an analog clock pointing at 6:30?" The final stage of the RT interaction was a word linking game, where the robot would ask a participant to pick any starting word and then the robot would respond with a word that starts with the last letter of that word. This sequence would be repeated between the robot and participant. The overall interaction finished with Brian 2.1 informing the user that all the games were finished.

2) Robot Restaurant Finder Interaction: The RRF interaction consisted of the robot assisting a participant in identifying and locating a suitable restaurant for dinner based on a participant's preferences. This interaction had four main stages: 1) greeting; 2) information gathering; 3) restaurant suggestion; and 4) providing directions. Similar to the RT interaction, Brian 2.1 would greet a participant and explain the objective of the interaction.

The information gathering stage consisted of the robot asking a participant his/her preferences with respect to the type of food he/she would want to eat, possible restaurant locations based on a list of local areas, and the price range for the meal. Utilizing these preferences, the robot would choose a potential restaurant (obtained from Urbanspoon.com). If the participant did not want to go to that particular restaurant, the robot would then suggest alternative restaurants. With a restaurant chosen, the robot would offer to provide directions to the restaurant.

C. Robot Behavior Design

The robot autonomously implemented its behavior types (using a combination of verbal and nonverbal modes) using finite state machines (FSMs). The input into the FSM from the participant for the robot's nonaccessibility-aware behavior was speech. Speech and accessibility levels were both used as inputs into the FSM for the robot's accessibility-aware behavior. An operator was utilized only for speech recognition during HRI in order to minimize reliability issues of current speech recognition software. The use of an operator for speech recognition is a commonly used approach in social HRI research, i.e., [53], [54]. A microphone placed on the robot supplied audio output to the operator, who was located at a remote location, away from the robot and participant. The FSMs for both behavior types then autonomously determined the robot's behavior based on the current state of the interaction scenario and the inputs from the participants.

A participant's verbal responses to Brian 2.1's behaviors are categorized as positive, negative, and no response. Positive responses include providing a correct answer during the RT interaction or providing the necessary information for the robot to select an appropriate restaurant during the RRF interaction. Negative responses include providing incorrect answers for the RT interaction and not providing the robot with the information needed during the RRF interaction.

1) Brian 2.1's Nonaccessibility-Aware Behaviors: With respect to Brian 2.1's nonaccessibility-aware behaviors, the robot replies to positive responses during the RT and the RRF interactions by verbally acknowledging the responses. For example, during the RT interaction one reply to a positive response is "Yes, that answer is correct." During the RRF interaction the robot replies to a positive response by repeating and confirming the information provided by the user. A negative response from a participant results in the robot providing assistance utilizing instructor error-correction techniques [55] in order for the participant to identify a positive response. This is achieved by giving an example answer during the RT interaction or by restating the question during the RRF interaction. To re-engage a participant who did not respond to the robot, for both interaction types, Brian 2.1 asks the participant if he/she would like it to repeat its previous statement. The behaviors

TABLE VI Nonaccessibility-Aware Robot Behaviors

User Speech Input	Robot Behavior	Example Speech
Positive Response	Verbally acknowledges positive response.	"Yes that answer is correct."
Negative Response	Provides verbal assistance.	"I am not familiar with that type of food. How about pasta?"
No Response	Offers to repeat its last statement.	"I can repeat the rules for you."



Fig. 6. Brian 2.1 providing verbal assistance while swaying its trunk during the nonaccessibility-aware behavior type.

of Brian 2.1 are displayed with a neutral facial expression and tone of voice; while the robot repetitively sways its torso from side to side. To initiate each interaction, Brian 2.1 greets a user by saying hello to the user by name. To end the interaction, Brian informs the user that the tasks have been completed and says goodbye. The robot's nonaccessibility-aware behaviors are summarized in Table VI. Fig. 6 shows a visual example of this robot behavior type.

2) Brian 2.1's Accessibility-Aware Behaviors: The goal of the accessibility-aware interactions is to detect a person's accessibility levels toward Brian 2.1 and utilize emotional robot behaviors to keep this person engaged and accessible to the robot, while also promoting desired responses from him/her. The robot reinforces high levels of accessibility by displaying positive valence emotional states and it decreases its level of displayed valence as participant accessibility levels also decrease. Responding to a participant's affective state with congruent emotional behaviors communicates empathy toward the participant [56], which is important for building rapport and trust between communicators [57]. Emotional behaviors displayed by the robot during social interaction can also improve user engagement [58] and affect [59], [60] as well as encourage correct responses during learning scenarios [61]. Namely, Brian 2.1 displays emotions with high positive valence for accessibility level IV, positive valence for level III, neutral valence for level II, and negative valence for level I.

Brian 2.1 displays high positive valence with a happy tone of voice and an open mouth smile. The happy voice is characterized by its faster speed and higher pitch compared to the neutral voice used by Brian 2.1 during its nonaccessibility-aware behaviors. An open mouth smile is used as it distinguishably conveys increased positive valence as compared to a closed mouth smile [62]. The robot displays positive valence using a closed mouth smile and a happy tone of voice. Neutral valence is simply displayed utilizing



Fig. 7. Brian 2.1's facial expressions. (a) High positive valence. (b) Positive valence. (c) Neutral valence. (d) Negative valence.

a combination of a neutral facial expression and tone of voice. Negative valence is displayed by Brian 2.1 using a sad facial expression and tone of voice, where the latter has a slower speed and lower pitch than the robot's neutral voice. Examples of Brian 2.1's facial expressions are shown in Fig. 7.

When a participant is in accessibility level IV, during the RT interaction, the robot encourages a positive (correct) response by verbally congratulating him/her, enthusiastically nodding its head and clapping its hands while displaying high positive valence [Fig. 8(a)]. Such behaviors have all been shown to convey positive emotions [63], [64] and positively reinforce desired behaviors in others [65]. During the RRF interaction, Brian 2.1 verbally acknowledges a positive response with high positive valence while nodding its head enthusiastically [Fig. 8(b)]. For a negative response, during both scenarios, the robot displays high positive valence while thanking the participant for responding and then offering assistance in order for the participant to state a positive response. When the participant does not respond, Brian 2.1 displays high positive valence while waiting for a response and offers to repeat its last statement. When the accessibility level is lower, the robot also adapts its valence and behaviors with respect to a person's positive, negative, or no response behavior. For participant behaviors displayed in accessibility level III, the robot displays positive valence without the nodding or clapping gestures. Removing such nonverbal gestures reduces the level of positive reinforcement [66].

When a participant is in accessibility level I, Brian 2.1 displays negative valence and waves its arm in a beckoning gesture [Fig. 8(c)]. The combination of the beckoning gesture and the sad facial expression is used to get the person's attention [67] and evoke sympathy which motivates a person to help the robot [68], [69]. In this scenario, this corresponds to engaging with the robot in order to respond to the robot's questions. When the participant is in accessibility level II, the robot responds to participant statements in a neutral emotional state. This is motivated by the fact that displays of neutral behaviors are neither reinforcing nor punishing with respect to another person's behaviors [70]. The robot does not respond to accessibility level II with negative valence behaviors as the user is somewhat accessible to the interaction (i.e., his/her accessibility is higher than level I). Furthermore, it does not respond with positive valence emotional behaviors, as previously mentioned, as it utilizes these to reinforce the more accessible levels (levels III and IV) of the user with respect to





(b)



(c)



(d)



Fig. 8. Example accessibility-aware robot behaviors. Brian 2.1 displaying (a) high positive valence while congratulating a user and clapping, (b) high positive valence while acknowledging a positive response and nodding, (c) negative valence while offering assistance and using a beckoning gesture, (d) positive valence while telling a joke and giggling, and (e) positive valence while saying goodbye and waving.

the robot. Overall, the robot behaviors for responding to each accessibility level are utilized to promote higher accessibility levels of the user toward the robot. In interactions where a user does not display any static poses, the robot will use its neutral valence behaviors, similar to the nonaccessibility-aware robot behaviors.

To begin each interaction with the accessibility-aware robot, Brian 2.1 greets the user while displaying positive valence by waving and saying hello to the user by name and telling a joke. We utilize humor, herein, in addition to the emotional

TABLE VII Accessibility-Aware Robot Behaviors

User Input				
Acces- sibility Level	Speech Input	Robot Example Spee Behavior		
	Positive Response	Acknowledges/ Congratulates with high positive valence, head nodding and/or clapping.	"Italian food sounds great!"	
IV	Negative Response	Thanks participant for responding and offers assistance with high positive valence.	"Thanks for the information, but I am not familiar with that location. How about a place downtown?"	
	No Response	Waits 10 seconds for response from participant then offers to repeat its last statement with a high positive valence.	"I can repeat the rules of the game for you if you would like."	
III N	Positive Response	Acknowledges/ Congratulates with positive valence.	"Great job! How about we try a different game now."	
	Negative Response	Thanks participant for responding and offers assistance with positive valence.	"Thanks for responding, but I am not familiar with that kind of food. Would you like to eat pasta?"	
	No Response	Offers to repeat its last statement with positive valence.	"Why don't I repeat the rules for you?"	
	Positive Response	Acknowledges/ Congratulates with neutral valence.	"Good job. Let's move on to the next game."	
II	Negative Response	Offers assistance with neutral valence.	"That answer is not correct, another example is apples but not oranges."	
	No Response	Offers to repeat its last statement with neutral valence.	"Why don't I repeat the last question?"	
	Positive Response	Acknowledges response with negative valence and a beckoning gesture.	"Sushi it is!"	
Ι	Negative Response	Offers assistance with negative valence and a beckoning gesture.	"That is incorrect. Please try to find how many degrees are between two numbers on a clock."	
	No Response	Repeats previous statement with negative valence and a beckoning gesture.	"Excuse me, did you hear the question? Do you need directions to the restaurant?"	

displays, to promote emotional engagement during the interaction. Previous research with automated dialogue systems have shown greater emotional bonds were generated between users and a system that told jokes [71]. Telling jokes has also been shown to improve users' task enjoyment during HRI [72]. After delivering the punchline of the joke, the robot lifts its hand to cover its mouth while giggling, as seen in Fig. 8(d). At the end of the interaction, Brian 2.1, while displaying positive valence, waves goodbye and thanks the user for participating [Fig. 8(e)]. A summary of the robot's accessibility-aware behaviors, based on verbal responses and accessibility levels of participants are presented in Table VII. A video featuring examples of accessibility-aware robot behaviors can be found at http://asblab.mie.utoronto.ca/research-areas/socialand-personal-robots.

3) Post-Interaction Questionnaire: After each interaction scenario with the robot, the participants completed a questionnaire about the robot. The questionnaire incorporates the constructs from the social behavior questionnaire (SBQ) [73]. The SBQ was developed specifically to measure user perceptions of a robot's social intelligence with varying types of social behaviors [73]. Cronbach's alpha has determined the SBQ to be 0.7–0.9 [73], which is defined as substantial to excellent. The validity of the scale has been verified by its ability to obtain statistically significant results, p < 0.05, indicating that participants give socially intelligent agents significantly higher ratings for all the constructs of the SBQ than nonsocially intelligent agents [73]. The constructs used in our question-naire include: altruism, assertiveness, competence, dutifulness, empathy, helpfulness, modesty, responsibility, sociability, sympathy, and trust. The detailed questions that we have used for these constructs are provided as supplementary material. Responses to the questionnaire were obtained by each participant indicating his/her agreement with each statement using a five-point Likert scale (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, and 5 = strongly agree).

D. Results and Discussion

The results of the interaction experiments were analyzed to determine the performance of the automated accessibility classification system as well as the influence of the two robot behavior types on the accessibility levels of the participants. Questionnaire responses were also utilized to determine if the participants perceived one of Brian 2.1's behavior types to be more socially intelligent than the other.

1) Accessibility Classification: We compared the most frequent accessibility levels of the participants obtained by the robot during each stage of the interactions for both behavior types with a self-study report from the participants. The comparison was used to analyze the performance of the robot's ability to detect the participants' accessibility levels during HRI. For the self-study, each participant, via playback video, was asked during each of the stages of interaction to identify if he/she was either feeling open to the interaction with the robot, somewhat open or not open to the interaction, where openness is defined by his/her level of comfort and engagement. A three level scale was created to correlate these three levels of openness to the accessibility levels of NISA. Level 1 of the self-study was associated with accessibility level I of NISA. Level 2 of the self-study was associated with levels II and III of NISA. Level 3 of the self-study was associated with level IV of NISA. This three level scale was utilized because the participants themselves were not knowledgeable of NISA or how it classifies accessibility levels and hence, it would be difficult for them as untrained users to distinguish between accessibility levels II and III.

Overall the multimodal static body pose estimation and accessibility classification system appropriately matched 75% of the self-reported levels for all the interactions for both behavior types. Namely, 73% of the self-reported level 3 ratings were matched with NISA accessibility level IV classifications of the automated system. No poses during these interactions were classified as NISA accessibility level III. Eighty-five percent of the self-reported level 2 ratings were matched with NISA accessibility level 1 ratings were matched with NISA accessibility level 1 from the automated system. Forty-eight percent of the self-reported level 1 ratings were matched with NISA accessibility level 1 from the automated system. It should be noted that overall only seven participants self-reported a small number of their poses as level 1. The poses that were not identified as NISA accessibility

Accessibility Level	Number of Poses for Nonaccessibility-Aware Behavior Type	Number of Poses for Accessibility-Aware Behavior Type
IV	210 (29.0%)	401 (52.1%)
III	0 (0%)	0 (0%)
П	477 (65.9%)	348 (45.2%)
I	37 (5.1%)	21 (2.7%)
Total	724 (100%)	770 (100%)

TABLE VIII Participant Accessibility Levels

level I by the automated system were instead classified as level II. Further investigation of these latter level 1 poses found that the majority of them included neutral or toward lower and upper trunk orientations with crossed arms. NISA identifies these poses as higher accessibility levels due to the importance of the trunk orientations over the finer-scaling arm orientations. A two-tailed Wilcoxon signed rank test showed that no statistically significant difference exists between the accessibility levels of the automated system and the openness levels of the self-study report, z = 0.393 and p = 0.695.

2) Comparison of Robot Behavior Types: In total, 1494 different static poses were obtained and classified by the robot during the interactions using the multimodal pose estimation technique, with 724 poses obtained during the nonaccessibility-aware interactions and 770 poses obtained during the accessibility-aware interactions. Static poses were obtained for every participant during both types of interactions. Table VIII summarizes the number of static poses identified for each accessibility level and robot behavior type. For the nonaccessibility-aware robot interactions, 29.0% of the poses were classified as accessibility level IV, 0% as level III, 65.9% as level II, and 5.1% as level I. Whereas for the accessibilityaware robot interactions, 52.1% of the poses were classified as level IV, 0% as level III, 45.2% as level II, and 2.7% as level I. On average, the participants interacted for 11 min with the nonaccessibility-aware robot (6 min during the RT interaction and 5 min during RRF interaction) and 12 min with the accessibility-aware robot (7 min during the RT interaction and 5 min during RRF interaction).

We hypothesized that the participants' accessibility levels would be higher during interactions with the accessibilityaware robot than during interactions with the nonaccessibility aware robot. A two-tailed Wilcoxon signed rank test was utilized to test this hypothesis. The results showed that the accessibility levels of the participants were statistically higher during interactions with the accessibility-aware robot, z = 4.0, p < 0.001. Sixteen participants had a most frequent accessibility level of II when interacting with the nonaccessibility-aware robot, however, when they interacted with the accessibilityaware robot, they had a most frequent accessibility level of IV. Seven participants had the same most frequent accessibility level of II and one participant had the same most frequent accessibility level of IV for both robot behavior types. These results show that, in general, the participants were more accessible toward the social robot when it had the capability to both recognize and respond to their accessibility levels.

TABLE IX MEAN QUESTIONNAIRE CONSTRUCT RESULTS

Construct	Mean Rating for Nonaccessibility- Aware Behavior (Cronbach's Alpha)	Mean Rating for Accessibility-Aware Behavior (Cronbach's Alpha)	p value for 2-Tailed Wilcoxon Signed Rank Test
Altruism	3.8 (0.8)	4.4 (0.8)	0.010
Assertiveness	4.0 (0.7)	3.7 (0.7)	0.198
Competence	4.0 (0.7)	4.0 (0.7)	0.708
Dutifulness	3.6 (0.2)	3.9 (0.6)	NA
Empathy	3.3 (0.7)	3.9 (0.6)	0.004
Helpfulness	3.1 (0.7)	3.9 (0.6)	0.009
Modesty	4.0 (0.9)	4.2 (0.8)	0.333
Responsibility	4.0 (0.7)	4.2 (0.6)	0.267
Sociability	3.7 (0.7)	4.0 (0.7)	0.175
Sympathy	3.9 (0.7)	4.1 (0.8)	0.338
Trust	3.8 (0.7)	4.1 (0.8)	0.070
Overall	3.7 (0.9)	4.1 (0.9)	< 0.001

NA: Not Applicable due to low inter-reliability for the construct

3) Questionnaire Results: A summary of the mean participant ratings for the constructs of the post-interaction questionnaire are presented in Table IX. The inter-reliability of the statements in each construct were also calculated utilizing Cronbach's alpha. Construct reliability was improved by removing statistically weak statements [74]. All the constructs obtained alpha values of 0.6 or higher except for Dutifulness, which had an alpha value of 0.2 for the nonaccessibility-aware robot behavior type (Table IX). Therefore, this construct was removed from further analysis. Alpha values of 0.6 or higher are acceptable for constructs with a small number of items, i.e., 2 or 3 [75].

A Wilcoxon signed rank test was conducted to compare the overall results for the two robot behavior types. The results showed that the accessibility-aware robot behavior type was perceived to be significantly more socially intelligent than the nonaccessibility-aware robot behavior type, z = 4.332, p < 0.001. This result is similar to the study conducted by de Ruyter *et al.* [73] that found participants perceived a teleoperated iCat robot with social etiquette to be more socially intelligent than when the robot was socially neutral.

It is interesting to note that the competence and assertiveness constructs had the same or slightly higher mean ratings for the nonaccessibility-aware behavior type when compared to the accessibility-aware behavior type. With respect to competence, the same mean rating may have been obtained, since for both behavior types the robot had the knowledge to complete the necessary interaction tasks, which is an indicator of competence [66]. Namely, the robot was always able to identify correct or incorrect participant responses to questions during the RT interaction and find a restaurant during the RRF interaction. Assertiveness may have been rated a bit lower for the accessibility-aware behavior type due to it displaying more body movements/gestures. In [76], it was found that an increased amount of body movements was an indicator of nonassertiveness during human-human social interaction. However, in general, assertiveness is linked to having the capability to express emotions and recognize an interaction partner's affective state [65].

As evident from the questionnaire results for both the accessibility-aware robot and the nonaccessibility-aware robot behaviors, the participants rated the robot behaviors as either neutral or positive, and did not have negative attitudes toward the robot. We postulate that this supports the lack of accessibility level III poses identified during the experiments. Namely, backwards leans and leans away from an interactant (the robot in our case) have been found as indicators of a negative attitude toward the interactant [27]–[29].

4) Discussion: People involved in social interactions generally display a series of static poses defined as resting poses which can define natural units of behaviors, affect, and rapport. In order to be considered a resting pose, a person needs to hold the pose. Through various clinical and research observations by Davis and Hadiks [30] and Davis [34] this has been defined to be at least 4 s. This was also verified in the presented experiments, where participants assumed 1494 different static poses during interactions with Brian 2.1. In future work, these static poses that represent a person's accessibility level can be combined with dynamic arm and hand gestures in order to determine other affective states that may be present during HRI.

The presented human body pose identification technique utilizes skin color information and 3-D data of a person to generate an indirect ellipsoid model. Namely, a new ellipsoid model is created for each new pose. This technique allows the size and shape of the ellipsoid model to accurately estimate the poses of people of various sizes and shapes automatically, without relying on large amounts of training data. Even though the technique requires that the lower arms of a user be exposed, none of the participants commented on this constraint as a limitation for their interaction. As an alternative approach, future work could consider generating 3-D human kinematics models (see [77]), with the appropriate body part centroids and joints defined to determine accessibility.

During these experiments, the participants stood approximately 1.2–1.8 m from the robot while interacting with it. The robot is capable of identifying each participant's distance utilizing the Kinect sensor. This was within our sensing technique range of 1.0–3.5 m and also consistent with the social distance determined for interpersonal one-on-one communication by Hall [78] in his work on proxemics. If Brian 2.1 is mounted on a mobile platform, it can also actively maintain the distance range for social interaction.

The scenarios presented in this paper are specifically designed for one-on-one social human-robot interaction with a static robot. Hence, the presented system only identifies the closest person as a user. Brian 2.1 can utilize the proposed automated static body language identification and classification system for a number of social interaction scenarios in which the robot can provide information to individuals such as at a help desk at a library, shopping mall, or museum, or at a reception desk in an office building. The robot can also be used in long-term care facilities to assist with activities of daily living, schools as a tutor, and private homes for various information providing and reminders tasks.

The human identification technique proposed herein, which uses the MOG, connected component, and head and shoulders contour technique for identifying people in a scene can be used to identify the static body language of multiple users. It can easily be extended to more than one person by detecting if multiple people are within a certain interaction distance from the robot. The connected component analysis and head and shoulders contour technique can be used to identify multiple people within this distance. Then microphone arrays can be used to localize which user is speaking [79]. Furthermore, the technique can deal with slowly changing background environments, since the MOG model is updated iteratively.

VI. CONCLUSION

In this paper, we implemented the first automated static body language identification and categorization system for designing an accessibility-aware robot that can identify and adapt its own behaviors to the accessibility levels of a person during one-on-one social HRI. We presented two sets of social HRI experiments. The first consisted of a performance comparison study which showed that our multimodal static body pose estimation approach is more robust and accurate in identifying a person's accessibility levels over a system which utilized the Kinect SDK joint locations. The second set of HRI experiments investigated how individuals interact with an accessibility-aware social robot, which determines its own behaviors based on the accessibility levels of a user toward the robot. The results indicated that the participants were more accessible toward an accessibility-aware robot over a nonaccessibility aware robot, and perceived the former to be more socially intelligent. Overall, our results show the potential of integrating an accessibility identification and categorization system into a social robot, allowing the robot to interpret, classify, and respond to adaptor style body language during social interactions.

Our results motivate future work to extend our technique to scenarios which may include interactions with more than one person and when individuals are sitting. Furthermore, we will consider extending the current system to an affectaware system which will consider the fusion of other modes of communication in addition to static body language, such as for example, head pose and facial expression as investigated in [80] as well as dynamic gestures.

ACKNOWLEDGMENT

The authors would like to thank A. Hong for his assistance with the experiments.

References

- M. A. Goodrich and A. C. Schultz, "Human-robot interaction," *Found. Trends Human Comput. Interact.*, vol. 1, no. 3, pp. 203–275, 2007.
- [2] A. Tapus, C. Tapus, and M. J. Mataric, "Hands-off therapist robot behavior adaptation to user personality for post-stroke rehabilitation therapy," in *Proc. IEEE Int. Conf. Robot. Autom.*, Rome, Italy, 2007, pp. 1547–1553.
- [3] H. I. Son *et al.*, "Human-centered design and evaluation of haptic cueing for teleoperation of multiple mobile robots," *IEEE Trans. Cybern.*, vol. 43, no. 2, pp. 597–609, Apr. 2013.
- [4] B. Doroodgar, Y. Liu, and G. Nejat, "A learning-based semi-autonomous controller for robotic exploration of unknown disaster scenes while searching for victims," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2719–2732, Dec. 2014.

- [5] R. Heliot, A. L. Orsborn, K. Ganguly, and J. M. Carmena, "System architecture for stiffness control in brain-machine interfaces," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 40, no. 4, pp. 732–742, Jul. 2010.
- [6] W. Sheng, A. Thobbi, and Y. Gu, "An integrated framework for humanrobot collaborative manipulation," *IEEE Trans. Cybern.*, vol. 45, no. 10, pp. 2030–2041, Oct. 2015.
- [7] G. Nejat and M. Ficocelli, "Can I be of assistance? The intelligence behind an assistive robot," in *Proc. IEEE Int. Conf. Robot. Autom.*, Pasadena, CA, USA, 2008, pp. 3564–3569.
- [8] D. McColl and G. Nejat, "Meal-time with a socially assistive robot and older adults at a long-term care facility," *J. Human Robot Interact.*, vol. 2, no. 1, pp. 152–171, 2013.
- [9] W. G. Louie, D. McColl, and G. Nejat, "Playing a memory game with a socially assistive robot: A case study at a long-term care facility," in *Proc. IEEE Int. Symp. Robot Human Interact. Commun.*, Paris, France, 2012, pp. 345–350.
- [10] D. McColl, J. Chan, and G. Nejat, "A socially assistive robot for mealtime cognitive interventions," *J. Med. Devices Trans. ASME*, vol. 6, no. 1, 2012, Art. ID 017559.
- [11] S. Gong, P. W. McOwan, and C. Shan, "Beyond facial expressions: Learning human emotion from body gestures," in *Proc. British Mach. Vis. Conf.*, Warwick, U.K., 2007, pp. 1–10.
- [12] J. Sundberg, S. Patel, E. Bjorkner, and K. R. Scherer, "Interdependencies among voice source parameters in emotional speech," *IEEE Trans. Affect. Comput.*, vol. 2, no. 3, pp. 162–174, Jul./Sep. 2011.
- [13] M. Song, D. Tao, Z. Liu, X. Li, and M. C. Zhou, "Image ratio features for facial expression recognition application," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 3, pp. 779–788, Jun. 2010.
- [14] K. L. Walters and R. D. Walk, "Perception of emotion from body posture," Bull. Psychon. Soc., vol. 24, no. 5, p. 329, 1986.
- [15] S. J. Schouwstra and J. Hoogstraten, "Head position and spinal position as determinants of perceived emotional state," *Percept. Motor Skills*, vol. 81, no. 2, pp. 673–674, 1995.
- [16] M. Karg, K. Kühnlenz, and M. Buss, "Recognition of affect based on gait patterns," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 4, pp. 1050–1061, Aug. 2010.
- [17] H. Gunes and M. Piccardi, "Automatic temporal segment detection and affect recognition from face and body display," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 1, pp. 64–84, Feb. 2009.
- [18] G. Castellano, S. Villalba, and A. Camurri, "Recognising human emotions from body movement and gesture dynamics," *Affective Computing and Intelligent Interaction*, (LNCS 4738). Berlin Heidelberg, Germany: Springer, 2007, pp. 71–82.
- [19] D. Bernhardt and P. Robinson, "Detecting affect from non-stylized body motions," *Affective Computing and Intelligent Interaction*, (LNCS 4738). Berlin Heidelberg, Germany: Springer, 2007, pp. 59–70.
- [20] A.-A. Samadani, R. Gorbet, and D. Kulic, "Affective movement recognition based on generative and discriminative stochastic dynamic models," *IEEE Trans. Human–Mach. Syst.*, vol. 44, no. 4, pp. 454–467, Aug. 2014.
- [21] A. Kleinsmith, N. Bianchi-Berthouze, and A. Steed, "Automatic recognition of non-acted affective postures," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 41, no. 4, pp. 1027–1038, Aug. 2011.
- [22] A. Kapoor, W. Burleson, and R. W. Picard, "Automatic prediction of frustration," *Int. J. Human Comput. Stud.*, vol. 65, no. 8, pp. 724–736, 2007.
- [23] A. S. Shminan, T. Tamura, and R. Huang, "Student awareness model based on student affective response and generic profiles," in *Proc. IEEE Int. Conf. Inf. Sci. Technol.*, Hubei, China, 2012, pp. 675–681.
- [24] L. Zhang and B. Yap, "Affect detection from text-based virtual improvisation and emotional gesture recognition," Adv. Human-Comput. Interact., vol. 2012, Jan. 2012, Art. ID 461247.
- [25] D. McColl, Z. Zhang, and G. Nejat, "Human body pose interpretation and classification for social human–robot interaction," *Int. J. Soc. Robot.*, vol. 3, no. 3, pp. 313–332, 2011.
- [26] D. McColl and G. Nejat, "Affect detection from body language during social HRI," in *Proc. IEEE Int. Symp. Robot Human Interact. Commun.*, Paris, France, 2012, pp. 1013–1018.
- [27] E. P. Bull, Posture and Gesture. Oxford, U.K.: Pergamon Press, 1987.
- [28] M. Coulson, "Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence," J. Nonverbal Behav., vol. 28, no. 2, pp. 117–139, 2004.
- [29] A. Mehrabian, "Significance of posture and position in the communication of attitude and status relationships," *Psychol. Bull.*, vol. 71, no. 5, pp. 359–372, 1969.

- [30] M. Davis and D. Hadiks, "Nonverbal aspects of therapist attunement," J. Clin. Psychol., vol. 50, no. 3, pp. 393–405, 1994.
- [31] M. Davis and D. Hadiks, "Nonverbal behavior and client state changes during psychotherapy," J. Clin. Psychol., vol. 46, no. 3, pp. 340–351, 1990.
- [32] N. Bianchi-Berthouze, P. Cairns, and A. L. Cox, "On posture as a modality for expressing and recognizing emotions," presented at Emotion HCI Workshop BCS HCI London, 2006, pp. 74–80.
- [33] P. Ekman and W. V. Friesen, "Head and body cues in the judgment of emotion: A reformulation," *Percept. Motor Skills*, vol. 24, no. 3, pp. 711–724, 1967.
- [34] M. Davis, "Guide to movement analysis methods," Pittsburgh, PA, USA: Behavioral Measurement Database Services, 1997.
- [35] P. Ekman and V. Friesen, "The repertoire of nonverbal behavior: Categories, origins, and coding," *Semiotica*, vol. 1, no. 1, pp. 49–98, 1969.
- [36] B. Burger, I. Ferrané, and F. Lerasle, "Multimodal interaction abilities for a robot companion," in *Proc. Int. Conf. Comput. Vis. Syst.*, 2008, pp. 549–558.
- [37] H. Park et al., "HMM-based gesture recognition for robot control," Pattern Recognit. Image Anal., (LNCS 3522). Berlin Heidelberg, Germany: Springer, 2005, pp. 607–614.
- [38] N. Koenig, S. Chernova, C. Jones, M. Loper, and O. Jenkins, "Handsfree interaction for human-robot teams," in *Proc. ICRA Workshop Soc. Interact. Intell. Indoor Robot.*, Pasadena, CA, USA, 2008, pp. 35–41.
- [39] J. Sanghvi et al., "Automatic analysis of affective postures and body motion to detect engagement with a game companion," in Proc. ACM/IEEE Int. Conf. Human Robot Interact., Lausanne, Switzerland, 2011, pp. 305–311.
- [40] T. Lourens, R. van Berkel, and E. Barakova, "Communicating emotions and mental states to robots in a real time parallel framework using Laban movement analysis," *Robot. Auton. Syst.*, vol. 58, no. 12, pp. 1256–1265, 2010.
- [41] B. Freedman, A. Shpunt, M. Machline, and Y. Arieli, "Depth mapping using projected patterns," U.S. Patent 2010 0 118 123 A1, May 13, 2010.
- [42] C. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer-Verlag, 2006.
- [43] M. M. Loper, N. P. Koenig, S. H. Chernova, C. V. Jones, and O. C. Jenkins, "Mobile human-robot teaming with environmental tolerance," in *Proc. ACM/IEEE Int. Conf. Human Robot Interact.*, La Jolla, CA, USA, 2009, pp. 157–164.
- [44] J. Satake and J. Miura, "Robust stereo-based person detection and tracking for a person following robot," in *Proc. ICRA Workshop Person Detect. Tracking*, Kobe, Japan, 2009, pp. 1–6.
- [45] G. Medioni, A. R. J. François, M. Siddiqui, K. Kim, and H. Yoon, "Robust real-time vision for a personal service robot," *Comput. Vis. Image Und. Archive*, vol. 108, nos. 1–2, pp. 196–203, 2007.
- [46] D. Chai and K. N. Ngan, "Face segmentation using skin-color map in videophone applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 4, pp. 551–564, Jun. 1999.
- [47] M. Hall et al., "The WEKA data mining software: An update," ACM SIGKDD Explor. Newslett., vol. 11, no. 1, pp. 10–18, 2009.
- [48] Y.-H. Kim, S.-Y. Hahn, and B.-T. Zhang, "Text filtering by boosting naive Bayes classifiers," in *Proc. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, Athens, Greece, 2000, pp. 168–175.
- [49] Microsoft. (2014). Kinect for Windows Programming Guide. [Online]. Available: http://msdn.microsoft.com/en-us/library/hh855348.aspx
- [50] J. Shotton *et al.*, "Real-time human pose recognition in parts from single depth images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Colorado Springs, CO, USA, 2011, pp. 1297–1304.
- [51] K. Khoshelham and S. O. Elberink, "Accuracy and resolution of kinect depth data for indoor mapping applications," *Sensors*, vol. 12, no. 2, pp. 1437–1454, 2012.
- [52] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.
- [53] T. Shiwa, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita, "How quickly should communication robots respond?" in *Proc. ACM/IEEE Conf. Human–Robot Interact.*, Amsterdam, The Netherlands, 2008, pp. 153–160.
- [54] M. K. Lee, S. Kiesler, J. Forlizzi, and P. Rybski, "Ripple effects of an embedded social agent: A field study of a social robot in the workplace," in *Proc. ACM SIGCHI Conf. Human Factors Comput. Syst.*, Austin, TX, USA, 2012, pp. 695–704.
- [55] P. A. D. Weeks, "Error-correction techniques and sequences in instructional settings: Toward a comparative framework," *Human Studies*, vol. 8, no. 3, pp. 195–233, 1985.

- [56] N.-T. Telle and H.-R. Pfister, "Not only the miserable receive help: Empathy promotes prosocial behaviour toward the happy," *Current Psychol.*, vol. 31, no. 4, pp. 393–413, 2012.
- [57] A. Comfort, Reality and Empathy: Physics, Mind and Science in the 21st Century. Albany, NY, USA: State Univ. New York Press, 1984.
- [58] A. Bruce, I. Nourbakhsh, and R. Simmons, "The role of expressiveness and attention in human-robot interaction," in *Proc. IEEE Int. Conf. Robot. Autom.*, vol. 4. Washington, DC, USA, 2002, pp. 4138–4142.
- [59] K. Hone, "Empathic agents to reduce user frustration: The effects of varying agent characteristics," *Interact. Comput.*, vol. 18, no. 2, pp. 227–245, 2006.
- [60] C. N. Moridis and A. A. Economides, "Affective learning: Empathetic agents with emotional facial and tone of voice expressions," *IEEE Trans. Affect. Comput.*, vol. 3, no. 3, pp. 260–272, Jul./Sep. 2012.
- [61] H. Maldonado et al., "We learn better together: Enhancing eLearning with emotional characters," in *Computer Supported Collaborative Learning: The Next 10 Years!*, T. Koschmann, D. Suthers, and T. W. Chan, Eds. Mahwah, NJ, USA: Lawernce Erlbaum Associates, 2005, pp. 408–417.
- [62] P. Ekman, W. Friesen, and S. Ancoli, "Facial signs of emotional experience," J. Pers. Social Psychol., vol. 39, no. 6, pp. 1125–1134, 1980.
- [63] A. Kapoor and R. W. Picard, "A real-time head nod and shake detector," in *Proc. ACM Workshop Perceptive User Interfaces*, Orlando, FL, USA, 2001, pp. 1–5.
- [64] J. Montepare *et al.*, "The use of body movements and gestures as cues to emotions in younger and older adults," *J. Nonverbal Behav.*, vol. 23, no. 2, pp. 133–152, 1999.
- [65] O. Hargie and D. Dickson, Skilled Interpersonal Communication: Research Theory and Practise. New York, NY, USA: Routledge, 2004.
- [66] T. Housel and C. Wheeler, "The effects of nonverbal reinforcement and interviewee–interviewer relationship on interviewee's verbal response," *J. Appl. Commun. Res.*, vol. 8, no. 2, pp. 120–127, 1980.
- [67] D. McNeill, Hand and Mind: What Gestures Reveal About Thought. Chicago, IL, USA: Univ. Chicago Press, 1992.
- [68] D. A. Small and N. M. Verrochi, "The face of need: Facial emotion expression on charity advertisements," J. Mark. Res., vol. 46, no. 6, pp. 777–787, 2009.
- [69] N. Eisenberg *et al.*, "Relation of sympathy and personal distress to prosocial behavior: A multimethod study," *J. Pers. Soc. Psychol.*, vol. 57, no. 1, pp. 55–66, 1989.
- [70] W. Furman and J. C. Masters, "Affective consequences of social reinforcement, punishment, and neutral behavior," *Develop. Psychol.*, vol. 16, no. 2, pp. 100–104, 1980.
- [71] M. De Boni, A. Richardson, and R. Hurling, "Humour, relationship maintenance and personality matching in automated dialogue," *Interact. Comput.*, vol. 20, no. 3, pp. 342–353, 2008.
- [72] A. Niculescu, B. van Dijk, A. Nijholt, H. Li, and S. Lan See, "Making social robots more attractive: The effects of voice pitch, humor and empathy," *Int. J. Soc. Robot.*, vol. 5, no. 2, pp. 171–191, 2013.
- [73] B. de Ruyter, P. Saini, P. Markopoulos, and A. van Breemen, "Assessing the effects of building social intelligence in a robotic interface for the home," *Interact. Comput.*, vol. 17, no. 5, pp. 522–541, 2005.
- [74] A. P. Field, Discovering Statistics Using SPSS. Los Angeles, CA, USA: Sage, 2009.
- [75] P. Sturmey, J. T. Newton, A. Cowley, N. Bouras, and G. Holt, "The PAS-ADD checklist: Independent replication of its psychometric properties in a community sample," *Brit. J. Psychiat.*, vol. 186, no. 4, pp. 319–323, 2005.
- [76] M. S. Mast, J. A. Hall, N. A. Murphy, and C. R. Colvin, "Judging assertiveness," *Facta Univ. Philos. Sociol. Psychol.*, vol. 2, no. 10, pp. 731–744, 2003.
- [77] L. V. Calderita, J. P. Bandera, P. Bustos, and A. Skiadopoulos, "Modelbased reinforcement of kinect depth data for human motion capture applications," *Sensors*, vol. 13, no. 7, pp. 8835–8855, 2013.

- [78] E. T. Hall, *The Hidden Dimension*. Garden City, NY, USA: Doubleday, 1966.
- [79] I. Markovic and I. Petrovic, "Speaker localization and tracking with a microphone array on a mobile robot using Von Mises distribution and particle filtering," *Robot. Auton. Syst.*, vol. 58, no. 11, pp. 1185–1196, 2010.
- [80] F. Cid, J. Moreno, P. Bustos, and P. Núñez, "Muecas: A multi-sensor robotic head for affective human robot interaction and imitation," *Sensors*, vol. 14, no. 5, pp. 7711–7737, 2014.



Derek McColl (S'11–M'15) received the B.Sc. (Eng.) and M.A.Sc. degrees in mechanical engineering from Queen's University, Kingston, ON, Canada, in 2007 and 2010, respectively, and the Ph.D. degree in mechanical engineering from the University of Toronto (UofT), Toronto, ON, Canada, in 2015.

He was a Graduate Research Assistant with the Autonomous Systems and Biomechatronics Laboratory, UofT. He is currently a Post-Doctoral Fellow with Defence Research and Development

Canada, Ottawa, ON, Canada. His current research interests include robotics, human-robot interaction, sensing, human machine interfaces, and intelligent adaptive systems.



Chuan Jiang received the B.A.Sc. degree in engineering science (aerospace), and the M.A.Sc. degree in mechanical engineering from the University of Toronto (UofT), Toronto, ON, Canada, in 2012 and 2014, respectively.

He was a Research Assistant with the Autonomous Systems and Biomechatronics Laboratory, UofT. He is currently a Business Technology Analyst in management consulting at Deloitte LLP, New York, NY, USA. His current research interests include robotic systems integra-

tion, systems control, sensing and artificial intelligence, business development with disrupting technologies, and technology strategy development and implementation.



Goldie Nejat (S'03–M'06) received the B.A.Sc. and Ph.D. degrees in mechanical engineering from the University of Toronto (UofT), Toronto, ON, Canada, in 2001 and 2005, respectively.

She is currently an Associate Professor and the Director of the Autonomous Systems and Biomechatronics Laboratory, Department of Mechanical and Industrial Engineering, UofT. She is also the Director of the Institute for Robotics and Mechatronics, UofT, and an Adjunct Scientist with the Toronto Rehabilitation Institute, Toronto,

ON, Canada. Her current research interests include sensing, human-robot interaction, semi-autonomous and autonomous control, intelligence of assistive/service robots for search and rescue, exploration, healthcare, and surveillance applications.